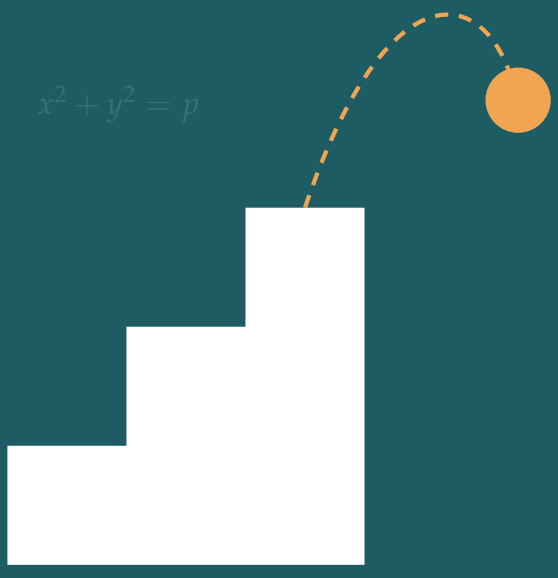
STEP 備試 · 大學數學

# Beyond STEP

## STEP I, II 及 III

1987 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**

## 1987 年詳細解答

Keith McNulty

**Beyond STEP：以歷屆 STEP 試題為門，走進大學數學**

**1987 年詳細解答**

1987 年 STEP 試卷（STEP I, II 及 III）全部 48 題的詳細解答。自樣本試卷以來每一道 STEP 題目的詳細解答、*Beyond STEP* 全書以及各章的單行本，均可在 `beyondstep.keithmcnulty.org` 免費閱讀及下載。

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

由機器從英文翻譯而成；尚未經審閱。

2026 年 11 月首次於網上發表。

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

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

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

# 目錄

## STEP I


第 1 題 曲線描繪 純數學 . . . 3

第 2 題 圓填充的面積 純數學 . . . 7

第 3 題 可分離變量微分方程 純數學 . . . 11

第 4 題 對數與等比級數 純數學 . . . 14

第 5 題 代換積分法 純數學 . . . 16

第 6 題 面積與反函數 純數學 . . . 20

第 7 題 三角級數 純數學 . . . 24

第 8 題 代換積分法 純數學 . . . 27

第 9 題 重心坐標 純數學 . . . 30

第 10 題 彈性繩 力學 . . . 34

第 11 題 圓周運動與能量 力學 . . . 38

第 12 題 拋體與恢復係數 力學 . . . 41

第 13 題 變力 力學 . . . 44

第 14 題 概率的等比級數 概率與統計 . . . 47

第 15 題 連續隨機變量 概率與統計 . . . 51

第 16 題 貝葉斯定理與正態近似 概率與統計 . . . 54


## STEP II


第 1 題 代數恆等式 純數學 . . . 57

第 2 題 三角恆等式 純數學 . . . 60

第 3 題 共軛根式 純數學 . . . 63

第 4 題 複數與幾何 純數學 . . . 67

第 5 題 拉格朗日反演級數 純數學 . . . 71

第 6 題 利用對稱性求定積分 純數學 . . . 75

第 7 題 梯形法則與階乘 純數學 79
第 8 題 三維空間中的向量軌跡 純數學 83
第 9 題 正交矩陣 純數學 87
第 10 題 龐公理 純數學 90
第 11 題 摩擦力與力矩 力學 93
第 12 題 衡與能量 力學 96
第 13 題 牆壁的斜碰撞 力學 100
第 14 題 錐上的彈性帶 力學 105
第 15 題 幾何概率 概率與統計 108
第 16 題 葉斯定理 概率與統計 112

## STEP III

第 1 題 階乘的整除性 純數學 115
第 2 題 三角函數歸約公式 純數學 118
第 3 題 弧角與梅欽公式 純數學 121
第 4 題 旋轉體的體積 純數學 125
第 5 題 模算術密碼 純數學 128
第 6 題 耦合微分方程 純數學 131
第 7 題 反正切的級數 純數學 135
第 8 題 平行四邊形面積不等式 純數學 138
第 9 題 群的直積 純數學 142
第 10 題 值努利多項式 純數學 146
第 11 題 相對速度的最優化 力學 149
第 12 題 變質量轉動 力學 153
第 13 題 涉及摩擦力的轉動桿 力學 157
第 14 題 彈力坍縮 力學 160
第 15 題 期望成本的最優化 概率與統計 164
第 16 題 順序統計量與泊松分佈 概率與統計 167

詳細解答

# STEP I 1987，第 1 題

純數學 • 曲線描繪

**STEP I 1987，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/87-s1-q1

## 關鍵思路

用乘積法則（product rule）求導後，會剩下一個永不為零的因子 $e^{ax}$，乘以 $\cos bx$ 與 $\sin bx$ 的某個組合。因此駐點（stationary points）全部來自同一條方程 $\tan bx =$ 常數，而這條方程的解是等距分佈的，相鄰兩解相距 $\pi/b$。從一個駐點移到下一個，指數部分乘以一個固定的因子，而餘弦部分則變號，這正正就是一個等比數列（geometric progression）。

## 解答

### 駐點

由乘積法則，

$$f'(x) = ae^{ax}\cos bx - be^{ax}\sin bx = e^{ax}(a\cos bx - b\sin bx).$$

由於對所有 $x$ 都有 $e^{ax} > 0$，所以 $f'(x) = 0$ 當且僅當 $a\cos bx = b\sin bx$。

? **想一想** 這裏可以放心除以 $\cos bx$ 嗎？

可以。若 $\cos bx = 0$，則 $\sin bx = \pm 1$，方程便變成 $0 = \pm b$，但因為 $b > 0$，這是不成立的。所以在任何駐點上都有 $\cos bx \neq 0$，兩邊相除得

$$\tan bx = \frac{a}{b}.$$

設 $\alpha = \arctan(a/b)$。因為 $a/b > 0$，所以 $0 < \alpha < \frac{\pi}{2}$。由於 tan 的週期（period）是 $\pi$，方程的解為 $bx = \alpha + n\pi$，因此駐點位於

$$x_n = \frac{\alpha + n\pi}{b}, \qquad n \in \mathbb{Z}, \qquad \text{其中 } \tan\alpha = \frac{a}{b}.$$

考慮兩條直角邊分別為 $a$ 和 $b$ 的直角三角形，可得 $\cos\alpha = \dfrac{b}{\sqrt{a^2+b^2}}$。由於 $\cos(\theta + n\pi) = (-1)^n \cos\theta$，f 在該處的值為

$$\begin{aligned} \mathrm{f}(x_n) &= \mathrm{e}^{a(\alpha+n\pi)/b}\cos(\alpha+n\pi) \\ &= (-1)^n \frac{b}{\sqrt{a^2+b^2}}\,\mathrm{e}^{a\alpha/b}\,\mathrm{e}^{na\pi/b}. \end{aligned}$$

**極大還是極小？** 再求導一次，

$$\mathrm{f}''(x) = \mathrm{e}^{ax}((a^2-b^2)\cos bx - 2ab\sin bx).$$

在駐點上 $b\sin bx = a\cos bx$，所以 $2ab\sin bx = 2a^2\cos bx$，於是

$$\mathrm{f}''(x_n) = \mathrm{e}^{ax_n}\cos bx_n\,(a^2-b^2-2a^2) = -(a^2+b^2)\,\mathrm{f}(x_n).$$

因此當 $\mathrm{f}(x_n) > 0$（$n$ 為偶數）時，$x_n$ 是極大點（maximum）；當 $\mathrm{f}(x_n) < 0$（$n$ 為奇數）時，$x_n$ 是極小點（minimum）。駐點交替出現：極大、極小、極大、...

**等比數列**

由於 $0 < \alpha < \pi$，$x_n > 0$ 當且僅當 $n \geq 0$。所以 $x > 0$ 的駐點是 $x_0 < x_1 < x_2 < \cdots$，而對每個 $n \geq 0$，

$$\frac{\mathrm{f}(x_{n+1})}{\mathrm{f}(x_n)} = \frac{\mathrm{e}^{a(\alpha+(n+1)\pi)/b}\cos(\alpha+(n+1)\pi)}{\mathrm{e}^{a(\alpha+n\pi)/b}\cos(\alpha+n\pi)} = \mathrm{e}^{a\pi/b}\cdot(-1) = -\mathrm{e}^{a\pi/b},$$

這裏用到 $\cos(\theta+\pi) = -\cos\theta$ 及 $\cos(\alpha+n\pi) \neq 0$。這個比值與 $n$ 無關，所以 $\mathrm{f}(x_0), \mathrm{f}(x_1), \ldots$ 組成一個等比數列，首項為 $\dfrac{b\,\mathrm{e}^{a\alpha/b}}{\sqrt{a^2+b^2}}$，公比（common ratio）為 $-\mathrm{e}^{a\pi/b}$，得證。

（同樣的計算對負數 $n$ 亦成立，所以事實上所有駐值組成一個向兩邊無限延伸的等比數列。限制 $x > 0$ 只是給它一個自然的首項。）

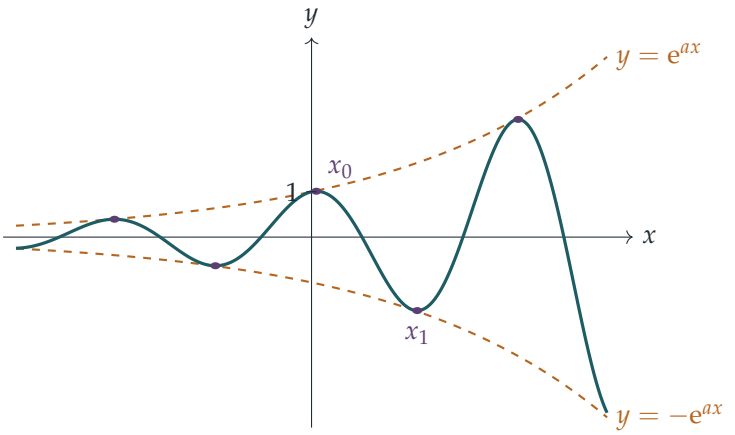
**草圖**

所需的事實：

- 由於 $-1 \leq \cos bx \leq 1$，圖像夾在曲線 $y = \mathrm{e}^{ax}$ 與 $y = -\mathrm{e}^{ax}$ 之間。在 $\cos bx = 1$（$bx = 2m\pi$）處，圖像與 $y = \mathrm{e}^{ax}$ 相接；在 $\cos bx = -1$（$bx = (2m+1)\pi$）處，則與 $y = -\mathrm{e}^{ax}$ 相接。在這些點上 $\sin bx = 0$，所以 $\mathrm{f}'(x) = \pm a\mathrm{e}^{ax}$，與包絡線（envelope）的斜率相同：曲線在該處與包絡線相切，而不是轉向。
- 零點位於 $bx = \frac{\pi}{2} + m\pi$，相鄰兩個相距 $\pi/b$，而且 $\mathrm{f}(0) = 1$。

- 每個轉向點（turning point）$bx = \alpha + n\pi$ 都位於對應的相接點 $bx = n\pi$ 稍後的位置，因為 $0 < \alpha < \frac{\pi}{2}$：不斷上升的包絡線把每個峯頂稍稍拉向右方。
- 當 $x \to \infty$ 時，振盪幅度無界地增大；當 $x \to -\infty$ 時，$\mathrm{e}^{ax} \to 0$，所以 $\mathrm{f}(x) \to 0$，但仍然不斷振盪。

草圖所示為 $a = 0.3$、$b = 2$ 的情況，包絡線 $y = \pm\mathrm{e}^{ax}$ 以虛線表示。




## 驗算

- 取 $a = b = 1$：此時 $\alpha = \frac{\pi}{4}$，$x_0 = \frac{\pi}{4}$，$\mathrm{f}(x_0) = \mathrm{e}^{\pi/4}/\sqrt{2} \approx 1.551$，而 $\mathrm{f}(x_1) = \mathrm{e}^{5\pi/4}\cos\frac{5\pi}{4} \approx -35.89$。比值約為 $\approx -23.14 = -\mathrm{e}^{\pi}$。
- 當 $a \to 0$ 時，函數變成 $\cos bx$：$\alpha \to 0$，轉向點移到 $bx = n\pi$，比值趨向 $-1$，得出大家熟悉的 $1, -1, 1, \ldots$
- 我們以數值方法對 500 組隨機的 $(a, b)$ 作驗證，確認了 $\mathrm{f}'(x_n) = 0$、$\mathrm{f}(x_n)$ 與上述公式相符、$\mathrm{f}''(x_n) = -(a^2 + b^2)\mathrm{f}(x_n)$、相鄰比值等於 $-\mathrm{e}^{a\pi/b}$，以及 $\mathrm{f}'$ 沒有其他零點。

## 常見錯誤

- 把轉向點放在曲線與包絡線相接之處，即 $bx = n\pi$。這些點的斜率是 $\pm a\mathrm{e}^{ax} \neq 0$；真正的轉向點稍稍偏右。
- 只取 $bx = \alpha$，忘記了 $\tan bx = a/b$ 的通解（general solution）$bx = \alpha + n\pi$。
- 漏掉比值中的負號：餘弦在相鄰兩個駐點之間變號，所以駐值交替為極大、極小。

## STEP 之外

函數 $e^{ax}\cos bx$ 是 $e^{(a+ib)x}$ 的實部（real part），而這類函數正正就是那些輔助方程（auxiliary equation）有複數根的線性微分方程（linear differential equation）$y'' + py' + qy = 0$ 的解。當 $a < 0$ 時，得到的是阻尼振動（damped oscillation），例如有摩擦力的擺；當 $a > 0$ 時(如本題)，則是一個振盪不斷增大的不穩定系統。相鄰峯值之比是常數，而這個比值的對數就是工程師所說的對數衰減率（logarithmic decrement）。Boyce 與 DiPrima 的 *Elementary Differential Equations*（《初等微分方程》）在二階線性方程一章中詳細發展了這個題材；Needham 的 *Visual Complex Analysis*（《直觀複分析》）則說明了為甚麼乘以 $e^{(a+ib)x}$ 會描出一條螺線。

詳細解答

# STEP I 1987，第 2 題

純數學 • 圓填充的面積

> **STEP I 1987，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/87-s1-q2

## 關鍵思路

把各個小圓的圓心連起來。這些圓心全都位於半徑為 $R$ 的圓上，即恰好在兩個大圓的正中間；它們構成一個正 $n$ 邊形（regular polygon），而這個多邊形的各邊都穿過相鄰小圓的切點。這個多邊形正是劃分 $Y$（在多邊形內）和 $X$（在多邊形外）兩部分的分界線。因此，每一塊面積都可以由三種簡單圖形得出：多邊形、圓盤，以及由多邊形頂點截出的小圓扇形（sector）。

## 解答

### $n$、$R$ 與 $r$ 的關係

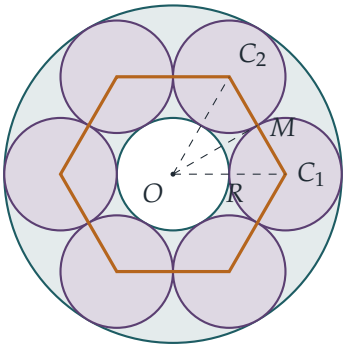
每個小圓都與兩個大圓相切，所以它的圓心與公共圓心 $O$ 的距離是 $R - r + r = R$：全部 $n$ 個圓心都在半徑為 $R$ 的圓上。由對稱性可知，這些圓心等距分佈，所以相鄰的圓心 $C_1$、$C_2$ 在 $O$ 處所張的角是 $2\pi/n$。

兩個相鄰的圓互相相切，所以 $C_1C_2 = 2r$。在等腰三角形（isosceles triangle）$OC_1C_2$ 中，$OC_1 = OC_2 = R$；從 $O$ 向 $C_1C_2$ 的中點 $M$ 作垂線，可得

$$r = C_1M = R\sin\frac{\pi}{n}.$$

這就是所求的關係。（我們需要 $n \geq 3$；若 $n = 2$，內圓的半徑會是 $R - r = 0$。）同一個三角形亦給出

$$OM = R\cos\frac{\pi}{n} = \sqrt{R^2 - r^2}.$$




上圖（以 $n = 6$ 繪畫）用琥珀色顯示由圓心構成的多邊形。它的各邊都穿過相鄰圓的切點，例如 $M$。

**面積 $Y$**

這個多邊形由 $n$ 個與 $OC_1C_2$ 全等（congruent）的三角形組成，每個的面積是 $\frac{1}{2} \cdot 2r \cdot \sqrt{R^2 - r^2}$。所以

$$\text{多邊形的面積} = nr\sqrt{R^2 - r^2}.$$

多邊形內有甚麼？它的內切圓半徑（inradius）是 $OM = \sqrt{R^2 - r^2}$，而 $R^2 - r^2 \geq (R - r)^2$，因為這個不等式（inequality）移項整理後就是 $2r(R - r) \geq 0$。所以半徑為 $R - r$ 的內圓盤位於多邊形之內。每個頂點處都有一個小圓的扇形。正 $n$ 邊形的內角（interior angle）是 $\pi - \frac{2\pi}{n}$，所以每個這樣的扇形面積是

$$\tfrac{1}{2}r^2\left(\pi - \frac{2\pi}{n}\right) = \pi r^2\left(\frac{1}{2} - \frac{1}{n}\right).$$

這 $n$ 個扇形和內圓盤彼此不重疊，因為小圓之間以及小圓與內圓之間都只是相切。多邊形餘下的部分，恰好就是緊貼內圓的 $n$ 個曲邊三角形，而它們合起來就是 $Y$：

$$Y = nr\sqrt{R^2 - r^2} - \pi(R - r)^2 - n\pi r^2\left(\frac{1}{2} - \frac{1}{n}\right),$$

正如所求。

**想一想** 為甚麼邊 $C_1C_2$ 正是把 $X$ 與 $Y$ 分開的恰當位置？

邊 $C_1C_2$ 由圓心連到圓心，並穿過切點 $M$，所以它永遠不會穿過圓與圓之間的空隙。所有緊貼內圓的空隙都在它靠 $O$ 的一側，而所有緊貼外圓的空隙都在另一側。因此沿着多邊形切開，便能乾淨地把兩類空隙分開；而且它沿兩條半徑切開每個小圓，這就是我們得到扇形的原因。

### 面積 $X$ 與 $Z$

在多邊形之外而外圓之內的區域，我們要除去每個小圓餘下的部分。餘下的部分是一個優角扇形（reflex sector），角度為 $2\pi-(\pi-\frac{2\pi}{n})=\pi+\frac{2\pi}{n}$，面積為 $\pi r^2\left(\frac{1}{2}+\frac{1}{n}\right)$。因此

$$X=\pi(R+r)^2-nr\sqrt{R^2-r^2}-n\pi r^2\left(\frac{1}{2}+\frac{1}{n}\right).$$

$n$ 個半徑為 $r$ 的圓盤給出

$$Z=n\pi r^2.$$

### 當 $n\to\infty$ 時的極限

兩式相加，多邊形的項互相抵消，而扇形的項合併為 $n\pi r^2$：

$$\begin{aligned}X+Y&=\pi(R+r)^2-\pi(R-r)^2-n\pi r^2\\&=4\pi Rr-n\pi r^2.\end{aligned}$$

這其實就是圓環（annulus）的面積減去各小圓的面積，理應如此。因此

$$\frac{X+Y}{Z}=\frac{4\pi Rr-n\pi r^2}{n\pi r^2}=\frac{4R}{nr}-1=\frac{4}{n\sin(\pi/n)}-1,$$

這裏用了 $r=R\sin(\pi/n)$。當 $n\to\infty$ 時，$\pi/n\to 0$ 且 $\dfrac{\sin(\pi/n)}{\pi/n}\to 1$，所以 $n\sin(\pi/n)\to\pi$，從而

$$\frac{X+Y}{Z}\to\frac{4}{\pi}-1\approx 0.273.$$

## 驗算

- 取 $n=3$、$R=1$，得 $r=\frac{\sqrt{3}}{2}$、$Y\approx 0.0646$ 及 $X\approx 3.7497$。在精細的極坐標網格上進行數值積分，把每一點分類為在某個圓內、在多邊形內或在多邊形外，結果在 $n=3,5,8,12$ 時均與公式吻合至四位有效數字。
- 比值 $\frac{4}{n\sin(\pi/n)}-1$ 在 $n=10$ 時為 $0.2944$，在 $n=100$ 時為 $0.27345$，在 $n=1000$ 時為 $0.273240$。它正逐漸逼近 $4/\pi-1=0.273240\ldots$。
- 直觀地（heuristically）看，當 $n$ 很大時，這些圓就像寬度為 $2r$ 的長條中一排直徑為 $2r$ 的圓盤。每個 $2r\times 2r$ 的正方形中，圓盤佔 $\pi r^2$，空隙佔 $4r^2-\pi r^2$，所以比值是 $(4-\pi)/\pi$。

## 常見錯誤

- 計算扇形時誤用多邊形的外角（exterior angle）$2\pi/n$，而不是內角 $\pi-2\pi/n$。

► 計算 $Y$ 時忘記除去內圓盤，或沒有驗證內圓盤確實位於多邊形之內。

## STEP 之外

最後的極限（limit）其實就是改頭換面的 $\lim_{x\to 0}\frac{\sin x}{x}=1$。正是同一事實，讓阿基米德（Archimedes）能夠用內接和外切多邊形的周長把 $\pi$ 夾在中間。這道題亦是圓填充密度問題（circle-packing density problem）的一個簡化版本，這裏問的是：這些圓能填滿一個狹窄圓環的多少比例？若想嚴格處理這類極限，可以閱讀 Abbott 的 *Understanding Analysis*（《理解分析》）。

詳細解答

# STEP I 1987，第 3 題

純數學 • 可分離變量微分方程

> **STEP I 1987，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/87-s1-q3

## 關鍵思路

這條方程以 $v$ 表示時看似棘手，但左邊其實很容易辨認：$x^3v' + x^2v$ 正是 $x^2$ 乘以 $xv$ 的導數（derivative）。因此，題目建議的代換（substitution）把左邊變成 $x^2y'$；而右邊改用 $y$ 表示後，變量可以乾淨利落地分離（separate）。之後只需運用部分分式（partial fractions），再小心選取正負號。

## 解答

### 變量代換

設 $y = xv$，由乘積法則（product rule）得 $y' = v + xv'$，因此

$$x^3\frac{\mathrm{d}v}{\mathrm{d}x} + x^2v = x^2\left(x\frac{\mathrm{d}v}{\mathrm{d}x} + v\right) = x^2\frac{\mathrm{d}y}{\mathrm{d}x}.$$

在右邊，$x^2v^2 = y^2$ 而 $v = y/x$，所以右邊為

$$\frac{1+y^2}{(1+x^2)\,y/x} = \frac{x(1+y^2)}{(1+x^2)\,y}.$$

方程變為

$$x^2\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{x(1+y^2)}{(1+x^2)\,y}, \qquad \text{即} \qquad \frac{y}{1+y^2}\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{1}{x(1+x^2)},$$

這是把兩邊除以 $x^3$（我們在 $x = 1$ 附近考慮，故 $x > 0$），再乘以 $y/(1+y^2)$ 所得。變量已經分離。

### 積分

左邊積分得 $\frac{1}{2}\ln(1+y^2)$。至於右邊，用部分分式可得

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

只要把右邊通分便可輕易驗證。因此，當 $x>0$ 時，

$$\tfrac{1}{2}\ln(1+y^2) = \ln x - \tfrac{1}{2}\ln(1+x^2) + C, \qquad \text{因此} \qquad 1+y^2 = \frac{Ax^2}{1+x^2}$$

其中 $A = \mathrm{e}^{2C}$ 為正常數。

當 $x=1$ 時 $v=1$，故 $y=1$，於是 $2 = A/2$，即 $A=4$。因此

$$y^2 = \frac{4x^2}{1+x^2} - 1 = \frac{3x^2-1}{1+x^2}, \qquad v^2 = \frac{y^2}{x^2} = \frac{3x^2-1}{x^2(1+x^2)}.$$

? **想一想** 應該取哪一個平方根？解在哪個範圍內成立？

原方程的分母含 $v$，所以解永遠不能經過 $v=0$；由於解是連續的，它會保持在 $x=1$ 處的符號，即正號。因此取正平方根：

$$\boxed{v(x) = \frac{\sqrt{3x^2-1}}{x\sqrt{1+x^2}}}$$

此解在 $x > 1/\sqrt{3}$ 時成立。當 $x$ 減小至 $1/\sqrt{3}$ 時，$v$ 降至 0，方程便不再成立，所以解不能延伸到該點以下。

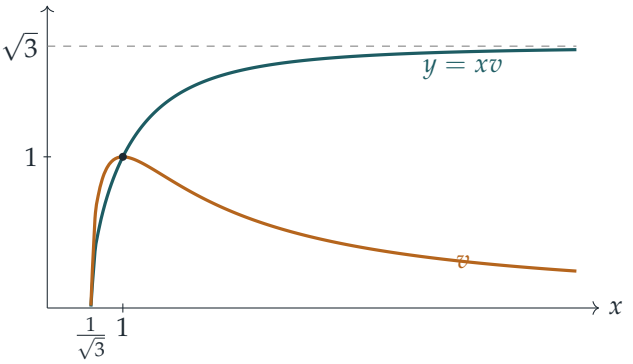
### $x$ 很大時

分子分母同除以 $x^2$，

$$v(x) = \frac{1}{x}\cdot\frac{\sqrt{3-1/x^2}}{\sqrt{1+1/x^2}} \approx \frac{\sqrt{3}}{x},$$

所以當 $x\to\infty$ 時 $v(x)\to 0$。等價地說，$y = xv \to \sqrt{3}$：$y$ 的圖像有水平漸近線（horizontal asymptote）$y=\sqrt{3}$，而 $v = y/x$ 則像 $\sqrt{3}/x$ 那樣衰減。

事實上，起點正是最高點：設 $u = x^2$，則 $(3u-1)/(u^2+u)$ 的導數的分子為 $-(3u+1)(u-1)$，所以 $v$ 上升至在 $x=1$ 處的最大值 1，然後遞減並趨向 0。




## 驗算

- 當 $x = 1$：$v = \sqrt{2}/(1 \cdot \sqrt{2}) = 1$，符合要求。
- 代回驗證：由 $1 + y^2 = 4x^2/(1 + x^2)$ 作隱函數微分（implicit differentiation）得 $2yy' = 8x/(1 + x^2)^2$，而 $(1 + y^2)/(x(1 + x^2)) = 4x/(1 + x^2)^2$，所以 $yy'/(1 + y^2) = 1/(x(1 + x^2))$，與預期相符。
- 數值上，以龍格—庫塔法（Runge–Kutta method）從 $(1, 1)$ 起逐步求解原方程，得 $v(10) \approx 0.172058$，與公式吻合；而在 $x = 10^4$ 時，公式給出 $xv \approx 1.73205 \approx \sqrt{3}$。

## 常見錯誤

- 把極限答成 $\sqrt{3}$：那是 $y$ 的極限，但題目問的是 $v$，它趨向 0。
- 錯誤地處理常數的指數：$\ln(1 + y^2) = \ln \frac{x^2}{1+x^2} + 2C$ 應得出 $1 + y^2 = Ax^2/(1 + x^2)$，即一個相乘的常數，而不是 $x^2/(1 + x^2) + 2C$。
- 開平方時寫上 $\pm$ 卻沒有作出取捨：條件 $v(1) = 1$ 及不容許 $v = 0$ 已確定要取正號。

## STEP 之外

這裏的代換之所以奏效，是因為左邊在整體相乘後成為一個恰當導數（exact derivative），這與積分因子（integrating factor）和齊次方程（homogeneous equation）背後的想法如出一轍。處理正負號和定義域時所需的謹慎，讓我們初步窺見存在性與唯一性理論（existence and uniqueness theory）：經過某一給定點的解會一直存在，直至遇到令方程失去意義的地方，在這裏就是 $v = 0$。Boyce 與 DiPrima 的 *Elementary Differential Equations*（《初等微分方程》）討論了可分離方程（separable equation），以及解能延伸多遠的問題。

詳細解答

# STEP I 1987，第 4 題

純數學 • 對數與等比級數

> **STEP I 1987，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/87-s1-q4

## 關鍵思路

以不便處理的數為底的對數（logarithm），最好先化成自然對數（natural logarithm），因為這樣底數就會以一個簡單的因子出現。這裏每個底數都是 2 的冪，而指數 $2^n$ 會降下來成為乘數。把每一項都用 $\ln 2$ 表示之後，這個級數（series）其實就是一個喬裝了的幾何級數（geometric series）。

## 解答

### 改寫通項

固定 $n \geq 0$，並設 $c = \log_{2^{2^n}} \mathrm{e}$。根據提示中的定義，這表示

$$\left(2^{2^n}\right)^c = \mathrm{e}.$$

兩邊取自然對數（兩邊均為正數），得

$$c \cdot 2^n \ln 2 = \ln \mathrm{e} = 1, \qquad \text{所以} \qquad \log_{2^{2^n}} \mathrm{e} = \frac{1}{2^n \ln 2}.$$

作為對下標的驗證，$n = 0, 1, 2$ 分別給出底數 2、4、16，與首三項吻合，其值為 $\frac{1}{\ln 2}$、$\frac{1}{2\ln 2}$、$\frac{1}{4\ln 2}$。

> ? **想一想** $2^n$ 是從哪裏來的？

來自法則 $\ln(a^k) = k \ln a$，取 $a = 2$ 及 $k = 2^n$：底數 $2^{2^n}$ 的自然對數是 $2^n \ln 2$。正是指數的「塔」令每個底數都是前一個的平方，而底數平方會令對數減半。

### 求級數之和

因此級數為

$$\sum_{n=0}^{\infty}(-1)^n \log_{2^{2^n}} \mathrm{e} = \frac{1}{\ln 2}\sum_{n=0}^{\infty}\left(-\frac{1}{2}\right)^n.$$

這是首項為 1、公比為 $r=-\frac{1}{2}$ 的幾何級數。由於 $|r|<1$，它收斂（converges），其和為 $\frac{1}{1-r}=\frac{1}{1+\frac{1}{2}}=\frac{2}{3}$。因此

$$\sum_{n=0}^{\infty}(-1)^n \log_{2^{2^n}} \mathrm{e} = \frac{2}{3\ln 2} = \frac{1}{\frac{3}{2}\ln 2} = \frac{1}{\ln\left(2^{3/2}\right)} = \frac{1}{\ln(2\sqrt{2})},$$

這裏再次用了 $\ln(a^k)=k\ln a$，以及 $2^{3/2}=2\cdot 2^{1/2}=2\sqrt{2}$。這就是所求的和。

## 驗算

數值上，$\frac{1}{\ln(2\sqrt{2})}\approx 0.96180$。級數的部分和（partial sums）為 1.4427、0.7213、1.0820、0.9017、0.9919、0.9468,...，在此值兩側來回擺動，而差距每次減半，正如交錯幾何級數應有的表現。取六十項時，結果與 0.961797 在機器精度內一致。

## 常見錯誤

- 把 $\log_{2^{2^n}}$ 誤讀為 $\log_{2^{2n}}$：這樣會得出底數 2、16、256,... 及錯誤的公比 $-\frac{1}{4}$。應該用首幾項來核對下標。
- 寫成 $\log_a \mathrm{e}=\ln a$，而不是 $\frac{1}{\ln a}$。換底公式（change-of-base rule）給出 $\log_a b=\frac{\ln b}{\ln a}$，而這裏 $\ln b=1$。

## STEP 之外

一個無窮和究竟是否「有值」，正是分析學（analysis）入門課程的核心問題。幾何級數是最基本的參照標準：許多收斂判別法，例如比值判別法（ratio test），都是透過把級數與幾何級數比較來運作的。Abbott 的 *Understanding Analysis*（《理解分析》）細緻而易讀地建立了這套理論。

詳細解答

# STEP I 1987，第 5 題

純數學 • 代換積分法

> **STEP I 1987，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/87-s1-q5

## 關鍵思路

這個代換（substitution）的設計，是要令根號下的兩個因子都變成完全平方（perfect square）：$x-\alpha$ 變成 $\sin^2\theta$ 的某個倍數，而 $\beta-x$ 則變成 $\cos^2\theta$ 的同一倍數。於是平方根恰好與 $\mathrm{d}x/\mathrm{d}\theta$ 互相抵消，第一個積分的被積函數（integrand）便化為常數。第二個積分多保留了一個因子 $1/x$，只要把分子分母同除以 $\cos^2\theta$，它就變成一個標準的 $\tan\theta$ 積分。

## 解答

**第一個積分，$\alpha<\beta$ 的情況**

設 $x=\alpha\cos^2\theta+\beta\sin^2\theta$。利用 $\cos^2\theta+\sin^2\theta=1$，

$$x-\alpha=\alpha\cos^2\theta+\beta\sin^2\theta-\alpha(\cos^2\theta+\sin^2\theta)=(\beta-\alpha)\sin^2\theta,$$
$$\beta-x=\beta(\cos^2\theta+\sin^2\theta)-\alpha\cos^2\theta-\beta\sin^2\theta=(\beta-\alpha)\cos^2\theta,$$

以及

$$\frac{\mathrm{d}x}{\mathrm{d}\theta}=-2\alpha\cos\theta\sin\theta+2\beta\sin\theta\cos\theta=2(\beta-\alpha)\sin\theta\cos\theta.$$

至於積分上下限：$\theta=0$ 時 $x=\alpha$，$\theta=\frac{\pi}{2}$ 時 $x=\beta$。由於在 $(0,\frac{\pi}{2})$ 上 $\mathrm{d}x/\mathrm{d}\theta>0$，當 $\theta$ 走遍 $[0,\frac{\pi}{2}]$ 時，$x$ 遞增地走遍 $[\alpha,\beta]$ 恰好一次。

在這個範圍內 $\sin\theta\geq0$ 及 $\cos\theta\geq0$，而且 $\beta-\alpha>0$，所以

$$\sqrt{(x-\alpha)(\beta-x)}=\sqrt{(\beta-\alpha)^2\sin^2\theta\cos^2\theta}=(\beta-\alpha)\sin\theta\cos\theta.$$

因此

$$\int_\alpha^\beta \frac{\mathrm{d}x}{\sqrt{(x-\alpha)(\beta-x)}} = \int_0^{\pi/2} \frac{2(\beta-\alpha)\sin\theta\cos\theta}{(\beta-\alpha)\sin\theta\cos\theta}\,\mathrm{d}\theta = \int_0^{\pi/2} 2\,\mathrm{d}\theta = \pi,$$

即為所求。(被積函數在兩端都趨於無窮大，所以嚴格來說這是一個反常積分 (improper integral)。代換顯示它是收斂 (converges) 的，因為變換後的被積函數是有界的常數 2。)

**想一想** 我們究竟在哪裏用到 $\alpha < \beta$？

只在開平方根的時候：$\sqrt{(\beta-\alpha)^2}$ 等於 $\beta-\alpha$，正正是因為 $\beta-\alpha>0$。一般而言它等於 $|\beta-\alpha|$，而這一點正是下一部分的關鍵。

### 第一個積分，$\alpha > \beta$ 的情況

現在積分由較大的數 $\alpha$ 往下積到較小的數 $\beta$。在兩者之間的區間上，$x-\alpha<0$ 且 $\beta-x<0$，所以乘積仍然是正數，被積函數依然有意義。

代換可以逐字照搬，唯一不同的是現在 $\sqrt{(\beta-\alpha)^2} = \alpha-\beta$，所以

$$\frac{2(\beta-\alpha)\sin\theta\cos\theta}{(\alpha-\beta)\sin\theta\cos\theta} = -2,$$

而積分為

$$\int_0^{\pi/2} (-2)\,\mathrm{d}\theta = -\pi.$$

還有一個更快的看法。把 $\alpha$ 與 $\beta$ 互換時，式子 $(x-\alpha)(\beta-x)$ 保持不變，所以被積函數與第一種情況是同一個函數，只是兩個字母的角色對調了。把上下限倒轉會改變正負號，所以

$$\int_\alpha^\beta = -\int_\beta^\alpha = -\pi,$$

這裏用了第一個結果，以 $\beta$ 作為較小的端點。

### 第二個積分，$0 < \alpha < \beta$ 的情況

用完全相同的代換，同樣地抵消後，只剩下因子 $1/x$：

$$\int_\alpha^\beta \frac{\mathrm{d}x}{x\sqrt{(x-\alpha)(\beta-x)}} = \int_0^{\pi/2} \frac{2\,\mathrm{d}\theta}{\alpha\cos^2\theta+\beta\sin^2\theta}.$$

由於 $\alpha,\beta>0$，分母在整個 $[0,\frac{\pi}{2}]$ 上都是正數，所以這是一個正常積分。要計算它，把分子分母同除以 $\cos^2\theta$（對 $0\le\theta<\frac{\pi}{2}$ 成立）：

$$\int_0^{\pi/2} \frac{2\sec^2\theta\,\mathrm{d}\theta}{\alpha+\beta\tan^2\theta}.$$

再設 $t = \tan\theta$，則 $\mathrm{d}t = \sec^2\theta\,\mathrm{d}\theta$，而當 $\theta$ 由 0 增至 $\frac{\pi}{2}$ 時，$t$ 由 0 走到 $\infty$：

$$\int_0^\infty \frac{2\,\mathrm{d}t}{\alpha+\beta t^2} = \frac{2}{\beta}\int_0^\infty \frac{\mathrm{d}t}{t^2+\alpha/\beta} = \frac{2}{\beta}\sqrt{\frac{\beta}{\alpha}}\left[\arctan\left(t\sqrt{\frac{\beta}{\alpha}}\right)\right]_0^\infty$$
$$= \frac{2}{\sqrt{\alpha\beta}}\cdot\frac{\pi}{2} = \frac{\pi}{\sqrt{\alpha\beta}},$$

即為所求。我們兩次用到 $\alpha > 0$：一是確保 $1/x$ 在 $[\alpha,\beta]$ 內沒有奇點（singularity），二是確保 $\sqrt{\beta/\alpha}$ 是實數。

**想一想** 上限變成了 $\infty$。這會造成問題嗎？

不會。關於 $\theta$ 的積分完全是普通的積分；無窮上限之所以出現，只是因為當 $\theta \to \frac{\pi}{2}$ 時 $\tan\theta \to \infty$。關於 $t$ 的積分是收斂的，因為 arctan 趨於 $\frac{\pi}{2}$；而它等於關於 $\theta$ 的積分，因為兩者都是在 $[0,\Theta]$ 上的積分當 $\Theta \to \frac{\pi}{2}$ 時的極限。

## 驗算

- 縮放：把 $x$、$\alpha$、$\beta$ 換成 $kx$、$k\alpha$、$k\beta$，第一個積分不變，第二個積分則除以 $k$。兩個答案 $\pi$ 和 $\pi/\sqrt{\alpha\beta}$ 都符合這個性質。
- 令 $\alpha \to \beta$。這時在整個（極短的）區間上 $1/x \approx 1/\beta$，所以第二個積分應約為 $\pi/\beta$；而 $\pi/\sqrt{\alpha\beta} \to \pi/\beta$，果然如此。
- 數值上，取 $\alpha = 2$、$\beta = 7$，數值積分（quadrature）給出第二個積分為 0.83963，與 $\pi/\sqrt{14} = 0.83963$ 吻合。隨機選取 $\alpha < \beta$，第一個積分全部得出 $\pi$；把端點互換則得出 $-\pi$。

## 常見錯誤

- 在 $\alpha > \beta$ 時仍然答 $\pi$。被積函數是正的，但上下限倒過來了，所以答案是 $-\pi$。
- 寫下 $\sqrt{(\beta-\alpha)^2\sin^2\theta\cos^2\theta} = (\beta-\alpha)\sin\theta\cos\theta$，卻沒有檢查三個因子的正負號。

## STEP 之外

這個代換有一個漂亮的幾何意義。由於 $x = \frac{\alpha+\beta}{2} - \frac{\beta-\alpha}{2}\cos 2\theta$，它把以 $[\alpha,\beta]$ 為直徑的半圓參數化（parametrises），而 $\sqrt{(x-\alpha)(\beta-x)}$ 正是該半圓在 $x$ 處的高度。第二個積分讓你初嘗以圍道積分（contour integration）計算積分的滋味：在複分析（complex analysis）中，兩個結果都可以用留數定理（residue theorem）在短短幾行內得出。Needham 的

*Visual Complex Analysis*（《直觀複分析》）對此有精彩的闡述，而反常積分的嚴謹處理則見於 Abbott 的 *Understanding Analysis*（《理解分析》）。

詳細解答

# STEP I 1987，第 6 題

純數學 • 面積與反函數

**STEP I 1987，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/87-s1-q6

## 關鍵思路

F$(s)$ 是曲線下方、介乎曲線與 $x$ 軸之間的面積；G$(t)$ 則是曲線左方、介乎曲線與 $y$ 軸之間的面積。把這兩個區域畫在同一幅圖上，它們合起來覆蓋以 $(0,0)$ 和 $(s,t)$ 為頂點的長方形，而且除非頂點 $(s,t)$ 剛好落在曲線上，否則還會有多出來的部分。在第二部分，我們選取 $s$ 使這種情況出現，於是較難的積分 $\int \sin^{-1} y \,\mathrm{d}y$ 便可以從容易的積分 $\int \sin x \,\mathrm{d}x$ 直接得出。

## 解答

**不等式** $(*)$

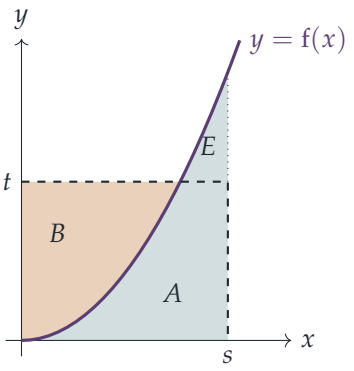
首先看看 f 的一個性質。它是連續（continuous）的，而且每個值最多只取一次，所以它是嚴格單調（strictly monotonic）的（根據介值定理（intermediate value theorem），一個先上升後下降的連續函數必定會把某個值取兩次）。由於 $\mathrm{f}(0) = 0$ 且 $\mathrm{f}(x) \geq 0$，它必定是嚴格遞增（strictly increasing）的。因此 g 亦是遞增的，並且對於 $0 \leq x \leq a$、$0 \leq y \leq \mathrm{f}(a)$，

$$y \leq \mathrm{f}(x) \iff \mathrm{g}(y) \leq x.$$

定義兩個區域：

$$A = \{(x,y) : 0 \leq x \leq s,\ 0 \leq y \leq \mathrm{f}(x)\}, \qquad \text{面積F}(s),$$
$$B = \{(x,y) : 0 \leq y \leq t,\ 0 \leq x \leq \mathrm{g}(y)\}, \qquad \text{面積G}(t),$$

並設 $R$ 為長方形 $0 \leq x \leq s$、$0 \leq y \leq t$，其面積為 $st$。$B$ 的面積是 G$(t)$，因為我們是把長度為 $\mathrm{g}(y)$ 的水平窄條積分起來（「對 $y$ 求面積」）。




$R$ 中每一點都位於 $A$ 或 $B$ 之內：若 $(x,y)\in R$，則要麼 $y\le \mathrm{f}(x)$，該點在 $A$ 內；要麼 $y>\mathrm{f}(x)$，即 $x<\mathrm{g}(y)$，該點在 $B$ 內。兩者的重疊部分 $A\cap B$ 只是曲線本身的一段弧，面積為零。因此

$$\mathrm{F}(s)+\mathrm{G}(t)=\text{面積}(A)+\text{面積}(B)\ge\text{面積}(A\cup B)\ge\text{面積}(R)=st,$$

這就是 $(*)$。

**想一想** 究竟多算了哪些不在長方形內的部分？

多出來的是 $A\cup B$ 中伸出 $R$ 之外的部分 $E$。若 $t<\mathrm{f}(s)$（如圖所示），$B$ 留在 $R$ 之內，但 $A$ 在區域 $\mathrm{g}(t)<x\le s$、$t<y\le\mathrm{f}(x)$ 中伸出了 $y=t$ 之上。由於對每個 $x>\mathrm{g}(t)$ 都有 $\mathrm{f}(x)>t$，而且 f 連續，這個區域的面積為正，所以不等式嚴格成立。若 $t>\mathrm{f}(s)$，情況相同，只是角色互換：$B$ 伸出了 $x=s$ 的右方。若 $t=\mathrm{f}(s)$，頂點 $(s,t)$ 位於曲線上，$A\cup B$ 恰好就是 $R$，沒有多出來的部分。

所以 $(*)$ **中等號成立當且僅當** $t=\mathrm{f}(s)$，即 $(s,t)$ 位於曲線上的時候。

### $\sin^{-1}y$ 的積分

在 $[0,\frac{1}{2}\pi]$ 上取 $\mathrm{f}(x)=\sin x$。它連續、遞增、通過原點，並且留在第一象限，所以上面所有結果都適用，其中 $a=\frac{1}{2}\pi$、$\mathrm{f}(a)=1$ 及 $\mathrm{g}(y)=\sin^{-1}y$。

給定 $t\in[0,1]$，選取 $s=\sin^{-1}t$，它位於 $[0,\frac{1}{2}\pi]$ 之內。於是 $t=\sin s=\mathrm{f}(s)$，所以 $(*)$ 中等號成立：

$$\int_0^t \sin^{-1}y\,\mathrm{d}y=\mathrm{G}(t)=st-\mathrm{F}(s).$$

接着計算容易的積分：

$$\mathrm{F}(s)=\int_0^s \sin x\,\mathrm{d}x=1-\cos s.$$

代入 $s=\sin^{-1}t$，

$$\int_0^t \sin^{-1}y\,\mathrm{d}y=t\sin^{-1}t-\bigl(1-\cos(\sin^{-1}t)\bigr),$$

正是所求。

### 以微分驗證

由於 $\sin^{-1} t \in [0, \frac{1}{2}\pi]$，它的餘弦非負，所以 $\cos(\sin^{-1} t) = \sqrt{1-t^2}$。因此右方為 $t\sin^{-1} t - 1 + \sqrt{1-t^2}$。對於 $0 \le t < 1$，利用 $\dfrac{\mathrm{d}}{\mathrm{d}t}\sin^{-1} t = \dfrac{1}{\sqrt{1-t^2}}$ 及乘積法則（product rule），

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\left(t\sin^{-1} t - 1 + \sqrt{1-t^2}\right) &= \sin^{-1} t + \frac{t}{\sqrt{1-t^2}} - \frac{t}{\sqrt{1-t^2}} \\ &= \sin^{-1} t,\end{aligned}$$

根據微積分基本定理（fundamental theorem of calculus），這正是左方的導數。兩方在 $t = 0$ 時都等於 0，所以它們在 $[0,1)$ 上相等，再由連續性，在 $t = 1$ 時亦相等。

## 驗算

- 當 $t = 1$：公式給出 $\frac{1}{2}\pi - 1 \approx 0.5708$。從幾何角度看，$y = \sin x$ 左方直至高度 1 的面積，等於長方形 $\frac{1}{2}\pi \times 1$ 減去拱形下方的面積 1。對 $\sin^{-1} y$ 在 $[0,1]$ 上作數值積分，得 0.57080。
- 當 $t = \frac{1}{2}$：公式給出 $\frac{\pi}{12} - 1 + \frac{\sqrt{3}}{2} \approx 0.12782$，與辛普森法則（Simpson's rule）的結果吻合。
- 對於 $(*)$，試取 $\mathrm{f}(x) = x^2$、$s = 1$、$t = 4$：$\mathrm{F} + \mathrm{G} = \frac{1}{3} + \frac{16}{3} = \frac{17}{3} > 4$；而取 $t = 1 = \mathrm{f}(1)$ 時，則得 $\frac{1}{3} + \frac{2}{3} = 1 = st$。

## 常見錯誤

- 寫成 $\cos(\sin^{-1} t) = \pm\sqrt{1-t^2}$ 而沒有選定符號：$\sin^{-1}$ 的值域決定了它取正值。
- 只說「這些區域覆蓋了長方形」，卻沒有解釋為何當 $t \ne \mathrm{f}(s)$ 時不等式嚴格成立；等號的問題正需要這一點。
- 忘記 f 必須是遞增的：整幅圖都依賴這一點。

## STEP 之外

不等式 $(*)$ 就是楊氏不等式（Young's inequality）。取 $\mathrm{f}(x) = x^{p-1}$，它便成為 $st \le \frac{s^p}{p} + \frac{t^q}{q}$，其中 $\frac{1}{p} + \frac{1}{q} = 1$；這是證明赫爾德不等式（Hölder's inequality）的關鍵步驟，

而赫爾德不等式是分析學中最常用的工具之一。F 與 G 的這種配對，讓我們初窺勒讓德變換（Legendre transform），它把拉格朗日力學（Lagrangian mechanics）與哈密頓力學（Hamiltonian mechanics）聯繫起來。如想了解這一切背後的分析學，可以閱讀 Bartle 與 Sherbert 的 *Introduction to Real Analysis*（《實分析導論》）。

詳細解答

# STEP I 1987，第 7 題

純數學 • 三角級數

> **STEP I 1987，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/87-s1-q7

## 關鍵思路

兩個級數（series）都是正弦之和，而且角度以相等的步長遞增。把 $\sin\big((2k+1)\theta\big)$ 這類項乘以 $2\sin\theta$ 或 $2\cos\theta$，再利用積化和差公式（product-to-sum formula），便可把它化為相鄰兩項的*差*或*和*，於是整個級數便會裂項相消（telescopes）。唯一要決定的是用哪個因子：$2\sin\theta$ 適用於普通的和，$2\cos\theta$ 則適用於交錯（alternating）的和。

## 解答

### 準備工作

設 $\theta = \dfrac{2\pi}{23}$。各角為 $2\pi/23,\ 6\pi/23,\ 10\pi/23, \ldots, 42\pi/23$，每次增加 $4\pi/23 = 2\theta$，所以它們是

$$(2k+1)\theta \qquad \text{其中} k = 0, 1, \ldots, 10,$$

因為 $21\theta = 42\pi/23$。這樣共有 11 項，而在交錯級數中，第 $k$ 項的符號是 $(-1)^k$（最後一項 $k = 10$ 的符號是 $+$，正合題意）。因此我們要求

$$S_1 = \sum_{k=0}^{10} \sin\big((2k+1)\theta\big), \qquad S_2 = \sum_{k=0}^{10} (-1)^k \sin\big((2k+1)\theta\big).$$

我們會用到 $22\theta = \dfrac{44\pi}{23} = 2\pi - \dfrac{2\pi}{23}$，因此

$$\cos 22\theta = \cos\theta, \qquad \sin 22\theta = -\sin\theta.$$

### 第一個級數

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

$$2\sin\theta\,\sin\big((2k+1)\theta\big)=\cos(2k\theta)-\cos\big((2k+2)\theta\big).$$

對 $k=0,\ldots,10$ 求和，每個 $1\le k\le 10$ 的 $\cos(2k\theta)$ 都以正負號各出現一次而互相抵消，剩下

$$2\sin\theta\,S_1=\cos 0-\cos 22\theta=1-\cos\theta.$$

由於 $0<\theta<\pi$，$\sin\theta\neq 0$，所以可以相除：

$$S_1=\frac{1-\cos\theta}{2\sin\theta}=\frac{2\sin^2(\theta/2)}{4\sin(\theta/2)\cos(\theta/2)}=\tfrac{1}{2}\tan\frac{\theta}{2},$$

這裏用了 $1-\cos\theta=2\sin^2(\theta/2)$ 和 $\sin\theta=2\sin(\theta/2)\cos(\theta/2)$，並約去 $\sin(\theta/2)\neq 0$。由於 $\theta/2=\pi/23$，

$$\boxed{\sin\tfrac{2\pi}{23}+\sin\tfrac{6\pi}{23}+\cdots+\sin\tfrac{42\pi}{23}=\tfrac{1}{2}\tan\tfrac{\pi}{23}.}$$

**想一想** 為甚麼處理交錯級數時要乘以 $2\cos\theta$，而不是再乘以 $2\sin\theta$？

若乘以 $2\sin\theta$，得到的是差 $\cos(2k\theta)-\cos((2k+2)\theta)$，而交錯的符號會令相鄰的項相加而非抵消。若乘以 $2\cos\theta$，得到的是和 $\sin((2k+2)\theta)+\sin(2k\theta)$，這時正是交錯的符號帶來抵消。

### 第二個級數

由 $2\cos A\sin B=\sin(B+A)+\sin(B-A)$，

$$2\cos\theta\,\sin\big((2k+1)\theta\big)=\sin\big((2k+2)\theta\big)+\sin(2k\theta).$$

乘以 $(-1)^k$ 再求和。對於 $1\le k\le 10$，$\sin(2k\theta)$ 一次來自第 $k-1$ 項，符號為 $(-1)^{k-1}$，另一次來自第 $k$ 項，符號為 $(-1)^k$，所以互相抵消。剩下的是來自 $k=0$ 的 $\sin 0$ 和來自 $k=10$ 的 $(-1)^{10}\sin 22\theta$：

$$2\cos\theta\,S_2=\sin 22\theta=-\sin\theta.$$

由於 $0<\theta<\pi/2$，$\cos\theta\neq 0$，所以

$$S_2=-\frac{\sin\theta}{2\cos\theta}=-\tfrac{1}{2}\tan\theta,$$

即

$$\boxed{\sin\tfrac{2\pi}{23}-\sin\tfrac{6\pi}{23}+\cdots-\sin\tfrac{38\pi}{23}+\sin\tfrac{42\pi}{23}=-\tfrac{1}{2}\tan\tfrac{2\pi}{23}.}$$

如果你喜歡看起來是正的形式，由於 $\tan(\pi-x)=-\tan x$，這也等於 $\frac{1}{2}\tan\frac{21\pi}{23}$。

## 驗算

- 以數值計算，$\frac{1}{2}\tan(\pi/23) \approx 0.068723$，而 $-\frac{1}{2}\tan(2\pi/23) \approx -0.140093$；直接把十一項相加，得到的值在 15 位有效數字（significant figures）內完全相同。
- 第二個答案是負數，這是合理的：部分和（partial sums）在約 $\pm 0.5$ 之間擺動，而最後加上的一項 $\sin(42\pi/23) \approx -0.27$ 本身就是負數。
- 同樣的方法可得出一般公式 $\sum_{k=0}^{n-1} \sin((2k+1)\theta) = \dfrac{\sin^2 n\theta}{\sin\theta}$ 和 $\sum_{k=0}^{n-1}(-1)^k \sin((2k+1)\theta) = (-1)^{n-1}\dfrac{\sin 2n\theta}{2\cos\theta}$；試取 $n = 1$，兩者都化為 $\sin\theta$。

## 常見錯誤

- 數錯項數：角度每次增加 $4\pi/23$，而不是 $2\pi/23$，所以共有 11 項，而不是 21 項。
- 以為兩個答案都是正數。大於 $\pi$ 的角的正弦是負的，而負號正是來自 $\sin(44\pi/23) = -\sin(2\pi/23)$。

## STEP 之外

這些和是幾何級數（geometric series）的虛部（imaginary part）：$S_1 = \operatorname{Im}\sum e^{i(2k+1)\theta}$，而 $S_2$ 也一樣，只是公比為 $-e^{2i\theta}$，這為兩個答案提供了第二條途徑。這類和，例如狄利克雷核（Dirichlet kernel），是傅里葉分析（Fourier analysis）的起點；在傅里葉分析中，函數由正弦和餘弦組合而成。Needham 的 *Visual Complex Analysis*（《直觀複分析》）借助旋轉向量的圖像，說明這類和為何有這樣的表現。

詳細解答

# STEP I 1987，第 8 題

純數學 • 代換積分法

**STEP I 1987，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/87-s1-q8

## 關鍵思路

代換（substitution）要成為合法的變數轉換，前提是當舊變數走過某個區間時，新變數也要平滑地、單向地走過一個區間。這裏 $t = 1/x$ 在 $x = 0$ 處趨於無窮，而 $x = 0$ 正好位於 $[-1, 1]$ 的中間，所以「新的上下限」$-1$ 和 $1$ 描述的是一組完全不同的 $t$ 值。只要看一看兩個被積函數（integrand）的正負號，就能確認出了問題。

## 解答

### 代換為何失效

形式上，設 $x = 1/t$，則 $\mathrm{d}x = -\frac{1}{t^2}\,\mathrm{d}t$，而且

$$\frac{1}{(1+x^2)^2} = \frac{1}{\left(1+\frac{1}{t^2}\right)^2} = \frac{t^4}{(1+t^2)^2}, \qquad \text{所以} \qquad \frac{\mathrm{d}x}{(1+x^2)^2} = \frac{-t^2}{(1+t^2)^2}\,\mathrm{d}t,$$

而 $x = \pm 1$ 對應 $t = \pm 1$。這就是「證明」兩個積分相等的做法。但這不可能是對的：第一個被積函數處處為正，所以第一個積分是正數；而第二個被積函數處處 $\leq 0$，所以第二個積分是負數。

**想一想** 當 $x$ 由 $-1$ 走到 $1$ 時，$t = 1/x$ 實際上取哪些值？

當 $x$ 由 $-1$ 增加到 $0^-$ 時，$t$ 由 $-1$ 減少到 $-\infty$；在 $x = 0$ 處根本沒有對應的 $t$；而當 $x$ 由 $0^+$ 走到 $1$ 時，$t$ 由 $+\infty$ 下降到 $1$。所以 $x$ 的區間 $[-1, 1]$ 對應的是 $t \in (-\infty, -1] \cup [1, \infty)$，即 $[-1, 1]$ 的外面，而不是 $[-1, 1]$ 本身。代換法則要求 $x = \phi(t)$，其中 $\phi$ 在新上下限之間的整個 $t$ 區間上可導，但 $\phi(t) = 1/t$ 在 $t = 0 \in [-1, 1]$ 處連定義都沒有。$(-1, 1)$ 內的 $t$ 值對應 $|x| > 1$，那些值根本在原來的範圍以外。

正確地使用這個代換，即在兩半上分別進行，例如可得

$$\int_0^1 \frac{\mathrm{d}x}{(1+x^2)^2} = \int_\infty^1 \frac{-t^2}{(1+t^2)^2}\,\mathrm{d}t = \int_1^\infty \frac{t^2}{(1+t^2)^2}\,\mathrm{d}t,$$

這是一個正確的命題（下面會加以驗證），但它對於 $t$ 在 $[-1,1]$ 上的積分毫無說明。

### 第一個積分

設 $x = \tan\theta$，則 $\mathrm{d}x = \sec^2\theta\,\mathrm{d}\theta$ 及 $1+x^2 = \sec^2\theta$；這個代換沒有問題，因為 tan 在 $[-\frac{\pi}{4}, \frac{\pi}{4}]$ 上平滑而且遞增。於是

$$\begin{aligned}\int \frac{\mathrm{d}x}{(1+x^2)^2} &= \int \frac{\sec^2\theta}{\sec^4\theta}\,\mathrm{d}\theta = \int \cos^2\theta\,\mathrm{d}\theta = \int \tfrac{1}{2}(1+\cos 2\theta)\,\mathrm{d}\theta \\ &= \tfrac{1}{2}\theta + \tfrac{1}{4}\sin 2\theta = \tfrac{1}{2}\theta + \tfrac{1}{2}\cdot\frac{\tan\theta}{1+\tan^2\theta} = \tfrac{1}{2}\arctan x + \frac{x}{2(1+x^2)},\end{aligned}$$

這裏用了 $\sin 2\theta = 2\sin\theta\cos\theta = 2\tan\theta\cos^2\theta$。因此

$$\int_{-1}^1 \frac{\mathrm{d}x}{(1+x^2)^2} = \left[\tfrac{1}{2}\arctan x + \frac{x}{2(1+x^2)}\right]_{-1}^1 = 2\left(\frac{\pi}{8}+\frac{1}{4}\right) = \boxed{\frac{\pi}{4}+\frac{1}{2}}.$$

### 第二個積分

把分子寫成 $-t^2 = 1-(1+t^2)$：

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

由上一部分，第一項在 $[-1,1]$ 上的積分是 $\frac{\pi}{4}+\frac{1}{2}$，而 $\int_{-1}^1 \frac{\mathrm{d}t}{1+t^2} = \left[\arctan t\right]_{-1}^1 = \frac{\pi}{2}$。所以

$$\int_{-1}^1 \frac{-t^2}{(1+t^2)^2}\,\mathrm{d}t = \frac{\pi}{4}+\frac{1}{2}-\frac{\pi}{2} = \boxed{\frac{1}{2}-\frac{\pi}{4}},$$

這是負數（$\approx -0.285$），正如被積函數的正負號所要求的；而第一個積分約為 1.285。

## 驗算

- 用辛普森法則（Simpson's rule）計算兩個積分，得 1.2853981… 和 −0.2853981…，與 $\frac{\pi}{4}+\frac{1}{2}$ 及 $\frac{1}{2}-\frac{\pi}{4}$ 吻合。
- 兩個答案之和是 1。這是有原因的：$\frac{1}{(1+x^2)^2} - \frac{x^2}{(1+x^2)^2} = \frac{1-x^2}{(1+x^2)^2}$，它正是 $\frac{x}{1+x^2}$ 的導數（derivative），而 $\left[\frac{x}{1+x^2}\right]_{-1}^1 = 1$。
- 正確的半區間命題也驗證無誤：$\int_1^\infty \frac{t^2}{(1+t^2)^2}\,\mathrm{d}t = \frac{\pi}{4} - \left(\frac{\pi}{4}-\frac{\pi}{8}-\frac{1}{4}\right) = \frac{\pi}{8}+\frac{1}{4} = \int_0^1 \frac{\mathrm{d}x}{(1+x^2)^2}$。

## 常見錯誤

- 只說「$1/t$ 在 $t=0$ 處無定義」，卻沒有解釋 $x$ 的範圍 $[-1,1]$ 對應的是 $|t|\geq 1$：真正的重點是兩個積分涵蓋的是不同的取值集合。
- 把 $\mathrm{d}x=-\frac{1}{t^2}\,\mathrm{d}t$ 的正負號弄錯，以致掩蓋了那個露出破綻的正負號矛盾。
- 積分 $\cos^2\theta$ 時漏掉 $\frac{1}{4}\sin 2\theta$ 一項。

## STEP 之外

代換法則 $\int_a^b f(\phi(t))\phi'(t)\,\mathrm{d}t = \int_{\phi(a)}^{\phi(b)} f(x)\,\mathrm{d}x$ 在大學第一門數學分析（analysis）課程中，會由鏈式法則（chain rule）和微積分基本定理（Fundamental Theorem of Calculus）證明，而證明過程清楚顯示所需的條件：$\phi$ 必須在整個 $[a,b]$ 上連續可微（continuously differentiable）。這裏它在一點上失效，而單單這一點就足以令結果崩潰。Abbott 的 *Understanding Analysis*（《理解分析》）細緻地建立了黎曼積分（Riemann integral）及這些定理。

詳細解答

# STEP I 1987，第 9 題

純數學 • 重心坐標

> **STEP I 1987，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/87-s1-q9

## 關鍵思路

從 $A$ 向 $B$ 移動，就是加上向量（vector）$\overrightarrow{AB} = \mathbf{b} - \mathbf{a}$ 的某個分數倍。做一次，得到的是線段（segment）上的點；做兩次，先沿 $AB$ 走，再從該點向 $C$ 走，就能掃遍整個三角形。至於作圖部分，竅門是把每種情況改寫成我們已經理解的形式：要麼是三角形情況的翻轉，要麼是點 $A$ 加上兩條邊向量的倍數。

## 解答

### $AB$ 上的點

線段 $AB$ 上的點 $P$，可以從 $A$ 出發，沿 $\overrightarrow{AB}$ 走全程的 $t$ 倍到達，其中 $0 \le t \le 1$（$t = 0$ 給出 $A$，$t = 1$ 給出 $B$）。因此

$$\mathbf{p} = \mathbf{a} + t(\mathbf{b} - \mathbf{a}) = (1 - t)\mathbf{a} + t\,\mathbf{b}.$$

設 $\rho = 1 - t$ 及 $\sigma = t$。由於 $0 \le t \le 1$，所以 $\rho \ge 0$ 及 $\sigma \ge 0$，而且 $\rho + \sigma = 1$。這正是所求的形式。反過來，取 $t = \sigma$，每一對這樣的 $\rho, \sigma$ 都給出線段上的一點。

### $PC$ 上的點，以及三角形內的點

用完全相同的方法，線段 $PC$ 上的點 $Q$ 的位置向量（position vector）為

$$\mathbf{q} = (1 - s)\mathbf{p} + s\,\mathbf{c}, \qquad 0 \le s \le 1.$$

代入 $\mathbf{p} = \rho\mathbf{a} + \sigma\mathbf{b}$，得

$$\mathbf{q} = (1 - s)\rho\,\mathbf{a} + (1 - s)\sigma\,\mathbf{b} + s\,\mathbf{c}.$$

設 $\lambda = (1-s)\rho$、$\mu = (1-s)\sigma$ 及 $\nu = s$。每一個都是非負（non-negative）數的乘積，所以 $\lambda, \mu, \nu \geq 0$。此外

$$\lambda + \mu + \nu = (1-s)(\rho+\sigma) + s = (1-s) + s = 1.$$

**想一想** 為甚麼三角形內的每一點都位於某條線段 $PC$ 上？

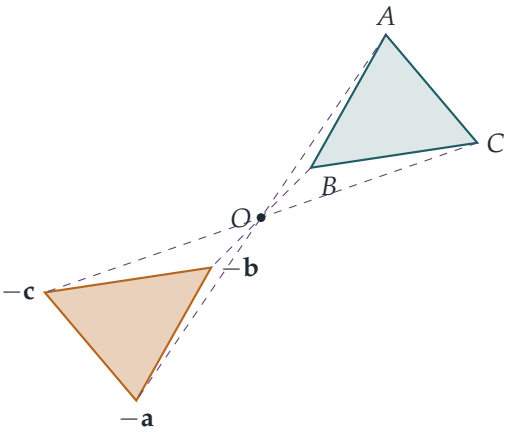
取三角形內（或其邊界上）的一點 $X$。若 $X = C$，則無需證明：取 $s = 1$ 即可。否則，從 $C$ 經過 $X$ 畫一條直線，並延伸越過 $X$。由於 $X$ 位於角 $ACB$ 之內，這條直線保持在 $CA$ 與 $CB$ 之間，因此必定與對邊 $AB$ 相交，設交點為 $P$。由於 $X$ 在三角形內，它位於 $C$ 與 $P$ 之間，所以 $X$ 在線段 $PC$ 上。根據上面的推導，$\mathbf{x} = \lambda\mathbf{a} + \mu\mathbf{b} + \nu\mathbf{c}$，其中 $\lambda, \mu, \nu \geq 0$ 且 $\lambda + \mu + \nu = 1$，命題得證。

### 第 (i) 部分

設 $\lambda = -\lambda'$、$\mu = -\mu'$ 及 $\nu = -\nu'$。條件變為 $\lambda', \mu', \nu' \geq 0$ 及 $\lambda' + \mu' + \nu' = 1$，而該點為

$$\lambda\mathbf{a} + \mu\mathbf{b} + \nu\mathbf{c} = -\left(\lambda'\mathbf{a} + \mu'\mathbf{b} + \nu'\mathbf{c}\right).$$

根據剛才所證，括號內的式子遍歷三角形 $ABC$ 的每一點（再加上第一部分的逆命題，每一個這樣的組合都在三角形內）。因此所求區域是所有點 $-\mathbf{x}$ 的集合，其中 $X$ 在三角形 $ABC$ 內。換言之，它是以 $-\mathbf{a}$、$-\mathbf{b}$、$-\mathbf{c}$ 為頂點的三角形連同其內部：即三角形 $ABC$ 繞原點 $O$ 旋轉半周（half-turn）後的像。




琥珀色陰影三角形就是 (i) 的答案。它的每個頂點都是 $ABC$ 的一個頂點經 $O$ 的點對稱像，如虛線所示。

**第 (ii) 部分**

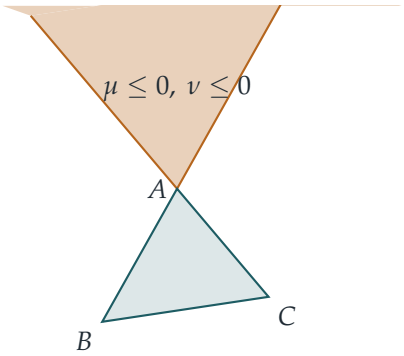
現在 $\lambda = 1 - \mu - \nu$，所以

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

我們從 $A$ 出發，加上 $\overrightarrow{AB}$ 和 $\overrightarrow{AC}$ 的倍數。設 $\mu = -m$ 及 $\nu = -n$，其中 $m, n \geq 0$，則該點為

$$\mathbf{a} + m(\mathbf{a} - \mathbf{b}) + n(\mathbf{a} - \mathbf{c}), \qquad m \geq 0,\ n \geq 0.$$

向量 $\mathbf{a} - \mathbf{b}$ 指向 $BA$ 越過 $A$ 的延長方向，而 $\mathbf{a} - \mathbf{c}$ 指向 $CA$ 越過 $A$ 的延長方向。兩個不平行向量的非負組合恰好填滿它們之間的角（由於 $\mathbf{b} - \mathbf{a}$ 與 $\mathbf{c} - \mathbf{a}$ 不平行，平面上每一點都可唯一地寫成 $\mathbf{a} + m(\mathbf{a} - \mathbf{b}) + n(\mathbf{a} - \mathbf{c})$，即 $m, n$ 恰有一對）。因此所求區域是以 $A$ 為頂點、以 $BA$ 和 $CA$ 越過 $A$ 的延長線為界的無限楔形區域（wedge），邊界射線包括在內。它是角 $BAC$ 的對頂角（vertically opposite angle）。注意題目沒有限制 $\lambda$，但在這裏它自動至少為 1。




琥珀色楔形區域（無限延伸）就是 (ii) 的答案。它的邊是邊 $BA$ 和 $CA$ 經 $A$ 延長所得。圖中青色三角形 $ABC$ 僅供參照。

## 驗算

- 頂點：在 (ii) 中，$(\lambda, \mu, \nu) = (1, 0, 0)$ 給出 $A$ 本身，即楔形的尖端；$(2, -1, 0)$ 給出 $2\mathbf{a} - \mathbf{b}$，即 $B$ 關於 $A$ 的對稱點，它位於其中一條邊上。在 (i) 中，$(-1, 0, 0)$ 給出 $-\mathbf{a}$，即對稱後三角形的一個頂點。
- 重心：在 (i) 中，$\lambda = \mu = \nu = -\frac{1}{3}$ 給出 $ABC$ 重心（centroid）的負值，這正是對稱後三角形的重心，與預期相符。
- 數值方面，我們用電腦隨機取了 20,000 個三角形，以及滿足各組條件的隨機係數。結果確認：(i) 的每一點都位於三角形 $(-\mathbf{a}, -\mathbf{b}, -\mathbf{c})$ 內；(ii) 的每一點從 $A$ 出發沿

$\mathbf{a}-\mathbf{b}$ 和 $\mathbf{a}-\mathbf{c}$ 的坐標都是非負的；而 $ABC$ 的每一點都可分解為某條線段 $PC$ 上的點。

## 常見錯誤

- 在 (i) 中，在 $ABC$ 附近畫出區域。係數之和是 $-1$ 而不是 1，所以該區域是 $ABC$ 繞原點旋轉半周的結果，並且取決於 $O$ 的位置。
- 在 (ii) 中，只畫出一個三角形。$\mu$ 和 $\nu$ 沒有下界，所以該區域是無界的。
- 把與 $BAC$ 相鄰的角塗上陰影（例如 $AB$ 與 $CA$ 延長線之間的角）。那需要 $\mu,\nu$ 其中之一為正數。

## STEP 之外

滿足 $\lambda+\mu+\nu=1$ 的數組 $(\lambda,\mu,\nu)$ 稱為一點相對於該三角形的重心坐標（barycentric coordinates）。當總和為 1 時，該點與原點的位置無關：把每個位置向量都平移 $\mathbf{d}$，$\lambda\mathbf{a}+\mu\mathbf{b}+\nu\mathbf{c}$ 也會平移 $(\lambda+\mu+\nu)\mathbf{d}=\mathbf{d}$。這就是為甚麼 (ii) 的答案是一個真正的幾何區域，而 (i) 則取決於 $O$。三角形本身是 $A$、$B$、$C$ 的凸包（convex hull），這個概念是線性規劃（linear programming）和凸優化（convex optimisation）的核心。重心坐標的正負號把平面分成七個區域，(ii) 是其中之一。若想從大學一年級的角度認識這些仿射（affine）概念，可參閱 Strang 的 *Introduction to Linear Algebra*（《線性代數導論》）。

詳細解答

# STEP I 1987，第 10 題

力學 • 彈性繩

**STEP I 1987，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/87-s1-q10

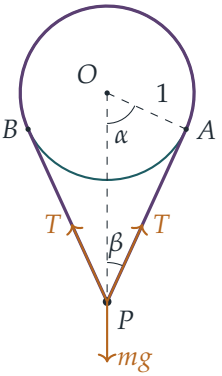
## 關鍵思路

質點把橡皮圈的底部從圓柱上拉下來，因此橡皮圈分成兩部分：一段仍然緊貼圓柱上方的圓弧，以及兩段直線部分；這兩段直線沿切線（*tangent*）離開圓柱，並在質點處相交。拉伸後的長度純粹是幾何問題：利用直角三角形，可以求出直線部分的長度，以及它們在軸上所對的角。由於圓柱是光滑（smooth）的，橡皮圈各處的張力（tension）都相同；在質點處分解力，就能把這張力與 $mg$ 聯繫起來，再用胡克定律（Hooke's law）便可求出 $\lambda$。

## 解答

### 模型

我們把橡皮圈視為一條自然長度（natural length）為 $2\pi$、遵從胡克定律的輕彈性繩，圓柱視為光滑，質點視為點質量。我們在一個豎直橫截面上分析：圓柱是一個單位圓，圓心 $O$ 在軸上；質點位於 $P$，在 $O$ 豎直下方距離 $x$ 處。橡皮圈要離開圓柱，必須有 $x > 1$。




由於圖形關於通過 $O$ 的豎直線對稱，橡皮圈在兩點 $A$ 和 $B$ 離開圓柱，兩邊各一點。一條拉緊的繩離開光滑曲線時，是沿切線離開的，因此 $PA$ 在 $A$ 點與圓相切，即 $OA \perp AP$。設 $\alpha = \angle AOP$，$\beta = \angle APO$。

**拉伸後的長度**

三角形 $OAP$ 在 $A$ 處是直角，$OA = 1$，斜邊 $OP = x$。因此

$$AP = \sqrt{x^2 - 1}, \qquad \cos\alpha = \frac{1}{x}, \qquad \sin\beta = \frac{1}{x}.$$

除了經底部從 $A$ 到 $B$ 的一段（在 $O$ 處所對的角為 $2\alpha$）之外，橡皮圈處處與圓柱接觸。所以接觸部分的圓弧所對的角是 $2\pi - 2\alpha$，在單位圓上其長度就是 $2\pi - 2\alpha$。加上兩段直線部分，拉伸後的長度為

$$\boxed{L = 2\sqrt{x^2 - 1} + 2\pi - 2\arccos\frac{1}{x}}$$

其中由於 $x > 1$，$\arccos(1/x)$ 介於 0 與 $\frac{\pi}{2}$ 之間。因此伸長量（extension）為

$$L - 2\pi = 2\sqrt{x^2 - 1} - 2\arccos\frac{1}{x}.$$

? **想一想** 橡皮圈繞着圓柱彎曲。為甚麼直線部分的張力與圓弧部分的張力相同？

對於與光滑表面接觸的一小段橡皮圈，圓柱施加的力只有垂直於表面的法向力，因此沒有任何力可以抵消沿橡皮圈方向的張力變化。所以張力 $T$ 在整條橡皮圈上都是常數，胡克定律可以應用於整條橡皮圈：

$$T = \frac{\lambda\,(L - 2\pi)}{2\pi}.$$

（如果圓柱是粗糙的，張力在圓弧上就可能有變化，這一步便不成立。）

**質點的平衡**

作用在質點上的力有重量 $mg$，以及沿 $PA$ 和 $PB$ 方向的張力 $T$，兩者與向上豎直方向的夾角均為 $\beta$。由對稱性，水平分量互相抵消。豎直方向分解，

$$2T\cos\beta = mg, \qquad \text{其中} \qquad \cos\beta = \frac{\sqrt{x^2-1}}{x},$$

所以

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

**$x = 2$ 的情況**

此時 $\cos\alpha = \frac{1}{2}$，故 $\alpha = \frac{\pi}{3}$，且 $\sqrt{x^2-1} = \sqrt{3}$。拉伸後的長度為

$$L = 2\sqrt{3} + 2\pi - \frac{2\pi}{3} = 2\sqrt{3} + \frac{4\pi}{3}, \qquad \text{所以} \qquad L - 2\pi = 2\sqrt{3} - \frac{2\pi}{3}.$$

張力為 $T = \dfrac{2mg}{2\sqrt{3}} = \dfrac{mg}{\sqrt{3}}$。由胡克定律得

$$\lambda = \frac{2\pi T}{L-2\pi} = \frac{2\pi mg/\sqrt{3}}{2\sqrt{3} - 2\pi/3} = \frac{2\pi mg}{6 - 2\pi/\sqrt{3}} = \frac{\pi mg}{3 - \pi/\sqrt{3}},$$

分子分母同乘以 $\sqrt{3}$，

$$\boxed{\lambda = \frac{\sqrt{3}\,\pi\, mg}{3\sqrt{3} - \pi} \approx 2.65\, mg.}$$

## 驗算

- 當 $x \to 1^+$ 時，$\arccos(1/x) \to 0$ 且 $\sqrt{x^2-1} \to 0$，所以 $L \to 2\pi$：橡皮圈剛好碰到圓柱底部而沒有被拉伸，這正合理。同時 $T \to \infty$，這與以下事實吻合：兩條接近水平的繩子只有在張力極大時才能承托重物。
- 當 $x = 2$ 時，$L = 2\sqrt{3} + 4\pi/3 \approx 7.653$，比 $2\pi \approx 6.283$ 稍大；而 $\lambda$ 與 $mg$ 一樣具有力的單位。
- 另外，我們取 $\lambda \approx 2.6484\, mg$，以數值方法求總勢能（potential energy）$-mgx + \lambda(L - 2\pi)^2/(4\pi)$ 的最小值；最小值出現在 $x = 2$，這同時證實了幾何部分和張力部分的結果。

## 常見錯誤

- 把直線部分當作從 $P$ 連到水平直徑的兩個端點。橡皮圈是沿切線離開的，所以它在軸的高度以下離開圓柱，位置與向下豎直方向成角 $\alpha = \arccos(1/x)$。

► 在胡克定律中只用一段直線部分的伸長量。張力是均勻的，所以應該用整條橡皮圈的伸長量 $L - 2\pi$，除以整條的自然長度 $2\pi$。

## STEP 之外

「在光滑表面上張力處處相同」這一說法，是絞盤方程（capstan equation）在無摩擦情況下的特例：若摩擦系數（coefficient of friction）為 $\mu$，繞過角度 $\theta$ 後，張力最多可以改變 $e^{\mu\theta}$ 倍。這就是為甚麼一條在柱上繞幾圈的繩子能夠繫住一艘船。上面的能量驗算，也讓你初步領略力學的變分觀點（variational view）：平衡位置就是勢能的駐點（stationary point）。若想看大學一年級程度、以這種思路處理力、摩擦和勢能的教材，可讀 Kleppner 和 Kolenkow 的 *An Introduction to Mechanics*（《力學導論》）。

詳細解答

# STEP I 1987，第 11 題

力學 • 圓周運動與能量

**STEP I 1987，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/87-s1-q11

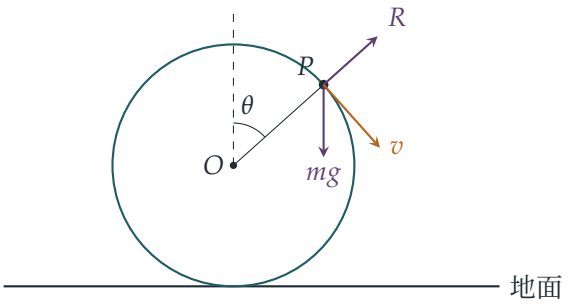
## 關鍵思路

質點留在球面上時作圓周運動（circular motion），因此由兩條方程支配：能量守恆（conservation of energy）給出其速率，而沿半徑方向應用牛頓（Newton）第二定律則給出法向反作用力（normal reaction）。當反作用力降至零的一刻，質點便離開球面。此後質點成為拋射體（projectile），所以其水平速度不再改變；把這一點與由頂部一直到地面的能量守恆結合，便可得出鉛直速度，而毋須求出軌跡。

## 解答

### 模型

把質點視為一個質點（point mass），置於一個光滑、固定的球面上，球心為 $O$，位於地面之上高度 $r$ 處（球體擱在地面上）。「輕輕移動」的意思是質點從頂部出發時速率可忽略不計。設 $P$ 為質點所在位置，$\theta$ 為 $OP$ 與向上鉛垂線之間的角，$v$ 為其速率。作用於質點的力只有重量 $mg$ 及沿 $OP$ 向外的法向反作用力 $R$。



### 何處離開球面

質點已下降高度 $r - r\cos\theta$，因此由能量守恆（反作用力與運動方向垂直，故不作功）得

$$\tfrac{1}{2}mv^2 = mgr(1-\cos\theta), \qquad \text{所以} \qquad v^2 = 2gr(1-\cos\theta).$$

質點與球面接觸時沿半徑為 $r$ 的圓運動，因此其指向 $O$ 的加速度為 $v^2/r$。沿指向 $O$ 的方向分解：

$$mg\cos\theta - R = \frac{mv^2}{r} = 2mg(1-\cos\theta), \qquad \text{所以} \qquad R = mg(3\cos\theta - 2).$$

在頂部時 $R = mg > 0$，而 $R$ 隨 $\theta$ 增大而減少。當 $R = 0$ 時質點離開球面，即當

$$\cos\theta = \tfrac{2}{3}, \qquad v^2 = \tfrac{2}{3}gr.$$

**想一想** 速度與水平線所成的角，是哪一個角？

速度沿圓的切線（tangent）方向，而切線與半徑 $OP$ 垂直。半徑與鉛垂線成角 $\theta$，所以切線與水平線所成的角同樣是 $\theta$（把兩條線都轉一個直角即可）。因此，質點離開球面的一刻，正以水平線之下

$$\cos^{-1}\tfrac{2}{3} \approx 48.2^\circ$$

的角度向下運動。

### 撞擊地面

離開球面的一刻，速度的水平分量為

$$v\cos\theta = \sqrt{\tfrac{2}{3}gr}\cdot\tfrac{2}{3}, \qquad \text{所以} \qquad (v\cos\theta)^2 = \tfrac{8}{27}gr.$$

此後唯一的作用力是重力，所以這個水平分量直至質點到達地面都保持不變。（質點不會再碰到球面：一旦 $R$ 會變成負數，球面向外彎離的速度便快於質點下墮的速度；而質點落地點與通過 $O$ 的鉛垂線的水平距離約為 $1.46r$，遠離球體。）

現在對整個運動應用能量守恆，由頂部（高度 $2r$）靜止開始，一直到地面。質點以速率 $V$ 撞擊地面，其中

$$\tfrac{1}{2}mV^2 = mg(2r), \qquad \text{所以} \qquad V^2 = 4gr.$$

設 $V_y$ 為撞擊時速度的鉛直分量，則 $V^2 = (v\cos\theta)^2 + V_y^2$，所以

$$V_y^2 = 4gr - \tfrac{8}{27}gr = \tfrac{100}{27}gr, \qquad V_y = \frac{10}{3}\sqrt{\frac{gr}{3}}.$$

因此，質點撞擊地面時動量（momentum）的鉛直分量的大小為

$$mV_y = \frac{10m}{3}\sqrt{\frac{gr}{3}} = \frac{10m\sqrt{3gr}}{9},$$

方向向下。

## 驗算

- 直接用拋射體運動計算：質點在高度 $r + r\cos\theta = \frac{5}{3}r$ 處離開，向下的鉛直速率為 $v\sin\theta$，其中 $\sin\theta = \frac{\sqrt{5}}{3}$，所以 $(v\sin\theta)^2 = \frac{10}{27}gr$。再下墮 $\frac{5}{3}r$ 會增加 $2g \cdot \frac{5}{3}r = \frac{90}{27}gr$，再次得到 $V_y^2 = \frac{100}{27}gr$。
- $V_y^2 = \frac{100}{27}gr \approx 3.70gr$ 比 $V^2 = 4gr$ 略小，這是必然的，因為有少量能量留在水平運動之中。
- 以數值方法逐步模擬運動（取 $g = 9.81$、$r = 1.3$、$m = 2$），得出在 48.19° 處離開球面，鉛直動量為 13.745，與 $\frac{10m\sqrt{3gr}}{9}$ 吻合；軌跡一直留在球體之外。

## 常見錯誤

- 由地面而非由球頂量度下降高度：在球面上，下降的高度是 $r(1-\cos\theta)$。
- 忘記離開球面後水平速度固定不變，誤把 $V_y^2 = 4gr$ 寫出來，彷彿全部動能都屬於鉛直方向。
- 誤以為速度與水平線成角 $90° - \theta$；那是半徑的角度，而非切線的角度。

## STEP 之外

「反作用力消失時便離開接觸面」這個條件，是約束力(constraint force)的第一個例子：約束力會自行調節，使質點留在某個曲面上，而且只能推、不能拉。在拉格朗日力學（Lagrangian mechanics）中，這類約束從一開始便納入方程，而反作用力則以拉格朗日乘數（Lagrange multiplier）的形式出現。Kleppner 與 Kolenkow 合著的 *An Introduction to Mechanics*（《力學導論》）是在更高層次上練習圓周運動和能量方法的絕佳教材。

詳細解答

# STEP I 1987，第 12 題

力學 • 拋體與恢復係數

**STEP I 1987，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/87-s1-q12

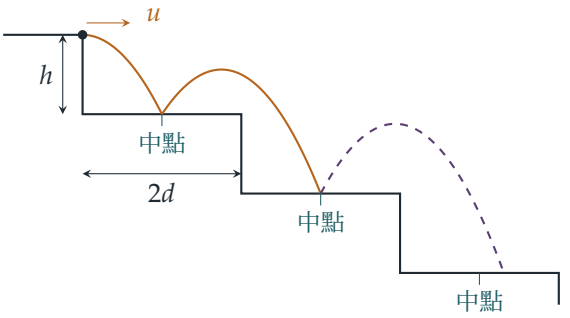
## 關鍵思路

在兩次反彈之間，粒子是一個拋射體（projectile）；而（台階光滑時）每次反彈只會把速度的鉛直分量反向並按比例縮小，所以水平速率始終不變。這意味著水平距離決定了每段飛行的時間，而飛行時間加上一級台階的落差又決定了鉛直速率。所有量都可以用第一次下落的時間來表示，令代數運算非常簡短。

## 解答

### 模型

把粒子視為質點（point mass），忽略空氣阻力，並假設台階光滑且水平，因此每次撞擊不改變水平速度，而鉛直速度則乘以 $-e$。設粒子被踢出時的水平速率為 $u$，並設 $t_1$ 為第一段飛行的時間。




實線弧是首兩段飛行；虛線弧是接下來實際發生的情況（見最後一部分）。

### 恢復係數

**第一段飛行**。粒子開始時沒有鉛直速度，下落 $h$，並水平移動 $d$（半個台階寬度，即到達第一級台階的中點）。因此

$$h = \tfrac{1}{2}gt_1^2, \qquad d = ut_1.$$

粒子以向下的鉛直速率 $gt_1$ 撞擊第一級台階，然後以向上的鉛直速率 $egt_1$ 離開，水平速率仍是 $u$。

**第二段飛行**。由第一級台階的中點到第二級台階的中點，水平距離為 $2d$。由於水平速率仍是 $u = d/t_1$，這段飛行歷時

$$T = \frac{2d}{u} = 2t_1.$$

在這段時間內，粒子最終必須下降高度 $h$。取向上為正，

$$\begin{aligned} -h &= (egt_1)(2t_1) - \tfrac{1}{2}g(2t_1)^2 \\ &= (2e-2)\,gt_1^2 \\ &= (2e-2)(2h), \end{aligned}$$

其中用了 $gt_1^2 = 2h$。兩邊除以 $h$ 得 $-1 = 4e - 4$，所以

$$e = \tfrac{3}{4}.$$

（在這段飛行中，粒子越過第一級台階邊緣時，位於該台階上方 $egt_1 \cdot t_1 - \frac{1}{2}gt_1^2 = \frac{1}{2}h$ 的高度，所以它確實能越過邊緣。）

### 能否一直持續下去？

**想一想** 如果這個規律持續下去，之後每一段飛行必須是怎樣的？

每段由一級台階中點飛到下一級台階中點的飛行，都以速率 $u$ 水平移動 $2d$，所以每段都歷時 $2t_1$，而且每段都必須恰好下降 $h$。由上面的計算可知，每次反彈後瞬間向上的鉛直速率 $w$ 必須滿足

$$-h = w(2t_1) - 2gt_1^2 = 2wt_1 - 4h, \qquad \text{所以} \qquad w = \frac{3h}{2t_1} = \tfrac{3}{4}gt_1.$$

因此每次反彈後，粒子都必須以鉛直速率 $\frac{3}{4}gt_1$ 離開；由於 $e = \frac{3}{4}$，它到達每一級台階時的鉛直速率都必須恰好是 $gt_1$。

這在第一級台階成立，但在第二級台階並不成立。經過歷時 $2t_1$ 的飛行到達第二級台階時，其鉛直速度為

$$\tfrac{3}{4}gt_1 - g(2t_1) = -\tfrac{5}{4}gt_1,$$

所以它以向下速率 $\frac{5}{4}gt_1$（而非 $gt_1$）撞擊第二級台階，並以向上速率 $\frac{3}{4}\cdot\frac{5}{4}gt_1=\frac{15}{16}gt_1>\frac{3}{4}gt_1$ 反彈。所以這是**不可能的**：粒子不能擊中每一級台階的中點。事實上，它會越過第三級台階的中點。設第三段飛行歷時 $st_1$，則

$$-h=\tfrac{15}{16}gt_1(st_1)-\tfrac{1}{2}gs^2t_1^2 \quad\Longrightarrow\quad s^2-\tfrac{15}{8}s-1=0 \quad\Longrightarrow\quad s=\frac{15+\sqrt{481}}{16}\approx 2.31,$$

所以它水平移動約 $2.31d$，落在第三級台階上、中點以外約 $0.31d$ 處（仍在該台階上，因為遠端邊緣在 $3d$ 之外）。

## 驗算

- $e=\frac{3}{4}$ 與 $g$、$h$ 或 $d$ 無關，這是理所當然的：把整道樓梯按比例放大或縮小，並不會改變路徑的形狀。
- 以 $g=9.8$、$h=d=1$ 及 $e=0.75$ 進行數值模擬，粒子落在 $x=1.000$ 及 $x=3.000$（首兩級台階的中點），然後落在 $x\approx 5.308$，與上面預測的 $3+2.308$ 吻合。
- 若沒有反彈（$e=0$），第二段飛行在時間 $2t_1$ 內會下降 $4h$；反彈必須補回其中的 $3h$，即 $2t_1\cdot w=3h$，與所得結果一致。

## 常見錯誤

- 把第一段水平距離當成 $2d$ 而不是 $d$：粒子由邊緣出發，落在第一級台階的中點。
- 在反彈時改變了水平速率。在光滑表面上，衝量（impulse）是鉛直的，所以只有鉛直分量會改變。
- 因為每段飛行看似相同而得出「可以」的結論：到達時的速率會不斷增大，所以反彈也會隨之增大。

## STEP 之外

彈跳的粒子是*離散動力系統*（discrete dynamical system）的一個簡單例子：某次反彈後瞬間的狀態，經由一個固定的映射決定下一次反彈後的狀態；而問運動能否重複，就是問這個映射的不動點（fixed point）。在這裏，不動點（反彈速率為 $\frac{6}{7}gt_1$，只會落低 $\frac{4}{7}h$）與這道樓梯並不相容，所以沒有任何周期運動（periodic motion）能符合。Kleppner 與 Kolenkow 的 *An Introduction to Mechanics*（《力學導論》）對碰撞與衝量有細緻的討論。

詳細解答

# STEP I 1987，第 13 題

力學 • 變力

> **STEP I 1987，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/87-s1-q13

## 關鍵思路

力的規律在 $x = 0$ 處改變，因此運動分成幾個階段，每個階段都由一條簡單的線性微分方程（differential equation）支配。在正方向的一側，力是與位移成正比的回復力（restoring force），即簡諧運動（simple harmonic motion）；在負方向的一側，力是與速度成正比的阻力（resistance），令速率呈指數衰減（exponential decay）。一個階段結束時的位置和速度，就是下一個階段的初始條件（initial conditions）。真正的問題在於共有多少個階段：質點究竟會不會回來？

## 解答

### 模型

把物體視為 $x$ 軸上只受所給的力作用的質點，並記 $\dot{x} = \mathrm{d}x/\mathrm{d}t$。牛頓第二定律（Newton's second law）$m\ddot{x} = \mathrm{F}$ 給出

$$\ddot{x} = -\mu^2 x \quad (x \geq 0), \qquad\qquad \ddot{x} = -\kappa\dot{x} \quad (x < 0).$$

質量互相抵消，因此與答案無關。

### 第一階段：$x \geq 0$ 時的簡諧運動

由於 $v_0 > 0$，質點立即進入 $x > 0$ 的區域，在那裏 $\ddot{x} = -\mu^2 x$。這是角頻率（angular frequency）為 $\mu$ 的簡諧運動，其通解（general solution）為

$$x = A\cos\mu t + B\sin\mu t.$$

條件「$t=0$ 時 $x=0$」給出 $A=0$，而「$t=0$ 時 $\dot{x}=B\mu=v_0$」給出 $B=v_0/\mu$。因此

$$x=\frac{v_0}{\mu}\sin\mu t,\qquad \dot{x}=v_0\cos\mu t.$$

只要 $x\geq 0$，即 $\sin\mu t\geq 0$，這個解便成立，也就是 $0\leq t\leq \pi/\mu$。質點在 $t=\pi/(2\mu)$ 時到達最大位移 $v_0/\mu$，並在

$$t=T=\frac{\pi}{\mu},\qquad \text{此時速度為}\qquad \dot{x}=v_0\cos\pi=-v_0.$$

時返回原點。它回來時的速率與出發時相同，因為在這一側能量守恆（力 $-m\mu^2x$ 來自位能（potential energy）$\frac{1}{2}m\mu^2x^2$，而位能在原點為零）。

### 第二階段：$x<0$ 時的受阻運動

質點以速度 $-v_0$ 進入 $x<0$ 的區域，在那裏 $\ddot{x}=-\kappa\dot{x}$。記 $u=\dot{x}$，這就是一階方程（first-order equation）$\mathrm{d}u/\mathrm{d}t=-\kappa u$，滿足「$t=T$ 時 $u=-v_0$」的解為

$$\dot{x}=-v_0\,\mathrm{e}^{-\kappa(t-T)}.$$

利用「$t=T$ 時 $x=0$」積分：

$$x=\int_T^t -v_0\,\mathrm{e}^{-\kappa(s-T)}\,\mathrm{d}s=-\frac{v_0}{\kappa}\left(1-\mathrm{e}^{-\kappa(t-T)}\right).$$

**想一想** 質點會不會回到 $x\geq 0$，開始第三個階段？

不會。對每個 $t>T$，速度 $-v_0\mathrm{e}^{-\kappa(t-T)}$ 都是負數，所以在整個階段中 $x$ 不斷減少；對所有 $t>T$，它都嚴格為負，上面的公式永遠成立。阻力令質點減速，卻永遠不會令它反向。當 $t\to\infty$ 時，$\mathrm{e}^{-\kappa(t-T)}\to 0$，因此質點趨近點 $x=-v_0/\kappa$，但永遠不會到達。

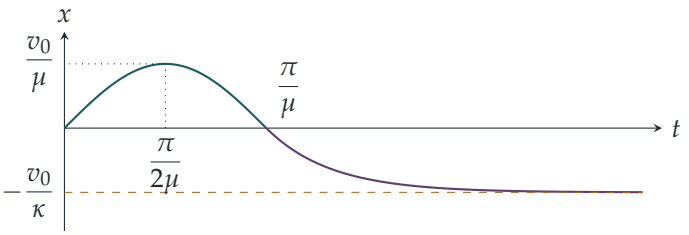
### 答案

$$x=\begin{cases}\dfrac{v_0}{\mu}\sin\mu t, & 0\leq t\leq\dfrac{\pi}{\mu},\\[2ex] -\dfrac{v_0}{\kappa}\left(1-\mathrm{e}^{-\kappa(t-\pi/\mu)}\right), & t>\dfrac{\pi}{\mu}.\end{cases}$$

兩段在 $t=\pi/\mu$ 處都有 $x=0$ 和 $\dot{x}=-v_0$，所以圖像在該處是光滑的（斜率連續），儘管 $\ddot{x}$ 從 0 跳到 $\kappa v_0$。

### 草圖

圖像先是軸上方半個正弦曲線的拱形，接着是一條以相同斜率 $-v_0$ 離開原點的曲線，並逐漸趨平，趨近漸近線（asymptote）$x=-v_0/\kappa$。（草圖取 $\mu=1$、$\kappa=1$。）




一般而言，峯高 $v_0/\mu$ 與漸近線的深度 $v_0/\kappa$ 互不相關：阻力較弱（$\kappa$ 較小）時，質點會向左漂移很遠。

## 驗算

- 單位：$\mu$ 和 $\kappa$ 的單位都是 $\mathrm{s}^{-1}$，所以 $v_0/\mu$ 和 $v_0/\kappa$ 都是長度，理應如此。
- 極限情況：當 $\kappa \to 0$ 時沒有阻力，而確實 $-\frac{v_0}{\kappa}(1-\mathrm{e}^{-\kappa s}) \to -v_0 s$，即以速率 $v_0$ 作勻速運動。
- 數值驗證：取 $\mu = 1.7$、$\kappa = 0.9$、$v_0 = 2.2$，用龍格－庫塔法（Runge–Kutta method）逐步求解運動方程，在 $0 \le t \le 15$ 範圍內與公式的誤差在 $3 \times 10^{-5}$ 以內；質點的峯值為 $x = 1.2941 = v_0/\mu$，最終停在 $x = -2.4444 = -v_0/\kappa$ 附近。

## 常見錯誤

- 第二階段的時間由 $t = 0$ 而不是由 $t = \pi/\mu$ 起計：指數必須是 $-\kappa(t - \pi/\mu)$。
- 忘記檢查質點會否重新進入 $x \ge 0$；題目要求的是其後的整個運動，所以「沒有第三階段」這個結論需要一句理由支持。

## STEP 之外

這裏的力在 $x = 0$ 兩側各是一個不同的線性表達式，屬於分段線性系統（piecewise-linear system）；把各段的解拼接起來、在切換點匹配位置和速度的技巧，是微分方程入門課程處理這類系統（以及碰撞和摩擦問題）的標準方法。若同時加入兩項，得 $\ddot{x} + \kappa\dot{x} + \mu^2 x = 0$，這就是阻尼諧振子（damped harmonic oscillator），它的欠阻尼（underdamped）、臨界阻尼（critically damped）和過阻尼（overdamped）三種情況，是任何力學入門課程的核心例子。Kleppner 和 Kolenkow 的 *An Introduction to Mechanics*（《力學導論》）對阻尼振子有細緻的討論。

詳細解答

# STEP I 1987，第 14 題

概率與統計 • 概率的等比級數

**STEP I 1987，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/87-s1-q14

## 關鍵思路

第一局打完之後，一場仍未分勝負的錦標賽便完全沒有自由度可言。每局的勝者要對上該局的裁判（referee）；若錦標賽要繼續未分勝負，裁判就必須勝出，因此下一局的勝者是被迫決定的。所以一切只取決於誰贏第 1 局，以及某位選手首次連勝兩局的那一刻。之後各概率便化為以週期（period） 3 循環的幾何級數（geometric series）。

## 解答

在整個過程中，每局都由上一局的勝者對上上一局的裁判。上一局的負方在旁擔任裁判，因此不可能贏當前這一局。

### 第 (i) 部分

第 1 局：$A$ 擊敗 $B$，由 $B$ 擔任第 2 局的裁判。第 2 局是 $A$ 對 $C$，而 $C$ 勝出，這是可能的。第 3 局是 $C$ 對 $B$，而 $B$ 勝出，這同樣可能。但沒有人連勝兩局，所以在 $ACB$ 之後錦標賽仍未分勝負，必須繼續。因此 $ACB$ **不是**一場完整錦標賽的紀錄（它只是一場錦標賽的開頭）。

### 第 (ii) 部分

$A$ 贏了第 1 局後，第 2 局是 $A$ 對 $C$，由 $B$ 擔任裁判。所以 $B$ 不可能贏第 2 局，$ABB$ 是**不可能**的。

### 第 (iii) 部分

與 (i) 一樣，首三局 $ACB$ 是可能的，而第 3 局是 $C$ 對 $B$，由 $A$ 擔任裁判。所以第 4 局是 $B$ 對 $A$，而 $B$ 可以贏這一局。這時 $B$ 連贏了第 3 和第 4 局，錦標賽結束。因此 $ACBB$ **是**一場完整的錦標賽，由 $B$ 在四局內勝出。

### 5 局後仍未分勝負

設 $X$ 贏第 1 局，其中 $X$ 是 $A$ 或 $B$，並設 $Y$ 為另一人。若錦標賽要保持未分勝負，每局都必須由不是上一局勝者的選手勝出，而這只能是上一局的裁判。這就迫使紀錄為

$$X\,C\,Y\,X\,C\,Y\,X\,\ldots$$

因為每位新的勝者都是上一局的裁判。具體來說，第 2 局是 $X$ 對 $C$（由 $Y$ 擔任裁判），所以 $C$ 必須勝出。第 3 局是 $C$ 對 $Y$（由 $X$ 擔任裁判），所以 $Y$ 必須勝出。第 4 局是 $Y$ 對 $X$，所以 $X$ 必須勝出，如此類推，週期為 3。

第 1 局誰勝都可以。之後，每多一局保持未分勝負的概率都是 $\frac{1}{2}$，因為各局互相獨立 (independent)，而且每位選手的勝算均等。所以

$$P(n \text{ 局後仍未分勝負}) = \left(\tfrac{1}{2}\right)^{n-1}, \qquad P(5 \text{ 局後仍未分勝負}) = \tfrac{1}{16}.$$

### 誰勝出，何時勝出

錦標賽在第 $n$ 局（對 $n \geq 2$）結束，當且僅當 $n-1$ 局後仍未分勝負，而第 $n-1$ 局的勝者再次勝出。其概率為 $\left(\frac{1}{2}\right)^{n-2} \cdot \frac{1}{2} = \left(\frac{1}{2}\right)^{n-1}$。冠軍就是第 $n-1$ 局的勝者，即被迫序列中的第 $(n-1)$ 個字母。記 $k = n-1$，則：

| $k \bmod 3$ | 1 | 2 | 0 |
|---|---|---|---|
| 若 $A$ 贏第 1 局，冠軍為 | $A$ | $C$ | $B$ |
| 若 $B$ 贏第 1 局，冠軍為 | $B$ | $C$ | $A$ |

所以錦標賽在第 $k+1$ 局結束的概率為 $\left(\frac{1}{2}\right)^{k}$。若 $k \equiv 2 \pmod 3$，這整個概率都歸 $C$。否則它由 $A$ 和 $B$ 平分，各得 $\left(\frac{1}{2}\right)^{k+1}$。

**想一想** 為甚麼在 $k \equiv 2$ 時 $C$ 得到全部概率，而在其他時候只得一半？

無論誰贏第 1 局，$C$ 永遠是被迫序列中的第二個字母。所以兩個分支都在同樣的時刻讓 $C$ 勝出。$A$ 和 $B$ 在兩個分支之間互換角色，所以他們各自只能取得自己佔據該位置的那一個分支。

### 在 5 局或以內勝出

這裏 $n = 2,3,4,5$，所以 $k = 1,2,3,4$：

$$\begin{aligned} P(A \text{ 勝出，局數} \le 5) &= \underbrace{\tfrac{1}{4}}_{k=1} + \underbrace{\tfrac{1}{16}}_{k=3} + \underbrace{\tfrac{1}{32}}_{k=4} = \tfrac{11}{32}, \\ P(B \text{ 勝出，局數} \le 5) &= \tfrac{11}{32} \quad \text{（由 } A \text{ 與 } B \text{ 之間的對稱性），} \\ P(C \text{ 勝出，局數} \le 5) &= \underbrace{\tfrac{1}{4}}_{k=2}. \end{aligned}$$

驗算一下，$\frac{11}{32} + \frac{11}{32} + \frac{8}{32} = \frac{15}{16} = 1 - \frac{1}{16}$，與 5 局後仍未分勝負的概率吻合。

### 最終勝者

由於 $P(n \text{ 局後仍未分勝負}) = \left(\frac{1}{2}\right)^{n-1} \to 0$，錦標賽以概率 1 結束。所以我們只需對所有 $k \ge 1$ 求和。每個剩餘類（residue class）都給出一個公比為 $\frac{1}{8}$ 的幾何級數：

$$\begin{aligned} P(C \text{ 勝出}) &= \tfrac{1}{4} + \tfrac{1}{32} + \tfrac{1}{256} + \cdots = \frac{1/4}{1 - 1/8} = \frac{2}{7}, \\ P(A \text{ 勝出}) &= \underbrace{\left(\tfrac{1}{4} + \tfrac{1}{32} + \cdots\right)}_{k \equiv 1} + \underbrace{\left(\tfrac{1}{16} + \tfrac{1}{128} + \cdots\right)}_{k \equiv 0} = \frac{1/4}{7/8} + \frac{1/16}{7/8} = \frac{2}{7} + \frac{1}{14} = \frac{5}{14}, \end{aligned}$$

正如所求。由對稱性，$P(B \text{ 勝出}) = \frac{5}{14}$ 亦然，而且確實 $\frac{5}{14} + \frac{5}{14} + \frac{2}{7} = 1$。

所以第一局在旁輪空的 $C$ 處於劣勢：$\frac{2}{7} \approx 0.286$，而其餘兩人各有 $\frac{5}{14} \approx 0.357$。

## 驗算

- 三個在 5 局內勝出的概率之和為 $\frac{15}{16}$，而最終勝出的概率之和為 1。
- 以分數對所有可能的對局序列作精確枚舉，確認了 $\frac{1}{16}$、$\frac{11}{32}$、$\frac{11}{32}$、$\frac{1}{4}$。計算至 60 局，得出 $A$ 和 $B$ 的值為 0.357142…，$C$ 的值為 0.285714…。
- 求 $C$ 的另一個快捷方法：設 $c$ 為 $C$ 勝出的概率。第 1 局後，$C$ 立即勝出的概率為 $\frac{1}{4}$。若三局之後局面回到相同的情況（概率為 $\frac{1}{8}$），便重新開始，所以 $c = \frac{1}{4} + \frac{1}{8}c$，得 $c = \frac{2}{7}$。

## 常見錯誤

- 說 $ACB$ 是有效的紀錄：其中每一局都是可能的，但沒有人連勝兩局，所以錦標賽還未結束。

- 忘記上一局的負方擔任裁判，因而不能贏下一局（這排除了 $ABB$）。
- 在 $k \equiv 2$ 時只給 $C$ 一半的概率：$C$ 在兩個分支中都在這些時刻勝出。

## STEP 之外

這個錦標賽是一條馬可夫鏈（Markov chain）：每局之後的狀態就只是「上一局誰贏、誰當裁判」，而當鏈被吸收（absorbed）時冠軍便決定了。「驗算」一節中的首步分析（first-step argument），即透過以下一步發生的事為條件，為某個概率寫出方程，是求吸收概率（absorption probability）的標準方法，而且無須對級數求和。Grimmett 和 Welsh 的 *Probability: An Introduction*（《概率導論》）在大學一年級程度介紹馬可夫鏈與吸收。

詳細解答

# STEP I 1987，第 15 題

概率與統計 • 連續隨機變量

**STEP I 1987，第 15 題**

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

step.maths.org/questions/87-s1-q15

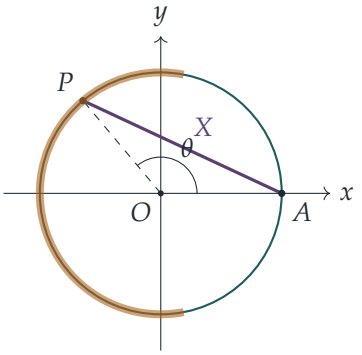
## 關鍵思路

「在圓上均勻分佈（uniformly distributed）」的意思是 $P$ 在圓心所成的角是均勻分佈的，所以自然的做法是設 $P = (\cos\theta, \sin\theta)$，其中 $\theta$ 在 $[0, 2\pi)$ 上均勻分佈。這樣，與 $(1,0)$ 的距離就是單位圓的一條弦（chord），而半角公式（half-angle formula）可把它化成單一個正弦。之後，平均值（mean）只需一個積分，方差（variance）幾乎完全不用積分，而所求概率則變成一個關於角度區間的問題。

## 解答

### 設定

設 $\theta$ 為由正 $x$ 軸到 $OP$ 的角，其中 $O$ 為原點。在圓上均勻地選取 $P$，即是均勻地選取弧長（arc length），所以 $\theta$ 在 $[0, 2\pi)$ 上均勻分佈，概率密度（density）為 $\frac{1}{2\pi}$。記 $A = (1,0)$ 及 $P = (\cos\theta, \sin\theta)$。



於是

$$X^2 = (\cos\theta - 1)^2 + \sin^2\theta = 2 - 2\cos\theta = 4\sin^2\tfrac{\theta}{2},$$

這裏用了 $1 - \cos\theta = 2\sin^2\frac{\theta}{2}$。由於 $0 \le \frac{\theta}{2} < \pi$，我們有 $\sin\frac{\theta}{2} \ge 0$，所以取正平方根是穩妥的：

$$X = 2\sin\tfrac{\theta}{2}.$$

### 平均值

$$\mathrm{E}(X) = \frac{1}{2\pi}\int_0^{2\pi} 2\sin\tfrac{\theta}{2}\,\mathrm{d}\theta = \frac{1}{2\pi}\Big[-4\cos\tfrac{\theta}{2}\Big]_0^{2\pi} = \frac{1}{2\pi}(4+4) = \frac{4}{\pi}.$$

所以 $\mathrm{E}(X) = \dfrac{4}{\pi} \approx 1.273$。

### 方差

**想一想** 要求 $\mathrm{E}(X^2)$，我們需要對 $4\sin^2\frac{\theta}{2}$ 積分嗎？

不需要：未化成平方的形式更容易處理。$X^2 = 2 - 2\cos\theta$，而 $\cos\theta$ 在整整一圈上的平均值為零，因為 $\int_0^{2\pi}\cos\theta\,\mathrm{d}\theta = 0$。因此 $\mathrm{E}(X^2) = 2$，而

$$\mathrm{Var}(X) = \mathrm{E}(X^2) - \big(\mathrm{E}(X)\big)^2 = 2 - \frac{16}{\pi^2} \approx 0.379.$$

### $X$ 大於其平均值的概率

我們需要 $2\sin\frac{\theta}{2} > \frac{4}{\pi}$，即 $\sin\frac{\theta}{2} > \frac{2}{\pi}$。留意 $0 < \frac{2}{\pi} < 1$，所以此不等式有解。設 $\alpha = \arcsin\frac{2}{\pi}$，它位於 $(0, \frac{\pi}{2})$ 內。當 $\frac{\theta}{2}$ 取遍 $[0,\pi)$ 時，正弦值由 0 升至 1 再回落，而它大於 $\sin\alpha$ 當且僅當

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

這就是圖中加粗的圓弧：圓上位於 $A$ 對面一側的點。由於 $\theta$ 均勻分佈，所求概率就是此區間的長度除以 $2\pi$：

$$\mathrm{P}\big(X > \mathrm{E}(X)\big) = \frac{2\pi - 4\alpha}{2\pi} = \boxed{1 - \frac{2}{\pi}\arcsin\frac{2}{\pi}} \approx 0.561.$$

## 驗算

- $X$ 的取值範圍是 $[0,2]$，而平均值 $4/\pi \approx 1.27$ 位於中點 1 之上。這與圖像吻合：在圓上頗大的一部分，弦長都接近其最大值（當 $P$ 經過對面一側時，$X$ 變化緩慢），所以分佈偏向較大的值。

- 基於同樣原因，大於平均值的概率應略大於 $\frac{1}{2}$，而 0.561 正是如此。
- 在數值上，以中點法則（midpoint rule）對 $2 \times 10^5$ 個等距角度取平均，得到平均值 1.273240、方差 0.378861 及 $\mathrm{P} = 0.56066$；而一個 $10^6$ 點的模擬（simulation）在抽樣誤差範圍內亦與此一致。

## 常見錯誤

- 取 $\theta \in (-\pi, \pi]$ 而仍寫 $X = 2\sin\frac{\theta}{2}$。此時 $\sin\frac{\theta}{2}$ 可以是負數，你需要寫 $X = 2\left|\sin\frac{\theta}{2}\right|$；使用 $[0, 2\pi)$ 就可避免絕對值（modulus）。
- 把 $P$ 的 $x$ 坐標當作在 $[-1, 1]$ 上均勻分佈。在圓上均勻分佈是指在角度（弧長）上均勻，而不是在 $x$ 上均勻。

## STEP 之外

這是隨機變量變換（transformation of a random variable）的第一個例子：$X$ 的概率密度函數是 $[0, 2]$ 上的 $f(x) = \dfrac{2}{\pi\sqrt{4 - x^2}}$，屬於反正弦（arcsine）類型的分佈，而上面每一個答案都可以由它重新推導出來。這道題亦暗示了「隨機」一詞可以多麼難以捉摸：以不同的自然方式隨機選取圓的一條弦，會得出不同的答案（貝特朗悖論（Bertrand's paradox））。這兩個概念在 Blitzstein 與 Hwang 的 *Introduction to Probability*（《概率導論》）中都有進一步發展。

詳細解答

# STEP I 1987，第 16 題

概率與統計 • 貝葉斯定理與正態近似

**STEP I 1987，第 16 題**

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

step.maths.org/questions/87-s1-q16

## 關鍵思路

第 (i) 部分其實是喬裝了的貝葉斯定理（Bayes' theorem）。兩次投票立場相同，是判斷該議員屬於哪個政黨的證據（evidence），因為守成黨（Preservative）議員必定如此，而進步黨（Progressive）議員只是有時如此。所以我們先更新該議員是守成黨議員的概率，然後才預測第三次投票。在第 (ii) 部分，守成黨議員的投票是固定的數目，所以所有隨機性都來自那 40 位進步黨議員。他們投贊成票的人數服從二項分佈（binomial distribution），而我們要用正態分佈（normal distribution）來近似的，正是這個分佈。

## 解答

### 第 (i) 部分

設 $R$ 為「被選中的議員是守成黨議員」這事件，故 $P(R) = \frac{60}{100} = \frac{3}{5}$，而 $P(R') = \frac{2}{5}$。設 $S_2$ 為「該議員首兩次投票立場相同」這事件，$S_3$ 為「第三次投票的立場亦相同」這事件。

守成黨議員從不改變主意，所以 $P(S_2 \mid R) = 1$。進步黨議員則隨機投票，每次投票都獨立（independently）地以概率 $\frac{1}{2}$ 投贊成或反對。無論第一次投了甚麼，第二次與之相同的概率都是 $\frac{1}{2}$，所以 $P(S_2 \mid R') = \frac{1}{2}$。因此

$$P(S_2) = \tfrac{3}{5} \cdot 1 + \tfrac{2}{5} \cdot \tfrac{1}{2} = \tfrac{4}{5}, \qquad P(R \mid S_2) = \frac{P(R)\,P(S_2 \mid R)}{P(S_2)} = \frac{3/5}{4/5} = \frac{3}{4}.$$

所以，根據我們觀察所得，該議員是守成黨議員的概率為 $\frac{3}{4}$，是進步黨議員的概率為 $\frac{1}{4}$。

? **想一想** 既然該議員是隨機選出的，為甚麼我們不能直接說 $P(R \mid S_2) = \frac{3}{5}$？

因為資訊 $S_2$ 對守成黨議員而言比對進步黨議員更有可能出現，所以它會改變勝算比

（odds）。觀察之前，勝算比是 $3:2$，偏向守成黨議員。觀察到 $S_2$ 之後，勝算比變成 $3\cdot 1 : 2\cdot \frac{1}{2} = 3:1$。

現在，在已知該議員所屬政黨的條件下，第三次投票與首兩次投票無關，首兩次投票的作用只在於透露該議員屬於哪個政黨。守成黨議員必定重複同一立場，進步黨議員則以概率 $\frac{1}{2}$ 投出相同立場。所以

$$P(S_3 \mid S_2) = \tfrac{3}{4}\cdot 1 + \tfrac{1}{4}\cdot\tfrac{1}{2} = \frac{7}{8}.$$

（等價地，$P(S_3 \mid S_2) = \dfrac{P(S_2 \cap S_3)}{P(S_2)} = \dfrac{\frac{3}{5} + \frac{2}{5}\cdot\frac{1}{4}}{\frac{4}{5}} = \dfrac{7/10}{4/5} = \dfrac{7}{8}$。）

**第 (ii) 部分**

考慮就一項守成黨政策進行的投票，並假設全部 100 位議員都投票。60 位守成黨議員中有比例 $\alpha$，即 $60\alpha$ 位議員，總是投反對票，其餘 $60(1-\alpha)$ 位則總是投贊成票。設 $X$ 為投贊成票的進步黨議員人數。每位進步黨議員獨立地以概率 $\frac{1}{2}$ 投贊成票，所以

$$X \sim \mathrm{B}\left(40, \tfrac{1}{2}\right), \qquad E(X) = 20, \qquad \mathrm{Var}(X) = 40\cdot\tfrac{1}{2}\cdot\tfrac{1}{2} = 10.$$

在 100 票中取得簡單多數（simple majority），即至少有 51 票贊成：

$$60(1-\alpha) + X \geq 51 \iff X \geq 60\alpha - 9.$$

由於 $n = 40$ 相當大，而 $p = \frac{1}{2}$ 令二項分佈對稱，我們用 $Y \sim \mathrm{N}(20, 10)$ 來近似 $X$。這裏 $60\alpha - 9$ 是整數（它是議員人數），所以加上連續性校正（continuity correction）後，

$$P(\text{過半數}) = P(X \geq 60\alpha - 9) \approx P(Y > 60\alpha - 9.5) = P\left(Z > \frac{60\alpha - 29.5}{\sqrt{10}}\right),$$

其中 $Z \sim \mathrm{N}(0,1)$。

這個概率隨 $\alpha$ 增加而遞減，所以恰好當 $\alpha$ 大得足以把它壓到 0.8 以下時，便會舉行選舉。查表得 $P(Z > -0.8416) = 0.8$，所以我們需要

$$\frac{60\alpha - 29.5}{\sqrt{10}} > -0.8416 \iff 60\alpha > 29.5 - 0.8416\sqrt{10} = 26.84$$

$$\iff \alpha > 0.447.$$

由於 $60\alpha$ 必須是整數（議員人數），最小的可能值是 $60\alpha = 27$。所以導致選舉的 $\alpha$ 最小值為

$$\alpha = \frac{27}{60} = 0.45.$$

驗證這個答案：有 26 位倒戈議員時，取得過半數的近似概率為 $P(Z > -1.107) \approx 0.866$；有 27 位倒戈議員時，則為 $P(Z > -0.791) \approx 0.785 < 0.8$。

## 驗算

- 在 (i) 中，$\frac{7}{8}$ 介乎 $\frac{1}{2}$（已知是進步黨議員）與 1（已知是守成黨議員）之間，理應如此。以 400,000 位隨機議員進行模擬，得出 0.875。
- 在 (ii) 中，精確的二項概率為 $P(X \geq 17) = 0.8659$（26 位倒戈議員）及 $P(X \geq 18) = 0.7852$（27 位倒戈議員）。這些數值與正態近似值在三位小數內一致，並確認分界點位於 26 與 27 位倒戈議員之間，即 $\alpha = 0.45$。

## 常見錯誤

- 在 (i) 中答 $\frac{3}{5} \cdot 1 + \frac{2}{5} \cdot \frac{1}{2} = \frac{4}{5}$。這忽略了「該議員已經兩次投出相同立場」這項資訊。
- 把簡單多數當作「至少 50 票」。50 對 50 打和並不算過半數，所以需要 51 票。
- 漏了連續性校正。沒有它的話，臨界值會變成 $\alpha > 0.456$，從而得出 28 位倒戈議員而不是 27 位；根據精確的二項計算，這是錯的。

## STEP 之外

第 (i) 部分是貝葉斯更新（Bayesian updating）的一個小例子：數據改變我們對某個隱藏「類型」的信念，然後預測就是對更新後的信念取平均（即後驗預測分佈（posterior predictive distribution））。第 (ii) 部分的正態近似是中央極限定理(central limit theorem)的一個例子，而這裏近乎完美的吻合並非偶然，因為當 $p = \frac{1}{2}$ 時二項分佈是對稱的。這兩個概念在 Blitzstein 與 Hwang 的 *Introduction to Probability*（《概率導論》）中都有仔細的闡述。

詳細解答

# STEP II 1987，第 1 題

純數學 • 代數恆等式

**STEP II 1987，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/87-s2-q1

## 關鍵思路

不要急於代入消去其中一個字母，而應該把右邊展開（expand）。平方項 $a^2$、$b^2$、$c^2$ 在兩邊都出現，可以互相抵消，剩下的是 $x$ 乘以一個線性表達式（linear expression），而這個表達式恰好是已知條件的某個倍數。一旦看清數字 7 和 11 從何而來，一般結果便水到渠成。

## 解答

先一次過做一般情況的展開會比較方便。對任意數 $p$、$q$、$r$，

$$(px-a)^2+(qx-b)^2+(rx-c)^2=(p^2+q^2+r^2)x^2-2x(pa+qb+rc)+a^2+b^2+c^2,$$

所以

$$(px-a)^2+(qx-b)^2+(rx-c)^2-(a^2+b^2+c^2)=x\big[(p^2+q^2+r^2)x-2(pa+qb+rc)\big]. \qquad (*)$$

### 第 (i) 部分

這裡 $(p,q,r)=(1,2,3)$，所以 $p^2+q^2+r^2=14$，而 $(*)$ 變成

$$(x-a)^2+(2x-b)^2+(3x-c)^2-(a^2+b^2+c^2)=x\big[14x-2(a+2b+3c)\big].$$

若 $a+2b+3c=7x$，方括號內便是 $14x-14x=0$。因此兩邊相等，得證。

### 第 (ii) 部分

現在 $(p,q,r)=(2,3,3)$，所以 $p^2+q^2+r^2=4+9+9=22$，而

$$(2x-a)^2+(3x-b)^2+(3x-c)^2-(a^2+b^2+c^2)=x\big[22x-2(2a+3b+3c)\big].$$

若 $2a+3b+3c=11x$，方括號內便是 $22x-22x=0$，這就證明了 (ii)。

> **想一想** 為甚麼 (i) 中是 7，(ii) 中是 11？

它們都是 $p^2+q^2+r^2$ 的一半：$7=\frac{1}{2}(1+4+9)$，$11=\frac{1}{2}(4+9+9)$。這就告訴我們一般的命題應該是甚麼。

### 一般結果

恆等式（identity）$(*)$ 直接證明了以下結果。

*對任意實數 $p,q,r,a,b,c,x$，若*

$$2(pa+qb+rc)=(p^2+q^2+r^2)\,x,$$

則

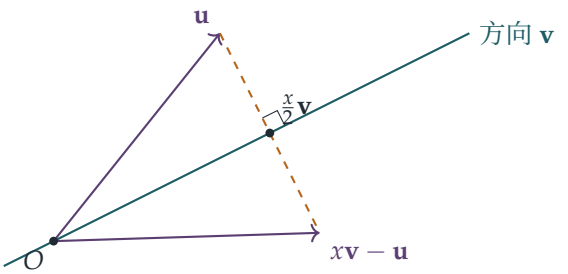
$$a^2+b^2+c^2=(px-a)^2+(qx-b)^2+(rx-c)^2.$$

第 (i) 部分是 $(p,q,r)=(1,2,3)$ 的情況，第 (ii) 部分是 $(p,q,r)=(2,3,3)$ 的情況。

同樣的展開對任意多個變量都適用。若 $2\sum_{i=1}^n p_ia_i=\big(\sum_{i=1}^n p_i^2\big)x$，則 $\sum a_i^2=\sum(p_ix-a_i)^2$。由 $(*)$ 我們還得到一個逆命題（converse）：若 $x\neq 0$ 且恆等式成立，則條件必定成立。

### 幾何上的意義

記 $\mathbf{u}=(a,b,c)$ 及 $\mathbf{v}=(p,q,r)$。條件可寫成 $\mathbf{u}\cdot\mathbf{v}=\frac{x}{2}|\mathbf{v}|^2$，意思是 $\frac{x}{2}\mathbf{v}$ 是從 $\mathbf{u}$ 到經過 $O$、方向為 $\mathbf{v}$ 的直線的垂足（foot of the perpendicular）。於是向量 $x\mathbf{v}-\mathbf{u}=2\cdot\frac{x}{2}\mathbf{v}-\mathbf{u}$ 就是 $\mathbf{u}$ 關於該直線的反射（reflection）。反射保持長度不變，所以 $|x\mathbf{v}-\mathbf{u}|=|\mathbf{u}|$，而這正是那個恆等式。



## 驗算

以 $x=1$、$a=b=0$、$c=\frac{7}{3}$ 試驗 (i)。左邊是 $\frac{49}{9}$，右邊是 $1+4+\left(\frac{2}{3}\right)^2=5+\frac{4}{9}=\frac{49}{9}$。我們還用精確的有理數運算測試了兩部分及一般結果，每項各測試 2000 個隨機例子（一般情況最多六個變量），並確認了當 $x\neq 0$ 而條件不成立時，恆等式必定不成立。

## 常見錯誤

- 一開始就代入 $c=\frac{1}{3}(7x-a-2b)$。這樣做行得通，但代數運算很繁瑣，而且掩蓋了問題的結構。
- 把一般結果寫成「$pa+qb+rc=kx$」，卻沒有交代 $k$ 與 $p,q,r$ 的關係。恆等式要求 $k=\frac{1}{2}(p^2+q^2+r^2)$。

## STEP 之外

這個幾何圖像正是線性代數（linear algebra）的起點。映射 $\mathbf{u}\mapsto 2\frac{\mathbf{u}\cdot\mathbf{v}}{|\mathbf{v}|^2}\mathbf{v}-\mathbf{u}$ 是一個反射，屬於正交線性映射（orthogonal linear map），而正交映射恰好就是保持長度不變的線性映射。像這樣的投影（projection）和反射，是最小二乘法（least squares）以及數值計算中所用的 Householder 變換（Householder transformations）的基礎。Strang 的 *Introduction to Linear Algebra*（《線性代數導論》）以同樣的幾何風格發展這些概念。

詳細解答

# STEP II 1987，第 2 題

純數學 • 三角恆等式

> **STEP II 1987，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/87-s2-q2

## 關鍵思路

這個式子只涉及 $\sin^2 A$、$\sin^2 B$ 和 $\sin^2 C$，而且它是 $\sin^2 C$ 的二次式（quadratic）。關於 $A \pm B \pm C$ 的條件表示 $C$ 與 $A+B$ 或 $A-B$ 有關，因此自然會猜想這個二次式的兩個根（roots）就是 $\sin^2(A+B)$ 和 $\sin^2(A-B)$。驗證兩根之和與兩根之積便能證實這個猜想，結果隨之而出。

## 解答

設 $x = \sin^2 A$、$y = \sin^2 B$、$z = \sin^2 C$，並把左邊記作 $E$。於是

$$E = x^2 + y^2 + z^2 - 2yz - 2zx - 2xy + 4xyz.$$

按 $z$ 的冪次整理，

$$\begin{aligned} E &= z^2 - 2z(x + y - 2xy) + (x^2 - 2xy + y^2) \\ &= z^2 - 2(x + y - 2xy)\,z + (x - y)^2. \end{aligned}$$

### 二次式的根

我們斷言

$$E = \big(z - \sin^2(A+B)\big)\big(z - \sin^2(A-B)\big).$$

由於兩邊都是 $z$ 的首一（monic）二次式，只需驗證 $\sin^2(A+B)$ 與 $\sin^2(A-B)$ 的和與積正確即可。

和。展開 $\sin(A \pm B) = \sin A \cos B \pm \cos A \sin B$ 再平方，相加時交叉項互相抵消：

$$\begin{aligned}\sin^2(A+B) + \sin^2(A-B) &= 2\sin^2 A\cos^2 B + 2\cos^2 A \sin^2 B \\ &= 2x(1-y) + 2(1-x)y \\ &= 2(x+y-2xy).\end{aligned}$$

積。利用平方差（difference of two squares），

$$\sin(A+B)\sin(A-B) = \sin^2 A\cos^2 B - \cos^2 A \sin^2 B = x(1-y) - (1-x)y = x - y,$$

所以 $\sin^2(A+B)\sin^2(A-B) = (x-y)^2$。

兩者都與 $E$ 的係數吻合，斷言得證。

> **想一想** 為甚麼只需知道 $\sin^2 C$ 等於 $\sin^2(A+B)$ 或 $\sin^2(A-B)$ 就足夠？

因為這樣 $z$ 就是兩個根之一，於是 $E$ 的其中一個因式為零。餘下要證明的，就是角的條件恰好給出這一點。

### 運用條件

假設對某個整數 $k$ 及某些符號 $\varepsilon_1, \varepsilon_2 \in \{1, -1\}$，有 $A + \varepsilon_1 B + \varepsilon_2 C = k\pi$。那麼

$$C = \varepsilon_2\big(k\pi - (A + \varepsilon_1 B)\big).$$

函數 $\sin^2$ 是偶函數（even function），且週期（period）為 $\pi$，因此符號 $\varepsilon_2$ 和 $k\pi$ 都不會改變它的值：

$$\sin^2 C = \sin^2(A + \varepsilon_1 B).$$

所以 $z = \sin^2(A+B)$（當 $\varepsilon_1 = 1$）或 $z = \sin^2(A-B)$（當 $\varepsilon_1 = -1$）。無論哪種情況，$E$ 都有一個因式為零，因此

$$E = 0,$$

即為所求。■

### 更簡潔的形式（及逆命題）

每個因式都是兩個正弦平方之差，而 $\sin^2 P - \sin^2 Q = \sin(P+Q)\sin(P-Q)$（即上面的積公式取 $A = P$、$B = Q$）。所以

$$E = \sin(A+B+C)\,\sin(C-A-B)\,\sin(C+A-B)\,\sin(C-A+B).$$

這裏的四個角在不計正負號下，正好是 $A+B+C$、$A+B-C$、$A-B+C$ 和 $A-B-C$。這令結果一目了然：$E = 0$ 當且僅當四個角 $A \pm B \pm C$ 之一是 $\pi$ 的倍數。這也給出了逆命題（converse）。

## 驗算

- 取 $A=\frac{\pi}{6}$、$B=\frac{\pi}{3}$、$C=\frac{\pi}{2}$，於是 $A+B-C=0$。那麼 $x=\frac{1}{4}$、$y=\frac{3}{4}$、$z=1$，且 $E=\frac{1}{16}+\frac{9}{16}+1-\frac{3}{2}-\frac{1}{2}-\frac{3}{8}+\frac{3}{4}=0$。
- 取 $A=0.3$、$B=0.7$、$C=1.1$（沒有特殊關係），$E$ 和四個正弦之積的值都是 0.05538，可見在條件不成立時，這個式子確實不為零。
- 對 $10^5$ 組隨機三元組的測試證實了這個因式分解（factorisation），並證實每當強制 $A\pm B\pm C$ 之一為 $\pi$ 的倍數時，$E=0$ 的誤差在 $10^{-14}$ 以內。

## 常見錯誤

- 只處理 $A+B+C=\pi$（三角形）的情況。題目容許四種符號組合中的任何一種，以及 $\pi$ 的任何倍數；$\sin^2$ 的偶性和以 $\pi$ 為週期的性質可一併處理所有情況。
- 用公式解二次方程，結果被 $\sqrt{xy(1-x)(1-y)}$ 纏住。驗證所猜測的根的和與積，就能完全避開開方和正負號的煩惱。

## STEP 之外

式子 $x^2+y^2+z^2-2yz-2zx-2xy$ 在數學中隨處可見：把 $x,y,z$ 換成 $a^2,b^2,c^2$，它就是邊長為 $a$、$b$、$c$ 的三角形面積平方的 $-16$ 倍（海倫公式（Heron's formula）），它也是一個類似判別式（discriminant）的量，決定三條長度能否構成三角形。把對稱多項式（symmetric polynomial）看成其中一個變量的二次式，並猜出它的根，是一種在代數及不變量理論（invariant theory）中反覆出現的技巧。想做更多同類題目，可參閱 Siklos 的 *Advanced Problems in Mathematics*（《高等數學問題》）。

詳細解答

# STEP II 1987，第 3 題

純數學 • 共軛根式

> **STEP II 1987，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/87-s2-q3

## 關鍵思路

設 $\alpha = a + \sqrt{a^2 - b}$ 及 $\beta = a - \sqrt{a^2 - b}$。兩者互為共軛（conjugate）：它們的和 $2a$ 與積 $b$ 都是整數，因此 $\alpha^n + \beta^n$ 必定是整數。若 $b$ 所滿足的條件迫使較小的共軛數 $\beta$ 嚴格介乎 0 與 1 之間，那麼 $\beta^n$ 正是我們所需的小數部分（fractional part），其餘結論便隨之而來。

## 解答

在整個解答中，設

$$\alpha = a + \sqrt{a^2 - b}, \qquad \beta = a - \sqrt{a^2 - b}, \qquad \text{所以} \qquad \alpha + \beta = 2a, \quad \alpha\beta = a^2 - (a^2 - b) = b.$$

**$\beta$ 的位置。** 首先，這個平方根是實數。由於 $b < 2a - 1$，

$$a^2 - b > a^2 - 2a + 1 = (a - 1)^2 \geq 0.$$

對這條嚴格不等式（inequality）兩邊取正平方根，得 $\sqrt{a^2 - b} > a - 1$。另外 $b > 0$，所以 $a^2 - b < a^2$，從而 $\sqrt{a^2 - b} < a$。綜合兩者，

$$a - 1 < \sqrt{a^2 - b} < a \quad \Longrightarrow \quad 0 < \beta < 1.$$

（這個條件是最佳的：若 $b = 2a - 1$，則 $\sqrt{a^2 - b} = a - 1$，而 $\beta = 1$ 恰好成立。）

### 第 (i) 部分

由於 $0 < \beta < 1$，對每個正整數 $n$ 都有 $0 < \beta^n < 1$。現在 $M$ 是整數，且 $\beta^n = M + s$，其中 $0 \le s < 1$，所以 $M$ 是 $\beta^n$ 的整數部分（integer part），即 0。因此

$$M = 0 \qquad 及 \qquad s = \beta^n.$$

### 第 (ii) 部分

設 $S_n = \alpha^n + \beta^n$。我們斷言 $S_n$ 是整數。設 $d = \sqrt{a^2 - b}$，用二項式定理（binomial theorem）展開兩個冪，

$$S_n = \sum_{k=0}^{n} \binom{n}{k} a^{n-k}\left(d^k + (-d)^k\right) = 2 \sum_{k\ 偶數} \binom{n}{k} a^{n-k}(a^2 - b)^{k/2},$$

因為 $k$ 為奇數的項互相抵消。餘下的每一項都是整數，所以 $S_n \in \mathbb{Z}$。（等價地，$\alpha, \beta$ 是 $t^2 - 2at + b = 0$ 的根，所以 $S_n = 2aS_{n-1} - bS_{n-2}$，其中 $S_0 = 2$ 及 $S_1 = 2a$。）

因此

$$\alpha^n = S_n - \beta^n, \qquad S_n \in \mathbb{Z}, \quad 0 < \beta^n < 1.$$

? **想一想** 為甚麼這就迫使 $N = S_n$ 及 $r = \beta^n$？

把 $\alpha^n = N - r$ 表示成 $N$ 為整數且 $0 \le r < 1$ 的形式，這種表示是唯一的：$N$ 必須是不小於 $\alpha^n$ 的最小整數。若 $N' - r' = N - r$，且 $r, r' \in [0, 1)$，則 $N - N' = r - r'$ 是一個嚴格介乎 $-1$ 與 1 之間的整數，所以它等於 0。我們的表達式 $S_n - \beta^n$ 正是這種形式，所以

$$N = S_n = \alpha^n + \beta^n, \qquad r = \beta^n = s.$$

### 第 (iii) 部分

利用 $r = \beta^n$、$N = \alpha^n + \beta^n$ 及 $(\alpha\beta)^n = b^n$，

$$\begin{aligned} r^2 - Nr + b^n &= \beta^{2n} - (\alpha^n + \beta^n)\beta^n + \alpha^n\beta^n \\ &= \beta^{2n} - \alpha^n\beta^n - \beta^{2n} + \alpha^n\beta^n \\ &= 0, \end{aligned}$$

得證。換句話說，$r$ 是 $t^2 - Nt + b^n = 0$ 較小的根，而這方程的兩根是 $\alpha^n$ 和 $\beta^n$。

### 最後部分

取 $a = 8$ 及 $a^2 - b = 63$，所以 $b = 1$，且 $\sqrt{63} = 3\sqrt{7}$。條件 $b < 2a - 1 = 15$ 成立，所以以上所有結果都適用於 $\alpha = 8 + 3\sqrt{7}$：

$$\left(8 + 3\sqrt{7}\right)^n = N - r, \qquad N \in \mathbb{Z},$$

而 $\alpha^n$ 比整數 $N$ 小 $r$。在第 (iii) 部分中取 $b=1$，得 $r(N-r)=1$，即

$$r=\frac{1}{N-r}=\frac{1}{\left(8+3\sqrt{7}\right)^n}=\left(8-3\sqrt{7}\right)^n.$$

特別地，$r\neq 0$，所以這個冪本身永遠不是整數，而且 $r\to 0$。現在 $3\sqrt{7}=\sqrt{63}$ 略小於 $\sqrt{64}=8$，所以

$$8+3\sqrt{7}\approx 16=2^4,\qquad \text{因此}\qquad r=\left(8+3\sqrt{7}\right)^{-n}\approx 16^{-n}=2^{-4n}.$$

所以當 $n$ 很大時，$\left(8+3\sqrt{7}\right)^n$ 與某個整數相差大約 $2^{-4n}$，並且略小於該整數。

這個「大約」有多準確？由於 $8+3\sqrt{7}\approx 15.937$，我們有 $r=2^{-cn}$，其中 $c=\log_2(8+3\sqrt{7})\approx 3.994$。這個指數與 $4n$ 的差距在 0.2% 以內，估算正是在這個意義下成立。比值 $r/2^{-4n}=\left(16/(8+3\sqrt{7})\right)^n$ 緩慢上升，由 $n=1$ 時約 1.004 升至 $n=10$ 時的 1.04。

## 驗算

- $n=1$：$8+3\sqrt{7}\approx 15.93725$，即 $16-0.06275$，而 $8-3\sqrt{7}\approx 0.06275$，與預測相符。此外，$r^2-16r+1=0$ 較小的根是 $8-\sqrt{63}$。
- $n=2$：$(8+3\sqrt{7})^2=127+48\sqrt{7}\approx 253.99606$，所以 $N=254=S_2=2\cdot 8\cdot 16-2$，而 $r\approx 0.003937$，對比 $2^{-8}\approx 0.003906$。
- 以電腦對數以千計滿足 $b<2a-1$ 的隨機數對 $(a,b)$ 及 $n\leq 7$ 進行驗算，確認了 $0<\beta^n<1$、$N-\alpha^n=\beta^n$ 及 $r^2-Nr+b^n=0$。對 $a=8$、$b=1$、$n\leq 10$ 進行的高精度驗算確認了 $N-\alpha^n=(8-3\sqrt{7})^n$；當 $n=10$ 時，此值為 $9.46\times 10^{-13}$，對比 $2^{-40}\approx 9.09\times 10^{-13}$。

## 常見錯誤

- 只證明了 $\beta<1$，卻忘記 $\beta>0$（這需要 $b>0$）。兩者都是 $M=0$ 所必需的。
- 斷言 $N=\alpha^n+\beta^n$，卻沒有說明為甚麼 $N-r$ 中的整數部分與小數部分是唯一的。
- 以為 $r/2^{-4n}\to 1$。事實並非如此：近似只是在指數上成立，因為 $8+3\sqrt{7}$ 接近 16，但不等於 16。

## STEP 之外

數 $\alpha^n+\beta^n$ 是跡（trace）的一個例子：$\alpha$ 是一個代數整數（algebraic integer），而它的共軛數位於單位區間之內，所以它的冪以指數速度趨近整數。這類數稱為 Pisot 數（Pisot

numbers）；黃金比例（golden ratio）是另一個例子，這就是為甚麼 $\varphi^n$ 接近盧卡斯數（Lucas number）$L_n$。這裏的情況亦與佩爾方程（Pell's equation）密切相關：$8^2 - 7 \cdot 3^2 = 1$，所以 $(8,3)$ 是 $x^2 - 7y^2 = 1$ 的最小正整數解。Davenport 的 *The Higher Arithmetic*（《高等算術》）對佩爾方程及連分數（continued fractions）有精彩的論述。

詳細解答

# STEP II 1987，第 4 題

純數學 • 複數與幾何

> **STEP II 1987，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/87-s2-q4

## 關鍵思路

乘以 $e^{i\theta}$ 就是一個旋轉（rotation）。把三角形的一條邊繞某個頂點旋轉 $60°$，得到的正是等邊三角形（equilateral triangle）。這樣，幾何條件就化為關於 $\alpha, \beta, \gamma$ 的線性方程（linear equation）。若容許兩種定向（orientation），條件就變成兩個線性表達式中有一個為零；兩者之積正是題目中的對稱（symmetric）二次式。由於它是對稱的，所以可以用三次方程（cubic）的係數（coefficients）來表示。

## 解答

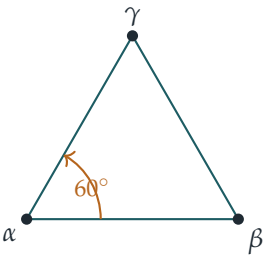
全文中設 $\omega = e^{2\pi i/3}$，於是 $\omega^3 = 1$、$\bar{\omega} = \omega^2$ 及 $1 + \omega + \omega^2 = 0$。

### 乘以 $e^{i\theta}$

若 $z = re^{i\phi}$，則 $ze^{i\theta} = re^{i(\phi+\theta)}$：模（modulus）不變，輻角（argument）增加 $\theta$。因此，點 $ze^{i\theta}$ 就是點 $z$ 繞原點逆時針（anticlockwise）旋轉角 $\theta$ 後的像。

更一般地，$\gamma - \alpha = e^{i\theta}(\beta - \alpha)$ 表示由 $\alpha$ 到 $\gamma$ 的向量，等於由 $\alpha$ 到 $\beta$ 的向量逆時針旋轉 $\theta$ 所得；換言之，$\gamma$ 是 $\beta$ 繞 $\alpha$ 旋轉 $\theta$ 後的像。

### 逆時針方向的等邊三角形




相異的點 $\alpha, \beta, \gamma$ 按逆時針次序構成等邊三角形，當且僅當 $\gamma$ 是 $\beta$ 繞 $\alpha$ 逆時針旋轉 $\frac{\pi}{3}$ 所得，即

$$\gamma - \alpha = \mathrm{e}^{\mathrm{i}\pi/3}(\beta - \alpha).$$

（必要性（necessity）：在這樣的三角形中，$|\alpha\gamma| = |\alpha\beta|$，而且在 $\alpha$ 處的角為 $\frac{\pi}{3}$，由 $\alpha\beta$ 逆時針量度至 $\alpha\gamma$。充分性（sufficiency）：該方程給出 $|\alpha\gamma| = |\alpha\beta|$，兩者夾角為 $\frac{\pi}{3}$，所以三角形是頂角為 60° 的等腰（isosceles）三角形，因而是等邊三角形，而且定向是逆時針的。）

把條件寫成對稱的形式會更簡潔。由於 $\mathrm{e}^{\mathrm{i}\pi/3} = -\omega^2$（因為 $-\omega^2 = \mathrm{e}^{\mathrm{i}\pi}\mathrm{e}^{4\pi\mathrm{i}/3} = \mathrm{e}^{\mathrm{i}\pi/3}$，這裏用了 $\mathrm{e}^{7\pi\mathrm{i}/3} = \mathrm{e}^{\mathrm{i}\pi/3}$），條件變為

$$\begin{aligned} & \gamma - \alpha = -\omega^2\beta + \omega^2\alpha \\ \iff \quad & \gamma + \omega^2\beta - (1+\omega^2)\alpha = 0 \\ \iff \quad & \gamma + \omega^2\beta + \omega\alpha = 0, \end{aligned}$$

其中用了 $1 + \omega^2 = -\omega$。兩邊乘以 $\omega^2$（它不為零，所以這一步可逆），得到等價條件

$$\alpha + \omega\beta + \omega^2\gamma = 0.$$

簡單驗證一下：$1, \omega, \omega^2$ 是一個等邊三角形按逆時針次序的頂點，而 $1 + \omega \cdot \omega + \omega^2 \cdot \omega^2 = 1 + \omega^2 + \omega = 0$。

### 任一定向

任何三個構成等邊三角形的相異點，若不是按 $\alpha, \beta, \gamma$ 的次序為逆時針，就是按 $\alpha, \gamma, \beta$ 的次序為逆時針（循環重排不會改變定向）。由上一部分，對相異的 $\alpha, \beta, \gamma$，

$$\begin{aligned} \text{等邊} & \iff \alpha + \omega\beta + \omega^2\gamma = 0 \text{ 或 } \alpha + \omega\gamma + \omega^2\beta = 0 \\ & \iff (\alpha + \omega\beta + \omega^2\gamma)(\alpha + \omega^2\beta + \omega\gamma) = 0. \end{aligned}$$

**想一想** 怎樣看出這個乘積正是我們想要的表達式？

把它展開就可以了。平方項給出 $\alpha^2+\omega^3\beta^2+\omega^3\gamma^2=\alpha^2+\beta^2+\gamma^2$。交叉項為

$$\begin{aligned}\alpha\beta(\omega^2+\omega)+\alpha\gamma(\omega+\omega^2)+\beta\gamma(\omega^2+\omega^4) &= \alpha\beta(\omega+\omega^2)+\gamma\alpha(\omega+\omega^2)+\beta\gamma(\omega^2+\omega)\\ &= -\alpha\beta-\gamma\alpha-\beta\gamma,\end{aligned}$$

其中用了 $\omega^4=\omega$ 及 $\omega+\omega^2=-1$。因此

$$(\alpha+\omega\beta+\omega^2\gamma)(\alpha+\omega^2\beta+\omega\gamma)=\alpha^2+\beta^2+\gamma^2-\beta\gamma-\gamma\alpha-\alpha\beta,$$

所以相異的點 $\alpha,\beta,\gamma$（按某個次序）構成等邊三角形，當且僅當

$$\alpha^2+\beta^2+\gamma^2-\beta\gamma-\gamma\alpha-\alpha\beta=0,$$

得證。

值得留意的是，若點不一定相異，情況會怎樣：例如若 $\alpha=\beta$，表達式變為 $(\alpha-\gamma)^2$，只有當 $\gamma=\alpha$ 時才為零。所以表達式為零，恰好是在三點構成等邊三角形或三點重合的時候。下文會用到這一點。

### 三次方程

設 $z^3+az^2+bz+c=0$ 的根為 $\alpha,\beta,\gamma$。則

$$\alpha+\beta+\gamma=-a,\qquad \beta\gamma+\gamma\alpha+\alpha\beta=b,\qquad \alpha\beta\gamma=-c.$$

該表達式是對稱的，所以可以用這些量來表示：

$$\begin{aligned}\alpha^2+\beta^2+\gamma^2-(\beta\gamma+\gamma\alpha+\alpha\beta) &= (\alpha+\beta+\gamma)^2-3(\beta\gamma+\gamma\alpha+\alpha\beta)\\ &= a^2-3b.\end{aligned}$$

根據上面的說明，$a^2=3b$ 恰好在各根構成等邊三角形或全部相等時成立。三個根都等於 $t$，即 $z^3+az^2+bz+c=(z-t)^3$，所以 $a=-3t$、$b=3t^2$、$c=-t^3$；也就是 $b=\frac{a^2}{3}$ 及 $c=\frac{a^3}{27}$。因此，各根位於一個（真正的）等邊三角形的頂點，當且僅當

$$\boxed{a^2=3b\quad\text{且}\quad 27c\neq a^3}$$

若願意把三點重合視為退化的等邊三角形，條件就只是 $a^2=3b$。

有一個漂亮的方法可以直接看出這一點。代入 $z=w-\frac{a}{3}$ 可消去 $z^2$ 項；當 $a^2=3b$ 時，一次項亦會消失，方程變為 $w^3=\frac{a^3}{27}-c$。它的根是一個非零數的三個立方根，構成以 $w=0$ 為中心的等邊三角形，即以 $z=-\frac{a}{3}$ 為中心，而這正是各根的形心（centroid）。

## 驗算

- $z^3 - 1 = 0$（$a = b = 0$，$c = -1$）的根為 $1, \omega, \omega^2$：構成等邊三角形，而確實 $a^2 = 0 = 3b$ 且 $27c = -27 \neq 0$。
- $z^3 - z = 0$ 的根為 $0, \pm 1$，三點共線（collinear）；此時 $a^2 - 3b = 3 \neq 0$。
- 以數值方法，對 1000 個隨機三角形 $\alpha$、$\beta$、$\gamma = \alpha + e^{i\pi/3}(\beta - \alpha)$，我們驗證了三角形為逆時針方向的等邊三角形，$\alpha + \omega\beta + \omega^2\gamma$ 及該二次表達式（在每種次序下）都為零，而且對隨機複數，乘積恆等式及 $a^2 - 3b$ 公式在 $10^{-8}$ 的誤差內成立。

## 常見錯誤

- 弄錯正負號或定向：$e^{i\pi/3}$ 是 $-\omega^2$，而不是 $\omega$；繞一個頂點乘以 $\omega$ 旋轉（即旋轉 120°）並不會得到等邊三角形。
- 忘記三重根的情況：單憑 $a^2 = 3b$ 亦容許 $z^3 + az^2 + bz + c = (z + \frac{a}{3})^3$，此時三個「頂點」重合。

## STEP 之外

條件 $\alpha + \omega\beta + \omega^2\gamma = 0$ 是三元組 $(\alpha, \beta, \gamma)$ 的離散傅立葉變換（discrete Fourier transform）的一個分量；而消去 $z^2$ 項的技巧，是以根式解（solving by radicals）三次方程的第一步，由此可通往伽羅瓦理論（Galois theory）。從幾何角度看，這個結果表明：各根構成等邊三角形，恰好是導數（derivative）$3z^2 + 2az + b$ 在形心處有二重根（double root）的時候；這是關於導數之根的馬登定理（Marden's theorem）的一個特例。Needham 的 *Visual Complex Analysis*（《可視化複分析》）是進一步探索複數乘法幾何意義的絕佳讀物。

詳細解答

# STEP II 1987，第 5 題

純數學 • 拉格朗日反演級數

> **STEP II 1987，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/87-s2-q5

## 關鍵思路

拉格朗日(Lagrange)恆等式(Lagrange's identity)就像一部機器：把函數 $\phi(y) = y - \mathrm{f}(y)$ 放進去，取其 $n$ 次方，求導 $n-1$ 次，再把結果加起來。所有功夫都在於把問題設定好，使 $\phi$ 盡量簡單，最好是 $y$ 的單一冪次或單一指數函數，因為這樣 $\phi^n$ 的 $(n-1)$ 階導數（derivative）就能一下子寫出來。在每一部分中，我們都選取適當的 f，使所要解的方程寫成 $\mathrm{f}(x) = y$ 的形式，其中 $y$ 取一個方便的值。

## 解答

全文設 $\phi(y) = y - \mathrm{f}(y)$，於是恆等式可寫成

$$\mathrm{f}^{-1}(y) = y + \sum_{n=1}^{\infty} \frac{1}{n!} \frac{\mathrm{d}^{n-1}}{\mathrm{d}y^{n-1}} \left[\phi(y)^n\right].$$

注意 $n = 1$ 那一項涉及的是零階導數，也就是 $\phi(y)$ 本身。

我們會反覆用到一個事實：對整數 $m \geq n-1$，

$$\frac{\mathrm{d}^{n-1}}{\mathrm{d}y^{n-1}} y^m = \frac{m!}{(m-n+1)!} y^{m-n+1},$$

因為每求導一次，就把當時的冪次乘到前面，並把冪次減一。

### 第 (i) 部分

取 $\mathrm{f}(x) = \alpha x$，則 $\phi(y) = (1-\alpha)y$，所以 $\phi(y)^n = (1-\alpha)^n y^n$。在上式中取 $m = n$，求導 $n-1$ 次得

$$\frac{\mathrm{d}^{n-1}}{\mathrm{d}y^{n-1}} \left[(1-\alpha)^n y^n\right] = (1-\alpha)^n \, n! \, y.$$

$n!$ 與 $\frac{1}{n!}$ 互相抵消，所以恆等式的右方為

$$y + \sum_{n=1}^{\infty} (1-\alpha)^n y = y \sum_{n=0}^{\infty} (1-\alpha)^n.$$

這是一個公比為 $1-\alpha$ 的等比級數（geometric series）。由於 $0 < \alpha < 2$，有 $-1 < 1-\alpha < 1$，因此級數收斂（converges），其和為

$$y \cdot \frac{1}{1-(1-\alpha)} = \frac{y}{\alpha}.$$

另一方面，由 $y = \alpha x$ 直接可得 $x = y/\alpha$，所以 $\mathrm{f}^{-1}(y) = y/\alpha$。兩者一致，於是恆等式在這情況下得到驗證。條件 $0 < \alpha < 2$ 正正就是級數收斂的條件。

### 第 (ii) 部分

方程為 $\mathrm{f}(x) = \frac{1}{2}$，其中

$$\mathrm{f}(x) = x - \tfrac{1}{4}x^3, \qquad \text{所以} \qquad \phi(y) = y - \mathrm{f}(y) = \tfrac{1}{4}y^3.$$

於是 $\phi(y)^n = 4^{-n} y^{3n}$，在導數公式中取 $m = 3n$（它不小於 $n-1$），得

$$\frac{1}{n!} \frac{\mathrm{d}^{n-1}}{\mathrm{d}y^{n-1}} \left[ 4^{-n} y^{3n} \right] = \frac{(3n)!}{n!\,(2n+1)!\,4^n} y^{2n+1}.$$

因此拉格朗日恆等式給出

$$\mathrm{f}^{-1}(y) = y + \sum_{n=1}^{\infty} \frac{(3n)!}{n!\,(2n+1)!\,4^n} y^{2n+1}.$$

開頭的 $y$ 恰好就是通項在 $n = 0$ 時的值，因為 $\frac{0!}{0!\,1!\,4^0} y^1 = y$。所以可以把它併入總和，由 $n = 0$ 開始：

$$\mathrm{f}^{-1}(y) = \sum_{n=0}^{\infty} \frac{(3n)!}{n!\,(2n+1)!\,4^n} y^{2n+1}.$$

現在代入 $y = \frac{1}{2}$。由於 $4^n \cdot 2^{2n+1} = 2^{2n} \cdot 2^{2n+1} = 2^{4n+1}$，得

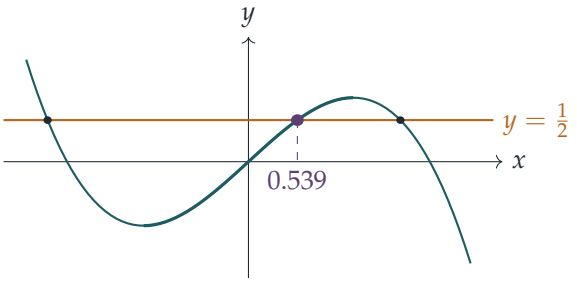
$$x = \mathrm{f}^{-1}\left(\tfrac{1}{2}\right) = \sum_{n=0}^{\infty} \frac{(3n)!}{n!\,(2n+1)!\,2^{4n+1}},$$

又因 $\mathrm{f}(x) = \frac{1}{2}$，這個 $x$ 是 $\frac{1}{2} = x - \frac{1}{4}x^3$ 的一個根，得證。

**想一想** 這個三次方程有三個實根。級數給出的是哪一個？

函數 f 在整條實數線上並非一一對應（one-to-one），所以「$\mathrm{f}^{-1}$」指的是級數所描述的那一支。這個級數是 $y$ 的冪級數（power series），以 $y + \cdots$ 開始，所以它給出 $\mathrm{f}^{-1}(0) = 0$，而且當 $y$ 很小時 $\mathrm{f}^{-1}(y) \approx y$。它是 f 限制在通過原點的遞增部分上的反函數，在那裏

$\mathrm{f}'(x) = 1 - \frac{3}{4}x^2 > 0$，即 $|x| < 2/\sqrt{3}$。首幾項為 $\frac{1}{2} + \frac{1}{32} + \frac{3}{512} + \cdots \approx 0.537$，而總和準確至四位小數為 0.5392：這是最小的正根，已在下圖標出。另外兩個根約為 1.675 和 $-2.214$，位於曲線的遞減部分。




曲線上較粗的一段是通過原點的遞增分支；級數給出的就是這分支上的根，以梅紅色標示。

**第 (iii) 部分**

這裏的竅門是把整條方程化成 $\mathrm{f}(x) = y$ 的形式，並使 $\phi$ 簡單。把 $x = \mathrm{e}^{\lambda x}$ 改寫為

$$x - \mathrm{e}^{\lambda x} = 0,$$

我們取

$$\mathrm{f}(x) = x - \mathrm{e}^{\lambda x}, \qquad \phi(y) = y - \mathrm{f}(y) = \mathrm{e}^{\lambda y},$$

而所求的是 $x = \mathrm{f}^{-1}(0)$。此時 $\phi(y)^n = \mathrm{e}^{n\lambda y}$，每求導一次就乘以 $n\lambda$，所以

$$\frac{\mathrm{d}^{n-1}}{\mathrm{d}y^{n-1}}\,\mathrm{e}^{n\lambda y} = (n\lambda)^{n-1}\mathrm{e}^{n\lambda y}.$$

（當 $n = 1$ 時，這給出 $\mathrm{e}^{\lambda y}$，正確無誤。）拉格朗日恆等式給出

$$\mathrm{f}^{-1}(y) = y + \sum_{n=1}^{\infty} \frac{(n\lambda)^{n-1}}{n!}\,\mathrm{e}^{n\lambda y},$$

代入 $y = 0$，

$$x = \sum_{n=1}^{\infty} \frac{n^{n-1}}{n!}\,\lambda^{n-1} = 1 + \lambda + \frac{3}{2}\lambda^2 + \frac{8}{3}\lambda^3 + \frac{125}{24}\lambda^4 + \cdots.$$

等價地，把指標平移，可得 $x = \sum_{n=0}^{\infty} \frac{(n+1)^{n-1}}{n!}\lambda^n$。

這個級數在 $\lambda$ 取甚麼值時收斂？相鄰兩項之比為

$$\frac{(n+1)^n\lambda^n/(n+1)!}{n^{n-1}\lambda^{n-1}/n!} = \lambda\left(1 + \frac{1}{n}\right)^{n-1} \to \mathrm{e}\lambda,$$

所以由比值檢驗法（ratio test），級數在 $|\lambda| < 1/\mathrm{e}$ 時收斂。這是一個自然的界線：$y = x$ 與 $y = \mathrm{e}^{\lambda x}$ 相切，恰好發生在 $\lambda = 1/\mathrm{e}$ 時（切點在 $x = \mathrm{e}$），而當 $\lambda > 1/\mathrm{e}$ 時根本沒有實根。當 $0 < \lambda < 1/\mathrm{e}$ 時有兩個正根，級數給出較小的那一個，即當 $\lambda \to 0$ 時趨向 1 的那個根。

## 驗算

- 在 (ii) 中，首四項為 $\frac{1}{2}$、$\frac{1}{32}$、$\frac{3}{512}$ 和 $\frac{3}{2048}$。以數值方法加總 200 項得 0.539189，而對 $x^3 - 4x + 2 = 0$ 使用牛頓法（Newton's method）得根 0.5391888…，兩者在機器精度內一致。
- 在 (iii) 中，代入 $\lambda = 0$：級數給出 $x = 1$，而確實 $x = \mathrm{e}^0 = 1$。取 $\lambda = 0.2$ 加總級數得 $x \approx 1.29586$，而 $\mathrm{e}^{0.2 \times 1.29586} = 1.29586$。取 $\lambda = -0.3$ 則得 0.78918，同樣滿足方程。當 $\lambda = 0.2$ 時，另一個根約為 12.71，與級數的值相距甚遠。
- 在 (i) 中，試取 $\alpha = 1$：此時 $\phi = 0$，總和中每一項都是零，於是 $\mathrm{f}^{-1}(y) = y$，理應如此。

## 常見錯誤

- 忘記 $n = 1$ 那一項不用求導（即零階導數），因而漏掉或弄錯這一項。
- 在 (ii) 中，沒有察覺單獨的 $y$ 正是通項公式在 $n = 0$ 時的項，因而未能湊成由 $n = 0$ 開始的總和。
- 在 (iii) 中，嘗試取 $\mathrm{f}(x) = x\mathrm{e}^{-\lambda x}$ 之類的函數。這樣做並無不可，但 $\phi$ 便成了兩項之差，導數會變得一團糟。選取 $\phi = \mathrm{e}^{\lambda y}$ 則令導數變得輕而易舉。

## STEP 之外

這就是*拉格朗日反演定理*（Lagrange inversion theorem），通常用複分析（complex analysis）（柯西（Cauchy）留數定理）或形式冪級數（formal power series）來證明。(iii) 中的係數 $(n+1)^{n-1}/n!$ 屬於 Lambert $W$ 函數，而 $n^{n-1}$ 是 $n$ 個頂點上的有根標號樹（rooted labelled trees）的數目。同樣地，(ii) 中的數 $\frac{(3n)!}{n!\,(2n+1)!}$ 計算的是三叉樹（ternary trees）的數目，與卡特蘭數（Catalan numbers）是近親。因此，拉格朗日反演是枚舉組合學（enumerative combinatorics）的核心工具。Needham 的 *Visual Complex Analysis*（《直觀複分析》）提供複分析方面的背景，而 Siklos 的 *Advanced Problems in Mathematics*（《高等數學問題》）則有更多關於級數的 STEP 練習。

詳細解答

# STEP II 1987，第 6 題

純數學 • 利用對稱性求定積分

> **STEP II 1987，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/87-s2-q6

## 關鍵思路

區間 $[-\frac{1}{2}\pi, \frac{1}{2}\pi]$ 是對稱的，所以反射 $\theta \mapsto -\theta$ 是自然的第一步：它把 $\sin\theta$ 變成 $-\sin\theta$，第二個形式便不費吹灰之力得到。把兩個形式取平均，可消去奇部分（odd part），剩下分母只含 $\sin^2\theta$，再稍作代數處理，就能拆成一個常數加上一個標準的 $\tan\theta$ 積分（integral）。最後一部分的重點在於一個看似無害的步驟：$\sqrt{\cos^2 2\alpha}$ 是 $|\cos 2\alpha|$，而不是 $\cos 2\alpha$。

## 解答

全文記 $s = \sin 2\alpha$ 及 $c = \cos 2\alpha$，因此 $s^2 + c^2 = 1$。當 $0 < \alpha < \frac{1}{4}\pi$ 時，有 $0 < s < 1$ 及 $0 < c < 1$。由於 $s < 1$，分母 $1 \mp s\sin\theta$ 在整個區間上都是正數，所以所有積分都是常義積分（proper integral）。

### $I$ 的第二個形式

代入 $\theta = -\phi$，於是 $\mathrm{d}\theta = -\mathrm{d}\phi$，而上下限互換：

$$I = \int_{\frac{1}{2}\pi}^{-\frac{1}{2}\pi} \frac{\cos^2(-\phi)}{1 - \sin(-\phi)\,s}\,(-\mathrm{d}\phi) = \int_{-\frac{1}{2}\pi}^{\frac{1}{2}\pi} \frac{\cos^2\phi}{1 + s\sin\phi}\,\mathrm{d}\phi,$$

這裏用了 $\cos(-\phi) = \cos\phi$ 及 $\sin(-\phi) = -\sin\phi$。把 $\phi$ 改名為 $\theta$，即得所求形式。

### 含 $\cot^2 2\alpha$ 的表達式

把 $I$ 的兩個表達式相加，

$$2I = \int_{-\frac{1}{2}\pi}^{\frac{1}{2}\pi} \cos^2\theta \left( \frac{1}{1 - s\sin\theta} + \frac{1}{1 + s\sin\theta} \right) \mathrm{d}\theta = \int_{-\frac{1}{2}\pi}^{\frac{1}{2}\pi} \frac{2\cos^2\theta}{1 - s^2\sin^2\theta}\,\mathrm{d}\theta,$$

所以

$$I = \int_{-\frac{1}{2}\pi}^{\frac{1}{2}\pi} \frac{\cos^2\theta}{1 - s^2\sin^2\theta}\,\mathrm{d}\theta.$$

現在利用 $1 = \cos^2\theta + \sin^2\theta$ 及 $1 - s^2 = c^2$ 改寫分母：

$$1 - s^2\sin^2\theta = \cos^2\theta + c^2\sin^2\theta.$$

? **想一想** 怎樣把 $\dfrac{\cos^2\theta}{\cos^2\theta + c^2\sin^2\theta}$ 拆成一個常數加上一個我們能積分的部分？

目標是把分子寫成分母的倍數加上一個餘項，而這個餘項除以 $\cos^2\theta$ 之後不再依賴 $\theta$。由於

$$(\cos^2\theta + c^2\sin^2\theta) - c^2(\cos^2\theta + \sin^2\theta) = (1 - c^2)\cos^2\theta = s^2\cos^2\theta,$$

兩邊除以 $s^2(\cos^2\theta + c^2\sin^2\theta)$（因為 $s \neq 0$，這是容許的），得

$$\frac{\cos^2\theta}{\cos^2\theta + c^2\sin^2\theta} = \frac{1}{s^2} - \frac{c^2}{s^2}\cdot\frac{1}{\cos^2\theta + c^2\sin^2\theta} = \frac{1}{s^2} - \frac{c^2}{s^2}\cdot\frac{\sec^2\theta}{1 + c^2\tan^2\theta},$$

其中最後一步把分子分母同除以 $\cos^2\theta$。在長度為 $\pi$ 的區間上積分，並利用 $c^2/s^2 = \cot^2 2\alpha$，得

$$I = \frac{\pi}{\sin^2 2\alpha} - \cot^2 2\alpha \int_{-\frac{1}{2}\pi}^{\frac{1}{2}\pi} \frac{\sec^2\theta}{1 + \cos^2 2\alpha\tan^2\theta}\,\mathrm{d}\theta,$$

此即所求。

### 求 $I$ 的值

把餘下的積分記為 $J$。代入 $t = \tan\theta$，於是 $\mathrm{d}t = \sec^2\theta\,\mathrm{d}\theta$；當 $\theta$ 由 $-\frac{1}{2}\pi$ 走到 $\frac{1}{2}\pi$ 時，$t$ 走遍整條實數線。於是，當 $c \neq 0$ 時，

$$J = \int_{-\infty}^{\infty} \frac{\mathrm{d}t}{1 + c^2t^2} = \left[ \frac{1}{|c|}\arctan\big(|c|\,t\big) \right]_{-\infty}^{\infty} = \frac{\pi}{|c|}.$$

（$|c|$ 正是由代換 $u = |c|t$ 產生的：它使上下限仍按正確次序由 $-\infty$ 到 $\infty$。）

當 $0 < \alpha < \frac{1}{4}\pi$ 時，$c > 0$，所以 $J = \pi/c$，而

$$I = \frac{\pi}{s^2} - \frac{c^2}{s^2}\cdot\frac{\pi}{c} = \frac{\pi(1 - c)}{1 - c^2} = \frac{\pi}{1 + c} = \frac{\pi}{1 + \cos 2\alpha} = \frac{\pi}{2\cos^2\alpha} = \tfrac{1}{2}\pi\sec^2\alpha,$$

這裏約去了 $1 - c \neq 0$，並用了 $1 + \cos 2\alpha = 2\cos^2\alpha$。這就是所求結果。

### $\frac{1}{4}\pi < \alpha < \frac{1}{2}\pi$ 的情況

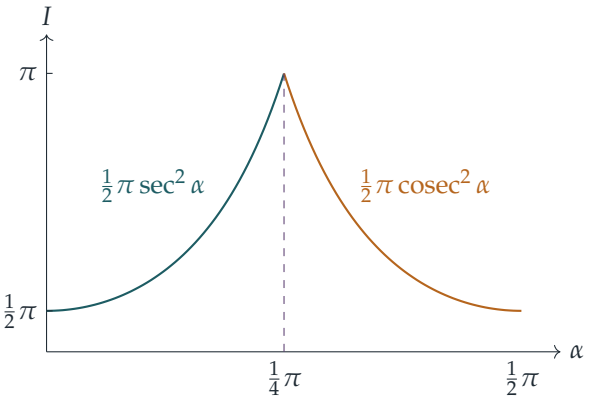
**想一想** 當 $\alpha$ 越過 $\frac{1}{4}\pi$ 時，論證中有沒有任何地方會改變？

幾乎沒有：我們仍然有 $0 < s < 1$，所以直至以 $J$ 表示 $I$ 的公式為止，每一步都照樣成立。唯一的分別是現在 $c = \cos 2\alpha < 0$，所以 $|c| = -c$，而 $J = -\pi/c$。因此

$$I = \frac{\pi}{s^2} + \frac{c^2}{s^2}\cdot\frac{\pi}{c} = \frac{\pi(1+c)}{1-c^2} = \frac{\pi}{1-c} = \frac{\pi}{2\sin^2\alpha} = \tfrac{1}{2}\pi\operatorname{cosec}^2\alpha,$$

這裏用了 $1 - \cos 2\alpha = 2\sin^2\alpha$。若盲目照搬第一個答案，便會得出錯誤的值。

$I$ 對 $\alpha$ 的圖像顯示兩條公式在 $\alpha = \frac{1}{4}\pi$ 處相接，那裏兩者都等於 $\pi$：




## 驗算

- 當 $\alpha = \frac{1}{4}\pi$ 時，$s = 1$，被積函數（integrand）為 $\cos^2\theta/(1-\sin\theta) = 1+\sin\theta$，其積分為 $\pi$。兩條公式在該處都給出 $\pi$，理應如此。
- 當 $\alpha = \frac{1}{6}\pi$ 時：$s = \frac{\sqrt{3}}{2}$，$c = \frac{1}{2}$，所以 $I = \pi/(1+c) = \frac{2}{3}\pi$，而 $\frac{1}{2}\pi\sec^2\frac{1}{6}\pi = \frac{1}{2}\pi\cdot\frac{4}{3} = \frac{2}{3}\pi$。
- 在 $\frac{1}{4}\pi$ 兩側隨機取十個 $\alpha$ 值，用辛普森法則（Simpson's rule）計算原積分，與公式的結果相差少於 $10^{-6}$；例如 $\alpha \approx 0.6308$ 時，兩種方法都得出 $I \approx 2.408759$。
- 當 $\alpha \to 0$ 時，被積函數趨向 $\cos^2\theta$，其積分為 $\frac{1}{2}\pi$，與 $\frac{1}{2}\pi\sec^2 0$ 吻合。

## 常見錯誤

- 寫成 $\int_{-\infty}^{\infty}\frac{\mathrm{d}t}{1+c^2t^2} = \frac{\pi}{c}$ 而漏了絕對值（modulus）。當 $c < 0$ 時，這是負數，而被積函數為正，這是不可能的。

- 遺漏了 $\mathrm{d}\theta = -\mathrm{d}\phi$ 中的負號，或忘記了它會因上下限互換而抵消。
- 擔心 $t = \tan\theta$ 在 $\theta = \pm\frac{1}{2}\pi$ 處無定義：關於 $t$ 的積分是反常積分（improper integral），但它收斂（converges），而這已足夠。

## STEP 之外

答案中整齊的對稱 $\alpha \leftrightarrow \frac{1}{2}\pi - \alpha$ 並非巧合：$I$ 只透過 $\sin 2\alpha$ 依賴 $\alpha$，而它在這個互換下不變；但像 $\frac{1}{2}\pi\sec^2\alpha$ 這樣的公式卻會改變。默默地平方再開方的代數運算，正好會掩蓋這類問題。$\sin\theta$ 和 $\cos\theta$ 的有理函數在一個完整週期上的積分（例如我們取平均後的被積函數，其週期為 $\pi$），在大學裏通常用沿單位圓的圍道積分（contour integration）來計算，而留數定理（residue theorem）使這類計算幾乎變得機械化。如想看一個優美而富幾何直觀的入門介紹，可參閱 Needham 的 *Visual Complex Analysis*（《直觀複分析》）。

詳細解答

# STEP II 1987，第 7 題

純數學 • 梯形法則與階乘

> **STEP II 1987，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/87-s2-q7

## 關鍵思路

以單位步長對 $\ln x$ 求梯形和（trapezium sum），結果幾乎正好是 $\ln n!$，因為所加的高度正是 $\ln 1, \ln 2, \ldots, \ln n$。由於 $\ln x$ 是凹函數（concave），每一條弦都在曲線下方，所以梯形法則（trapezium rule）總是低估積分；單憑這一個事實，就能決定每條不等式的方向。至於最後一部分，技巧是把細網格上的誤差重新縮放，使之變成另一個區間上的單位步長誤差，從而可以直接與 $\mathrm{g}(kn)$ 背後的誤差比較。

## 解答

### 說明法則的依據

把 $[x_0, x_N]$ 分成 $N$ 條狹條 $[x_r, x_{r+1}]$，每條寬度為 $h$。在每條狹條上，以連接 $(x_r, \mathrm{f}(x_r))$ 與 $(x_{r+1}, \mathrm{f}(x_{r+1}))$ 的直線（弦（chord））代替曲線。弦下方的區域是一個梯形，面積為 $\frac{1}{2}h\{\mathrm{f}(x_r) + \mathrm{f}(x_{r+1})\}$。對 $r = 0, 1, \ldots, N-1$ 求和，每個內部縱坐標（ordinate）$\mathrm{f}(x_1), \ldots, \mathrm{f}(x_{N-1})$ 都在兩個相鄰梯形中出現，而兩端的縱坐標各出現一次，於是得到題中的公式。當 $h$ 很小時，弦與曲線相當接近，所以近似效果良好。

### 單位步長與 $n! \approx \mathrm{g}(n)$

取 $u = \ln x$、$\mathrm{d}v = \mathrm{d}x$ 作分部積分（integration by parts），

$$I_n = \int_1^n \ln x \, \mathrm{d}x = \Big[x \ln x\Big]_1^n - \int_1^n 1 \, \mathrm{d}x = n \ln n - n + 1.$$

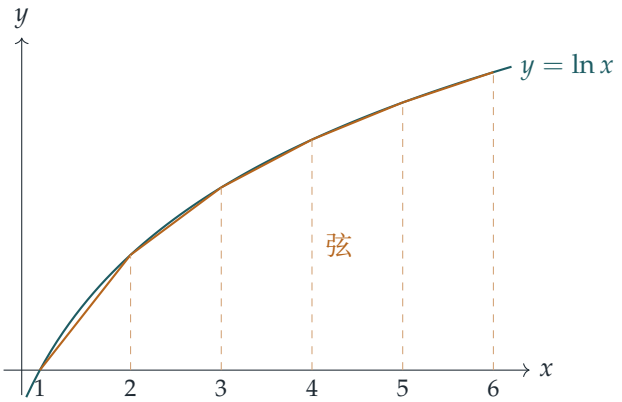
當 $h=1$ 時，縱坐標位於 $x=1,2,\ldots,n$，所以梯形和為

$$\begin{aligned} T_n &= \tfrac{1}{2}\{\ln 1 + 2\ln 2 + \cdots + 2\ln(n-1) + \ln n\} \\ &= \ln 2 + \ln 3 + \cdots + \ln n - \tfrac{1}{2}\ln n = \ln n! - \tfrac{1}{2}\ln n, \end{aligned}$$

這裏用了 $\ln 1 = 0$。令 $T_n \approx I_n$，

$$\ln n! \approx \left(n+\tfrac{1}{2}\right)\ln n + 1 - n, \qquad \text{所以} \qquad n! \approx n^{n+\frac{1}{2}}\mathrm{e}^{1-n} = \mathrm{g}(n).$$

**為何 $n! < \mathrm{g}(n)$**




由於 $\dfrac{\mathrm{d}^2}{\mathrm{d}x^2}\ln x = -\dfrac{1}{x^2} < 0$，曲線 $y=\ln x$ 是嚴格凹的：在每條狹條 $[m, m+1]$ 上，除端點外，弦都嚴格位於曲線下方（見草圖）。因此每個梯形的面積都嚴格小於該狹條上曲線下方區域的面積，把 $n-1 \geq 1$ 條狹條加起來，

$$T_n < I_n \quad \Longrightarrow \quad \ln n! - \tfrac{1}{2}\ln n < n\ln n - n + 1 \quad \Longrightarrow \quad \ln n! < \ln \mathrm{g}(n).$$

由於 exp 是遞增函數，所以 $n! < \mathrm{g}(n)$。

**寬度為 $1/k$ 的狹條**

現在 $h = 1/k$、$N = k(n-1)$，而對 $r = 0,1,\ldots,k(n-1)$，有 $x_r = 1 + \frac{r}{k} = \frac{k+r}{k}$。最後一個縱坐標是 $x_N = n$。把梯形和寫成「$h$ 乘以所有縱坐標之和，再減去兩端縱坐標各一半」，並利用 $\ln x_0 = 0$，

$$T_n^{(k)} = \frac{1}{k}\left\{\sum_{r=0}^{k(n-1)} \ln\frac{k+r}{k} - \tfrac{1}{2}\ln n\right\}.$$

分子 $k+r$ 依次取 $k, k+1, \ldots, kn$，而分母中共有 $k(n-1)+1$ 個因子 $k$，所以

$$\sum_{r=0}^{k(n-1)} \ln\frac{k+r}{k} = \ln\frac{(kn)!}{(k-1)!} - \big(k(n-1)+1\big)\ln k = \ln\frac{(kn)!}{k!} - k(n-1)\ln k,$$

因為 $(k-1)!\,k = k!$。令 $kT_n^{(k)} \approx kI_n = kn\ln n - kn + k$ 並整理，

$$\begin{aligned}\ln(kn)! &\approx \ln k! + k(n-1)\ln k + \tfrac{1}{2}\ln n + kn\ln n + k(1-n)\\ &= \ln k! + \left(kn+\tfrac{1}{2}\right)\ln n + k(1-n)(1-\ln k),\end{aligned}$$

即

$$(kn)! \approx k!\, n^{kn+\frac{1}{2}} \left(\frac{\mathrm{e}}{k}\right)^{k(1-n)} =: A,$$

正如所求。

**哪一個更接近？**

對整數 $a < b$，記誤差 (error) $\displaystyle\int_a^b \ln t\,\mathrm{d}t -$ (單位步長梯形和，區間為$[a,b]$) 為 $E(a,b)$。根據上面的凹性論證，$E(a,b) > 0$，而且相鄰區間上的誤差可以相加：$E(1,b) = E(1,a) + E(a,b)$。

推導 $A$ 的過程中，除了以 $I_n$ 代替 $T_n^{(k)}$ 之外，其餘都是精確的，所以

$$\ln A - \ln(kn)! = k\big(I_n - T_n^{(k)}\big).$$

同理，把單位步長的推導中的 $n$ 換成 $kn$，得 $\ln \mathrm{g}(kn) - \ln(kn)! = E(1,kn)$。

> ? **想一想** 寬度為 $1/k$ 的網格上的誤差，怎樣才能與寬度為 1 的網格上的誤差作比較？

代入 $x = t/k$。縱坐標 $x_r = (k+r)/k$ 變成 $t = k, k+1, \ldots, kn$，而 $\mathrm{d}x = \mathrm{d}t/k$，所以

$$k\big(I_n - T_n^{(k)}\big) = \int_k^{kn} \ln\frac{t}{k}\,\mathrm{d}t \;-\; \big(\text{單位步長梯形和，函數為}\ln\tfrac{t}{k}\text{，區間為}[k,kn]\big).$$

由於 $\ln\frac{t}{k} = \ln t - \ln k$，而梯形法則對常數 $\ln k$ 是精確的，常數項互相抵消，所以上式正是 $E(k,kn)$。因此

$$\ln \mathrm{g}(kn) - \ln(kn)! = E(1,kn) = E(1,k) + E(k,kn) = E(1,k) + \big(\ln A - \ln(kn)!\big).$$

$A$ 和 $\mathrm{g}(kn)$ 都大於 $(kn)!$，而當 $k \geq 2$ 時 $E(1,k) > 0$，所以

$$(kn)! < A < \mathrm{g}(kn) \qquad (k \geq 2).$$

所以無論以差還是以比來衡量，新的近似值 $A$ 都比 $\mathrm{g}(kn)$ 更接近 $(kn)!$。當 $k = 1$ 時兩者相同，因為 $A = 1!\,n^{n+\frac{1}{2}}\mathrm{e}^{1-n} = \mathrm{g}(n)$。換句話說，$A$ 對首 $k$ 個因子使用了*精確*值 $k!$，只對其餘部分作近似，因此避開了曲線起始處的誤差 $E(1,k)$，而那正是 ln 彎曲得最厲害的地方。

## 驗算

- 在最終公式中取 $k=1$，得回 $\mathrm{g}(n)$，理應如此。
- 取 $n=5$：$5!=120$，而 $\mathrm{g}(5)=5^{5.5}\mathrm{e}^{-4}\approx 128.0$，所以確實 $n!<\mathrm{g}(n)$。
- 取 $k=3$、$n=5$：$15!\approx 1.3077\times 10^{12}$，$A\approx 1.3369\times 10^{12}$，$\mathrm{g}(15)\approx 1.4102\times 10^{12}$，證實 $(kn)!<A<\mathrm{g}(kn)$。
- 比值 $\mathrm{g}(kn)/A=\exp E(1,k)$ 只取決於 $k$，與 $n$ 無關；當 $k=3$ 時它等於 $\mathrm{g}(3)/3!\approx 1.0548$，與上面的數值吻合。

## 常見錯誤

- 數錯寬度為 $1/k$ 的和中因子 $k$ 的個數：縱坐標共有 $k(n-1)+1$ 個，而多出來的那個 $\ln k$ 正是把 $(k-1)!$ 變成 $k!$ 的關鍵。
- 用一幅弦位於曲線上方的草圖去論證「$n!<\mathrm{g}(n)$」。對凹函數而言，弦位於曲線下方，所以梯形法則會低估積分。
- 忘記 $\ln A$ 的誤差是細網格誤差的 $k$ 倍，而不是細網格誤差本身。

## STEP 之外

這裏的 $\mathrm{g}(n)$ 是斯特靈公式（Stirling's formula）$n!\sim\sqrt{2\pi n}\,n^n\mathrm{e}^{-n}$ 的粗略形式。由於誤差 $E(1,n)$ 隨 $n$ 遞增且有界（每條狹條的誤差最多是某常數乘以 $1/m^2$），它們收斂（converge），這就解釋了為何 $n!/\mathrm{g}(n)$ 趨向常數 $\sqrt{2\pi}/\mathrm{e}\approx 0.922$。歐拉－麥克勞林公式（Euler–Maclaurin formula）把這一點精確化，並給出完整的漸近級數（asymptotic series）。Bartle 與 Sherbert 的 *Introduction to Real Analysis*（《實分析導論》）講述了黎曼積分（Riemann integral）以及支撐上述內容的數值積分法則的誤差估計。

詳細解答

# STEP II 1987，第 8 題

純數學 • 三維空間中的向量軌跡

> **STEP II 1987，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/87-s2-q8

## 關鍵思路

只要把每條方程改寫成與位置向量（position vector）為 $\mathbf{a}$ 和 $\mathbf{b}$ 的點 $A$、$B$ 之間的距離，就能認出它們各自代表甚麼。至於等價（equivalence）的部分，一切都可以只用兩個數表達：$P = |\mathbf{r}-\mathbf{a}|^2$ 及 $Q = |\mathbf{r}-\mathbf{b}|^2$。這樣，(i) 和 (ii) 就變成兩條關於 $P$ 和 $Q$ 的簡單線性方程，(iii) 和 (iv) 亦然。

## 解答

以下各處，$A$、$B$ 和 $R$ 分別是位置向量為 $\mathbf{a}$、$\mathbf{b}$ 和 $\mathbf{r}$ 的點，而 $M$ 是 $AB$ 的中點(midpoint)，其位置向量為 $\mathbf{m} = \frac{1}{2}(\mathbf{a}+\mathbf{b})$。我們假設 $\mathbf{a} \neq \mathbf{b}$（$\mathbf{a} = \mathbf{b}$ 的情況留到最後處理），並記 $d = |\mathbf{a}-\mathbf{b}| = AB$。

### 第 (i) 部分

展開，

$$\begin{aligned}|\mathbf{r}-\mathbf{a}|^2 - |\mathbf{r}-\mathbf{b}|^2 &= \left(|\mathbf{r}|^2 - 2\mathbf{a}\cdot\mathbf{r} + |\mathbf{a}|^2\right) - \left(|\mathbf{r}|^2 - 2\mathbf{b}\cdot\mathbf{r} + |\mathbf{b}|^2\right) \\ &= -2\Big[(\mathbf{a}-\mathbf{b})\cdot\mathbf{r} - \tfrac{1}{2}\left(|\mathbf{a}|^2 - |\mathbf{b}|^2\right)\Big].\end{aligned}$$

所以 (i) 成立當且僅當 $|\mathbf{r}-\mathbf{a}| = |\mathbf{r}-\mathbf{b}|$，即 $RA = RB$。其軌跡（locus）是 $AB$ **的垂直平分面（perpendicular bisector plane）**：經過 $M$ 而法向量（normal）為 $\mathbf{a}-\mathbf{b}$ 的平面。（我們也可以直接看出它是一個平面，因為 (i) 的形式是 $\mathbf{n}\cdot\mathbf{r} =$ 常數，其中 $\mathbf{n} = \mathbf{a}-\mathbf{b} \neq \mathbf{0}$；而 $\mathbf{r} = \mathbf{m}$ 滿足此方程，因為 $(\mathbf{a}-\mathbf{b})\cdot(\mathbf{a}+\mathbf{b}) = |\mathbf{a}|^2 - |\mathbf{b}|^2$。）

### 第 (ii) 部分

方程 (ii) 表示 $\overrightarrow{RA}$ 與 $\overrightarrow{RB}$ 互相垂直，或者 $R$ 就是 $A$ 或 $B$。要辨認這個軌跡，可以配方（complete the square）：

$$\begin{aligned}|\mathbf{r}-\mathbf{m}|^2 &= |\mathbf{r}|^2 - (\mathbf{a}+\mathbf{b})\cdot\mathbf{r} + \tfrac{1}{4}|\mathbf{a}+\mathbf{b}|^2 \\ &= (\mathbf{a}-\mathbf{r})\cdot(\mathbf{b}-\mathbf{r}) - \mathbf{a}\cdot\mathbf{b} + \tfrac{1}{4}|\mathbf{a}+\mathbf{b}|^2 \\ &= (\mathbf{a}-\mathbf{r})\cdot(\mathbf{b}-\mathbf{r}) + \tfrac{1}{4}|\mathbf{a}-\mathbf{b}|^2.\end{aligned}$$

所以 (ii) 等價於 $|\mathbf{r}-\mathbf{m}| = \frac{1}{2}d$。其軌跡是**以 $AB$ 為直徑（diameter）的球面（sphere）**：球心 $M$，半徑 $\frac{1}{2}AB$。這正是「半圓上的圓周角是直角」的三維版本。

### 第 (iii) 及 (iv) 部分

方程 (iii) 表示 $RA = d/\sqrt{2}$。其軌跡是**以 $A$ 為球心、$AB/\sqrt{2}$ 為半徑的球面**。同樣地，(iv) 的軌跡是**以 $B$ 為球心、$AB/\sqrt{2}$ 為半徑的球面**。

### 等價性

設 $P = |\mathbf{r}-\mathbf{a}|^2$ 及 $Q = |\mathbf{r}-\mathbf{b}|^2$。第 (i) 部分已證明

$$\text{(i)} \iff P = Q.$$

至於 (ii)，寫成 $\mathbf{a}-\mathbf{b} = (\mathbf{a}-\mathbf{r}) - (\mathbf{b}-\mathbf{r})$ 再取平方：

$$d^2 = |\mathbf{a}-\mathbf{r}|^2 + |\mathbf{b}-\mathbf{r}|^2 - 2(\mathbf{a}-\mathbf{r})\cdot(\mathbf{b}-\mathbf{r}) = P + Q - 2(\mathbf{a}-\mathbf{r})\cdot(\mathbf{b}-\mathbf{r}).$$

因此

$$\text{(ii)} \iff P + Q = d^2.$$

方程 (iii) 和 (iv) 分別表示 $P = \frac{1}{2}d^2$ 及 $Q = \frac{1}{2}d^2$。現在

$$\left(P = Q \text{ 且} P + Q = d^2\right) \iff P = Q = \tfrac{1}{2}d^2.$$

正方向成立，因為代入 $Q = P$ 得 $2P = d^2$。反方向方面，若 $P = Q = \frac{1}{2}d^2$，則顯然 $P = Q$ 且 $P + Q = d^2$。以上每一步都是「當且僅當」，所以 (i) 與 (ii) 同時成立，等價於 (iii) 與 (iv) 同時成立，得證。

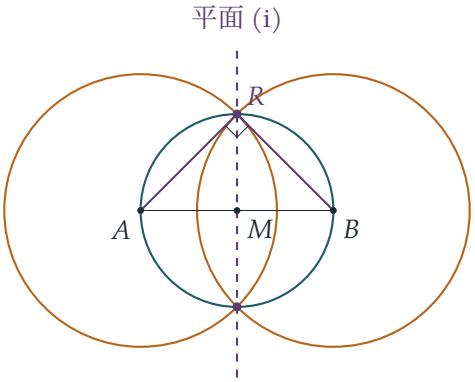
**想一想** 一個平面與一個球面相交，怎會跟兩個球面相交得出同一個集合？

兩組方程描述的都是同一個圓。平面 (i) 經過球面 (ii) 的球心 $M$，所以它把該球面截出一個大圓（great circle）：圓心 $M$，半徑 $\frac{1}{2}AB$，位於 $AB$ 的垂直平分面上。球面 (iii) 和 (iv) 的半徑同為 $AB/\sqrt{2}$，兩者之和大於 $AB$，所以它們確實相交。由對稱性，它們相交於垂直平分面上一個以 $M$ 為圓心的圓。在三角形 $AMR$ 中運用畢氏定理（Pythagoras' theorem），其半徑為 $\sqrt{\frac{1}{2}AB^2 - \frac{1}{4}AB^2} = \frac{1}{2}AB$，所以是同一個圓。

### 幾何意義

點 $R$ 位於這個圓上，當且僅當三角形 $ARB$ 是等腰直角三角形，且直角在 $R$。(i) 與 (ii) 這一組是透過三角形的性質來描述：$RA = RB$（等腰）及 $\angle ARB = 90°$（直角）。(iii) 與 (iv) 這一組則透過邊長來描述：在斜邊（hypotenuse）為 $AB$ 的等腰直角三角形中，兩條直角邊的長度都是 $AB/\sqrt{2}$。上面的代數運算，其實就是畢氏定理及其逆定理（converse）的向量形式。恆等式 $d^2 = P + Q - 2(\mathbf{a} - \mathbf{r}) \cdot (\mathbf{b} - \mathbf{r})$ 就是三角形 $ARB$ 中的餘弦定理（cosine rule）。

下圖展示任何一個包含 $A$ 和 $B$ 的平面所截出的截面。平分面顯示為虛線，球面 (ii) 顯示為青色圓，球面 (iii) 和 (iv) 則顯示為琥珀色圓。四者相交於兩點，就是解集的圓穿過紙面之處。




### 退化情況

若 $\mathbf{a} = \mathbf{b}$，則 (i) 變成 $0 = 0$，對所有 $\mathbf{r}$ 都成立，而 (ii)、(iii) 和 (iv) 都表示 $\mathbf{r} = \mathbf{a}$。等價性仍然成立，但每個軌跡都退化（degenerate）為單一點 $A$（對 (i) 而言則是整個空間）。

## 驗算

取 $\mathbf{a} = (1,0,0)$ 及 $\mathbf{b} = (-1,0,0)$。那麼 (i) 是平面 $x = 0$，(ii) 是 $x^2 + y^2 + z^2 = 1$。方程 (iii) 和 (iv) 是 $(x \mp 1)^2 + y^2 + z^2 = 2$，兩式相減又得到 $x = 0$。四者相交於單位圓 $y^2 + z^2 = 1, x = 0$。例如 $R = (0,1,0)$ 給出 $RA = RB = \sqrt{2}$ 及 $\overrightarrow{RA} \cdot \overrightarrow{RB} = -1 + 1 = 0$。我們亦以有理數運算，對 5000 組隨機選取的 $\mathbf{a}$、$\mathbf{b}$、$\mathbf{r}$ 精確驗證了上述所有恆等式，並確認預測的圓上的隨機點都滿足全部四條方程。

## 常見錯誤

- 把 (ii) 說成「一個圓」。在三維空間中它是球面，而 (i) 的軌跡是平面，不是直線。
- 只證明了一個方向。題目要求證明等價，所以要確保每一步都可逆，或者把兩個方向都證明。
- 只說兩組方程「都得出一個圓」，卻沒有核實那是同一個圓：同一平面、同一圓心、同一半徑。

## STEP 之外

這些軌跡每一個都是 $\mathbf{r}$ 的某個二次函數的等值集（level set），而把兩條球面方程相減得出一個平面的技巧，正是兩個球面的根面（radical plane）這個概念。更廣泛地說，用不同的方程組描述同一個幾何對象，是代數幾何（algebraic geometry）的起點。在這裏，空間中的一個圓由兩條方程截出，而有許多不同的方程組合都能做到。Strang 的 *Introduction to Linear Algebra*（《線性代數導論》）是鞏固向量代數的好去處，特別是點積（dot product）、長度與角度之間的關係。

詳細解答

# STEP II 1987，第 9 題

純數學 • 正交矩陣

> **STEP II 1987，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/87-s2-q9

## 關鍵思路

把 $\mathbf{B}$ 轉置（transpose）之後，條件 $\mathbf{B}^{\mathrm{T}}\mathbf{B} = \mathbf{I}$ 的最左和最右兩端各有一個逆矩陣（inverse matrix）。由於這些逆矩陣本身可逆（invertible），我們可以左乘 $\mathbf{I} - \mathbf{A}^{\mathrm{T}}$、右乘 $\mathbf{I} - \mathbf{A}$，而不會損失任何資訊。餘下的是一條關於 $\mathbf{A}$ 和 $\mathbf{A}^{\mathrm{T}}$ 的多項式恆等式，幾乎立刻就能化簡。要小心的是每一步都必須可以逆推，因為我們要證明的是「若且僅若」（if and only if）。

## 解答

我們會用到以下標準結果：

- $(\mathbf{XY})^{\mathrm{T}} = \mathbf{Y}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}$，而且轉置是線性（linear）的，所以 $(\mathbf{I} \pm \mathbf{A})^{\mathrm{T}} = \mathbf{I} \pm \mathbf{A}^{\mathrm{T}}$（因為 $\mathbf{I}^{\mathrm{T}} = \mathbf{I}$）；
- 若 $\mathbf{X}$ 是非奇異（non-singular）的，則 $\mathbf{X}^{\mathrm{T}}$ 也是，並且 $(\mathbf{X}^{-1})^{\mathrm{T}} = (\mathbf{X}^{\mathrm{T}})^{-1}$。

**第 1 步：$\mathbf{B}$ 的轉置。** 由於 $\mathbf{I} - \mathbf{A}$ 是非奇異的，它的轉置 $\mathbf{I} - \mathbf{A}^{\mathrm{T}}$ 也是非奇異的。把乘積的次序倒轉，

$$\mathbf{B}^{\mathrm{T}} = \left((\mathbf{I} - \mathbf{A})^{-1}\right)^{\mathrm{T}}(\mathbf{I} + \mathbf{A})^{\mathrm{T}} = (\mathbf{I} - \mathbf{A}^{\mathrm{T}})^{-1}(\mathbf{I} + \mathbf{A}^{\mathrm{T}}).$$

因此

$$\mathbf{B}^{\mathrm{T}}\mathbf{B} = (\mathbf{I} - \mathbf{A}^{\mathrm{T}})^{-1}\,(\mathbf{I} + \mathbf{A}^{\mathrm{T}})(\mathbf{I} + \mathbf{A})\,(\mathbf{I} - \mathbf{A})^{-1}.$$

**第 2 步：以可逆推的方式消去逆矩陣。** 設 $\mathbf{P} = \mathbf{I} - \mathbf{A}^{\mathrm{T}}$、$\mathbf{Q} = \mathbf{I} - \mathbf{A}$ 及 $\mathbf{M} = (\mathbf{I} + \mathbf{A}^{\mathrm{T}})(\mathbf{I} + \mathbf{A})$，於是 $\mathbf{B}^{\mathrm{T}}\mathbf{B} = \mathbf{P}^{-1}\mathbf{M}\mathbf{Q}^{-1}$。我們斷言

$$\mathbf{P}^{-1}\mathbf{M}\mathbf{Q}^{-1} = \mathbf{I} \quad \Longleftrightarrow \quad \mathbf{M} = \mathbf{P}\mathbf{Q}.$$

（⇒）左乘 $\mathbf{P}$、右乘 $\mathbf{Q}$：$\mathbf{M} = \mathbf{PIQ} = \mathbf{PQ}$。
（⇐）左乘 $\mathbf{P}^{-1}$、右乘 $\mathbf{Q}^{-1}$：$\mathbf{P}^{-1}\mathbf{M}\mathbf{Q}^{-1} = \mathbf{P}^{-1}\mathbf{P}\mathbf{Q}\mathbf{Q}^{-1} = \mathbf{I}$。

> ? **想一想** 為甚麼不能只是「兩邊同乘」就了事？

把一條等式乘以一個矩陣，得到的必定仍是正確的等式，但這一步一般不能還原：由 $\mathbf{XY} = \mathbf{XZ}$ 不能推斷 $\mathbf{Y} = \mathbf{Z}$，除非 $\mathbf{X}$ 可逆。這裏反方向之所以成立，全因 $\mathbf{P}$ 和 $\mathbf{Q}$ 都有逆矩陣，而這正是用上「$\mathbf{I} - \mathbf{A}$ 是非奇異的」（因而 $\mathbf{I} - \mathbf{A}^{\mathrm{T}}$ 也是）這個假設的地方。還要留意，$\mathbf{P}$ 必須保持在左邊，$\mathbf{Q}$ 必須保持在右邊，因為矩陣乘法不符合交換律（commutative）。

**第 3 步：展開。** 把 $\mathbf{M} = \mathbf{PQ}$ 的兩邊展開，並小心保持各因子的次序，

$$(\mathbf{I} + \mathbf{A}^{\mathrm{T}})(\mathbf{I} + \mathbf{A}) = \mathbf{I} + \mathbf{A} + \mathbf{A}^{\mathrm{T}} + \mathbf{A}^{\mathrm{T}}\mathbf{A},$$
$$(\mathbf{I} - \mathbf{A}^{\mathrm{T}})(\mathbf{I} - \mathbf{A}) = \mathbf{I} - \mathbf{A} - \mathbf{A}^{\mathrm{T}} + \mathbf{A}^{\mathrm{T}}\mathbf{A}.$$

$\mathbf{I}$ 和 $\mathbf{A}^{\mathrm{T}}\mathbf{A}$ 這兩項在兩邊都出現，所以相減後得

$$\mathbf{M} = \mathbf{PQ} \iff 2(\mathbf{A} + \mathbf{A}^{\mathrm{T}}) = \mathbf{O} \iff \mathbf{A} + \mathbf{A}^{\mathrm{T}} = \mathbf{O}.$$

以上每一步都可以逆推：兩邊同時加上或減去同一個矩陣，以及乘以非零純量（scalar）2 或 $\frac{1}{2}$。

**結論。** 把第 1 至第 3 步串連起來，每一環都是等價關係，

$$\mathbf{B}^{\mathrm{T}}\mathbf{B} = \mathbf{I} \iff (\mathbf{I} + \mathbf{A}^{\mathrm{T}})(\mathbf{I} + \mathbf{A}) = (\mathbf{I} - \mathbf{A}^{\mathrm{T}})(\mathbf{I} - \mathbf{A}) \iff \mathbf{A} + \mathbf{A}^{\mathrm{T}} = \mathbf{O},$$

這正是我們要證明的結果。用文字表達：$\mathbf{B}$ 是正交矩陣（orthogonal matrix），當且僅當 $\mathbf{A}$ 是斜對稱矩陣（skew-symmetric matrix）。

## 驗算

- 取 $\mathbf{A} = \begin{pmatrix} 0 & t \\ -t & 0 \end{pmatrix}$，它是斜對稱的。這時 $\det(\mathbf{I} - \mathbf{A}) = 1 + t^2 \neq 0$，而

  $$\mathbf{B} = \frac{1}{1+t^2}\begin{pmatrix} 1 - t^2 & 2t \\ -2t & 1 - t^2 \end{pmatrix},$$

  這是一個旋轉矩陣（rotation matrix）（代入 $t = \tan\frac{\theta}{2}$），所以一如所料 $\mathbf{B}^{\mathrm{T}}\mathbf{B} = \mathbf{I}$。
- $1 \times 1$ 的情況：$\mathbf{A} = (a)$，其中 $a \neq 1$，得 $\mathbf{B} = \left(\frac{1+a}{1-a}\right)$，而 $\mathbf{B}^2 = 1$ 迫使 $1 + a = \pm(1 - a)$，即 $a = 0$（另一個符號不可能成立）。這與 $2a = 0$ 吻合。

- 我們以精確的有理數運算測試了約 2000 個大小由 1 至 5 的隨機矩陣：每一個斜對稱的 $\mathbf{A}$ 都準確地給出 $\mathbf{B}^{\mathrm{T}}\mathbf{B} = \mathbf{I}$，其餘的則全部不成立。

## 常見錯誤

- 沒有倒轉次序就寫 $\mathbf{B}^{\mathrm{T}} = (\mathbf{I} + \mathbf{A}^{\mathrm{T}})(\mathbf{I} - \mathbf{A}^{\mathrm{T}})^{-1}$。（這條式子碰巧是對的，因為 $\mathbf{I} + \mathbf{A}^{\mathrm{T}}$ 與 $(\mathbf{I} - \mathbf{A}^{\mathrm{T}})^{-1}$ 可交換，但這一點需要另行說明理由。）
- 展開 $(\mathbf{I} + \mathbf{A}^{\mathrm{T}})(\mathbf{I} + \mathbf{A})$ 時寫成 $\mathbf{A}\mathbf{A}^{\mathrm{T}}$ 而不是 $\mathbf{A}^{\mathrm{T}}\mathbf{A}$。在這一題裏兩種寫法都會抵銷，但前提是兩邊的展開方式必須一致。
- 只證明了一個方向。題目要求的是「若且僅若」，所以要說明每一步為何可以逆推。

## STEP 之外

映射 $\mathbf{A} \mapsto (\mathbf{I} + \mathbf{A})(\mathbf{I} - \mathbf{A})^{-1}$ 稱為凱萊變換（Cayley transform）。它把構成一個向量空間（vector space）的斜對稱矩陣，變成構成一個羣（group）的正交矩陣，從而給出旋轉的有理參數化（rational parametrisation）。（可以證明，當 $\mathbf{A}$ 是實斜對稱矩陣時，$\mathbf{I} - \mathbf{A}$ 必定可逆，而且所得到的正交矩陣恰好是那些沒有特徵值（eigenvalue）$-1$ 的正交矩陣。）這讓我們初次窺見李羣（Lie group）與它的李代數（Lie algebra）之間的關係。想進一步了解正交矩陣和斜對稱矩陣，Strang 的 *Introduction to Linear Algebra*（線性代數入門）是很好的讀物。

詳細解答

# STEP II 1987，第 10 題

純數學 • 群公理

> **STEP II 1987，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/87-s2-q10

## 關鍵思路

一切都發生在下標（subscripts）之中，所以把每個運算轉化為下標 $1, 2, \ldots, N$ 上的普通算術：第一個運算是「取最大值」，第二個運算是「取兩者之差的絕對值，再加一」。然後逐一檢驗四條公理（axioms）：封閉性（closure）、結合律（associativity）、單位元（identity）、逆元（inverses）。要說明某條公理不成立，只需一個具體的反例（counterexample）；要說明某條公理成立，則需要一般性的論證。

## 解答

全文中，以 $j \circ k$ 表示 $a_j \circ a_k$ 的下標，$*$ 亦同理。兩個元素相等，當且僅當它們的下標相等。

**運算**。

這裏 $a_j \circ a_k = a_{\max(j,k)}$。

**封閉性成立**。若 $1 \le j, k \le N$，則 $\max(j,k)$ 是 $j$ 和 $k$ 其中之一，所以它屬於 $\{1, \ldots, N\}$，因此 $a_j \circ a_k \in S$。

**結合律成立**。對任意 $j, k, l$，

$$(a_j \circ a_k) \circ a_l = a_{\max(\max(j,k),l)} = a_{\max(j,k,l)} = a_{\max(j,\max(k,l))} = a_j \circ (a_k \circ a_l),$$

因為兩種加括號的方式都得出三個下標中最大的一個。

**單位元存在：它就是** $a_1$。每個下標都至少是 1，所以 $\max(1,k) = \max(k,1) = k$，因此對每個 $k$ 都有 $a_1 \circ a_k = a_k \circ a_1 = a_k$。

**逆元不成立**。$a_j$ 的逆元應是某個 $a_k$，使得 $a_j \circ a_k = a_1$，即 $\max(j,k) = 1$。但 $\max(j,k) \geq j$，所以當 $j \geq 2$ 時這是不可能的。例如，對每個 $k$，$a_2 \circ a_k$ 都是 $a_2$ 或排在其後的元素，永遠不會是 $a_1$。只有 $a_1$ 有逆元（就是它自己）。

所以 $S$ 在 $\circ$ 下滿足封閉性、結合律和單位元，但不滿足逆元公理，因此不是羣（group）。

> **想一想** 會不會有其他元素也是單位元，以致逆元的論證需要重做？

不會：單位元是唯一的。若 $e$ 和 $e'$ 都是單位元，則 $e = e \circ e' = e'$。所以「$a_j \circ a_k$ 等於單位元」中的單位元只可能是 $a_1$，上面的論證是完整的。

**運算 $*$**

這裏 $a_j * a_k = a_{|j-k|+1}$。當 $N = 4$ 時，下標的運算表為

| $*$ | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1 | 1 | 2 | 3 | 4 |
| 2 | 2 | 1 | 2 | 3 |
| 3 | 3 | 2 | 1 | 2 |
| 4 | 4 | 3 | 2 | 1 |

表中可見的規律對每個 $N$ 都成立。

**封閉性成立**。若 $1 \leq j,k \leq N$，則 $0 \leq |j-k| \leq N-1$，所以 $1 \leq |j-k|+1 \leq N$，因此 $a_j * a_k \in S$。

**單位元存在：它就是 $a_1$**。由於 $k \geq 1$，$|1-k| = k-1$，所以 $a_1 * a_k = a_k$；同樣地 $a_k * a_1 = a_k$（因為 $|j-k| = |k-j|$，這個運算是可交換的（commutative））。

**逆元成立**。$a_j * a_j = a_{0+1} = a_1$，即單位元，所以每個元素都是自己的逆元。（這就是表中對角線上的一列 1。）

**結合律不成立**。取 $a_2, a_2, a_3$（這需要 $N \geq 3$）：

$$(a_2 * a_2) * a_3 = a_1 * a_3 = a_3,$$
$$a_2 * (a_2 * a_3) = a_2 * a_{|2-3|+1} = a_2 * a_2 = a_1.$$

兩者不同，所以 $*$ 不滿足結合律。

所以 $S$ 在 $*$ 下滿足封閉性、單位元和逆元，但不滿足結合律，因此同樣不是羣。

> **想一想** 哪裏用到了 $N > 2$？

只在結合律的反例中用到，因為它用了 $a_3$。當 $N = 2$ 時，運算表就是 2 階羣的運算表（$a_1$ 為單位元，$a_2 * a_2 = a_1$），而 $*$ 是滿足結合律的；至於 $\circ$，逆元的論證只需要 $a_2$，所

以 $\circ$ 在 $N = 2$ 時同樣不成羣。

## 驗算

用窮舉法（brute force）檢查小的情況：對 2 至 9 的每個 $N$，我們列出所有三元組來檢驗結合律，並直接搜尋單位元和逆元。對於 $\circ$：封閉、滿足結合律、單位元為 $a_1$，且只有 $a_1$ 可逆。對於 $*$：封閉、單位元為 $a_1$、每個元素都是自己的逆元，而結合律對每個 $N \geq 3$ 都不成立（包括在 $(a_2, a_2, a_3)$ 處），但在 $N = 2$ 時成立。你可以用一個 $3 \times 3$ 的運算表，一分鐘內親手完成 $N = 3$ 的情況。

## 常見錯誤

- 以為「取最大值」不像乘法，便斷言 $\circ$ 沒有單位元。應檢查最小的元素：$a_1$ 正是單位元。
- 嘗試一般地證明 $*$ 滿足結合律，或者只檢驗含有 $a_1$ 的三元組（這些三元組必然滿足結合律，因為 $a_1$ 是單位元）。真正需要的只是一個具體的反例。
- 忘記檢驗封閉性。封閉性很容易被想當然，但這裏它需要用到上界 $|j - k| \leq N - 1$。

## STEP 之外

一個帶有滿足結合律的運算和單位元的集合，例如 $S$ 在 $\circ$ 下，稱為麼半羣（monoid）；全序集（totally ordered set）上的「取最大值」是一個標準例子，它同時也是一個半格（semilattice）。一個滿足封閉性、單位元和逆元但不滿足結合律的集合，例如 $S$ 在 $*$ 下，說明結合律是一條真正獨立的公理：它不能從其餘三條推導出來。若想了解羣的各條公理如何互相配合，可參閱 Allenby 的 *Rings, Fields and Groups*（《環、域與羣》）。

詳細解答

# STEP II 1987，第 11 題

力學 • 摩擦力與力矩

**STEP II 1987，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/87-s2-q11

## 關鍵思路

支點的作用力是未知的，我們也不需要求它，所以我們對支點 $O$ 取力矩（moment），因為該力對 $O$ 沒有力矩。桿是處於三維空間中的，但圓環所在的平面給出兩個自然的方向：垂直於平面的方向（法向反作用力（normal reaction）沿此方向作用）和平面內的方向（摩擦力（friction）在此方向作用）。把重力也按同樣方式分解，一條力矩方程給出法向反作用力，另一條給出摩擦力。兩者之比就是 $\mu$。

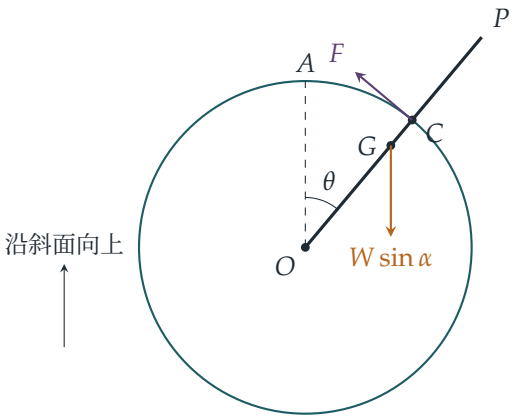
## 解答

**模型。** 桿 $OP$ 是一條重量為 $W$ 的均勻桿，其質心（centre of mass）$G$ 位於中點，所以 $OG = \frac{1}{2}b$。桿以光滑支點（pivot）固定於圓環的圓心 $O$。桿橫跨圓環，擱在圓周上的一點 $C$，所以 $OC = a$。由於 $O$ 和 $C$ 都在圓環所在的平面上，桿也位於該平面內。設 $A$ 為圓環的最高點，於是 $OA$ 沿最大斜率線（line of greatest slope）直指向上，並設 $\angle AOC = \theta$。

**作用力。** 作用於桿上的力有：

- 作用於 $G$ 的重力 $W$；
- 光滑支點 $O$ 處的作用力，此力未知，但對 $O$ 沒有力矩；
- $C$ 處的接觸力。桿與圓環的金屬線在 $C$ 相交。桿沿半徑方向，金屬線沿切線方向，兩者都在平面內，所以它們的公共法線垂直於平面。因此法向反作用力 $R$ 垂直於圓環所在的平面。摩擦力 $F$ 位於平面內。桿唯一可能的運動是繞 $O$ 轉動，這會令 $C$ 沿圓環的切線移動。所以摩擦力沿該切線作用，與桿垂直。

把重力分解為垂直於平面（指向平面內）的分量 $W\cos\alpha$，以及在平面內沿最大斜率線向下的分量 $W\sin\alpha$。下圖是垂直俯視該平面所見的情況，$A$ 在最上方。$R$ 在 $C$ 處指出紙面。




**對 $O$ 取力矩。** 由於支點光滑，其餘各力對經過 $O$ 的每一條軸的總力矩都是零。其中兩條軸能給出有用的方程。

*在平面內、經過 $O$ 並垂直於桿的軸。*只有垂直於平面的力對這條軸有力矩，即距離為 $\frac{1}{2}b$ 的 $W\cos\alpha$ 和距離為 $a$ 的 $R$。平面內的力 $F$ 和 $W\sin\alpha$ 要麼與這條軸相交，要麼與它平行，所以對它沒有力矩。因此

$$R\,a = W\cos\alpha\cdot\tfrac{1}{2}b, \qquad \text{所以} \qquad R = \frac{Wb\cos\alpha}{2a}.$$

*垂直於平面、經過 $O$ 的軸。*現在只有平面內的力才起作用。桿與沿斜面向上的方向成 $\theta$ 角，所以 $W\sin\alpha$ 垂直於桿的分量是 $W\sin\alpha\sin\theta$，作用於距離 $\frac{1}{2}b$ 處。摩擦力本身已垂直於桿，作用於距離 $a$ 處。因此

$$F\,a = W\sin\alpha\sin\theta\cdot\tfrac{1}{2}b, \qquad \text{所以} \qquad F = \frac{Wb\sin\alpha\sin\theta}{2a}.$$

重力在平面內的分量傾向令桿擺離 $A$，向圓環底部擺下去，而摩擦力則阻止這個運動。

**想一想** 摩擦力會不會也有沿桿方向的分量？

沿桿方向的力經過 $O$，所以對 $O$ 沒有力矩。它既不能促進也不能阻礙轉動，而支點可以提供沿桿方向的任何力。唯一可能的運動是繞 $O$ 轉動，所以對極限平衡（limiting equilibrium）而言，有關的摩擦力是沿切線方向的分量，也就是 $F$。我們把摩擦力視為完全沿切線作用。

**極限平衡**。 在極限平衡時 $F = \mu R$：

$$\frac{Wb\sin\alpha\sin\theta}{2a} = \mu\,\frac{Wb\cos\alpha}{2a} \qquad \Longrightarrow \qquad \boxed{\mu = \tan\alpha\,\sin\theta}$$

（我們取 $0 \le \theta \le \pi$）。只要 $\tan\alpha\sin\theta \le \mu$，桿便能在角 $\theta$ 處保持平衡。留意 $a$、$b$ 和 $W$ 全都消去了。

## 驗算

- 若 $\theta = 0$ 或 $\theta = \pi$，桿沿最大斜率線放置。此時重力在平面內的分量正好沿桿方向拉，不需要摩擦力：$\mu = 0$。當 $\alpha = 0$ 時亦然，因為圓環水平時重力在平面內沒有分量。
- 當 $\theta = \frac{1}{2}\pi$，即桿處於水平時，所需的摩擦力最大。此時 $\mu = \tan\alpha$，與傾角為 $\alpha$ 的斜面上的木塊情況吻合。
- 我們做了一次數值驗算：對 $\alpha, \theta, a, b, W$ 隨機取 2000 組值，把各力寫成三維向量（vector），解對 $O$ 的向量力矩方程求 $R$ 和 $F$，發現 $F/R$ 與 $\tan\alpha\sin\theta$ 的差距在 $4 \times 10^{-15}$ 以內。例如 $\alpha = 30°$、$\theta = 60°$ 給出 $\mu = \frac{1}{2}$。

## 常見錯誤

- 以為 $C$ 處的法向反作用力沿半徑方向，就像桿靠在鉛垂平面中的圓環內側那樣。這裏桿是橫跨在金屬線上的，所以法線垂直於圓環所在的平面。
- 用分解力的方法而不取力矩。這樣會引入支點作用力的三個未知分量，而題目並不需要它們。

## STEP 之外

這裏我們對經過 $O$ 的兩條互相垂直的軸取力矩。這是單一向量方程 $\sum \mathbf{r} \times \mathbf{F} = \mathbf{0}$ 的捷徑，該方程表示對 $O$ 的總轉矩（torque）為零。把轉矩視為向量，以及剛體（rigid body）在三維空間中的平衡條件，都是大學一年級力學的標準內容，並會進一步引向剛體繞固定點的轉動。Kleppner 和 Kolenkow 的 *An Introduction to Mechanics*（《力學導論》）正是以這種方式，把轉矩和角動量（angular momentum）作為向量來發展。

詳細解答

# STEP II 1987，第 12 題

力學 • 平衡與能量

**STEP II 1987，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/87-s2-q12

## 關鍵思路

「在每個位置都處於平衡（equilibrium）」其實是關於勢能（potential energy）的陳述：如果珠子（bead）可以停在線上的任何位置，那麼珠子沿線滑動時，系統的勢能就不會改變，所以勢能在線上必須是常數。用珠子的高度及它與環的距離寫出這個能量，便立即得到曲線的方程。同一事實也令第二部分變得簡潔：勢能既然是常數，能量守恆（conservation of energy）便告訴我們珠子與質點的總動能（kinetic energy）是常數，餘下的只是一些把兩個速率聯繫起來的幾何。

## 解答

我們把珠子和懸掛的質量視為質點，繩視為輕而不可伸長（inextensible），環和線視為光滑。把環放在原點 $O$，$y$ 垂直向上量度。設珠子位於極坐標（polar coordinates）$(r,\theta)$，其中 $\theta$ 由水平方向逆時針量度，因此珠子相對於 $O$ 的高度是 $y = r\sin\theta$。設 $L$ 為繩的長度。

### 曲線的方程

由 $O$ 到珠子的一段繩長 $r$，所以懸掛的質點位於環下方距離 $L - r$ 處。取 $O$ 為零勢能水平，系統的勢能是

$$V = mgr\sin\theta - mg(L - r) = mg\left(r + r\sin\theta\right) - mgL.$$

線是光滑的，所以它對珠子的反作用力垂直於線，能使系統沿線移動的力只來自 $V$。系統在某位置處於平衡，當且僅當 $V$ 在該處對沿線的移動取駐值。在每個位置都平衡，意

味着沿整條線都有 $\mathrm{d}V/\mathrm{d}s = 0$（其中 $s$ 為弧長（arc length)），所以 $V$ 在線上是常數：

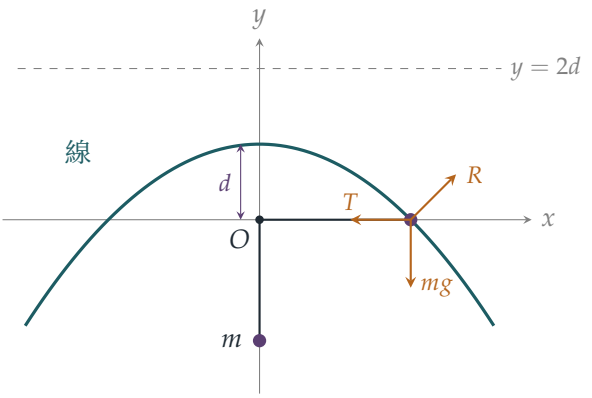
$$r(1+\sin\theta) = c \quad \text{對某常數} c.$$

用分解力的方法也可以看到這一點。平衡時，懸掛的質點令整條繩的張力（tension）為 $mg$（環是光滑的）。在珠子上，指向 $O$ 的張力 $mg$ 的切向分量是 $-mg\frac{\mathrm{d}r}{\mathrm{d}s}$，而重力的切向分量是 $-mg\frac{\mathrm{d}y}{\mathrm{d}s}$。兩者必須抵消，所以 $\frac{\mathrm{d}}{\mathrm{d}s}(r+y) = 0$，這正是同一條件。

求 $c$：由於 $1+\sin\theta \le 2$，我們有 $r = c/(1+\sin\theta) \ge c/2$，等號在 $\theta = \frac{\pi}{2}$ 時成立（即線上位於環正上方的點）。所以環到線的最短距離是 $c/2$，由此得 $c = 2d$，而曲線是

$$\boxed{r(1+\sin\theta) = 2d} \qquad \text{即} \qquad r = \frac{2d}{1+\sin\theta}.$$

（如果改為由向下的鉛垂線量度 $\theta$，同一曲線寫成 $r(1-\cos\theta) = 2d$。）由於 $r = 2d - y$，線上每一點與 $O$ 的距離都等於它與水平線 $y = 2d$ 的距離。所以這條線是一條拋物線（parabola），焦點（focus）在環上，準線（directrix）在環上方高度 $2d$ 處；以直角坐標表示為 $x^2 = 4d(d-y)$，頂點為 $(0, d)$。




圖中珠子與環處於同一水平；張力 $T = mg$、重力 $mg$ 和法向反作用力（normal reaction）$R$ 互相平衡，因為拋物線在該處的斜率恰好是 $-1$。

### 珠子的速率

現在珠子以速率 $u$ 沿線移動。繩保持拉緊，長度不變，所以懸掛的質點仍位於環下方距離 $L - r$ 處。

**想一想** 懸掛的質點移動得多快？

它在環下方的深度是 $L - r$，所以它的速度是垂直的，速率為 $|\dot{r}|$。注意 $\dot{r}$ 只是珠子速度的徑向部分，而不是它的整個速率 $u$。

一旦開始運動，張力便不再是 $mg$，但它對系統不做淨功：繩不可伸長而環光滑，所以張力對珠子所做的功與它對懸掛質點所做的功互相抵消。線的反作用力垂直於珠子的運動，不做功。所以能量守恆：

$$\tfrac{1}{2}mu^2 + \tfrac{1}{2}m\dot{r}^2 + V = \text{常數}.$$

但由第一部分，在線上 $V = mg\cdot 2d - mgL$ 是常數。因此

$$u^2 + \dot{r}^2 = \text{常數}. \qquad (*)$$

接着把 $\dot{r}$ 與 $u$ 聯繫起來。對 $r = 2d/(1+\sin\theta)$ 求導，

$$\frac{\mathrm{d}r}{\mathrm{d}\theta} = -\frac{2d\cos\theta}{(1+\sin\theta)^2} = -\frac{r\cos\theta}{1+\sin\theta}, \qquad \text{所以} \qquad \dot{r}^2 = r^2\dot{\theta}^2\,\frac{\cos^2\theta}{(1+\sin\theta)^2} = r^2\dot{\theta}^2\,\frac{1-\sin\theta}{1+\sin\theta},$$

這裏用了 $\cos^2\theta = (1-\sin\theta)(1+\sin\theta)$。由曲線方程，$1+\sin\theta = 2d/r$ 及 $1-\sin\theta = 2-2d/r = 2(r-d)/r$，所以

$$\frac{1-\sin\theta}{1+\sin\theta} = \frac{r-d}{d}, \qquad \text{得} \qquad \dot{r}^2 = \frac{r-d}{d}\,r^2\dot{\theta}^2.$$

在極坐標中，速率滿足 $u^2 = \dot{r}^2 + r^2\dot{\theta}^2$，所以

$$u^2 = r^2\dot{\theta}^2\left(\frac{r-d}{d}+1\right) = \frac{r}{d}\,r^2\dot{\theta}^2, \qquad \text{因此} \qquad \dot{r}^2 = \frac{r-d}{r}\,u^2.$$

**想一想** 當 $r = d$ 時，$\dot{r}$ 是多少？

當 $r = d$ 時，珠子位於頂點，$\theta = \frac{\pi}{2}$，此處 $r$ 取最小值，且 $\mathrm{d}r/\mathrm{d}\theta = 0$。所以該處 $\dot{r} = 0$（公式 $\dot{r}^2 = \frac{r-d}{r}u^2$ 也這樣說），懸掛的質點瞬間靜止。因此 $(*)$ 中的常數是 $v^2$：

$$u^2 + \frac{r-d}{r}\,u^2 = v^2 \quad\Longrightarrow\quad u^2\,\frac{2r-d}{r} = v^2 \quad\Longrightarrow\quad u = \left(\frac{r}{2r-d}\right)^{\frac{1}{2}} v,$$

由於 $u$ 和 $v$ 是速率，取正根。這就是所求的結果。

## 驗算

- 當 $r = d$ 時，公式給出 $u = v$，理應如此。
- 當 $r \to \infty$ 時，$u \to v/\sqrt{2}$。在很遠處，拋物線的切線幾乎正好沿着半徑方向，所以 $\dot{r} \approx u$，而動能 $\frac{1}{2}mv^2$ 由珠子與質點平均分享。
- 在環的水平上的平衡：點 $(2d, 0)$ 在曲線上，該處張力 $mg$（水平）與重力 $mg$（垂直）合成一個與水平成 $45^\circ$ 的力，垂直於拋物線在該處斜率為 $-1$ 的切線。
- 數值上，取 $d = 1.3$ 及 $v = 2$，我們由頂點出發，沿拋物線對珠子與質點系統的運動方程進行數值積分。在 $r = 2.8905$ 處，珠子的速率是 1.606309，與 $\sqrt{r/(2r-d)}\,v$ 相差約 $10^{-8}$。我們亦在線上 1000 個隨機點驗證了靜止珠子所受的切向力為零。

## 常見錯誤

- 忘記懸掛質點的動能，結果對所有 $r$ 都得出 $u = v$。勢能是常數，但動能是由兩者分享的。
- 把質點的速率當作 $u$ 而不是 $|\dot{r}|$。只有珠子運動的徑向部分才會把繩拉過環。
- 由另一條線量度 $\theta$，卻照搬 $r(1 + \sin\theta) = 2d$。任何正確的極坐標形式都可以，但要說明 $\theta$ 是由哪條線量度的。

## STEP 之外

系統在勢能取駐值之處處於平衡，這個想法就是*虛功原理*（principle of virtual work），它引向拉格朗日力學（Lagrangian mechanics）。在拉格朗日力學中，動能以一個廣義坐標（generalised coordinate）（這裏是 $\theta$）表示，運動方程由此直接得出，完全不必求張力。由於 $V$ 在這條線上是常數，珠子和質點的運動就好像完全沒有受力一樣，而我們的關係式不過是動能 $\frac{1}{2}m r^2(2r - d)\dot{\theta}^2/d$ 守恆。Kleppner 與 Kolenkow 的 *An Introduction to Mechanics*（《力學導論》）以這種精神講解能量方法和受約束的運動。

詳細解答

# STEP II 1987，第 13 題

力學 • 與牆壁的斜碰撞

> **STEP II 1987，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/87-s2-q13

## 關鍵思路

光滑（smooth）的檯邊只能改變垂直於它的速度分量（component）。這個分量會反向並乘以 $e$，而沿檯邊方向的分量則保持不變。兩條相鄰的檯邊分別作用於兩個不同的分量，所以每個分量各被反向並乘以 $e$ 一次。至於閉合路徑，我們用坐標追蹤圓盤由一條檯邊到另一條檯邊的運動，並選取唯一可自由調節的量，即初始斜率（slope），使圓盤回到出發點。

## 解答

**模型**。圓盤視為一個質點（particle），在設有光滑檯邊的光滑水平檯面上滑動，因此兩次碰撞之間沒有任何力作用於它，它以等速沿直線運動。碰撞時，檯邊施加的衝量（impulse）垂直於檯邊（因為檯邊光滑），所以速度的平行分量不變。根據牛頓恢復定律（Newton's law of restitution），垂直分量反向並乘以 $e$。我們設 $0 < e \leq 1$。

沿檯面的兩條邊設坐標軸，使檯面為 $0 \leq x \leq L$、$0 \leq y \leq B$。

### 第 (i) 部分：兩條相鄰的檯邊

設圓盤的速度為 $(u, v)$。相鄰的檯邊是一條形如 $x =$ 常數 的檯邊和一條形如 $y =$ 常數 的檯邊，碰撞次序不限。撞擊 $x$ 檯邊使 $(u, v) \mapsto (-eu, v)$，撞擊 $y$ 檯邊使 $(u, v) \mapsto (u, -ev)$。兩者各撞一次後，無論次序如何，速度均為

$$(-eu,\ -ev) = -e\,(u, v).$$

這是原速度的純量倍數，所以最後的路徑與原路徑平行，正如所求。此時圓盤沿相反方向運動，速率減為原來的 $e$ 倍。

### 第 (ii) 部分：從檯邊出發的閉合路徑

設圓盤靜止於檯邊 $y=0$ 上的點 $P=(a,0)$，且 $a:(L-a)=z:1$，於是

$$a=\frac{zL}{1+z},\qquad L-a=\frac{L}{1+z},\qquad 0<a<L.$$

從上方俯視檯面，$x$ 向右、$y$ 向上，由檯邊 $y=0$ 開始順時針方向，依次遇到 $x=0$、$y=B$，然後是 $x=L$。若比例是從另一端量度，則把 $z$ 換成 $1/z$，論證不變。以速度 $(-u,v)$ 擊出圓盤，其中 $u,v>0$，並記斜率為 $k=v/u$。四段路程上的速度依次為

$$(-u,v)\xrightarrow{x=0}(eu,v)\xrightarrow{y=B}(eu,-ev)\xrightarrow{x=L}(-e^2u,-ev)\xrightarrow{y=0}\cdots$$

- 第 1 段，由 $(a,0)$ 至 $x=0$：水平走過距離 $a$，上升 $ak$，所以撞擊左檯邊於 $W_1=(0,\,ak)$。
- 第 2 段，斜率為 $v/(eu)=k/e$：上升 $B-ak$，因此橫向移動 $e(B-ak)/k$。它在 $W_2=\big(e(B/k-a),\,B\big)$ 碰到頂邊。
- 第 3 段，斜率為 $-k$：橫向移動 $L-x_{W_2}$，下降 $k(L-x_{W_2})$，所以在高度 $y_{W_3}=B-kL+e(B-ak)$ 處碰到右檯邊。
- 第 4 段，斜率為 $k/e$：下降 $y_{W_3}$，向回移動 $e\,y_{W_3}/k$，到達 $y=0$ 時位於

$$x_{\text{end}}=L-\frac{e\,y_{W_3}}{k}=(1+e)L+e^2a-\frac{e(1+e)B}{k}.$$

我們要求 $x_{\text{end}}=a$。由此得 $e(1+e)B/k=(1+e)L-(1-e^2)a=(1+e)\big(L-(1-e)a\big)$，所以

$$k=\frac{eB}{L-(1-e)a}=\frac{B}{L}\cdot\frac{e(1+z)}{1+ez}.$$

由於 $L-(1-e)a>L-a>0$，對每個 $z>0$，這都是正數。

> **想一想** 我們完成了嗎？我們假設了圓盤依左、上、右、下的次序碰到檯邊。

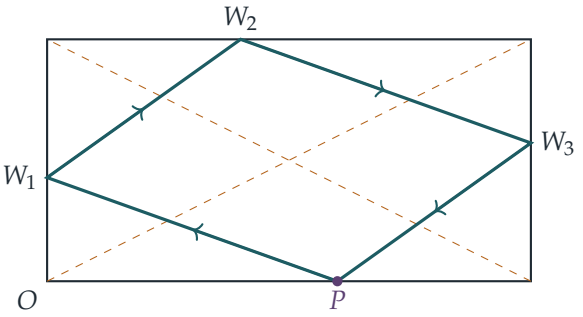
還未完成。我們仍須驗證每個碰撞點確實落在我們假設的那條邊上。記 $D=L-(1-e)a=L-a+ea$，則 $k=eB/D$。於是

$$\begin{aligned}
y_{W_1}&=ak=\frac{eaB}{D},\\
x_{W_2}&=e\Big(\frac{D}{e}-a\Big)=L-a,\\
y_{W_3}&=B(1+e)-k(L+ea)=\frac{B\big[(1+e)(L-a+ea)-eL-e^2a\big]}{D}=\frac{B(L-a)}{D}.
\end{aligned}$$

由於 $0<a<L$ 及 $e>0$，我們有 $0<ea<D$ 及 $0<L-a<D$。所以 $0<y_{W_1}<B$、$0<x_{W_2}<L$ 及 $0<y_{W_3}<B$。每次碰撞都落在我們假設的那條邊（不含端點）上，不會

撞到角位。因此對每個 $z$，以上述斜率 $k$ 擊出的圓盤會依次經過左、上、右檯邊，然後回到 $P$。

由第 (i) 部分，第 1 段與第 3 段平行，第 2 段與第 4 段亦平行，所以這條閉合路徑是內接於檯面的平行四邊形（parallelogram）$PW_1W_2W_3$。




虛線是對角線（diagonal）。圖中路徑取 $L = 2B$、$e = \frac{1}{2}$ 及 $z = 1.5$。

**第 (iii) 部分：從任意內點出發的兩種擊法**

每一種這樣的擊法都是第 $(ii)$ 部分的路徑之一。設圓盤從內點（interior point）$Q$ 以速度 $\mathbf{w}$ 出發，依順時針次序在四條檯邊上的 $V_1, V_2, V_3, V_4$ 反彈，然後回到 $Q$。應用第 (i) 部分兩次：四次反彈後速度為 $(-e)^2\mathbf{w} = e^2\mathbf{w}$，與原速度方向相同。因此最後一段 $V_4Q$ 與第一段 $QV_1$ 位於同一直線上，且方向相同。所以 $V_4 \to Q \to V_1$ 是同一段直線路程，圓盤的路徑是一條從檯邊上的 $V_4$ 出發、順時針在其餘三條檯邊反彈後回到 $V_4$ 的閉合路徑。第 (ii) 部分的推導顯示，這種路徑的斜率由其起點唯一決定。每條這樣的路徑在每條邊上各有一個頂點，所以它們全都是第 (ii) 部分中的平行四邊形 $\Pi_a$（$0 < a < L$），而 $Q$ 位於 $\Pi_a$ 的某條邊上。反之，若 $Q$ 位於某個 $\Pi_a$ 的一條邊上，從 $Q$ 沿該邊按行進方向擊出即可。因此我們要數出通過 $Q$ 的（平行四邊形，邊）組合的數目。

當 $z$ 改變時，各邊如何掃過檯面。當 $z$ 由 0 增至 $\infty$，$a$ 由 0 增至 $L$。由上面的公式，$y_{W_1}$ 由 0 增至 $B$，$x_{W_2} = L - a$ 由 $L$ 減至 0，而 $y_{W_3} = B/\big(1 + ea/(L-a)\big)$ 由 $B$ 減至 0。所以 $\Pi_a$ 起初是通過 $O$ 的對角線（當 $a \to 0$ 時，各頂點聚集到角位 $O$ 和 $(L, B)$），最後變成另一條對角線。現在考慮其中一條邊，例如 $PW_1$。它在角位 $O$ 處截出一個三角形，兩條直角邊分別是沿底邊的 $a$ 和沿左邊向上的 $y_{W_1}$。兩條直角邊都隨 $a$ 嚴格遞增，所以這些三角形嚴格地一個套一個，其斜邊永不相交。它們由無到有連續地增長，直至成為檯面在由 $(L, 0)$ 至 $(0, B)$ 的對角線以下的一半。準確地說，若 $Q = (q_1, q_2)$ 位於這一半之中，則函數

$$f(a) = \frac{q_1}{a} + \frac{q_2}{y_{W_1}(a)}$$

是連續（continuous）且嚴格遞減（strictly decreasing）的。當 $a \to 0$ 時它趨於 $\infty$，當

$a \to L$ 時它趨於 $q_1/L + q_2/B < 1$，所以恰好有一個 $a$ 使 $f(a) = 1$（即 $Q$ 位於 $PW_1$ 上）。若 $Q$ 不在這一半之中，則對每個 $a$ 都有 $f(a) > 1$，$PW_1$ 永不通過 $Q$。同樣的論證適用於每條邊：

- $PW_1$ 掃過由 $(L, 0)$ 至 $(0, B)$ 的對角線以下的半個檯面一次；
- $W_2W_3$ 掃過該對角線以上的半個檯面一次；
- $W_1W_2$ 掃過由 $O$ 至 $(L, B)$ 的對角線以上的半個檯面一次；
- $W_3P$ 掃過該對角線以下的半個檯面一次。

一個不在檯邊上、也不在對角線上的點 $Q$，恰好位於每條對角線的某一側。所以恰好有兩條邊通過它，而且分屬兩個不同的平行四邊形。這就是那兩種擊法：每一種都使圓盤沿該邊出發，順時針繞經四條檯邊後回到 $Q$。（若 $Q$ 位於對角線上，其中一種擊法會退化為射向角位的一擊。）

## 驗算

- 當 $e = 1$ 時，對每個 $z$ 斜率都是 $k = B/L$，所以所有邊都與對角線平行。這正是對完全彈性檯邊使用常見的鏡像展開（unfolding）方法所得的結果。
- 取 $L = 2$、$B = 1$、$e = \frac{1}{2}$、$z = 1$：則 $a = 1$、$k = \frac{1}{3}$，碰撞點為 $(0, \frac{1}{3})$、$(1, 1)$ 及 $(2, \frac{2}{3})$。最後一段有 $\mathrm{d}x/\mathrm{d}y = e/k = \frac{3}{2}$，所以下降 $\frac{2}{3}$ 時向回移動 1，到達 $(1, 0) = P$。
- 在一個逐次計算圓盤由一邊到另一邊的數值模擬中，三個部分的結論都成立：隨機選取的檯面、$e$ 值和 $z$ 值都給出預測的碰撞點，而隨機選取的內點都恰好位於兩條閉合路徑上。

## 常見錯誤

- 使用鏡像（展開）技巧：當 $e < 1$ 時，反射角不等於入射角，所以這技巧失效。
- 由 $x_{\text{end}} = a$ 求得 $k$ 後，沒有驗證每次碰撞是否落在預期的邊上。
- 在第 (iii) 部分中，假設從 $Q$ 出發的擊法必須沿某一特定的邊開始。正是第 (i) 部分的論證把第一段與最後一段連成同一段直線路程。

## STEP 之外

這類路徑稱為撞球（billiards）問題。當 $e = 1$ 時，它本身就是一個完整的學科，以反射和展開來研究，並通向動力系統（dynamical systems）。這裏的檯邊會損失能量，而「一

圈前的速度 $\mapsto$ 一圈後的速度」這個映射純粹是乘以 $e^2$ 的縮放。這就是為何對每個 $e$ 都存在閉合軌道，儘管圓盤每繞一圈都會減速。Kleppner 與 Kolenkow 的 *An Introduction to Mechanics*（《力學導論》）對衝量和恢復有仔細的講解。Siklos 的 *Advanced Problems in Mathematics*（《數學進階難題》）收錄了更多同類型的題目。

詳細解答

# STEP II 1987，第 14 題

力學 • 圓錐上的彈性帶

> **STEP II 1987，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/87-s2-q14

## 關鍵思路

由於橡皮圈在圓錐上始終保持為一個水平圓，它的整個狀態只需一個數便可描述：它沿母線（generator）量度與頂點（vertex）的距離 $s$。因此我們只需追蹤一小段元素。張力（tension）產生的向內拉力有一個沿圓錐面的分量，推動該元素沿母線移動；另有一個垂直於圓錐面的分量，被光滑的圓錐抵消。把 $s$ 換成長度 $x$，便得到簡諧運動（simple harmonic motion），而橡皮圈在四分之一個週期（period）後變得鬆弛（slack）。

## 解答

**模型。** 我們把橡皮圈視為一條質量為 $m$、沿圓周均勻分佈的細彈性弦。圓錐是光滑的，所以它對每一小段的反作用力都垂直於圓錐面；我們亦忽略重力。由對稱性，每一小段的運動方式都相同，所以橡皮圈始終是一個以軸為圓心的圓。設它沿母線量度與頂點的距離為 $s(t)$。由於每條母線與軸成角 $\alpha$，它的半徑便是 $s\sin\alpha$，因此

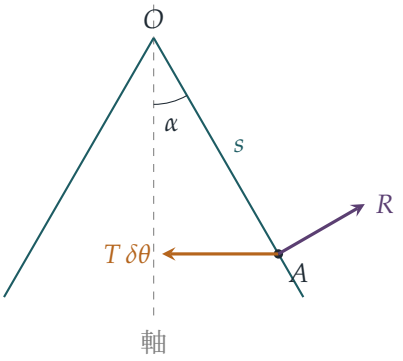
$$x = 2\pi s \sin\alpha.$$

當橡皮圈被拉伸時（$x > l$），胡克定律（Hooke's law）給出張力 $T = \dfrac{\lambda(x-l)}{l}$。

**作用於一小段的力。** 取一小段在軸上所對的角為 $\delta\theta$。它佔整條橡皮圈的 $\delta\theta/2\pi$，所以其質量為 $\dfrac{m\,\delta\theta}{2\pi}$。作用於它的力有兩個：

- 題目給出、由張力產生的力 $T\,\delta\theta$，垂直於軸並指向軸；
- 圓錐的法向反作用力（normal reaction）$R$，垂直於經過該小段的母線。

下圖顯示包含軸和該小段的平面。




該小段被限制在圓錐面上，所以只能沿母線 $OA$ 移動。我們沿這個方向及其垂直方向分解力。力 $T\,\delta\theta$ 垂直於軸，而母線與軸成角 $\alpha$，所以 $T\,\delta\theta$ 與方向 $\overrightarrow{AO}$ 之間的夾角是 $90^\circ - \alpha$。因此：

- 沿 $\overrightarrow{AO}$（指向頂點）的方向，$T\,\delta\theta$ 的分量是 $T\,\delta\theta\cos(90^\circ - \alpha) = T\,\delta\theta\sin\alpha$，而 $R$ 在此方向沒有分量；
- 在垂直於圓錐面的方向，由於該方向沒有運動，$R = T\,\delta\theta\cos\alpha$。

由於 $R \geq 0$，橡皮圈確實一直緊貼圓錐。

**想一想** 為甚麼我們不能直接說向內的力 $T\,\delta\theta$ 令半徑縮小？

該小段不能直接向軸移動，因為圓錐擋住了去路。只有 $T\,\delta\theta$ 沿圓錐面的部分會產生加速度，其餘部分被反作用力抵消。因子 $\sin\alpha$ 正是由此而來；這與光滑斜面上的珠子沿斜面只受 $mg\sin\theta$ 的道理相同。

**運動方程**。 對該小段沿母線應用牛頓第二定律（Newton's second law），並取 $s$ 以遠離頂點為正方向，得

$$\frac{m\,\delta\theta}{2\pi}\frac{\mathrm{d}^2 s}{\mathrm{d}t^2} = -T\,\delta\theta\sin\alpha \qquad \Longrightarrow \qquad \frac{\mathrm{d}^2 s}{\mathrm{d}t^2} = -\frac{2\pi T\sin\alpha}{m}.$$

$\delta\theta$ 被消去，所以每一小段都服從同一條方程，這證實了橡皮圈保持圓形。由於 $x = 2\pi s\sin\alpha$ 而 $\alpha$ 是常數，所以

$$\frac{\mathrm{d}^2 x}{\mathrm{d}t^2} = 2\pi\sin\alpha\frac{\mathrm{d}^2 s}{\mathrm{d}t^2} = -\frac{4\pi^2 T\sin^2\alpha}{m} = -\frac{4\pi^2\lambda}{ml}(x-l)\sin^2\alpha,$$

其中用了 $T = \lambda(x-l)/l$。此式在橡皮圈拉緊時成立，即對 $0 < t < t_0$ 成立，整理後便得所求方程。 □

**求 $t_0$**。 設 $y = x - l$ 及 $\omega^2 = \dfrac{4\pi^2\lambda\sin^2\alpha}{ml}$，則 $\ddot{y} = -\omega^2 y$，即繞 $y = 0$ 的簡諧運動。設 $x_0 > l$ 為初始長度（橡皮圈是被推上去的，所以開始時處於拉伸狀態）。橡皮圈由靜止釋

放，所以 $\dot{y}(0) = 0$，故

$$x - l = (x_0 - l)\cos\omega t.$$

當 $x$ 首次降至 $l$，即 $\cos\omega t = 0$ 首次成立時，橡皮圈變得鬆弛，所以 $\omega t_0 = \frac{\pi}{2}$。由於 $\omega = 2\pi\sin\alpha\sqrt{\lambda/(ml)}$，

$$t_0 = \frac{\pi}{2\omega} = \frac{1}{4\sin\alpha}\sqrt{\frac{ml}{\lambda}}.$$

留意 $t_0$ 與橡皮圈被拉伸了多少無關，這是簡諧運動的典型特徵：無論振幅（amplitude）多大，四分之一週期都是一樣的。

## 驗算

- 單位：$\lambda$ 是一個力，所以 $ml/\lambda$ 的單位是 $\mathrm{kg\,m/N} = \mathrm{s}^2$，而 $t_0$ 是一個時間。
- 極限情況：當 $\alpha \to \frac{\pi}{2}$ 時，圓錐變平成為一個平面，橡皮圈直接向中心收縮，$t_0 = \frac{1}{4}\sqrt{ml/\lambda}$。當 $\alpha \to 0$ 時，圓錐變成圓柱，張力沒有沿圓柱面的分量，而 $t_0 \to \infty$，正合預期。
- 數值驗證：取 $\alpha = 30°$、$m = 1$、$l = 1$ 及 $\lambda = 4$，公式給出 $t_0 = \frac{1}{4}$。對隨機選取的參數，以數值方法逐步求解該小段的運動方程（並另外在三維空間中直接計算一條精細多邊形橡皮圈所受的淨張力），均能準確到多個小數位地重現因子 $\sin\alpha$ 及 $t_0$。

## 常見錯誤

- 忘記圓錐的反作用力，對半徑 $r$ 寫出 $\frac{m\,\delta\theta}{2\pi}\ddot{r} = -T\,\delta\theta$。這樣會完全漏掉因子 $\sin^2\alpha$。
- 把該小段的質量寫成 $m$，而不是 $\frac{m\,\delta\theta}{2\pi}$。
- 以整個週期或半個週期作為 $t_0$：橡皮圈由最大拉伸開始，所以經過四分之一週期便回到自然長度（natural length）。

## STEP 之外

題目給出的力 $T\,\delta\theta$ 是一個普遍事實的小角度版本：一條張力為 $T$、曲率半徑（radius of curvature）為 $\rho$ 的弦，每單位長度向側面施加大小為 $T/\rho$ 的力。把同一想法推廣到曲面，便能解釋為何肥皂泡內的壓強大於外面的壓強（拉普拉斯定律（Laplace's law））。受限於曲面上的運動，最簡潔的處理方法是拉格朗日力學（Lagrangian mechanics），在其中反作用力根本無需出現。Kleppner 和 Kolenkow 所著的 *An Introduction to Mechanics*（《力學導論》）收錄了許多關於受張力的弦及受約束運動的題目。

詳細解答

# STEP II 1987，第 15 題

概率與統計 • 幾何概率

**STEP II 1987，第 15 題**

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

step.maths.org/questions/87-s2-q15

## 關鍵思路

先不要想位置，改為考慮時間。每輛車都會在一段時間區間內佔用交叉口，而碰撞發生的充要條件，正是這兩段時間區間有重疊。把這個重疊條件寫成兩條不等式（inequality），$(u_1, u_2)$ 平面上的碰撞區域就變成夾在兩條過原點直線之間的楔形區域。由於 $u_1, u_2$ 在 $(0, V)$ 上均勻分佈（uniform），點 $(u_1, u_2)$ 便在一個正方形上均勻分佈，所以所求概率（probability）就是正方形中落在楔形區域內的面積比例，而題目的三種情況，對應的正是兩條直線從正方形哪一邊穿出。

## 解答

### 碰撞條件

我們把火車和貨車看作以恆定速率 $u_1, u_2 > 0$ 沿兩條相交於一點的直路行駛（交叉口的寬度忽略不計，或已計入車長之內）。碰撞發生當且僅當在同一時刻，兩輛車各有某部分正處於交叉口。

火車車頭在時刻 $d_1/u_1$ 到達交叉口，車尾在時刻 $(d_1 + l_1)/u_1$ 駛離，因此火車佔用交叉口的時間為

$$\frac{d_1}{u_1} < t < \frac{d_1 + l_1}{u_1}.$$

同樣，貨車佔用交叉口的時間為 $d_2/u_2 < t < (d_2 + l_2)/u_2$。兩個開區間（open interval）$(a_1, b_1)$ 與 $(a_2, b_2)$ 有重疊，當且僅當每一個都在另一個結束之前開始：$a_1 < b_2$ 且 $a_2 < b_1$。所以碰撞發生當且僅當

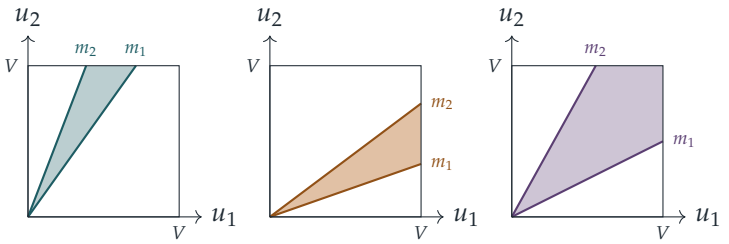
$$\frac{d_1}{u_1} < \frac{d_2 + l_2}{u_2} \quad \text{且} \quad \frac{d_2}{u_2} < \frac{d_1 + l_1}{u_1}.$$

兩邊乘以正數 $u_1u_2$ 再整理，即得

$$\boxed{\frac{d_2}{d_1+l_1}\,u_1 \;<\; u_2 \;<\; \frac{d_2+l_2}{d_1}\,u_1.}$$

（若某條不等式取等號，兩車只是車頭碰車尾地剛好接觸；這種情況的概率為零，所以算不算進去都沒有影響。）

記 $m_1 = \dfrac{d_2}{d_1+l_1}$ 及 $m_2 = \dfrac{d_2+l_2}{d_1}$。由於 $d_2 < d_2+l_2$ 且 $d_1+l_1 > d_1$，有 $m_1 < m_2$，因此碰撞區域就是第一象限中夾在直線 $y = m_1x$ 與 $y = m_2x$ 之間的楔形區域。下圖以陰影顯示它在正方形 $0 < x, y < V$ 內的部分，這正是下一步所需要的。




(A) $d_1+l_1 \le d_2$　(B) $d_2+l_2 \le d_1$　(C) 其他情況

### 火車車尾比貨車車頭更接近交叉口時的概率

由於 $u_1$ 與 $u_2$ 互相獨立（independent），且各自在 $(0, V)$ 上均勻分佈，數對 $(u_1, u_2)$ 在正方形 $(0, V)^2$ 上的聯合密度（joint density）為常數 $1/V^2$。所以碰撞的概率為

$$\frac{\text{楔形區域在正方形內的面積}}{V^2}.$$

**想一想** 這兩條直線分別從正方形的哪一邊穿出？

若 $m \ge 1$，直線 $y = mx$ 從正方形的上邊穿出（於 $(V/m, V)$）；若 $m \le 1$，則從右邊穿出（於 $(V, mV)$）。在這情況下 $d_1+l_1 < d_2$，所以 $m_1 = d_2/(d_1+l_1) > 1$，從而 $m_2 > m_1 > 1$：兩條直線都從上邊穿出，如圖 (A) 所示。

此時楔形區域在正方形內的部分，是頂點為 $(0,0)$、$(V/m_2, V)$ 和 $(V/m_1, V)$ 的三角形。它在上邊的底長為 $V/m_1 - V/m_2$，高為 $V$，因此

$$\begin{aligned}\mathrm{P}(\text{碰撞}) &= \frac{1}{V^2}\cdot\frac{1}{2}V\left(\frac{V}{m_1}-\frac{V}{m_2}\right) = \frac{1}{2}\left(\frac{d_1+l_1}{d_2}-\frac{d_1}{d_2+l_2}\right)\\ &= \frac{(d_1+l_1)(d_2+l_2)-d_1d_2}{2d_2(d_2+l_2)} = \frac{l_1l_2+l_2d_1+l_1d_2}{2d_2(l_2+d_2)},\end{aligned}$$

因為 $d_1d_2$ 項互相抵消。這就是所求的結果。

### 其餘兩種情況

**貨車車尾比火車車頭更接近交叉口，**$d_2 + l_2 < d_1$。 此時 $m_2 = (d_2 + l_2)/d_1 < 1$，所以 $m_1 < m_2 < 1$，兩條直線都從右邊穿出，如圖 (B) 所示。楔形區域是頂點為 $(0,0)$、$(V, m_1V)$、$(V, m_2V)$ 的三角形：在邊 $x = V$ 上的底長為 $(m_2 - m_1)V$，高為 $V$。所以

$$\mathrm{P}(\text{碰撞}) = \frac{1}{2}(m_2 - m_1) = \frac{1}{2}\left(\frac{d_2 + l_2}{d_1} - \frac{d_2}{d_1 + l_1}\right) = \frac{l_1 l_2 + l_1 d_2 + l_2 d_1}{2d_1(l_1 + d_1)}.$$

這正是把第一個答案中火車與貨車的角色對調（$l_1 \leftrightarrow l_2$，$d_1 \leftrightarrow d_2$）所得的結果；這是理所當然的，因為對調兩輛車就相當於把圖形沿 $y = x$ 作反射。

**兩輛車都不完全在對方前面，**$d_2 \le d_1 + l_1$ **且** $d_1 \le d_2 + l_2$。 此時 $m_1 \le 1 \le m_2$，所以直線 $y = m_1 x$ 從右邊穿出，而 $y = m_2 x$ 從上邊穿出，如圖 (C) 所示。最簡單的做法是從正方形中減去兩個不會碰撞的三角形：在 $y = m_1 x$ 下方的那個面積為 $\frac{1}{2}V \cdot m_1 V$，而在 $y = m_2 x$ 左方（即它與 $u_2$ 軸之間）的那個面積為 $\frac{1}{2}V \cdot V/m_2$。因此

$$\mathrm{P}(\text{碰撞}) = 1 - \frac{m_1}{2} - \frac{1}{2m_2} = 1 - \frac{d_2}{2(d_1 + l_1)} - \frac{d_1}{2(d_2 + l_2)}.$$

在各情況的交界處，答案互相吻合：當 $d_2 = d_1 + l_1$（即 $m_1 = 1$）時，第一條和第三條公式都給出 $\frac{1}{2} - \frac{d_1}{2(d_2 + l_2)}$；在 $d_1 = d_2 + l_2$ 處亦然。留意所有概率都與 $V$ 無關：把兩個速率按同一比例放大或縮小，並不會改變兩車是否相撞。

## 驗算

- 取 $l_1 = 1$、$l_2 = 2$、$d_1 = 1$、$d_2 = 5$（情況 A），公式給出 $\frac{9}{70} \approx 0.1286$；以 $4 \times 10^5$ 組隨機速率作模擬，得 0.1283。在情況 B 和 C 中的模擬（例如 $l_1 = l_2 = 1$、$d_1 = 1$、$d_2 = 1.5$，得 0.425）亦吻合至 0.002 以內。
- 在情況 A 中令 $l_1, l_2 \to 0$，概率為 0：兩個點只有在 $u_2/u_1 = d_2/d_1$ 恰好成立時才會在交叉口相遇。
- 在情況 C 中取 $d_1 = d_2$ 而兩車都很長，則 $m_1 \to 0$ 且 $m_2 \to \infty$，所以概率趨於 1，這是合理的。

## 常見錯誤

- 只寫一條不等式，例如「火車在貨車駛離交叉口之前到達」；兩個區間有重疊需要兩個條件同時成立。
- 在情況 C 中套用三角形面積公式，但那裏的區域其實是四邊形；應先確定每條直線從正方形哪一邊穿出。

## STEP 之外

這裏的方法是把關於獨立均勻隨機變量的概率轉化為面積，稱為幾何概率（geometric probability）；它是把 $\mathrm{P}((X,Y)\in A)$ 計算為聯合密度的二重積分（double integral）的最簡單例子。同一想法也是一些經典問題的基礎，例如布豐（Buffon）投針問題，以及兩人在隨機時刻到達的「相遇問題」。Grimmett 與 Welsh 所著的 *Probability: An Introduction*（《概率導論》）正是以這種方式發展聯合密度及其應用。

詳細解答

# STEP II 1987，第 16 題

概率與統計 • 貝葉斯定理

> **STEP II 1987，第 16 題**
>
> 繼續閱讀之前，請先看題目（原題為英文）。題目載於 STEP Support Programme 的試題資料庫：
>
> step.maths.org/questions/87-s2-q16

## 關鍵思路

這是貝葉斯定理（Bayes' theorem）最樸素的形式。兩個「是」的報告可以有兩種來源：說法屬實而兩位朋友都說真話，或者說法不實而兩人都說謊。我們按每條途徑的先驗概率（prior probability）和出現這對報告的概率為它加權，然後問總數中有多大比例來自「屬實」這條途徑。第二部分的做法相同。唯一的新要點是，朋友的回答只涉及 $A$，所以我們必須按 $A$ 是否在雪櫃裏，把三種可能情況分類。

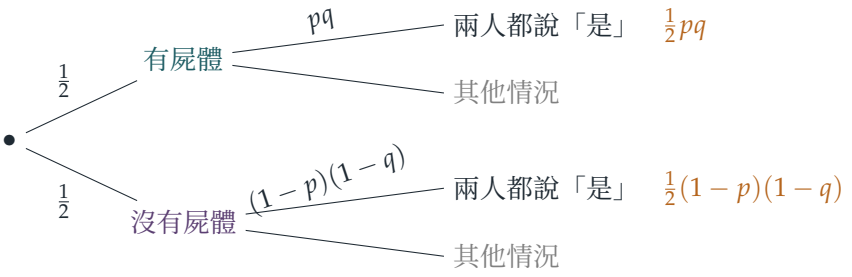
## 解答

### 第一個估計

設 $F$ 為雪櫃裏有屍體這一事件，$Y$ 為 $P$ 和 $Q$ 都說有屍體這一事件。今天早上 $\Pr(F) = \Pr(F') = \frac{1}{2}$。

我們把每位朋友看作回答一條是非題：說真話就是給出正確答案，說謊就是給出錯誤答案。朋友是否誠實與雪櫃裏有甚麼無關，而且 $P$ 和 $Q$ 的行為互相獨立（independent）。所以：

- 若有屍體，兩人都說「是」當且僅當兩人都說真話，所以 $\Pr(Y \mid F) = pq$；
- 若沒有屍體，兩人都說「是」當且僅當兩人都說謊，所以 $\Pr(Y \mid F') = (1-p)(1-q)$。




根據條件概率（conditional probability）的定義（貝葉斯定理），

$$\begin{aligned}\Pr(F \mid Y) &= \frac{\Pr(F)\Pr(Y \mid F)}{\Pr(F)\Pr(Y \mid F) + \Pr(F')\Pr(Y \mid F')} = \frac{\frac{1}{2}pq}{\frac{1}{2}pq + \frac{1}{2}(1-p)(1-q)} \\ &= \frac{pq}{pq + 1 - p - q + pq} = \frac{pq}{1 - p - q + 2pq},\end{aligned}$$

這就是所求的估計。由於先驗概率是平分的，因子 $\frac{1}{2}$ 互相抵消。

**想一想** 為甚麼答案不就是 $pq$，即兩人都說真話的概率？

因為我們已經知道一些資訊：他們的說法一致。恰好一人說謊的結果現在已不可能出現，因為那會產生一個「是」和一個「否」。我們必須在剩下的情況上重新按比例調整。在兩條剩下的途徑中，「兩人都說真話」的權重是 $pq$，「兩人都說謊」的權重是 $(1-p)(1-q)$。除以兩者之和，正是這個重新調整。（先驗概率 $\frac{1}{2}$ 也有作用。若先驗概率不平分，它們就不會抵消。）

### 第二個估計

現在肯定有屍體。三個等可能的情況，各自的先驗概率為 $\frac{1}{3}$，分別是

$$\text{只有 } A, \qquad \text{只有 } B, \qquad A \text{ 和 } B \text{ 都在.}$$

設 $Y_A$ 為兩位朋友都說 $A$ 在雪櫃裏這一事件。按照與之前相同的模型，他們的回答只取決於 $A$ 是否在雪櫃裏：

- 在「只有 $A$」和「兩者都在」的情況下，$A$ 在雪櫃裏，所以每種情況都有 $\Pr(Y_A) = pq$；
- 在「只有 $B$」的情況下，$A$ 不在，所以 $\Pr(Y_A) = (1-p)(1-q)$。

因此

$$\begin{aligned}\Pr(\text{兩者都在} \mid Y_A) &= \frac{\frac{1}{3}pq}{\frac{1}{3}pq + \frac{1}{3}(1-p)(1-q) + \frac{1}{3}pq} \\ &= \frac{pq}{2pq + 1 - p - q + pq} = \frac{pq}{1 - p - q + 3pq}.\end{aligned}$$

留意這些報告並沒有直接告訴我們任何關於 $B$ 的資訊。它們唯一的作用，是改變 $A$ 在場的概率。根據第一部分（以先驗概率 $\frac{2}{3}$ 代替 $\frac{1}{2}$），這個概率變成 $2pq/(1-p-q+3pq)$。已知 $A$ 在場時，「只有 $A$」和「兩者都在」仍然等可能，所以把它減半。這樣得出相同的答案。

## 驗算

- 若 $p=q=\frac{1}{2}$，朋友的話毫無用處，答案回到先驗概率：$\frac{1}{4}/\frac{1}{2}=\frac{1}{2}$ 及 $\frac{1}{4}/\frac{3}{4}=\frac{1}{3}$。
- 若 $p=q=1$，朋友絕不出錯，答案變成 1（有屍體）和 $\frac{1}{2}$（$A$ 在，所以 $B$ 是否也在是一半一半）。
- 當 $p=q=\frac{3}{4}$ 時，答案是 $\frac{9}{10}$ 和 $\frac{9}{19}$。對 200 組隨機的 $(p,q)$，逐一列舉所有狀態與說真話情況組合的精確計算均與答案吻合。以 $p=0.7$、$q=0.6$ 進行 $4\times10^5$ 次試驗的模擬，第二部分得出 0.4375，與 $\frac{0.42}{0.96}=0.4375$ 相符。

## 常見錯誤

- 答 $pq$，忘記了「兩人都說謊」同樣會產生兩個「是」。
- 在第二部分，分子只放「兩者都在」的情況，卻在分母中漏掉「只有 $A$」的情況。這兩種情況都與 $A$ 在場相符。

## STEP 之外

這就是貝葉斯更新（Bayesian updating）。先驗信念乘以證據的似然（likelihood），然後歸一化。一個簡潔的寫法是勝算（odds）形式：後驗勝算等於先驗勝算乘以似然比（likelihood ratio）$\frac{pq}{(1-p)(1-q)}$。這樣寫就清楚顯示，互相獨立的證人只需把各自的比率相乘。同樣的計算是醫學檢測、垃圾郵件過濾器以及法庭上綜合證據的基礎。Blitzstein 與 Hwang 的 *Introduction to Probability*（《概率導論》）有一章精彩地講解條件概率，收錄了許多正是這類風格的趣題。

詳細解答

# STEP III 1987，第 1 題

純數學 • 階乘的整除性

**STEP III 1987，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/87-s3-q1

## 關鍵思路

階乘（factorial）$(n-1)! = 1 \times 2 \times \cdots \times (n-1)$ 包含了每一個小於 $n$ 的正整數。如果 $n$ 可以寫成兩個小於 $n$ 的不同數的乘積，這兩個數都會在階乘中出現，於是 $n$ 整除這個階乘。所以問題的核心其實是：哪些 $n$ 不能這樣拆開？質數（prime）顯然不能，而質數的平方則要仔細看看。親手試算最初十來個數值，答案便呼之欲出。

## 解答

### 先做實驗

值得先列一個簡單的表：

| $n$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| $(n-1)!$ | 1 | 1 | 2 | 6 | 24 | 120 | 720 | 5040 |
| $n \mid (n-1)!$? | 是 | 否 | 否 | 否 | 否 | 是 | 否 | 是 |

不成立的情況是 $2, 3, 4, 5, 7$：即目前為止的質數，再加上 4。我們現在證明這個規律會一直延續下去。

**斷言。** $n$ 不整除 $(n-1)!$，當且僅當 $n$ 是質數或 $n = 4$。

$n = 1$

這時 $(n-1)! = 0! = 1$，而 1 整除 1。所以 $n = 1$ 不屬於這個集合。

### $n = p$ 為質數

$(p-1)!$ 的每個因子 $1, 2, \ldots, p-1$ 都是小於 $p$ 的正數，所以沒有一個能被 $p$ 整除。由於 $p$ 是質數，假如 $p$ 整除這個乘積，它就必須整除其中一個因子（歐幾里得引理，Euclid's lemma）。因此 $p \nmid (p-1)!$，即每個質數都屬於這個集合。

### $n$ 為合成數

設 $p$ 為 $n$ 的最小質因數，並寫成 $n = pm$。因為 $n$ 是合成數（composite），所以 $m > 1$，而且 $m$ 的每個質因數都不小於 $p$，故 $m \geq p$。以下分兩種情況。

**情況 1：** $m > p$。 這時 $1 < p < m < n$，所以 $p$ 和 $m$ 是數列 $1, 2, \ldots, n-1$ 中兩個不同的數。因此它們的乘積 $pm = n$ 整除 $(n-1)!$。

**情況 2：** $m = p$，**即** $n = p^2$。

> **想一想** 為甚麼情況 1 的論證在這裏行不通？可以用甚麼來代替？

因子 $p$ 在數列 $1, \ldots, n-1$ 中只出現一次，所以不能把它用兩次。我們改用 $p$ 和 $2p$。若 $p \geq 3$，則

$$2p < p \cdot p = n,$$

所以 $p$ 和 $2p$ 是數列 $1, 2, \ldots, n-1$ 中兩個不同的數。因此它們的乘積 $2p^2 = 2n$ 整除 $(n-1)!$，於是 $n$ 也整除它。

若 $p = 2$，這個論證便失效，因為 $2p = 4 = n$ 不在數列之中。事實上，$n = 4$ 時得 $3! = 6$，而 $4 \nmid 6$。所以 4 屬於這個集合。

### 結論

除 4 以外，每個合成數 $n$ 都整除 $(n-1)!$，而 $n = 1$ 亦然。因此所求的集合是

$$\{ n : n \text{ 是質數} \} \cup \{4\} = \{2, 3, 4, 5, 7, 11, 13, 17, \ldots\}.$$

## 驗算

上面的表在 $n \leq 8$ 時與斷言吻合。再試一個質數的平方：$n = 9$ 時得 $8! = 40320 = 9 \times 4480$，正如情況 2 所預測（用的是因子 3 和 6）。用電腦對每個 $n \leq 2000$ 直接計算 $(n-1)! \bmod n$，結果找不到任何違反斷言的例外。

## 常見錯誤

- 論證「寫成 $n = ab$，其中 $a, b < n$，所以 $ab \mid (n-1)!$」時，沒有堅持要求 $a \neq b$。這個論證會錯誤地「證明」$4 \mid 3!$。
- 忘記了 $n = 1$。由於 $0! = 1$，它不屬於這個集合。
- 說質數 $p$ 不整除 $(p-1)!$ 是「因為 $p$ 不是其中一個因子」。真正的原因是 $p$ 是質數，而且不整除任何一個因子；合成數即使不是乘積中的因子，也可以整除該乘積。

## STEP 之外

對於質數，有一個更精確的結果，稱為威爾遜定理（Wilson's theorem）：$(p-1)! \equiv -1 \pmod{p}$。它與本題合起來，便得出一個（非常慢的）質性測試（primality test）：$n > 1$ 是質數，當且僅當 $(n-1)! \equiv -1 \pmod{n}$。威爾遜定理的證明方法，是把每個數與它模 $p$ 的逆元（inverse）配對，這讓人初嘗整數模 $p$ 的羣（group）結構。Davenport 的 *The Higher Arithmetic* 對此有十分精彩的論述。

詳細解答

# STEP III 1987，第 2 題

純數學 • 三角函數歸約公式

> **STEP III 1987，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/87-s3-q2

## 關鍵思路

遞推公式（reduction formula）的來源，是先對「邊界項」$\cos^m x\cos nx$ 求導，再用複角公式（compound-angle formula）把 $\sin x\cos nx$ 化為含有 $\sin(n-1)x$ 的式子。有了這條公式之後，兩部分的關鍵都在於計算該邊界項在積分上下限的值：在 0 和 $\pi$ 處，它會互相抵消還是保留下來，取決於奇偶性（parity）；而在 $\frac{\pi}{2}$ 處，因子 $\cos^m x$ 令它變為零。

（題目印出的公式寫作 $-\cos^m\cos nx$；應為 $-\cos^m x\cos nx$。）

## 解答

### 遞推公式

用積法則（product rule）對 $\cos^m x\cos nx$ 求導：

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\cos^m x\cos nx\right) = -m\cos^{m-1}x\,\sin x\cos nx \;-\; n\cos^m x\sin nx.$$

第一項還不是我們想要的形式。由複角公式

$$\sin(n-1)x = \sin nx\cos x - \cos nx\sin x \quad \text{可得} \quad \sin x\cos nx = \cos x\sin nx - \sin(n-1)x.$$

代入後，

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}x}\left(\cos^m x\cos nx\right) &= -m\cos^m x\sin nx + m\cos^{m-1}x\sin(n-1)x - n\cos^m x\sin nx\\ &= -(m+n)\cos^m x\sin nx + m\cos^{m-1}x\sin(n-1)x.\end{aligned}$$

對兩邊積分（這裏需要 $m\geq 1$ 及 $n\geq 1$，使得 $m-1, n-1$ 均為非負），

$$\cos^m x\cos nx = -(m+n)I_{m,n} + mI_{m-1,n-1},$$

即

$$(m+n)I_{m,n} = -\cos^m x \cos nx + mI_{m-1,n-1},$$

這正是所需的結果（慣常的任意常數已併入不定積分（indefinite integral）之中）。

### 第 (i) 部分

設 $J_{m,n} = \int_0^\pi \cos^m x \sin nx \,\mathrm{d}x$。把公式應用於 0 與 $\pi$ 之間，並利用 $\cos^m \pi \cos n\pi = (-1)^m(-1)^n$，對 $m, n \geq 1$ 有：

$$(m+n)J_{m,n} = 1 - (-1)^{m+n} + mJ_{m-1,n-1}.$$

若 $m$ 與 $n$ 的奇偶性相同，則 $m+n$ 為偶數，邊界項消失，於是

$$J_{m,n} = \frac{m}{m+n} J_{m-1,n-1}.$$

此外，$m-1$ 與 $n-1$ 的奇偶性仍然相同，所以我們可以繼續遞推下去。

> **想一想** 這一連串的遞推會在哪裏停下來？為甚麼在那裏的值是零？

每一步都把兩個指標各減 1，所以經過 $\min(m,n)$ 步之後，其中一個指標為 0，而另一個指標因奇偶性相同，必為偶數。因此只需證明對每個偶數 $k \geq 0$，都有 $J_{k,0} = 0$ 及 $J_{0,k} = 0$。

- $J_{k,0} = \int_0^\pi \cos^k x \sin 0 \,\mathrm{d}x = 0$，因為被積函數（integrand）恆等於零。
- 對偶數 $k \geq 2$，$J_{0,k} = \int_0^\pi \sin kx \,\mathrm{d}x = \left[-\frac{1}{k}\cos kx\right]_0^\pi = \frac{1}{k}\left(1-(-1)^k\right) = 0$。

因此，對 $\min(m,n)$ 使用歸納法（induction），可知只要 $m, n$ 同為偶數或同為奇數，便有 $J_{m,n} = 0$。

*或用其他方法*：代換 $x = \pi - t$ 把 $\cos^m x$ 變為 $(-1)^m \cos^m t$，把 $\sin nx$ 變為 $-(-1)^n \sin nt$，所以 $J_{m,n} = (-1)^{m+n+1} J_{m,n}$，由此可見當 $m+n$ 為偶數時必有 $J_{m,n} = 0$。被積函數關於 $x = \frac{\pi}{2}$ 反對稱（antisymmetric）。

### 第 (ii) 部分

為了引入餘弦，寫 $\sin^2 x = 1 - \cos^2 x$：

$$\int_0^{\pi/2} \sin^2 x \sin 3x \,\mathrm{d}x = K_{0,3} - K_{2,3}, \qquad K_{m,n} = \int_0^{\pi/2} \cos^m x \sin nx \,\mathrm{d}x.$$

對 $m \geq 1$，邊界項為 $-\left[\cos^m x \cos nx\right]_0^{\pi/2} = 0 - (-1) = 1$，因為 $\cos\frac{\pi}{2} = 0$。所以

$$(m+n)K_{m,n} = 1 + mK_{m-1,n-1} \qquad (m, n \geq 1).$$

由 $K_{0,1} = \int_0^{\pi/2} \sin x \,\mathrm{d}x = 1$ 出發：

$$3K_{1,2} = 1 + K_{0,1} = 2, \qquad K_{1,2} = \tfrac{2}{3},$$
$$5K_{2,3} = 1 + 2K_{1,2} = \tfrac{7}{3}, \qquad K_{2,3} = \tfrac{7}{15}.$$

另外 $K_{0,3} = \left[-\frac{1}{3}\cos 3x\right]_0^{\pi/2} = \frac{1}{3}$，因為 $\cos\frac{3\pi}{2} = 0$。因此

$$\int_0^{\pi/2} \sin^2 x \sin 3x \,\mathrm{d}x = \frac{1}{3} - \frac{7}{15} = -\frac{2}{15}.$$

## 驗算

第 (ii) 部分可以直接用積化和差公式（product-to-sum formulae）計算：$\sin^2 x \sin 3x = \frac{1}{2}\sin 3x - \frac{1}{4}(\sin 5x + \sin x)$，它在 $[0, \frac{\pi}{2}]$ 上的積分為 $\frac{1}{6} - \frac{1}{4}\left(\frac{1}{5} + 1\right) = -\frac{2}{15}$。答案為負數是合理的：在 $(\frac{\pi}{3}, \frac{\pi}{2})$ 上 $\sin 3x < 0$，而這正是 $\sin^2 x$ 最大的地方。用辛普森法則（Simpson's rule）算得 $-0.133333$，並且對許多 $m, n$ 以數值方法驗證了遞推公式和第 (i) 部分。

## 常見錯誤

- 在 (ii) 中嘗試使用代換 $x = \frac{\pi}{2} - t$：它把 $\sin^2 x$ 變為 $\cos^2 t$，卻把 $\sin 3x$ 變為 $-\cos 3t$，所以該積分並不屬於 $I_{m,n}$ 的形式。
- 忘記遞推公式只適用於 $m, n \geq 1$，所以 (i) 中的一連串遞推需要另外處理有一個指標等於 0 的基本情況（base case）。

## STEP 之外

像這樣的遞推公式，只是一個更大的故事的開端：三角函數乘積在一整個週期上的積分，正是傅立葉級數（Fourier series）背後的正交關係（orthogonality relations）；在那裏，像 (i) 中那樣的對稱性論證，讓人一眼就能判斷哪些係數會消失。Boyce 與 DiPrima 所著的 *Elementary Differential Equations*（初等微分方程）對傅立葉級數及其在求解微分方程（differential equation）上的應用，有一段淺白易讀的介紹。

詳細解答

# STEP III 1987，第 3 題

純數學 • 輻角與梅欽公式

**STEP III 1987，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/87-s3-q3

本題在 *Beyond STEP* 第 5.1 節梅欽公式中有完整的引導式講解：https://beyondstep.keithmcnulty.org/zh-hant/ch05/machin.html

## 關鍵思路

兩部分都建基於複數（complex number）的極式（polar form）。在 (i) 中，點 $(|x|, |y|)$ 與原點的距離為 $|z|$，所以 $|x|\cos\alpha + |y|\sin\alpha$ 等於 $|z|$ 乘以一個餘弦值。在 (ii) 中，複數相乘時，輻角（argument）相加。因此，若由 $5-\mathrm{i}$ 和 $1+\mathrm{i}$ 這類數組成的乘積結果非常接近實數，它的輻角便會給出一條關於反正切（inverse tangent）的恆等式。唯一真正需要小心的，是輻角中 $2\pi$ 的不確定性。

## 解答

### 第 (i) 部分

設 $r = |z| = \sqrt{x^2+y^2}$。若 $r = 0$，則 $x = y = 0$，各項均為 0，故以下假設 $r > 0$。

*右邊的不等式。*兩邊均為非負數，所以只需比較兩邊的平方：

$$(|x| + |y|)^2 = x^2 + y^2 + 2|x||y| \geq x^2 + y^2 = |z|^2,$$

這是因為 $|x||y| \geq 0$。取非負平方根，得 $|z| \leq |x| + |y|$，等號成立當且僅當 $x = 0$ 或 $y = 0$。

*左邊的不等式。*點 $(|x|, |y|)$ 位於閉第一象限內，與原點的距離為 $r$，所以存在角 $\varphi \in [0, \frac{\pi}{2}]$ 使得

$$|x| = r\cos\varphi, \qquad |y| = r\sin\varphi.$$

於是，由複角公式（compound-angle formula），

$$|x|\cos\alpha + |y|\sin\alpha = r(\cos\varphi\cos\alpha + \sin\varphi\sin\alpha) = r\cos(\alpha-\varphi) \le r = |z|,$$

因為對任意實數 $\alpha$ 都有 $\cos(\alpha-\varphi) \le 1$。這就證明了兩條不等式。左邊的等號在 $\alpha = \varphi + 2k\pi$ 時成立，即 $(\cos\alpha, \sin\alpha)$ 與 $(|x|, |y|)$ 方向相同之時。

（左邊不等式的另一證法：由柯西–施瓦茲不等式（Cauchy–Schwarz inequality），$|x|\cos\alpha + |y|\sin\alpha \le \sqrt{x^2+y^2}\,\sqrt{\cos^2\alpha+\sin^2\alpha} = |z|$。）

**第 (ii) 部分：第一條公式**

逐步展開：

$$\begin{aligned}(5-\mathrm{i})^2 &= 25 - 10\mathrm{i} - 1 = 24 - 10\mathrm{i},\\ (5-\mathrm{i})^4 &= (24-10\mathrm{i})^2 = 576 - 480\mathrm{i} - 100 = 476 - 480\mathrm{i},\\ (5-\mathrm{i})^4(1+\mathrm{i}) &= 476 + 476\mathrm{i} - 480\mathrm{i} + 480 = 956 - 4\mathrm{i} = 4(239-\mathrm{i}).\end{aligned}$$

現在比較輻角。設 $\theta = \tan^{-1}\frac{1}{5}$。由於 $5-\mathrm{i}$ 位於第四象限，它的輻角為 $-\theta$，而 $1+\mathrm{i}$ 的輻角為 $\frac{\pi}{4}$。同理，$4(239-\mathrm{i})$ 的輻角為 $-\tan^{-1}\frac{1}{239}$。複數相乘時輻角相加，但可能相差 $2\pi$ 的整數倍，所以

$$\frac{\pi}{4} - 4\theta = -\tan^{-1}\frac{1}{239} + 2k\pi \quad \text{對於某個整數} k.$$

**想一想** 我們怎樣知道 $k = 0$？

我們估計兩邊的範圍。由於當 $0 < t < \frac{\pi}{2}$ 時 $\tan t > t$，我們有 $0 < \theta < \frac{1}{5}$，所以 $\frac{\pi}{4} - 4\theta$ 介乎 $\frac{\pi}{4} - \frac{4}{5} > -1$ 與 $\frac{\pi}{4} < 1$ 之間。另外，$-\tan^{-1}\frac{1}{239}$ 介乎 $-1$ 與 $0$ 之間。$(-1,1)$ 內的兩個數相差小於 $2 < 2\pi$，所以 $2k\pi = 0$，即 $k = 0$。整理後得

$$\frac{\pi}{4} = 4\tan^{-1}\frac{1}{5} - \tan^{-1}\frac{1}{239},$$

正是所要證明的結果。

**第 (ii) 部分：第二條公式**

按同樣的道理，我們需要一個乘積，其輻角為

$$\frac{\pi}{4} - 3\tan^{-1}\frac{1}{4} - \tan^{-1}\frac{1}{20} - \tan^{-1}\frac{1}{1985},$$

即 $(4-\mathrm{i})^3(20-\mathrm{i})(1985-\mathrm{i})(1+\mathrm{i})$，並希望它是一個正實數。把因子 $1985-\mathrm{i}$ 留到最後：

$$\begin{aligned}(4-\mathrm{i})^2 &= 16-8\mathrm{i}-1 = 15-8\mathrm{i},\\ (4-\mathrm{i})^3 &= (15-8\mathrm{i})(4-\mathrm{i}) = 60-15\mathrm{i}-32\mathrm{i}-8 = 52-47\mathrm{i},\\ (4-\mathrm{i})^3(20-\mathrm{i}) &= 1040-52\mathrm{i}-940\mathrm{i}-47 = 993-992\mathrm{i},\\ (4-\mathrm{i})^3(20-\mathrm{i})(1+\mathrm{i}) &= 993+993\mathrm{i}-992\mathrm{i}+992 = 1985+\mathrm{i}.\end{aligned}$$

所以到目前為止的乘積是 $1985+\mathrm{i}$，而最後一個因子 $1985-\mathrm{i}$ 正是它的共軛（conjugate）：整個乘積為 $1985^2+1$，是正實數。換一個等價的做法，不乘最後一個因子，直接比較輻角：

$$\frac{\pi}{4}-3\varphi-\psi = \tan^{-1}\frac{1}{1985}+2k\pi, \qquad \varphi=\tan^{-1}\frac{1}{4},\ \psi=\tan^{-1}\frac{1}{20}.$$

同樣有 $k=0$：由於 $0<\varphi<\frac14$ 及 $0<\psi<\frac{1}{20}$，左邊介乎 $\frac{\pi}{4}-\frac45>-1$ 與 $\frac{\pi}{4}<1$ 之間，而右邊減去 $2k\pi$ 後位於 $(0,1)$ 內。整理後得

$$\frac{\pi}{4} = 3\tan^{-1}\frac14+\tan^{-1}\frac{1}{20}+\tan^{-1}\frac{1}{1985}.$$

## 驗算

- 模（modulus）很適合用來檢查展開的結果。$|5-\mathrm{i}|^8\,|1+\mathrm{i}|^2 = 26^4\cdot 2 = 913\,952$，而 $956^2+4^2 = 913\,952$ 亦然。同樣地，$17^3\cdot 401\cdot 2 = 3\,940\,226 = 1985^2+1$。
- 數值上，$4\tan^{-1}\frac15\approx 0.789582$，減去 $\tan^{-1}\frac{1}{239}\approx 0.004184$ 得 $0.785398\approx\frac{\pi}{4}$。另外，$3\tan^{-1}\frac14+\tan^{-1}\frac{1}{20}\approx 0.784894$，加上 $\tan^{-1}\frac{1}{1985}\approx 0.000504$ 同樣得到 $\frac{\pi}{4}$。
- 在 (i) 中，試取 $z=3-4\mathrm{i}$：不等式變成 $3\cos\alpha+4\sin\alpha\le 5\le 7$，而 $3\cos\alpha+4\sin\alpha = 5\cos(\alpha-\varphi)$ 在 $\tan\alpha=\frac43$ 時取得 5。

## 常見錯誤

- 把 $\arg(zw)=\arg z+\arg w$ 當作精確的等式，卻從不檢查 $2\pi$ 的倍數。正是對各個反正切值的範圍估計，才令最後的等式成立。
- 以 $5+\mathrm{i}$ 代替 $5-\mathrm{i}$：這樣 $(5+\mathrm{i})^4(1+\mathrm{i}) = -4+956\mathrm{i}$，結果接近純虛數而非接近實數。正負號必須選得令乘積接近實數。
- 在 (i) 中，忘記 $\alpha$ 是任意的，所以 $\cos\alpha$ 或 $\sin\alpha$ 可能是負數。餘弦形式可一次過處理所有 $\alpha$。

## STEP 之外

第一條公式是梅欽公式（Machin's formula，1706 年）。級數 $\tan^{-1} x = x - \frac{x^3}{3} + \frac{x^5}{5} - \cdots$ 在 $x$ 很小時收斂得很快，前人把它與梅欽公式結合，以人手計算出 $\pi$ 的許多位小數。尋找這類公式，其實是關於高斯整數（Gaussian integers）$a + b\mathrm{i}$ 的因式分解問題，當中 $1 + \mathrm{i}$、$5 \pm \mathrm{i}$ 和 $4 \pm \mathrm{i}$ 的表現有如質數。這是代數數論（algebraic number theory）的初步體驗。Davenport 的 *The Higher Arithmetic* 討論了二平方和與高斯整數，而 Needham 的 *Visual Complex Analysis* 則展示了複數相乘的幾何意義。

詳細解答

# STEP III 1987，第 4 題

純數學 • 旋轉體的體積

**STEP III 1987，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：

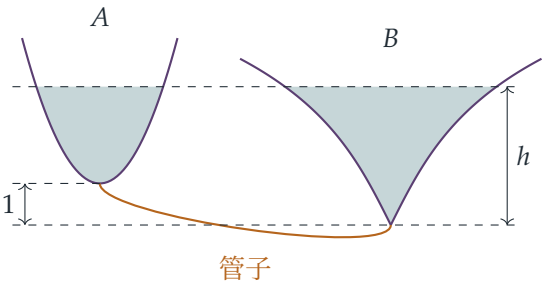
step.maths.org/questions/87-s3-q4

## 關鍵思路

水只是由一個漏斗流到另一個漏斗，所以水的體積守恆（conserved）。每個漏斗都是繞 $y$ 軸旋轉而成的旋轉體（solid of revolution），因此水深為 $d$ 時，水的體積是 $\int_0^d \pi x^2 \, \mathrm{d}y$，其中 $x^2$ 要以 $y$ 表示。求得兩個體積作為深度的函數之後，唯一還要小心的，就是兩個不同的基準高度：$A$ 和 $B$ 中的水深相差一個單位。

## 解答

我們假設管子的體積可以忽略不計，並把題目中第二個「closed」理解為「opened」（否則水根本無法流動）。由每個漏斗的底部起向上量度 $y$。



### 以深度表示的體積函數

對於 $A$，曲線為 $y = x^2$，所以在高度 $y$ 處的水平截面是一個圓盤，面積為 $\pi x^2 = \pi y$。$A$ 中水深為 $d$ 時，水的體積是

$$V_A(d) = \int_0^d \pi y \,\mathrm{d}y = \frac{\pi d^2}{2}.$$

對於 $B$，曲線 $y = 2\sinh^{-1} x$（其中 $x > 0$）的反函數（inverse）是 $x = \sinh \frac{y}{2}$。利用 $\sinh^2 \frac{y}{2} = \frac{1}{2}(\cosh y - 1)$，$B$ 中水深為 $d$ 時，水的體積是

$$V_B(d) = \int_0^d \pi \sinh^2 \frac{y}{2} \,\mathrm{d}y = \frac{\pi}{2} \int_0^d (\cosh y - 1) \,\mathrm{d}y = \frac{\pi}{2} (\sinh d - d).$$

### 水量守恆

開始時，$A$ 盛有深度為 4 的水，而 $B$ 是空的，所以總體積是

$$V_A(4) = \frac{16\pi}{2} = 8\pi.$$

最後，兩個水面都在 $B$ 的底部之上高度 $h$ 處。因此 $B$ 中的水深等於 $h$。$A$ 的底部比 $B$ 的底部高一個單位，所以 $A$ 中的水深是 $h - 1$。

> **想一想** 為甚麼兩個水面最終會處於同一高度？

兩個漏斗和管子構成一個連通而靜止的水體。水中的壓強只取決於高度，而兩個自由水面都處於大氣壓強之下，所以它們位於同一高度。題目已經告訴我們這一點，但這也解釋了為甚麼兩個水深之差恰好等於兩個底部的高度差。

水既沒有增加，也沒有減少，所以

$$\frac{\pi(h-1)^2}{2} + \frac{\pi}{2} (\sinh h - h) = 8\pi.$$

乘以 $2/\pi$ 並展開，得

$$\begin{aligned} (h-1)^2 + \sinh h - h &= 16 \\ h^2 - 2h + 1 + \sinh h - h &= 16 \\ h^2 - 3h + \sinh h &= 15, \end{aligned}$$

這就是所要求的方程。

中間一行的左方只在 $h \geq 1$ 時才是合理的模型，因為這時 $A$ 才真正盛有水。在這個範圍內，它的導數 $2(h-1) + \cosh h - 1$ 是正數，所以方程恰好有一個符合題意的根。數值上 $h \approx 3.33$，所以 $A$ 中的水深由 4 降至約 2.33。

## 驗算

- 以二分法（bisection）求得 $h \approx 3.3272$。於是 $V_A(h-1) + V_B(h) \approx 25.1327$，與 $8\pi \approx 25.1327$ 吻合。
- 對 $\int_0^d \pi \sinh^2(y/2)\,\mathrm{d}y$ 使用辛普森法則（Simpson's rule），所得結果在多個 $d$ 值下都與 $\frac{\pi}{2}(\sinh d - d)$ 一致（例如 $d = 1.7$ 時，兩種方法都得出 1.48540）。
- 合理性：$h - 1 < 4$，所以 $A$ 中的水位下降了；水從中流出，水位理應下降。

## 常見錯誤

- 在兩個漏斗中都用深度 $h$，忘記了 $A$ 的底部高出一個單位。這樣會得出 $h^2 - h + \sinh h = 16$。
- 對 $\pi y^2\,\mathrm{d}x$ 積分（即繞 $x$ 軸旋轉），而不是對 $\pi x^2\,\mathrm{d}y$ 積分。
- 把 $y = 2\sinh^{-1} x$ 的反函數寫成 $x = 2\sinh y$ 或 $x = \frac{1}{2}\sinh y$。正確的反函數是 $x = \sinh\frac{y}{2}$。

## STEP 之外

最後的方程把多項式和雙曲函數（hyperbolic function）混在一起，所以沒有閉合形式（closed-form）的解，必須用數值方法求解；而上面的單調性（monotonicity）論證，正好保證了二分法或牛頓–拉弗森法（Newton–Raphson）之類的方法能找到那個唯一具物理意義的根。大學一年級的數值分析（numerical analysis）會利用介值定理（intermediate value theorem）和收斂速度（convergence rate）把這一點嚴格化。至於背後的分析學，可參閱 Abbott 的 *Understanding Analysis*（一本分析學入門教科書）。

詳細解答

# STEP III 1987，第 5 題

純數學 • 模算術密碼

> **STEP III 1987，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/87-s3-q5

## 關鍵思路

這套密碼其實是兩個獨立的移位密碼（shift cipher）：偶數位置用一個（密鑰為 $n_0$），奇數位置用另一個（密鑰為 $n_1$），兩者都再加上額外的移位 $j$。先把已知的移位 $j$ 除去。餘下的，就是信息的兩半各自移了一個未知常數；而模（modulo）50 的移位是雙射（bijection），所以相等的數仍然相等，不同的數仍然不同。那三個 43 正是洩露 $n_0$ 的指紋。之後，取一個模 50 的和，便可確定 $n_1$。

## 解答

只有 $n_0$ 和 $n_1$ 模 50 的值才有關係，所以我們取 $0 \le n_0, n_1 \le 49$。每個 $a_k$ 都介乎 0 與 49 之間，所以只要知道 $a_k$ 模 50 的剩餘（residue），它的值便確定了。

### 除去已知的移位

對於偶數位置，設 $c_j = b_{2j} - j$，其中 $j = 1, \ldots, 6$。對於奇數位置，設 $d_j = b_{2j+1} - j$，其中 $j = 0, \ldots, 5$。於是

$$a_{2j} \equiv c_j - n_0, \qquad a_{2j+1} \equiv d_j - n_1 \pmod{50}.$$

由加密後的信息可得：

| $j$ | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| $b_{2j}$ | 27 | 36 | 35 | 40 | 37 | 48 |
| $c_j$ | 26 | 34 | 32 | 36 | 32 | 42 |

| $j$ | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| $b_{2j+1}$ | 35 | 2 | 15 | 8 | 40 | 24 |
| $d_j$ | 35 | 1 | 13 | 5 | 36 | 19 |

**求 $n_0$**

在模 50 之下減去一個固定的 $n$，是一對一（one-to-one）的運算。所以，奇偶性（parity）相同的兩個位置上的 $a$ 相等，當且僅當它們對應的 $c$（或 $d$）相等。

> **想一想** 那三個 43 可以放在哪裏？

六個 $d_j$ 各不相同，所以奇數位置上是六個不同的數，其中最多只有一個 43。因此最少有兩個 43 在偶數位置。在 $c_j$ 當中，唯一重複的是 32，它恰好出現兩次（$j = 3, 5$），不多也不少。所以偶數位置上恰好有兩個相等的數，而它們必定就是那兩個 43，因為 43 是信息中唯一出現多於一次的數。因此

$$32 - n_0 \equiv 43 \pmod{50} \quad \Longrightarrow \quad n_0 \equiv -11 \equiv 39 \pmod{50}.$$

現在 $a_{2j} \equiv c_j - 39 \equiv c_j + 11$，由此得

$$(a_2, a_4, a_6, a_8, a_{10}, a_{12}) = (37, 45, 43, 47, 43, 3),$$

這是因為 $42 + 11 = 53 \equiv 3$。這六個數全部都在信息之中，這是必然的。

**求 $n_1$**

從信息中取走這六個數，餘下的就是奇數位置上的數

$$S = \{1, 7, 23, 24, 39, 43\},$$

所以 $\{d_j - n_1\}$ 在模 50 之下必須等於 $S$。

> **想一想** 怎樣才可以不必把五十個值逐一嘗試，就求出 $n_1$？

把兩邊分別加起來。如果兩個集合相同，它們的和在模 50 之下也相同：

$$\sum d_j - 6n_1 \equiv \sum_{s \in S} s \pmod{50}, \qquad \text{即} \qquad 109 - 6n_1 \equiv 137.$$

所以 $6n_1 \equiv -28 \equiv 22 \pmod{50}$。全式除以 2（它同時整除模數），得 $3n_1 \equiv 11 \pmod{25}$。由於 $3 \times 17 = 51 \equiv 1$，我們得到 $n_1 \equiv 11 \times 17 = 187 \equiv 12 \pmod{25}$。所以 $n_1 = 12$ 或 $n_1 = 37$。

和的條件只是必要條件（necessary condition），並非充分條件（sufficient condition），所以兩個值都要檢驗。若 $n_1 = 37$，則 $d_0 - 37 = -2 \equiv 48$，而這個數不在信息之中，所以 $n_1 = 37$ 不成立。若 $n_1 = 12$，則

$$(a_1, a_3, a_5, a_7, a_9, a_{11}) \equiv (23, -11, 1, -7, 24, 7) \equiv (23, 39, 1, 43, 24, 7),$$

這正好就是 $S$。所以 $n_1 = 12$。

### 信息

把兩半交錯排列，原來的信息是

$$23,\ 37,\ 39,\ 45,\ 1,\ 43,\ 43,\ 47,\ 24,\ 43,\ 7,\ 3,$$

其中 $n_0 \equiv 39$ 及 $n_1 \equiv 12 \pmod{50}$。以上論證顯示這是唯一的答案：$n_0$ 的值別無選擇，而 $n_1$ 的兩個候選值中只有一個可行。

## 驗算

把答案重新加密。例如 $b_{12} \equiv a_{12} + 39 + 6 = 48$ 及 $b_7 \equiv a_7 + 12 + 3 = 58 \equiv 8$，兩者都正確。用電腦把全部 2500 對 $(n_0, n_1)$ 逐一嘗試，會發現恰好只有一對，即 $(39, 12)$，能解密出題目所給的那組數，而且得出的正是上面的信息。

## 常見錯誤

- 弄錯下標。奇數位置由 $b_1$ 開始，對應 $j = 0$；但偶數位置由 $b_2$ 開始，對應 $j = 1$。
- 把 $6n_1 \equiv 22 \pmod{50}$ 除以 6，彷彿 6 是可逆（invertible）的。其實不然，因為 $\gcd(6, 50) = 2$，而這個同餘式在模 50 之下有兩個解。
- 找到一個可行的 $n_1$ 值，卻沒有說明其他值都不可行。題目要求你解釋所用的方法，所以唯一性需要論證。

## STEP 之外

這是凱撒（Caesar）密碼，亦即移位密碼，而這裏的破解方法是頻率分析（frequency analysis）的一個小例子：移位會保留字母重複出現的規律。在大學的入門課程中，由 $6n \equiv 22$ 得出兩個解的那一步，屬於線性同餘式（linear congruence）$ax \equiv b \pmod{m}$ 的理論：這類同餘式有解，當且僅當 $\gcd(a, m)$ 整除 $b$。仿射密碼（affine cipher）$x \mapsto ax + b$，以及其後的 RSA，都是建基於同樣的想法。Davenport 的 *The Higher Arithmetic* 對同餘的講解十分精彩。

詳細解答

# STEP III 1987，第 6 題

純數學 • 耦合微分方程

> **STEP III 1987，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/87-s3-q6

## 關鍵思路

處理一對耦合的一階方程（coupled first-order equations），最好的做法是把它們化為只含一個未知函數的二階方程。利用第一條方程，可把 $y$ 用 $x$ 和 $\mathrm{d}x/\mathrm{d}t$ 表示，再代入第二條方程，便得到一條關於 $x$ 的受迫振子（forced oscillator）方程。答案的整體性質，就取決於這個振子的固有頻率（natural frequency）與外力項 $2\cos t$ 的頻率相比如何。

## 解答

以字母上方加點表示對 $t$ 的導數。

### 消去 $y$

由第一條方程，

$$y = \tfrac{1}{5}\left(\dot{x} + 2x\right), \qquad \text{所以} \qquad \dot{y} = \tfrac{1}{5}\left(\ddot{x} + 2\dot{x}\right).$$

把兩者代入第二條方程，再乘以 5，

$$\ddot{x} + 2\dot{x} + 5ax - 2\left(\dot{x} + 2x\right) = 10\cos t,$$
$$\ddot{x} + \left(5a - 4\right)x = 10\cos t.$$

含 $\dot{x}$ 的項互相抵消，所以沒有阻尼（damping）：這是一個純粹的受迫振子，固有頻率為 $\sqrt{5a-4}$（當 $5a > 4$ 時）。

### 轉換初始條件

題目給出 $x(0)=0$ 和 $\dot{y}(0)=0$，但關於 $x$ 的方程需要的是 $x(0)$ 和 $\dot{x}(0)$。把 $t=0$ 代入原來的第二條方程：

$$\dot{y}(0)+a\,x(0)-2y(0)=2 \quad\Longrightarrow\quad 0+0-2y(0)=2 \quad\Longrightarrow\quad y(0)=-1.$$

然後由第一條方程在 $t=0$ 的情況，得 $\dot{x}(0)=5y(0)-2x(0)=-5$。留意這兩步都沒有用到 $a$ 的值，所以

$$x(0)=0, \qquad \dot{x}(0)=-5 \qquad \text{對所有}a.$$

**想一想** 條件是關於 $\mathrm{d}y/\mathrm{d}t$ 的，而不是關於 $y$ 的。我們該怎樣運用它？

微分方程本身已把 $x$、$y$ 及它們的導數在任何一個時刻的值聯繫起來。在 $t=0$ 時，第二條方程把關於 $\dot{y}(0)$ 的資料轉化為關於 $y(0)$ 的資料，第一條方程再把它轉化為 $\dot{x}(0)$，而這正是關於 $x$ 的二階方程所需要的。

### $a=1$ 的情況

這時方程為

$$\ddot{x}+x=10\cos t.$$

餘函數（complementary function）是 $A\cos t+B\sin t$。外力項 $\cos t$ 本身就是齊次方程（homogeneous equation）的解（這就是共振（resonance）），所以試用 $C\cos t$ 是行不通的，我們改試 $x_p=t\,(C\cos t+D\sin t)$。求導兩次，

$$\ddot{x}_p+x_p=-2C\sin t+2D\cos t,$$

所以 $C=0$ 及 $D=5$，得 $x_p=5t\sin t$。因此

$$x=A\cos t+B\sin t+5t\sin t.$$

由條件 $x(0)=0$ 得 $A=0$。於是 $\dot{x}=B\cos t+5\sin t+5t\cos t$，所以 $\dot{x}(0)=B=-5$。因此

$$\boxed{x=5(t-1)\sin t}$$

至於 $y$，利用 $y=\frac{1}{5}(\dot{x}+2x)$ 以及 $\dot{x}=5\sin t+5(t-1)\cos t$：

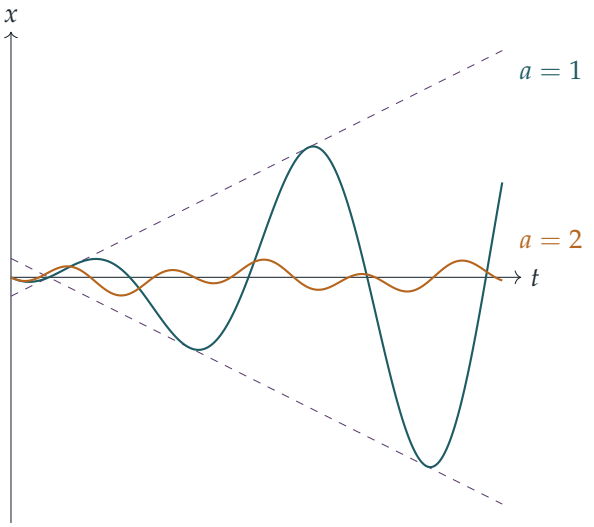
$$y=\sin t+(t-1)\cos t+2(t-1)\sin t, \qquad \text{即} \qquad \boxed{y=(t-1)\cos t+(2t-1)\sin t}$$

簡單驗算一下：$y(0)=-1$，而由 $\dot{y}=\cos t-(t-1)\sin t+2\sin t+(2t-1)\cos t$ 得 $\dot{y}(0)=1-1=0$，符合要求。

### $a > 1$ 的情況

這時 $5a - 4 > 1$，所以可寫 $\omega = \sqrt{5a-4}$，其中 $\omega > 1$。方程 $\ddot{x} + \omega^2 x = 10\cos t$ 不再處於共振狀態，因為外力頻率 1 與固有頻率 $\omega$ 不同。所以：

- 餘函數是 $A\cos\omega t + B\sin\omega t$，振動週期為 $2\pi/\omega$（比 $2\pi$ 短），而不再是 $2\pi$。
- 特積分（particular integral）只是 $\cos t$ 的倍數，沒有因子 $t$，所以不會出現 $t\sin t$ 這類項。
- 因此 $x$（以及 $y = \frac{1}{5}(\dot{x} + 2x)$）是頻率為 1 和 $\omega$ 的正弦型函數之和。兩者對所有 $t$ 都保持有界（bounded），而不像 $a = 1$ 時那樣，振幅隨 $t$ 線性增長。只有當 $\omega$ 是有理數時，運動才是週期性的；一般而言，它是兩個振動的疊加（superposition），兩者互相產生「拍」（beat）的現象。




## 驗算

- 把 $x = 5(t-1)\sin t$ 和 $y = (t-1)\cos t + (2t-1)\sin t$ 代回原來的兩條方程，兩者均恆等成立。逐步數值求解（由 $x = 0$、$y = -1$ 出發的四階龍格－庫塔法（Runge–Kutta））所得結果，在 $0 \le t \le 10$ 上與公式相差不超過 $10^{-10}$；例如 $x(10) = 45\sin 10 \approx -24.481$。
- 當 $a > 1$ 時，完整的解原來是 $x = \frac{2}{a-1}(\cos t - \cos\omega t) - \frac{5}{\omega}\sin\omega t$。當 $a \to 1$ 時，有 $\omega - 1 \approx \frac{5}{2}(a-1)$ 及 $\cos t - \cos\omega t \approx (\omega - 1)\, t\sin t$，由此重新得到 $5t\sin t - 5\sin t$。共振情況的答案正是非共振情況的極限（limit）。

## 常見錯誤

- 直接用 $y(0) = 0$，而沒有從 $\dot{y}(0)$ 的條件推出 $y(0) = -1$。
- 當 $a = 1$ 時試用 $C\cos t + D\sin t$ 作為特積分：它是餘函數的一部分，所以只會得出 $0 = 10\cos t$。
- 另外引入新的任意常數來獨立求解 $y$：一旦知道 $x$，$y$ 就由 $y = \frac{1}{5}(\dot{x} + 2x)$ 完全確定。

## STEP 之外

這類線性方程組其實就是 $\dot{\mathbf{v}} = M\mathbf{v} + \mathbf{f}(t)$，其中 $\mathbf{v} = (x, y)$，而 $M$ 是常數矩陣；$M$ 的特徵值（eigenvalue），在這裏是 $\pm\sqrt{4 - 5a}$，決定了自由運動是增長、衰減還是振動。當外力頻率與特徵值 $\pm\mathrm{i}$ 吻合時，便會出現共振，而這正好就是 $a = 1$。Boyce 與 DiPrima 所著的 *Elementary Differential Equations*（微分方程入門教材）既講解如何用矩陣處理方程組，也討論受迫振動以及共振和拍的現象。

詳細解答

# STEP III 1987，第 7 題

純數學 • 反正切的級數

**STEP III 1987，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/87-s3-q7

## 關鍵思路

$\tan^{-1} t$ 的級數（series）來自對 $1/(1+x^2)$ 求積分，而 $1/(1+x^2)$ 正是公比為 $-x^2$ 的等比級數（geometric series）之和。把等比級數在有限項之後截斷，會留下一個精確的餘項（remainder），再對這個餘項積分，便得到誤差項（error term）。在 $[0,1]$ 上，因子 $1/(1+x^2)$ 介乎 $\frac{1}{2}$ 與 1 之間，於是誤差便被夾在兩個容易計算的積分之間。

## 解答

### 恆等式

對任意實數 $x$，首項為 1、公比為 $-x^2$ 的有限等比級數給出

$$\sum_{r=0}^{n}(-x^2)^r = \frac{1-(-x^2)^{n+1}}{1+x^2},$$

此式成立，因為公比 $-x^2$ 永遠不等於 1。移項整理，得

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

這是一個精確的恆等式（identity），並非近似。把兩邊由 0 積分至 $t$。由於 $\int_0^t \frac{\mathrm{d}x}{1+x^2} = \tan^{-1} t$ 及 $\int_0^t x^{2r}\,\mathrm{d}x = \frac{t^{2r+1}}{2r+1}$，

$$\tan^{-1} t = t - \frac{t^3}{3} + \frac{t^5}{5} - \cdots + \frac{(-1)^n t^{2n+1}}{2n+1} + (-1)^{n+1}\int_0^t \frac{x^{2n+2}}{1+x^2}\,\mathrm{d}x,$$

這正是所求。此式對所有實數 $t$ 均成立。

### 上下界

記 $S_n(t) = \sum_{r=0}^{n} \frac{(-1)^r t^{2r+1}}{2r+1}$。由上述恆等式，

$$\left|\tan^{-1} t - S_n(t)\right| = \int_0^t \frac{x^{2n+2}}{1+x^2}\,\mathrm{d}x,$$

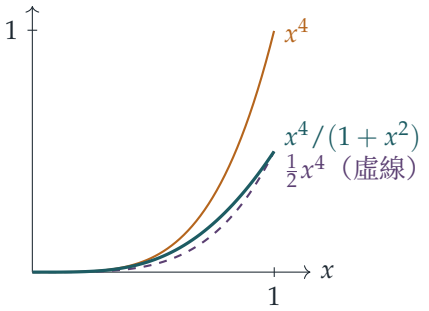
右邊的絕對值符號可以略去，因為當 $t \geq 0$ 時，被積函數（integrand）非負，而且積分是由小到大進行的。

> **想一想** 一個無法以閉合形式（closed form）求值的積分，我們可以怎樣估計？

用常數來界定被積函數中難處理的部分。當 $0 \leq x \leq 1$ 時，我們有 $1 \leq 1+x^2 \leq 2$，所以

$$\tfrac{1}{2}x^{2n+2} \leq \frac{x^{2n+2}}{1+x^2} \leq x^{2n+2}.$$

下圖顯示 $n = 1$ 的情況：被積函數（藍綠色）被夾在 $x^4$ 與 $\frac{1}{2}x^4$ 之間。




把這些不等式由 0 積分至 $t$（由於 $t \geq 0$，不等號方向保持不變），得

$$\frac{t^{2n+3}}{2(2n+3)} \leq \left|\tan^{-1} t - S_n(t)\right| \leq \frac{t^{2n+3}}{2n+3} \qquad (0 \leq t \leq 1).$$

這就是所要證明的結果。

### $\pi$ 的級數

取 $t = 1$，此時 $\tan^{-1} 1 = \frac{\pi}{4}$ 而 $S_n(1) = \sum_{r=0}^{n} \frac{(-1)^r}{2r+1}$。由上界（upper bound）可得

$$\left|\pi - 4\sum_{r=0}^{n} \frac{(-1)^r}{2r+1}\right| \leq \frac{4}{2n+3} \to 0 \quad \text{當} n \to \infty,$$

所以 $4\sum_{r=0}^{n} \frac{(-1)^r}{2r+1} \to \pi$。

> **想一想** 兩個界之中，哪一個告訴我們這個近似很差？

是下界（lower bound）：它說明誤差不可能很小。把 $t=1$ 時左邊的不等式乘以 4，

$$\left|\pi - 4\sum_{r=0}^{n}\frac{(-1)^r}{2r+1}\right| \geq \frac{4}{2(2n+3)} = \frac{2}{2n+3}.$$

若 $n \leq 98$，則 $2n+3 \leq 199 < 200$，所以 $\frac{2}{2n+3} > \frac{2}{200} = 10^{-2}$。因此對所有 $n \leq 98$，誤差至少為 $10^{-2}$（事實上嚴格大於此數）。由於誤差小於 $4/(2n+3)$，一百項只換來大約兩個小數位的準確度，所以用這個級數來計算 $\pi$ 是非常緩慢的。

## 驗算

- $n=0$，$t=1$：上下界給出 $\frac{1}{6} \leq \left|\frac{\pi}{4}-1\right| \leq \frac{1}{3}$。事實上 $1-\frac{\pi}{4} \approx 0.215$。
- 以數值計算，當 $n=98$ 時誤差為 $|\pi - 4S_{98}(1)| \approx 0.01010$，僅僅高於 $10^{-2}$。當 $n=99$ 時誤差約為 0.0099998，僅僅低於此值。所以 98 正好是分界點。我們亦以數值積分（quadrature）對隨機的 $t$ 和 $n$ 驗證了恆等式，並對數千個隨機的 $t \in [0,1]$ 驗證了雙邊的界。

## 常見錯誤

- 把 $1/(1+x^2)$ 展開成無窮級數，然後逐項積分。這樣做需要 $|x|<1$，還要說明為何可以交換求和與積分的次序。採用帶有精確餘項的有限形式，則兩者皆可避免。
- 用上界來論證誤差很大。只有下界才能做到這一點。

## STEP 之外

這其實是帶積分形式餘項的泰勒定理（Taylor's theorem），只是針對一個函數親手推導出來。$t=1$ 時的級數就是萊布尼茲（Leibniz）級數，又稱馬德哈瓦（Madhava）–格雷果里（Gregory）級數。它的誤差大約為 $1/(2n)$，這是各項遞減緩慢的交錯級數（alternating series）的典型表現。較快的公式，例如梅欽（Machin）公式 $\frac{\pi}{4} = 4\tan^{-1}\frac{1}{5} - \tan^{-1}\frac{1}{239}$，是把同一個級數用於較小的 $t$，此時 $t^{2n+3}$ 會迅速縮小。若想了解餘項與收斂（convergence）的嚴格處理，可參閱 Abbott 的 *Understanding Analysis*。

詳細解答

# STEP III 1987，第 8 題

純數學 • 平行四邊形面積不等式

> **STEP III 1987，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/87-s3-q8

## 關鍵思路

平行四邊形（parallelogram）的對角線（diagonal）互相平分（bisect），因此把它分成四個三角形，而這些三角形的面積可以用 $\frac{1}{2}ab\sin C$ 立即寫出。這樣便得到一條以對角線及其夾角表示面積的公式，題目開首的兩個結果都可由此直接推出。集合 $A$ 是兩條帶狀區域相交之處，所以是一個平行四邊形。只要求出它的面積，第二個結果便告訴我們它較長的對角線必定夠長，而這條對角線的兩個端點正是我們要找的點。

## 解答

### 以對角線表示面積

設兩條對角線的長度為 $p$ 和 $q$，相交所成的角為 $\theta$，其中 $0 < \theta < \pi$。平行四邊形的對角線互相平分，所以把它分成四個三角形，每個三角形都有兩條邊 $\frac{1}{2}p$ 和 $\frac{1}{2}q$ 在中心相交。其中兩個三角形在該處的角為 $\theta$，另外兩個則為 $\pi - \theta$。由於 $\sin(\pi - \theta) = \sin\theta$，每個三角形的面積都是 $\frac{1}{2} \cdot \frac{p}{2} \cdot \frac{q}{2}\sin\theta$，所以

$$\text{面積} = 4 \cdot \tfrac{1}{8}pq\sin\theta = \tfrac{1}{2}pq\sin\theta.$$

若 $p$ 和 $q$ 固定，則當 $\sin\theta$ 最大時面積最大，即 $\theta = \frac{\pi}{2}$ 之時。因此面積取最大值，當且僅當兩條對角線互相垂直，而最大面積為 $\frac{1}{2}pq$。

### 以較長對角線表示的上界

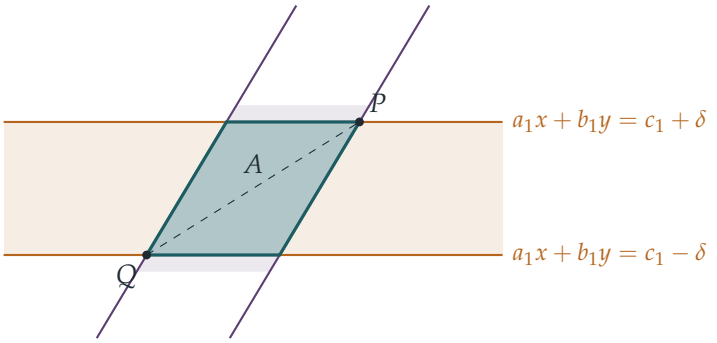
設 $p \geq q$，即 $p$ 是較長的對角線（或兩條等長對角線的其中一條）。先用 $\sin\theta \leq 1$，再用 $q \leq p$，得

$$\text{面積} = \tfrac{1}{2}pq\sin\theta \leq \tfrac{1}{2}pq \leq \tfrac{1}{2}p^2.$$

所以面積最多是較長對角線平方的一半。等號成立需要 $\sin\theta = 1$ 及 $q = p$，故只有正方形才能取等號。

### 集合 $A$ 是平行四邊形

假設 $\delta > 0$。（若 $\delta = 0$，則上界為 0，把 $A$ 中任意一點取兩次即可。）不等式（inequality）$|a_1x + b_1y - c_1| \leq \delta$ 的意思是 $c_1 - \delta \leq a_1x + b_1y \leq c_1 + \delta$。因此它描述的是兩條平行線 $a_1x + b_1y = c_1 \pm \delta$ 之間的閉帶狀區域（closed strip）。同理，第二條不等式給出第二條帶狀區域。條件 $a_1b_2 \neq a_2b_1$ 表示法向量（normal vector）$\mathbf{n}_1 = (a_1, b_1)$ 與 $\mathbf{n}_2 = (a_2, b_2)$ 不平行。所以兩條帶狀區域相交，而且 $\mathbf{n}_1$ 和 $\mathbf{n}_2$ 都不是零向量。它們的交集（intersection）$A$ 是一個平行四邊形，由兩對平行線圍成。




圖中顯示第一條帶狀區域（琥珀色）和第二條帶狀區域的一部分（紫紅色）。它們的交集 $A$（藍綠色）是一個平行四邊形，虛線是它較長的對角線。

### $A$ 的面積

平行線 $a_1x + b_1y = c_1 \pm \delta$ 之間的距離為

$$h_1 = \frac{2\delta}{\sqrt{a_1^2 + b_1^2}},$$

這裏用了點到直線的距離公式。同理，第二條帶狀區域的闊度為 $h_2 = 2\delta / \sqrt{a_2^2 + b_2^2}$。設 $\varphi$ 為兩組直線之間的夾角，其中 $0 < \varphi < \pi$。在 $A$ 的各邊之中，位於第二條帶狀區域邊界上的兩條邊以角 $\varphi$ 橫越第一條帶狀區域，所以每條的長度為 $h_1 / \sin\varphi$。這兩條邊之間

的垂直距離為 $h_2$，所以

$$\text{面積} A = \frac{h_1 h_2}{\sin\varphi}.$$

**想一想** 怎樣由各系數求出 $\sin\varphi$？

兩條直線之間的夾角等於它們法線之間的夾角，或該角的補角（supplement），而兩者的正弦值相同。對於向量 $\mathbf{n}_1 = (a_1, b_1)$ 和 $\mathbf{n}_2 = (a_2, b_2)$，它們所張成的平行四邊形面積為 $|\mathbf{n}_1||\mathbf{n}_2|\sin\varphi$。這個面積亦等於 $|a_1 b_2 - a_2 b_1|$，即 $2\times 2$ 行列式（determinant）的絕對值。因此

$$\sin\varphi = \frac{|a_1 b_2 - a_2 b_1|}{\sqrt{a_1^2 + b_1^2}\sqrt{a_2^2 + b_2^2}}.$$

代入後，平方根互相抵消：

$$\text{面積} A = \frac{2\delta}{\sqrt{a_1^2 + b_1^2}} \cdot \frac{2\delta}{\sqrt{a_2^2 + b_2^2}} \cdot \frac{\sqrt{a_1^2 + b_1^2}\sqrt{a_2^2 + b_2^2}}{|a_1 b_2 - a_2 b_1|} = \frac{4\delta^2}{|a_1 b_2 - a_2 b_1|}.$$

### 所求的兩點

設 $PQ$ 為 $A$ 較長的對角線，並記 $P = (x_1, y_1)$ 及 $Q = (x_2, y_2)$。這兩點是 $A$ 的頂點，而帶狀區域包含其邊界線，所以它們都在 $A$ 之內。由第一部分的上界，

$$\frac{4\delta^2}{|a_1 b_2 - a_2 b_1|} = \text{面積} A \le \tfrac{1}{2} PQ^2,$$

所以

$$(x_1 - x_2)^2 + (y_1 - y_2)^2 = PQ^2 \ge \frac{8\delta^2}{|a_1 b_2 - a_2 b_1|},$$

這正是所要證明的。

## 驗算

- 取 $a_1 = b_2 = 1$ 及 $a_2 = b_1 = 0$，則 $A$ 是正方形 $|x - c_1| \le \delta$，$|y - c_2| \le \delta$。它的對角線平方為 $(2\delta)^2 + (2\delta)^2 = 8\delta^2$，剛好等於上界。這是合理的，因為正方形正是第一部分中等號成立的情況。
- 縮放：若把 $a_1$、$b_1$ 和 $c_1$ 都加倍，第一條帶狀區域的闊度減半，$A$ 的面積減半，而 $|a_1 b_2 - a_2 b_1|$ 則加倍。公式的變化與此吻合。
- 我們用電腦測試了 $10^5$ 個隨機例子。我們解邊界方程求出 $A$ 的頂點，再用鞋帶公式（shoelace formula）計算面積，每次都與 $4\delta^2/|a_1 b_2 - a_2 b_1|$ 相符。（較長對角線）$^2$ 與上界之比的最小值為 1.0007，從未低於 1。

## 常見錯誤

- 把帶狀區域的闊度當作 $2\delta$。其實 $|a_1x + b_1y - c_1| \leq \delta$ 的闊度是 $2\delta/\sqrt{a_1^2 + b_1^2}$。
- 忘記說明為何所選的點在 $A$ 之內。原因是它們是頂點，而 $A$ 包含其邊界。
- 把 $A$ 的面積寫成 $h_1h_2$，這只在兩條帶狀區域互相垂直時才正確。

## STEP 之外

這個面積計算是一個基本事實的特例：矩陣為 $M$ 的線性映射（linear map）會把所有面積乘以 $|\det M|$。這裏的映射 $(x, y) \mapsto (a_1x + b_1y,\ a_2x + b_2y)$ 把 $A$ 映到一個面積為 $4\delta^2$ 的正方形，所以 $A$ 的面積為 $4\delta^2/|\det M|$。這就是在二重積分（double integral）中作變數代換時遇到的雅可比（Jacobian）因子。Strang 的 *Introduction to Linear Algebra*（線性代數入門）正是以這種方式，把行列式看作面積和體積來闡述。

詳細解答

# STEP III 1987，第 9 題

純數學 • 群的直積

> **STEP III 1987，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/87-s3-q9

## 關鍵思路

$G \times H$ 中的一切運算都是逐個坐標進行的，所以每一條羣公理(group axiom)以及 $G \times H$ 的每一項性質，都歸結為 $G$ 和 $H$ 各自具備同樣的性質。至於判斷真假的部分，答「真」需要給出一般的證明，而答「假」則只需找出一個確鑿的障礙。同構（isomorphic）的羣，凡是單靠乘法就能描述的性質都完全一致，例如羣的大小和各元素的階（order）。所以我們就比較這些性質。

## 解答

以下一律以 $e_G$ 和 $e_H$ 表示 $G$ 和 $H$ 的單位元（identity），並以 $g^{-1}$ 和 $h^{-1}$ 表示在 $G$ 和 $H$ 中的逆元（inverse）。

### $G \times H$ 是一個羣

我們逐一檢查四條公理。

封閉性（*closure*）。若 $g_1, g_2 \in G$ 且 $h_1, h_2 \in H$，則 $g_1 * g_2 \in G$ 且 $h_1 \circ h_2 \in H$，因為 $G$ 和 $H$ 都是封閉的。所以 $(g_1, h_1)(g_2, h_2) = (g_1 * g_2,\ h_1 \circ h_2) \in G \times H$。

結合律（*associativity*）。對任意三個元素，

$$\begin{aligned}\big((g_1, h_1)(g_2, h_2)\big)(g_3, h_3) &= \big((g_1 * g_2) * g_3,\ (h_1 \circ h_2) \circ h_3\big) \\ &= \big(g_1 * (g_2 * g_3),\ h_1 \circ (h_2 \circ h_3)\big) \\ &= (g_1, h_1)\big((g_2, h_2)(g_3, h_3)\big),\end{aligned}$$

其中中間一步用到 $G$ 和 $H$ 中的結合律。

單位元。序偶 $(e_G, e_H)$ 滿足

$$(e_G, e_H)(g,h) = (e_G * g,\ e_H \circ h) = (g,h),$$

同理 $(g,h)(e_G,e_H) = (g,h)$。所以它是單位元。

逆元。給定 $(g,h)$，序偶 $(g^{-1},h^{-1})$ 屬於 $G \times H$，並且

$$(g,h)(g^{-1},h^{-1}) = (g * g^{-1},\ h \circ h^{-1}) = (e_G, e_H),$$

而次序倒過來相乘，所得的積同樣是 $(e_G,e_H)$。所以每個元素都有逆元。

四條公理全部成立，所以 $G \times H$ 在這個乘法之下是一個羣。

**第 (i) 部分：真**

若 $G$ 和 $H$ 是阿貝爾羣（*abelian group*），則 $G \times H$ 也是。對任意兩個元素，

$$(g_1,h_1)(g_2,h_2) = (g_1 * g_2,\ h_1 \circ h_2) = (g_2 * g_1,\ h_2 \circ h_1) = (g_2,h_2)(g_1,h_1),$$

這裏用到 $G$ 和 $H$ 中的交換律（commutativity）。

若 $G \times H$ 是阿貝爾羣，則 $G$ 和 $H$ 也是。任取 $g_1, g_2 \in G$。由於 $G \times H$ 是阿貝爾羣，

$$(g_1 * g_2,\ e_H) = (g_1,e_H)(g_2,e_H) = (g_2,e_H)(g_1,e_H) = (g_2 * g_1,\ e_H).$$

兩個序偶相等，其第一個坐標必須相等，所以 $g_1 * g_2 = g_2 * g_1$，即 $G$ 是阿貝爾羣。以同樣方法使用序偶 $(e_G,h_1)$ 和 $(e_G,h_2)$，可證明 $H$ 是阿貝爾羣。

**想一想** 為甚麼在第二個方向的證明中需要用到單位元 $e_H$？

我們需要某個第二坐標，才能由 $g_1$ 和 $g_2$ 造出 $G \times H$ 的元素。任何固定的 $h \in H$ 都可以，因為比較時只有第一個坐標起作用。羣必定含有單位元，所以 $e_H$ 是一個穩妥的選擇，我們確知它存在。

**第 (ii) 部分：真**

設 $\widetilde{G} = \{(g,e_H) : g \in G\}$，並定義 $\phi\colon G \to \widetilde{G}$ 為 $\phi(g) = (g,e_H)$。

$\widetilde{G}$ 是 $G \times H$ 的子羣（subgroup）。它包含單位元 $(e_G,e_H)$。它對乘積封閉，因為 $(g_1,e_H)(g_2,e_H) = (g_1 * g_2, e_H)$。它對取逆元封閉，因為 $(g,e_H)^{-1} = (g^{-1},e_H)$。

$\phi$ 是一個同構（isomorphism）。按定義，它映滿 $\widetilde{G}$（即滿射，onto）。它是一對一的（one-to-one），因為 $(g_1,e_H) = (g_2,e_H)$ 必然推出 $g_1 = g_2$。它保持運算，因為

$$\phi(g_1 * g_2) = (g_1 * g_2,\ e_H) = (g_1,e_H)(g_2,e_H) = \phi(g_1)\phi(g_2).$$

所以 $G \times H$ 包含子羣 $\widetilde{G}$，而它與 $G$ 同構。

**第 (iii) 部分：假**

把兩個羣都寫成加法形式：$\mathbb{Z}_n = \{0, 1, \ldots, n-1\}$，運算為模 $n$ 加法。在 $\mathbb{Z}_2 \times \mathbb{Z}_2$ 中，每個元素 $x = (a, b)$ 都滿足

$$x + x = (2a, 2b) = (0, 0),$$

所以每個元素的階都是 1 或 2。相反，在 $\mathbb{Z}_4$ 中，$1 + 1 = 2 \neq 0$，所以 1 的階是 4。

假設存在同構 $\phi\colon \mathbb{Z}_4 \to \mathbb{Z}_2 \times \mathbb{Z}_2$。那麼

$$\phi(2) = \phi(1 + 1) = \phi(1) + \phi(1) = (0, 0) = \phi(0),$$

因為由上面的計算可知 $\phi(1) + \phi(1) = (0, 0)$。最後一個等號成立，是因為同構把單位元映到單位元。於是 $\phi(2) = \phi(0)$ 而 $2 \neq 0$，所以 $\phi$ 不是一對一的，這就產生矛盾。因此 $\mathbb{Z}_2 \times \mathbb{Z}_2$ 與 $\mathbb{Z}_4$ 不同構。（簡而言之，$\mathbb{Z}_4$ 是循環羣（cyclic group），而 $\mathbb{Z}_2 \times \mathbb{Z}_2$ 不是。）

**第 (iv) 部分：假**

同構是雙射（bijection），所以同構的羣有相同數目的元素。現在 $|S_n| = n!$，因為 $n$ 件物件共有 $n!$ 種排列；而 $|G \times H| = |G|\,|H|$，因為每個序偶都是由自由選取的 $g$ 和 $h$ 組成。所以

$$|S_2 \times S_3| = 2! \times 3! = 12, \qquad |S_6| = 6! = 720.$$

兩者大小不同，所以 $S_2 \times S_3$ 與 $S_6$ 不同構。

## 驗算

- 對一些細小的例子（例如 $\mathbb{Z}_2 \times \mathbb{Z}_3$、$\mathbb{Z}_2 \times S_3$、$S_3 \times S_3$ 和 $S_2 \times S_3$）逐一窮舉檢查，可見這個積滿足全部四條公理，單位元為 $(e_G, e_H)$。當且僅當兩個因子都是阿貝爾羣時，它才是阿貝爾羣，而 $g \mapsto (g, e_H)$ 是一個單射（injective）同態（homomorphism）。
- $\mathbb{Z}_2 \times \mathbb{Z}_2$ 中各元素的階是 $1, 2, 2, 2$，而 $\mathbb{Z}_4$ 中則是 $1, 2, 4, 4$，所以這兩個羣確實不同。
- 列出 $S_2 \times S_3$ 的元素，共有 12 個，而 $S_6$ 則有 720 個。

## 常見錯誤

- 在 (i) 中只證明了「當且僅當」的其中一個方向。逆命題（converse）才是需要技巧的部分，要把 $e_H$ 放在第二個坐標。
- 在 (ii) 中說「$G$ 是 $G \times H$ 的子羣」。嚴格來說，$G$ 連 $G \times H$ 的子集也不是。真正的子羣是 $G \times \{e_H\}$，它與 $G$ 同構。

► 在 (iii) 中指出兩個羣的階都是 4 便就此打住。大小相等是同構的必要條件，但並非充分條件，所以我們需要找出一個能區分這兩個羣的性質。

## STEP 之外

這就是*直積*（direct product）的構造，是由舊羣造出新羣的最簡單方法。大學第一門羣論課的一個重要結果是：每個有限阿貝爾羣都是循環羣的直積。與此相關的一個事實是：$\mathbb{Z}_m \times \mathbb{Z}_n \cong \mathbb{Z}_{mn}$ 當且僅當 $\gcd(m,n) = 1$，這正是中國剩餘定理（Chinese remainder theorem）在羣論中的面貌。第 (iii) 部分就是這個結論不成立的最小例子。想了解更多，可參閱 Allenby 的 *Rings, Fields and Groups*（環、域與羣）。

詳細解答

# STEP III 1987，第 10 題

純數學 • 伯努利多項式

> **STEP III 1987，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/87-s3-q10

## 關鍵思路

定義規則只透過導數（derivative）來確定 $P_n$，因此最自然的做法，是對差式 $P_n(x+1) - P_n(x) - nx^{n-1}$ 的導數運用歸納法（induction）：歸納假設令它變為零，於是只剩下一個常數，而積分條件就把這個常數確定下來。之後，所求的和是一個裂項和（telescoping sum）；至於立方和，我們只需要求出 $P_4$ 除常數項以外的部分，反正常數項會互相抵消。

## 解答

### 差分恆等式

對 $n \geq 1$，設

$$Q_n(x) = P_n(x+1) - P_n(x) - nx^{n-1}.$$

我們用歸納法證明：對所有 $x$，$Q_n(x) = 0$。

**基礎情況** $n = 1$。由於 $P_1' = 1 \cdot P_0 = 1$，故對某個常數 $c$ 有 $P_1(x) = x + c$。於是 $P_1(x+1) - P_1(x) = 1 = 1 \cdot x^0$，所以 $Q_1 \equiv 0$。（事實上，由 $\int_0^1 (x+c)\,\mathrm{d}x = 0$ 可得 $c = -\frac{1}{2}$，但我們用不着它。）

**歸納步驟**。假設 $n \geq 2$ 且 $Q_{n-1} \equiv 0$。求導，並利用 $P_n' = nP_{n-1}$，得

$$\begin{aligned} Q_n'(x) &= nP_{n-1}(x+1) - nP_{n-1}(x) - n(n-1)x^{n-2} \\ &= n\left[P_{n-1}(x+1) - P_{n-1}(x) - (n-1)x^{n-2}\right] \\ &= nQ_{n-1}(x) = 0. \end{aligned}$$

所以 $Q_n$ 是常數，並且等於它在 $x = 0$ 的值。因為 $n \geq 2$，$nx^{n-1}$ 這一項在 $x = 0$ 時為

零，所以

$$Q_n(0) = P_n(1) - P_n(0).$$

**想一想** 我們怎樣知道 $P_n(1) = P_n(0)$？積分條件說的是 $P_n$ 的積分，而不是它在個別點的取值。

把積分條件用在低一級的多項式上。根據微積分基本定理（fundamental theorem of calculus），

$$P_n(1) - P_n(0) = \int_0^1 P_n'(x)\,\mathrm{d}x = n\int_0^1 P_{n-1}(x)\,\mathrm{d}x = 0,$$

這是因為 $n-1 \geq 1$，所以條件 $\int_0^1 P_{n-1}\,\mathrm{d}x = 0$ 適用。（這正是 $n=1$ 的情況必須另外處理的原因：$\int_0^1 P_0\,\mathrm{d}x = 1$。）因此 $Q_n \equiv 0$，歸納法至此完成：

$$P_n(x+1) - P_n(x) = nx^{n-1} \qquad \text{對所有} n \geq 1.$$

### 冪和

在恆等式中代入 $x = m$，並對 $m = 0, 1, \ldots, k$ 求和。右邊逐項相消：

$$\begin{aligned} n\sum_{m=0}^{k} m^{n-1} &= \sum_{m=0}^{k}\big[P_n(m+1) - P_n(m)\big] \\ &= \big[P_n(1) - P_n(0)\big] + \big[P_n(2) - P_n(1)\big] + \cdots + \big[P_n(k+1) - P_n(k)\big] \\ &= P_n(k+1) - P_n(0), \end{aligned}$$

這正是所求。（當 $n=1$ 時，$m=0$ 的一項是 $0^0$，必須理解為 1；這樣兩邊都等於 $k+1$。）

### 立方和

取 $n = 4$，所以我們需要 $P_4$。由 $P_0 = 1$ 開始逐級向上，每次積分，並選取常數使得在 $[0,1]$ 上的積分為零：

$$\begin{aligned} P_1(x) &= x - \tfrac{1}{2}, \\ P_2(x) &= x^2 - x + \tfrac{1}{6} && \text{因為 } \textstyle\int_0^1 (x^2 - x)\,\mathrm{d}x = -\tfrac{1}{6}, \\ P_3(x) &= x^3 - \tfrac{3}{2}x^2 + \tfrac{1}{2}x && \text{因為 } \textstyle\int_0^1 (x^3 - \tfrac{3}{2}x^2 + \tfrac{1}{2}x)\,\mathrm{d}x = \tfrac{1}{4} - \tfrac{1}{2} + \tfrac{1}{4} = 0, \\ P_4(x) &= x^4 - 2x^3 + x^2 + c \end{aligned}$$

其中 $c$ 為某個常數。我們不需要求 $c$：它在 $P_4(k+1) - P_4(0)$ 中會抵消。記 $y = k+1$，則

$$P_4(y) - P_4(0) = y^4 - 2y^3 + y^2 = y^2(y-1)^2 = (k+1)^2k^2.$$

所以在求和公式中取 $n = 4$，得

$$\sum_{m=0}^{k} m^3 = \frac{k^2(k+1)^2}{4} = \left(\frac{k(k+1)}{2}\right)^2.$$

當 $k = 1000$ 時，$\frac{1}{2}k(k+1) = \frac{1}{2}\cdot 1000 \cdot 1001 = 500500$，所以

$$\sum_{m=0}^{1000} m^3 = (500500)^2,$$

這就是所要求的結果。

## 驗算

- 較小的 $k$：$k = 2$ 給出 $0+1+8 = 9 = \left(\frac{2\cdot 3}{2}\right)^2$，而 $k = 3$ 給出 $36 = 6^2$。
- 同樣的方法用於 $n = 2$，得 $2\sum m = P_2(k+1) - P_2(0) = (k+1)^2 - (k+1) = k(k+1)$，即大家熟悉的三角形數（triangular number）公式。留意立方和正好是它的平方，這是一個值得發現的巧妙之處。
- 用電腦以精確分數構造 $P_1, \ldots, P_{12}$，可在隨機選取的有理點上驗證差分恆等式，並對 $k \le 39$ 驗證求和公式。直接把立方相加，得 $250\,500\,250\,000 = 500500^2$。（我們略過的那個常數，原來是 $c = -\frac{1}{30}$。）

## 常見錯誤

- 試圖把 $Q_n$ 在 $[0,1]$ 上積分，藉此確定 $Q_n$ 這個常數。這樣會引入 $\int_1^2 P_n$，而我們對它一無所知。在 $x = 0$ 取值，再運用關於 $P_{n-1}$ 的條件，要簡潔得多。
- 忘記了 $P_n(1) = P_n(0)$ 這一步需要 $n \ge 2$，所以基礎情況必須是 $n = 1$，並要親手驗證。
- 花時間去求 $P_4$ 的常數項，而它其實會抵消。

## STEP 之外

伯努利多項式（Bernoulli polynomials）有生成函數（generating function）$\dfrac{te^{xt}}{e^t - 1} = \sum_{n\ge 0} P_n(x)\dfrac{t^n}{n!}$，而它們在 0 的值就是伯努利數（Bernoulli numbers）$1, -\frac{1}{2}, \frac{1}{6}, 0, -\frac{1}{30}, \ldots$。它們是歐拉—麥克勞林公式（Euler–Maclaurin formula）的核心；這條公式把一個和與一個積分及其誤差項作比較。它們也出現在歐拉（Euler）對 $\sum 1/n^{2k}$ 的求值之中。想進一步了解這些想法，可參閱 Graham、Knuth 與 Patashnik 合著的 *Concrete Mathematics*（《具體數學》），書中對冪和與伯努利數有深入的論述。

詳細解答

# STEP III 1987，第 11 題

力學 • 相對速度的最優化

**STEP III 1987，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：

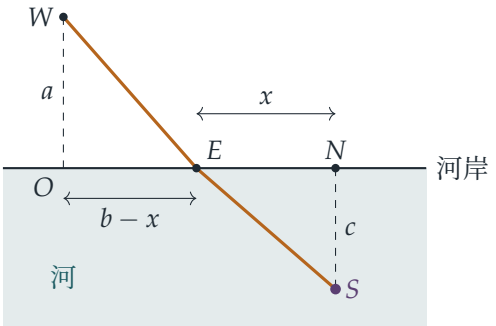
step.maths.org/questions/87-s3-q11

## 關鍵思路

把總時間寫成單一變數 $x$ 的函數，然後令其導數（derivative）等於零。在靜水中，這其實就是光學中的折射（refraction）問題，答案是斯涅耳定律（Snell's law）。有水流時，唯一要多做的工夫是計算游泳時間：水流速度與她的游泳速度大小相同，所以兩者之和是一個菱形（rhombus）的對角線，游泳時間因而成為 $x$ 的一個出奇地簡單的函數。

## 解答

把這名女子視為質點（particle），她在陸地上以恆定速率 $u$ 運動，在水中則以相對於水的恆定速率 $v$ 運動，兩段路程都沿直線進行。取河岸為 $X$ 軸，以下游為正方向，田野為 $Y > 0$，並設女子位於 $W = (0, a)$。於是樹樁位於 $S = (b, -c)$，河岸上離樹樁最近的點是 $N = (b, 0)$，而她在 $E = (b - x, 0)$ 下水，此點在 $N$ 的上游，距離為 $x$。




她跑步的距離是 $WE = \sqrt{a^2 + (b-x)^2}$，游泳的距離是 $ES = \sqrt{x^2 + c^2}$。

### 靜水

總時間為

$$T(x) = \frac{\sqrt{a^2 + (b-x)^2}}{u} + \frac{\sqrt{x^2 + c^2}}{v},$$

所以

$$T'(x) = -\frac{b-x}{u\sqrt{a^2 + (b-x)^2}} + \frac{x}{v\sqrt{x^2 + c^2}}.$$

函數 $s \mapsto s/\sqrt{k^2 + s^2}$ 是遞增的，所以 $T'$ 的兩項都隨 $x$ 增加而增加。因此 $T'$ 嚴格遞增，$T$ 恰好有一個駐點（stationary point），而這個駐點就是它的最小值點。由於 $T'(0) < 0 < T'(b)$，這一點位於 $0 < x < b$ 之內。所求的方程是

$$\boxed{\frac{b-x}{u\sqrt{a^2 + (b-x)^2}} = \frac{x}{v\sqrt{x^2 + c^2}}}$$

或等價地 $v(b-x)\sqrt{x^2 + c^2} = ux\sqrt{a^2 + (b-x)^2}$。若以 $WE$ 和 $ES$ 與河岸法線（normal）所成的角 $\alpha$ 和 $\beta$ 表示，這就是 $\dfrac{\sin\alpha}{u} = \dfrac{\sin\beta}{v}$，亦即折射的斯涅耳定律。

### 流動的河水

現在河水以速率 $v$ 向下游流動，而她相對於水的游泳速率也是 $v$。要沿直線 $ES$ 前進，她必須選擇一個游向（heading），與下游方向成向下的角 $\theta$，使她相對於地面的速度沿 $ES$ 的方向。

**想一想** 她沿 $ES$ 實際移動得有多快？為甚麼她必須在樹樁的上游下水？

她相對於地面的速度是水流速度加上她的游泳速度：

$$v\begin{pmatrix}1\\0\end{pmatrix} + v\begin{pmatrix}\cos\theta\\-\sin\theta\end{pmatrix} = 2v\cos\tfrac{\theta}{2}\begin{pmatrix}\cos\frac{\theta}{2}\\-\sin\frac{\theta}{2}\end{pmatrix},$$

這裏用了 $1 + \cos\theta = 2\cos^2\frac{\theta}{2}$ 和 $\sin\theta = 2\sin\frac{\theta}{2}\cos\frac{\theta}{2}$。從幾何上看，兩個速度的長度相等，所以它們構成一個菱形，而兩者之和平分它們之間的夾角。因此她相對於地面的路徑與下游方向成向下的角 $\phi = \theta/2$。因為 $0 \le \theta \le \pi$，我們有 $0 \le \phi \le \frac{\pi}{2}$，所以她只能向下游漂移：樹樁必須在 $E$ 的下游，即 $x > 0$。要使路徑沿着 $ES$，我們需要

$$\tan\phi = \frac{c}{x}, \qquad \text{所以} \qquad \cos\phi = \frac{x}{\sqrt{x^2 + c^2}},$$

而她沿 $ES$ 的速率是 $2v\cos\phi = \dfrac{2vx}{\sqrt{x^2 + c^2}}$。游泳時間等於距離除以速率：

$$\frac{\sqrt{x^2 + c^2}}{2vx/\sqrt{x^2 + c^2}} = \frac{x^2 + c^2}{2vx}.$$

所以總時間為

$$T(x) = \frac{\sqrt{a^2+(b-x)^2}}{u} + \frac{x^2+c^2}{2vx} = \frac{\sqrt{a^2+(b-x)^2}}{u} + \frac{1}{2v}\left(x + \frac{c^2}{x}\right), \qquad x > 0,$$

而

$$T'(x) = -\frac{b-x}{u\sqrt{a^2+(b-x)^2}} + \frac{1}{2v}\left(1 - \frac{c^2}{x^2}\right).$$

當 $x > 0$ 時，兩項同樣隨 $x$ 嚴格遞增。此外，當 $x \to 0^+$ 時 $T'(x) \to -\infty$，而當 $x$ 很大時 $T'(x) > 0$，所以恰好有一個駐點，並且它給出最短時間。令 $T'(x) = 0$，再乘以 $2vx^2u\sqrt{a^2+(b-x)^2}$（此式為正），得

$$2vx^2(b-x) = u(x^2-c^2)\left[a^2+(b-x)^2\right]^{\frac{1}{2}},$$

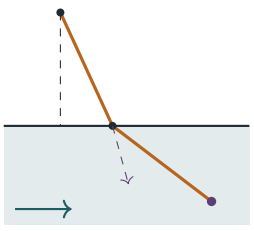
這正是所要求的結果。

### 三幅草圖

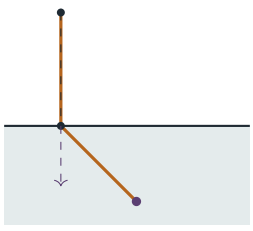
不必解這個方程，也能從中看出 $x$ 的位置。因子 $[a^2+(b-x)^2]^{1/2}$ 是正數，所以 $b-x$ 和 $x-c$ 必須同號（或同時為零）。

- $b > c$：這時 $c < x < b$。假如 $x \le c$，便會有 $b-x > 0 \ge x-c$；假如 $x \ge b$，便會有 $b-x \le 0 < x-c$；兩者都不可能。她在 $b-x > 0$ 處下水，位於 $O$ 的下游，而該點是河岸上離她最近的點。由於 $x > c$，她的游泳路徑與河岸所成的角 $\phi < 45°$，而她的游向 $\theta = 2\phi < 90°$ 偏向下游。
- $b = c$：這時按同樣的論證可得 $x = c = b$。她直接跑到 $O$，然後正對着對岸橫渡（$\theta = 90°$），水流則帶着她沿一條與河岸成 45° 的路徑前進。
- $b < c$：這時 $b < x < c$，所以 $b-x < 0$。她要向上游跑去下水，沿一條比 45° 更陡的路徑游泳，而她的游向 $\theta > 90°$ 部分指向上游。

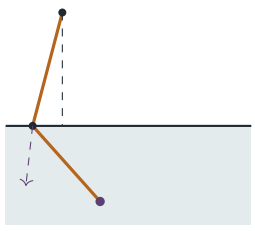
以下草圖按比例繪畫，取 $a = 3$、$c = 2$、$u = 2$、$v = 1$，以及 $b = 4, 2, 1$。數值上，$x \approx 2.62$、$x = 2$ 和 $x \approx 1.79$。水流向右；$E$ 處的虛線箭頭表示她相對於水的游向。



$b > c$



$b = c$



$b < c$

## 驗算

- 若 $b = c$，則 $x = c$ 使題目所給方程的兩邊都等於零，與上面所得一致。
- 正對着對岸橫渡（$\theta = 90°$）：她每橫渡 $v$，便向下游漂移 $v$，所以路徑成 45°。公式給出速率 $2v\cos 45° = \sqrt{2}\,v$，路徑長度為 $\sqrt{2}\,c$，所以時間是 $c/v$，這顯然是應有的結果。
- 根據算術平均—幾何平均不等式（AM–GM），游泳時間 $(x^2 + c^2)/(2vx)$ 在 $x = c$ 時最小。這是合理的：假如跑步不花任何時間，她會在 $N$ 上游距離 $c$ 處下水，然後正對着對岸橫渡。
- 數值上，對 $a, b, c, u, v$ 的 2000 組隨機取值直接求 $T$ 的最小值，所得的最小值點每次都滿足兩個方程，而且在第二種情況下都位於 $b$ 與 $c$ 之間。逐步模擬她在水流中游泳的過程，亦驗證了游泳時間。例如，$a = 3$、$b = 4$、$c = 2$、$u = 2$、$v = 1$ 給出 $x \approx 2.621$。

## 常見錯誤

- 把水流速率直接加到她的速率上，彷彿她是對準樹樁直游過去。她必須偏着游，使合速度（resultant velocity）指向樹樁。
- 忘記了有水流時只有 $x > 0$ 才可能：由於兩個速率相等，她根本無法逆流而上。
- 過早把關於 $x$ 的方程兩邊平方。這樣會失去正負號的資訊，而正是這些資訊把 $x$ 定在 $b$ 與 $c$ 之間。

## STEP 之外

第一個答案是斯涅耳定律，它源自費馬原理（Fermat's principle）：光沿所需時間最短的路徑行進。選擇整條路徑以使時間最短，正是變分法（calculus of variations）所研究的課題。與此密切相關的是策梅洛（Zermelo）導航問題：水流因地而異，如何駕駛一艘船穿越其中，才能最快到達目的地。Kleppner 與 Kolenkow 合著的 *An Introduction to Mechanics*（力學入門教科書）對相對速度和運動參考系有細緻的討論。

詳細解答

# STEP III 1987，第 12 題

力學 • 變質量轉動

> **STEP III 1987，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/87-s3-q12

## 關鍵思路

這是一道變質量（variable mass）問題，所以不能盲目套用 $F = ma$（或它在轉動中的對應公式）。穩妥的做法是回到基本原理：比較時刻 $t$ 與時刻 $t + \delta t$ 繞樞軸（pivot）的角動量（angular momentum），其間有少量已燃燒的燃料離開了系統。樞軸是光滑的，而重力沿鉛垂方向，所以沒有任何力對通過樞軸的鉛垂軸產生力矩（moment），因此該角動量守恆。

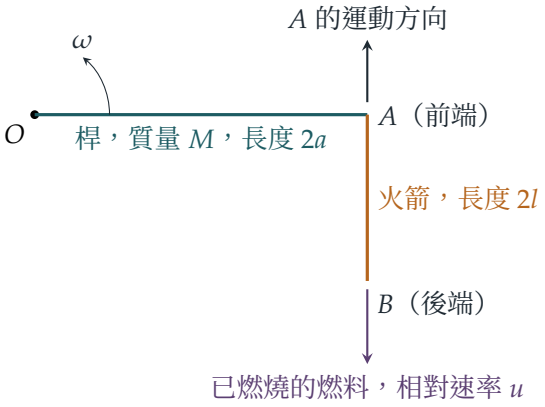
## 解答

### 建立模型

把桿和火箭都看成均勻剛體桿，在水平面上繞通過樞軸 $O$ 的鉛垂軸轉動。設桿為 $OA$，其中 $OA = 2a$。取單位向量 $\mathbf{e}_r$ 沿 $OA$ 方向，$\mathbf{e}_\theta$ 與它垂直並指向轉動的方向，又設 $\mathbf{k}$ 為鉛垂向上的單位向量，於是 $\mathbf{e}_r \times \mathbf{e}_\theta = \mathbf{k}$。

火箭的前端在 $A$，並指向前方，即 $A$ 的運動方向。因此火箭沿 $\mathbf{e}_\theta$ 方向放置，由 $A$ 向後延伸至其後端 $B$，其中

$$\overrightarrow{OB} = 2a\,\mathbf{e}_r - 2l\,\mathbf{e}_\theta.$$




**轉動慣量**

桿繞其一端的轉動慣量（moment of inertia）為 $\frac{1}{3}M(2a)^2 = \frac{4}{3}Ma^2$。

火箭的中心位於 $2a\,\mathbf{e}_r - l\,\mathbf{e}_\theta$，與 $O$ 的距離為 $\sqrt{4a^2+l^2}$。它繞自身中心的轉動慣量為 $\frac{1}{3}ml^2$，所以由平行軸定理（parallel axis theorem），它的貢獻是 $\frac{1}{3}ml^2 + m(4a^2+l^2)$。總和為

$$I = \tfrac{4}{3}Ma^2 + m\left(4a^2 + \tfrac{4}{3}l^2\right).$$

**$I$ 如何變化**

由於 $m = 2\alpha l$ 及 $\dfrac{\mathrm{d}m}{\mathrm{d}t} = -\alpha\beta$，我們有 $\dfrac{\mathrm{d}l}{\mathrm{d}t} = -\frac{1}{2}\beta$。用積法則（product rule）對 $I$ 求導，

$$\begin{aligned}\frac{\mathrm{d}I}{\mathrm{d}t} &= \frac{\mathrm{d}m}{\mathrm{d}t}\left(4a^2 + \tfrac{4}{3}l^2\right) + m\cdot\tfrac{8}{3}l\,\frac{\mathrm{d}l}{\mathrm{d}t}\\ &= -\alpha\beta\left(4a^2 + \tfrac{4}{3}l^2\right) + 2\alpha l\cdot\tfrac{8}{3}l\cdot\left(-\tfrac{1}{2}\beta\right) = -\alpha\beta\,(4a^2+4l^2).\end{aligned}$$

留意 $4a^2 + 4l^2 = OB^2$：轉動慣量下降的速率，恰好等於從後端 $B$ 移走質量時所得的速率。這是合理的，因為一枚均勻的火箭若要保持均勻，而前端又固定在 $A$，就只能從後端縮短。

**噴出燃料的角動量**

在由 $t$ 到 $t+\delta t$ 的短時間內，有質量為 $\delta\mu = \alpha\beta\,\delta t$ 的已燃燒燃料從 $B$ 離開。燃燒之前，它隨火箭一起運動，速度為 $\mathbf{v}_B = \omega\mathbf{k}\times\overrightarrow{OB}$。燃燒之後，它的速度是 $\mathbf{v}_B - u\,\mathbf{e}_\theta$，因為它相對於火箭是向正後方運動的。這時它繞 $O$ 的角動量為

$$\begin{aligned}\delta\mu\,\overrightarrow{OB}\times(\mathbf{v}_B - u\,\mathbf{e}_\theta) &= \delta\mu\left[\omega\,OB^2\,\mathbf{k} - (2a\,\mathbf{e}_r - 2l\,\mathbf{e}_\theta)\times u\,\mathbf{e}_\theta\right]\\ &= \delta\mu\left[\omega(4a^2+4l^2) - 2au\right]\mathbf{k},\end{aligned}$$

其中用到 $\mathbf{e}_\theta \times \mathbf{e}_\theta = \mathbf{0}$，以及當 $\mathbf{r}$ 垂直於 $\mathbf{k}$ 時 $\mathbf{r} \times (\omega\mathbf{k} \times \mathbf{r}) = \omega r^2 \mathbf{k}$。

**想一想** 為甚麼在排氣的那一項中，$\overrightarrow{OB}$ 只有 $2a\,\mathbf{e}_r$ 這部分起作用？

排氣的相對速度沿 $\mathbf{e}_\theta$ 方向，與火箭在同一直線上。所以 $\overrightarrow{OB}$ 沿火箭方向的分量不產生力矩。無論燃料在火箭上哪一點離開，由 $O$ 到排氣（以及推力）作用線的垂直距離都只是 $OA = 2a$。

### 角動量守恆

樞軸是光滑的，所以它的反作用力對通過 $O$ 的鉛垂軸沒有力矩；重力沿鉛垂方向，所以也沒有力矩。因此整個系統（煙花連同在 $\delta t$ 內燃燒的燃料）繞該軸的角動量在 $t$ 與 $t+\delta t$ 兩個時刻相同：

$$I(t+\delta t)\,\omega(t+\delta t) + \delta\mu\left[\omega(4a^2+4l^2) - 2au\right] = I(t)\,\omega(t),$$

誤差只有 $(\delta t)^2$ 階的項。兩邊除以 $\delta t$，並令 $\delta t \to 0$，其中 $\delta\mu/\delta t = \alpha\beta$，得

$$\frac{\mathrm{d}}{\mathrm{d}t}(I\omega) + \alpha\beta\,\omega(4a^2+4l^2) - 2\alpha\beta au = 0.$$

展開 $\dfrac{\mathrm{d}}{\mathrm{d}t}(I\omega) = I\dfrac{\mathrm{d}\omega}{\mathrm{d}t} + \omega\dfrac{\mathrm{d}I}{\mathrm{d}t}$，並代入 $\dfrac{\mathrm{d}I}{\mathrm{d}t} = -\alpha\beta(4a^2+4l^2)$，兩個含 $\omega$ 的項恰好互相抵消，餘下

$$I\frac{\mathrm{d}\omega}{\mathrm{d}t} = 2\alpha\beta au.$$

這正是「轉動慣量 × 角加速度 = 推力的力矩」：推力（thrust）為 $\alpha\beta u$，沿火箭方向作用，與 $O$ 的垂直距離為 $2a$。最後，

$$\frac{\mathrm{d}\omega}{\mathrm{d}t} = \frac{2\alpha\beta au}{\frac{4}{3}Ma^2 + 2\alpha l\left(4a^2 + \frac{4}{3}l^2\right)} = \frac{6\alpha\beta au}{4Ma^2 + 8\alpha l(3a^2+l^2)}$$
$$= \frac{3\alpha\beta au}{2\left[Ma^2 + 2\alpha l(3a^2+l^2)\right]},$$

正如所求。只要還有燃料剩下，即 $l > 0$，這個結果就成立。

## 驗算

- 沒有火箭時（$l \to 0$），結果變成 $\frac{4}{3}Ma^2\,\dot{\omega} = 2a \cdot \alpha\beta u$：推力乘以力臂，與預期相符。
- 量綱：$\alpha\beta u$ 是一個力（$\alpha$ 是每單位長度的質量，$\beta$ 是一個速率），所以分子是力 × 長度，而分母是轉動慣量。
- 數值上，取 $M = 1.3$、$a = 0.7$、$\alpha = 0.9$、$\beta = 0.4$、$u = 2.5$、$l(0) = 1.1$ 及 $\omega(0) = 0.3$，我們把運動分成多步來模擬：從後端逐一噴出 $2 \times 10^5$ 小塊，每次都令角動量守恆。在 $t = 4.95$ 時，這樣得出 $\omega \approx 2.39817$，而對公式積分則得出 2.39816。

## 常見錯誤

- 只對煙花本身寫出 $\frac{\mathrm{d}}{\mathrm{d}t}(I\omega) =$ 力矩。煙花並不是封閉系統，除非你同時計及燃料帶走的角動量，否則這樣會得出錯誤的答案。
- 把排氣的絕對速度當作向後的 $u$，而不是相對於運動中的火箭的 $u$。
- 忘記 $l$ 會改變，以致忽略了火箭的轉動慣量是透過 $m$ 和 $l$ 兩者而隨時間變化的。

## STEP 之外

這裏的抵消是一個普遍的事實。若物體失去質量時，離開的質量帶有當地的剛體速度再加上一個相對排氣速度，則運動方程就是瞬時物體的普通運動方程，另加一個由相對速度產生的「推力」。這就是火箭方程(rocket equation)的轉動版本。Kleppner 與 Kolenkow 的 *An Introduction to Mechanics*（力學入門教科書）仔細討論了變質量系統和角動量，程度正好與本題相若。

詳細解答

# STEP III 1987，第 13 題

力學 • 涉及摩擦力的轉動桿

> **STEP III 1987，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/87-s3-q13

## 關鍵思路

只要質點仍緊貼在桿上不滑動，桿與質點便作為一個剛體（rigid body）一同繞鉸鏈轉動，因此以整體的轉動慣量（moment of inertia）應用能量守恆（conservation of energy），即可立刻求得 $\dot{\theta}^2$。要找出何時開始滑動，我們再單獨考慮質點：它繞鉸鏈作圓周運動，所以在徑向和橫向應用牛頓第二定律（Newton's second law），便能準確知道桿須提供多大的摩擦力（friction）和法向反作用力（normal reaction）。當所需的摩擦力達到法向反作用力的 $\frac{1}{2}$ 時，滑動便開始。

## 解答

### 運動方程

把桿視為繞光滑鉸鏈 $O$ 轉動的均勻剛性桿，並把質點視為與 $O$ 相距 $a$ 的質點。在質點滑動之前，它一直與 $O$ 保持距離 $a$，並隨桿一同轉動。

系統對 $O$ 的轉動慣量為

$$I = \tfrac{4}{3}(3m)a^2 + ma^2 = 4ma^2 + ma^2 = 5ma^2,$$

這裏用了長度為 $2l$ 的均勻桿對其一端的轉動慣量 $\frac{4}{3}Ml^2$。

鉸鏈是光滑的，而質點與桿之間的摩擦力和法向反作用力都是系統的內力，所以它們所作的淨功為零。當桿轉過角度 $\theta$ 時，桿的中心和質點（兩者與 $O$ 的距離均為 $a$）各下降了 $a\sin\theta$。由於系統從靜止開始運動，能量守恆給出

$$\tfrac{1}{2}(5ma^2)\dot{\theta}^2 = 3mg\,a\sin\theta + mg\,a\sin\theta = 4mga\sin\theta,$$

即

$$5a\dot{\theta}^2 = 8g\sin\theta,$$

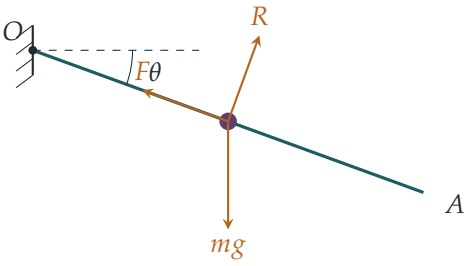
這正是所要證明的結果。對 $t$ 求導得 $10a\dot{\theta}\ddot{\theta} = 8g\cos\theta\,\dot{\theta}$，所以（當 $\dot{\theta} \neq 0$ 時，而運動一旦開始這便成立）

$$a\ddot{\theta} = \tfrac{4}{5}g\cos\theta.$$

（對 $O$ 取力矩也可得到同樣的結果：$5ma^2\ddot{\theta} = 4mga\cos\theta$。）

### 質點何時滑動

現在單獨考慮質點。作用在它身上的力有：重量 $mg$、垂直於桿的法向反作用力 $R$，以及沿桿方向的摩擦力 $F$。由於質點在以 $O$ 為圓心、半徑為 $a$ 的圓上運動，它的加速度有一個沿桿指向 $O$ 的分量 $a\dot{\theta}^2$，以及一個垂直於桿、沿 $\theta$ 增加方向的分量 $a\ddot{\theta}$。




我們取 $F$ 沿桿指向 $O$，稍後會看到這正是質點所需的方向。沿桿方向分解，以指向 $O$ 為正：重量有一個背離 $O$ 的分量 $mg\sin\theta$，所以

$$F - mg\sin\theta = ma\dot{\theta}^2 = \tfrac{8}{5}mg\sin\theta \quad\Longrightarrow\quad F = \tfrac{13}{5}mg\sin\theta.$$

垂直於桿方向分解，以 $\theta$ 增加的方向（即由桿向下）為正：重量的分量為 $mg\cos\theta$，所以

$$mg\cos\theta - R = ma\ddot{\theta} = \tfrac{4}{5}mg\cos\theta \quad\Longrightarrow\quad R = \tfrac{1}{5}mg\cos\theta.$$

**想一想** 重力看來是把質點沿斜面向下拉，為甚麼質點反而需要指向鉸鏈的摩擦力？

起作用的不只是重力：質點要繼續繞 $O$ 作圓周運動，就需要一個指向 $O$ 的淨力 $ma\dot{\theta}^2$。重力沿桿的分量卻指向相反方向，即背離 $O$，所以摩擦力既要提供向心力（centripetal force），又要抵銷重力的分量。這就是為甚麼 $F$ 算出來是指向 $O$ 為正，並且隨 $\sin\theta$ 增長；而 $R$ 很小，是因為桿正在質點下方「墜離」（在 $\theta = 0$ 時，質點有向下的加速度 $\frac{4}{5}g$，所以桿只承托它重量的 $\frac{1}{5}$）。

當 $0 < \theta < \frac{\pi}{2}$ 時，$F$ 和 $R$ 均為正，所以質點一直與桿保持接觸。只要 $F \le \frac{1}{2}R$，質點便相對於桿保持靜止，即

$$\tfrac{13}{5}mg\sin\theta \le \tfrac{1}{2}\cdot\tfrac{1}{5}mg\cos\theta \iff \tan\theta \le \tfrac{1}{26}.$$

在 $\theta = 0$ 時，我們有 $F = 0 < \frac{1}{2}R$，而 $F/R = 13\tan\theta$ 隨 $\theta$ 遞增，所以質點最初開始滑動（向外，背離鉸鏈）是在

$$\tan\theta = \tfrac{1}{26},$$

之時，這正是所要證明的結果。

## 驗算

- 在 $\theta = 0$ 時，力矩方程給出角加速度（angular acceleration）$\frac{4g}{5a}$；所以質點的加速度為 $\frac{4}{5}g$，小於 $g$，這與 $R = \frac{1}{5}mg > 0$ 一致。
- 以數值方法逐步計算剛體運動（取 $g = 9.81$，$a = 0.7$），所得結果與 $5a\dot\theta^2 = 8g\sin\theta$ 吻合，誤差在 $10^{-12}$ 以內；而直接由質點的計算加速度求得的摩擦力和反作用力，亦與 $\frac{13}{5}mg\sin\theta$ 和 $\frac{1}{5}mg\cos\theta$ 相符。比值 $F/R$ 最初達到 $\frac{1}{2}$ 時，$\theta \approx 0.03845$ 弧度，即 $\arctan\frac{1}{26} \approx 2.20^\circ$。

## 常見錯誤

- 忘記向心項，把質點當作在斜面上處於平衡來分解力；這樣不但得出錯誤的角度，連摩擦力的方向也會弄錯。
- 正確地使用了 $\frac{1}{3}(3m)(2a)^2$，卻忘記把質點的 $ma^2$ 加進轉動慣量。
- 像在固定斜面上那樣取 $R = mg\cos\theta$：桿正在加速，所以反作用力要小得多。

## STEP 之外

這道題讓我們初嘗在轉動參考系（rotating frame）中處理問題的滋味：從桿上觀察，質點受到一個向外的「離心力」（centrifugal force）$ma\dot\theta^2$ 和一個「歐拉力」（Euler force）$-ma\ddot\theta$，而這些虛擬力（fictitious forces）正是我們求得的那些項。大學本科的力學課程會有系統地發展這套理論，並會討論質點開始滑動後出現的科里奧利力（Coriolis force）。Kleppner 與 Kolenkow 合著的 *An Introduction to Mechanics*（力學入門）對剛體轉動和非慣性參考系（non-inertial frames）有十分清晰的講解。

詳細解答

# STEP III 1987，第 14 題

力學 • 引力坍縮

**STEP III 1987，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/87-s3-q14

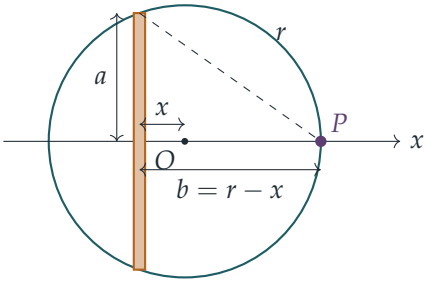
## 關鍵思路

實心球體可視為一疊薄圓盤，因此整個球體所施的力，就是把題目給出的圓盤公式沿着穿過質點（particle）的軸線積分。那個平方根看似麻煩，但只要留意到，當質點位於球面上時，根號內的式子可以因式分解，問題便迎刃而解。之後，關於雲團的部分就是牛頓第二定律（Newton's second law）配合鏈式法則（chain rule）$\frac{\mathrm{d}v}{\mathrm{d}t} = v\frac{\mathrm{d}v}{\mathrm{d}r}$，而代換（substitution）$r = R\cos^2\alpha$ 的用意，正是要整理第二次積分時出現的平方根。

## 解答

### 球面上的力

把球心置於原點，質點 $P$ 置於 $x$ 軸上 $x = r$ 之處。把球體切成與軸線垂直的圓盤。位於 $x$ 處的薄片（其中 $-r \le x \le r$）厚度為 $\delta x$，半徑為 $a$，其中 $a^2 = r^2 - x^2$，而 $P$ 位於該薄片的軸線上，與它相距 $b = r - x \ge 0$。



關鍵的化簡在於平方根：

$$a^2+b^2=(r^2-x^2)+(r-x)^2=2r^2-2rx=2r(r-x),$$

所以

$$\frac{b}{(a^2+b^2)^{1/2}}=\frac{r-x}{\sqrt{2r(r-x)}}=\sqrt{\frac{r-x}{2r}}.$$

每個圓盤都沿軸線把 $P$ 拉向球心，因此各個力的方向相同，直接相加即可。令 $\delta x\to 0$，合力為

$$\begin{aligned}F&=2\pi mk\rho\int_{-r}^{r}\left(1-\sqrt{\frac{r-x}{2r}}\right)\mathrm{d}x\\&=2\pi mk\rho\left(2r-\frac{1}{\sqrt{2r}}\int_0^{2r}u^{1/2}\,\mathrm{d}u\right)\qquad(u=r-x)\\&=2\pi mk\rho\left(2r-\frac{1}{\sqrt{2r}}\cdot\tfrac{2}{3}(2r)^{3/2}\right)=2\pi mk\rho\left(2r-\tfrac{4}{3}r\right)=\tfrac{4}{3}\pi mk\rho\, r.\end{aligned}$$

球體的質量為 $M=\frac{4}{3}\pi r^3\rho$，所以 $\frac{4}{3}\pi\rho=M/r^3$，於是

$$F=\frac{kmM}{r^2},$$

方向指向球心，正如題目所求。

### 雲團的運動方程

把雲團中的質點看成一個質量為 $M$ 的連續均勻球體，並考慮球面上一個質量為 $m$、與球心相距 $r$ 的質點。根據上面的結果，它所受的唯一的力是指向球心的 $kmM/r^2$。取 $r$ 為坐標（以向外為正），由牛頓第二定律得

$$m\frac{\mathrm{d}^2r}{\mathrm{d}t^2}=-\frac{kmM}{r^2}.$$

現在 $\frac{\mathrm{d}r}{\mathrm{d}t}=-v$，所以 $\frac{\mathrm{d}^2r}{\mathrm{d}t^2}=-\frac{\mathrm{d}v}{\mathrm{d}t}$，而根據鏈式法則，

$$\frac{\mathrm{d}v}{\mathrm{d}t}=\frac{\mathrm{d}v}{\mathrm{d}r}\frac{\mathrm{d}r}{\mathrm{d}t}=-v\frac{\mathrm{d}v}{\mathrm{d}r}.$$

因此 $\frac{\mathrm{d}^2r}{\mathrm{d}t^2}=v\frac{\mathrm{d}v}{\mathrm{d}r}$，約去 $m$ 後得

$$v\frac{\mathrm{d}v}{\mathrm{d}r}=-\frac{kM}{r^2}.$$

**想一想** 圓盤的結果是針對固定的球體而得出的。為甚麼可以把它用於正在坍縮的雲團？

任何一刻的力，只取決於該刻質量的分佈，所以我們需要雲團始終保持為均勻球體，而且球面以內的質量始終是 $M$。事實正是如此：一個從半徑 $\lambda R$ 處由靜止開始運動的質點（其中 $0 < \lambda \leq 1$），其內側的質量為 $\lambda^3 M$，而只要 $r(t)$ 滿足 $\ddot{r} = -kM/r^2$，$\lambda r(t)$ 便滿足 $\ddot{s} = -k\lambda^3 M/s^2$。所以每個質點與球心的距離，始終是球面半徑的同一個比例：雲團按均勻的比例因子收縮，保持均勻，而且沒有質點會越過球面上的質點。（質點外側的質量只會對內部的質點有影響，而球面之外並無任何質點。）

### 由靜止開始的坍縮

分離變數（separate the variables），並由 $r = R$（該處 $v = 0$）起積分：

$$\tfrac{1}{2}v^2 = \int_R^r -\frac{kM}{s^2}\,\mathrm{d}s = kM\left(\frac{1}{r} - \frac{1}{R}\right) = \frac{kM(R-r)}{rR}.$$

加速度始終向內，而雲團由靜止開始，所以 $r$ 一直遞減，全程 $v \geq 0$。

現在代入 $r = R\cos^2\alpha$，當 $r$ 由 $R$ 向 0 遞減時，$\alpha$ 由 $t = 0$ 時的 0 向 $\frac{\pi}{2}$ 遞增。這時 $R - r = R\sin^2\alpha$，所以

$$v^2 = \frac{2kM\,R\sin^2\alpha}{R^2\cos^2\alpha} = \frac{2kM}{R}\tan^2\alpha, \qquad v = \left(\frac{2kM}{R}\right)^{1/2}\tan\alpha,$$

這裏取正根，因為 $v \geq 0$，而且當 $0 \leq \alpha < \frac{\pi}{2}$ 時 $\tan\alpha \geq 0$。

> ? **想一想** 怎樣由 $v$ 作為 $\alpha$ 的函數，得出 $t$ 作為 $\alpha$ 的函數？

利用 $v = -\frac{\mathrm{d}r}{\mathrm{d}t}$ 和 $\frac{\mathrm{d}r}{\mathrm{d}\alpha} = -2R\cos\alpha\sin\alpha$：

$$\frac{\mathrm{d}t}{\mathrm{d}\alpha} = -\frac{1}{v}\frac{\mathrm{d}r}{\mathrm{d}\alpha} = \frac{2R\sin\alpha\cos\alpha}{(2kM/R)^{1/2}\tan\alpha} = \left(\frac{R^3}{2kM}\right)^{1/2} 2\cos^2\alpha = \left(\frac{R^3}{2kM}\right)^{1/2}(1+\cos 2\alpha).$$

積分，並利用 $\alpha = 0$ 時 $t = 0$，得

$$t = \left(\frac{R^3}{2kM}\right)^{1/2}\left(\alpha + \tfrac{1}{2}\sin 2\alpha\right),$$

這就是 $r = R\cos^2\alpha$ 的時刻，正如題目所求。

## 驗算

- 時間很短時：當 $\alpha$ 很小，$t \approx 2\alpha(R^3/2kM)^{1/2}$ 而 $R - r \approx R\alpha^2$，所以 $R - r \approx \frac{1}{2}(kM/R^2)\,t^2$：即由靜止開始、加速度為常數 $kM/R^2$ 的運動，與起初階段的預期完全吻合。
- 令 $\alpha \to \frac{\pi}{2}$，得出整個坍縮所需的時間為 $\frac{\pi}{2}(R^3/2kM)^{1/2}$；即使 $v \to \infty$，這個時間仍是有限的。

- 數值上，把 200 000 個薄圓盤的力相加，所得結果與 $kmM/r^2$ 的相對誤差約為 $2 \times 10^{-9}$。取 $k = 1.7$、$M = 2.3$、$R = 1.9$，對 $\ddot{r} = -kM/r^2$ 逐步作數值計算，直至 $\alpha = 0.8$ 所對應的時刻，得 $r = 0.922260$，與 $R\cos^2 0.8 = 0.922260$ 吻合。

## 常見錯誤

- 由球心而非由質點量度 $b$：應為 $b = r - x$，而不是 $x$。
- 忘記 $v$ 是向內的速率，這會令 $\frac{\mathrm{d}^2 r}{\mathrm{d}t^2} = v\frac{\mathrm{d}v}{\mathrm{d}r}$ 中的正負號相反。
- 為 $v$ 取了負平方根，結果算出的時間是負數。

## STEP 之外

第一部分是牛頓（Newton）球殼定理（shell theorem）的特殊情況：在球對稱物體之外，引力的作用就如全部質量都集中在球心一樣。大學課程會用引力的高斯定律（Gauss's law）$\nabla \cdot \mathbf{g} = -4\pi G\rho$ 簡潔地證明它。你求得的坍縮時間，就是天體物理學家所說的**自由落體時間**（free-fall time），用來估算氣體雲坍縮形成恆星有多快；而以 $\alpha$ 作參數的表示法，也正是描述徑向開普勒（Kepler）軌道的那一個。Kleppner 與 Kolenkow 的 *An Introduction to Mechanics*（力學入門）正是以這種方式處理有心力（central force）和球殼定理。

詳細解答

# STEP III 1987，第 15 題

概率與統計 • 期望成本的最優化

**STEP III 1987，第 15 題**

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

step.maths.org/questions/87-s3-q15

## 關鍵思路

與其直接求 $E_n$ 的最小值，不如看看多做一次測試會帶來甚麼改變。比較「做完 $n$ 次便停」與「做完 $n+1$ 次便停」這兩個策略時，只有至今已得到 $n$ 次陽性結果的病人才有分別。對這些病人來說，多做一次測試要花 £10，而若病人患的是 Steppe 綜合症而測試結果呈陰性，便可省下 £100 的治療費。因此一切取決於 $E_{n+1} - E_n$ 的正負號，而這個正負號最多只會改變一次。

## 解答

設 $F$ 為病人患上 Fenland 熱病的事件 (event)，故 $\mathrm{P}(F) = p$，並設 $S = F'$ 為患上 Steppe 綜合症。我們假設 Steppe 病人的各次測試結果互相獨立 (independent)，每次呈陽性的概率為 $q$，其中 $0 < q < 1$。不論誰接受治療，治療費都是 £100。

### 第 (i) 部分

求 $E_n$。若病人患的是 $F$，每次測試都呈陽性。因此 $n$ 次測試全部都要做，然後再接受治療，費用為 $10n + 100$。

若病人患的是 $S$，第 $k$ 次測試（其中 $1 \le k \le n$）只會在首 $k-1$ 次測試全部呈陽性時才進行，其概率為 $q^{k-1}$。因此測試次數的期望值（expected value）為

$$\sum_{k=1}^{n} q^{k-1} = \frac{1-q^n}{1-q},$$

而病人只會在 $n$ 次測試全部呈陽性時才接受治療，其概率為 $q^n$。根據期望值的線性性質

（linearity of expectation），

$$E_n = p\,(10n + 100) + (1 - p)\left(\frac{10(1 - q^n)}{1 - q} + 100q^n\right).$$

由此亦得 $E_0 = 100$，這正合預期，因為此時所有人都接受治療。

*差值。* $F$ 的部分增加了 10。在 $S$ 的部分，$\dfrac{1 - q^{n+1}}{1 - q} - \dfrac{1 - q^n}{1 - q} = q^n$，所以

$$\begin{aligned} E_{n+1} - E_n &= 10p + (1 - p)\big(10q^n + 100q^{n+1} - 100q^n\big) \\ &= 10p + 10(1 - p)q^n(1 + 10q - 10) \\ &= 10p - 10(1 - p)q^n(9 - 10q), \end{aligned}$$

這正是所要證明的結果。

解讀這個結果：有 $p$ 的概率，第 $n+1$ 次測試白白用在熱病病人身上。有 $(1 - p)q^n$ 的概率，這次測試用在 Steppe 病人身上，要花 £10，但它有 $1 - q$ 的概率可省下 £100。期望淨節省為 $100(1 - q) - 10 = 10(9 - 10q)$。

*代入數值。* 記 $D_n = E_{n+1} - E_n$。當 $p = 10^{-4}$ 及 $q = 10^{-2}$ 時，$9 - 10q = 8.9$ 而 $1 - p = 0.9999$，所以

$$D_n = 10^{-3} - 89 \times 0.9999 \times 10^{-2n}.$$

**想一想** 為甚麼只需找出 $D_n$ 在何處變號便已足夠？

這裏 $9 - 10q > 0$，而 $q^n$ 隨 $n$ 增大而嚴格遞減（strictly decreasing），所以 $D_n$ 隨 $n$ 嚴格遞增。因此當 $D_n < 0$ 時數列（sequence）$E_n$ 下降，一旦 $D_n > 0$ 便上升，而最小值出現在第一個使 $D_n > 0$ 的 $n$。現在

$$\begin{aligned} D_2 &= 10^{-3} - 0.0088991 < 0, \\ D_3 &= 10^{-3} - 0.000088991 = 0.000911\ldots > 0. \end{aligned}$$

所以 $E_0 > E_1 > E_2 > E_3$ 且 $E_3 < E_4 < E_5 < \cdots$，醫生應選 $n = 3$。（此時期望費用約為 £10.11，而立即治療則要 £100。）

### 第 (ii) 部分

取 $q_0 = \frac{9}{10}$，並假設 $q > \frac{9}{10}$。那麼 $9 - 10q < 0$，所以

$$D_n = 10p + 10(1 - p)q^n(10q - 9).$$

對任意 $p \in [0, 1]$，這兩項都是非負的。它們不可能同時為零：若 $p > 0$，第一項為正；若 $p = 0$，第二項為 $10q^n(10q - 9) > 0$。因此對任意 $n \geq 0$ 都有 $D_n > 0$，即

$$E_0 < E_1 < E_2 < \cdots.$$

不論 $p$ 取何值，醫生都應選 $n = 0$，立即治療。

這個 $q_0$ 的值不能再調低。若 $q < \frac{9}{10}$，則當 $p$ 足夠小時，$D_0 = 10p - 10(1-p)(9-10q)$ 為負，所以至少測試一次會比 $n = 0$ 更好。用文字來說，當 $q > \frac{9}{10}$ 時，Steppe 病人的測試呈陰性的機會不足十分之一，所以即使未把熱病病人計算在內，一次 £10 的測試平均省下的治療費也不足 £10。

## 驗算

- 直接計算 $n = 1$：$E_1 = 10 + 100\big(p + (1-p)q\big)$。取 $p = 0$ 得 $E_1 - E_0 = 10 - 100(1-q)$，與公式給出的 $-10q^0(9-10q)$ 吻合。
- 當 $p = 10^{-4}$ 及 $q = 10^{-2}$ 時，公式給出 $E_0 = 100$、$E_1 \approx 11.010$、$E_2 \approx 10.1210$、$E_3 \approx 10.11309$ 及 $E_4 \approx 10.11400$。最小值出現在 $n = 3$，與上面所得一致。
- 以分數作精確枚舉，已就數百組隨機的 $(p, q)$ 及 $n \le 8$ 驗證了 $E_n$ 的公式和差值恆等式；另外以 $p = 0.3$、$q = 0.5$、$n = 2$ 進行模擬，得出平均費用為 63.98，而精確值為 64。

## 常見錯誤

- 向 Steppe 病人收取全部 $n$ 次測試的費用。整個過程在第一次出現陰性結果時便停止，所以第 $k$ 次測試只有 $q^{k-1}$ 的概率會進行。
- 找出第一個使 $D_n > 0$ 的 $n$，卻沒有說明 $D_n$ 為何是單調的（monotonic）。少了這一步，數列 $E_n$ 原則上有可能在之後再次下降。
- 在第 (ii) 部分忘記了 $p = 0$ 的情況。正因為這個情況，條件才是 $q > q_0$ 而不是 $q \ge q_0$：當 $q = \frac{9}{10}$ 且 $p = 0$ 時，每個 $E_n$ 都等於 100。

## STEP 之外

這是最優停止（optimal stopping）和決策論（decision theory）的一個小例子：一連串帶有成本的行動、一個隨機的狀態，以及一條決定何時停止的法則。比較相鄰的策略（即「邊際分析」(marginal analysis)），並證明差值只變號一次，是處理這類問題的標準方法。這些想法會進一步通往貝葉斯決策論（Bayesian decision theory）和動態規劃（dynamic programming）。Grimmett and Welsh, *Probability: An Introduction*（概率論入門教材）涵蓋了支撐這些理論的條件期望（conditional expectation）工具。

詳細解答

# STEP III 1987，第 16 題

概率與統計 • 順序統計量與泊松分佈

**STEP III 1987，第 16 題**

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

step.maths.org/questions/87-s3-q16

## 關鍵思路

在第 (i) 部分，事件「$A \le k$ 且 $B \ge 1-k$」直接處理起來頗為棘手，但它的補集（complement）卻是兩個簡單事件的併集：「所有 $X_i$ 都大於 $k$」和「所有 $X_i$ 都小於 $1-k$」。餘下的工作交給容斥原理（inclusion–exclusion）便可。在第 (ii) 部分，得獎人數服從二項分佈（binomial distribution），其中 $n$ 極大而 $p$ 極小，這正是泊松分佈（Poisson distribution）大派用場的地方。因此我們先求出恰有一位或兩位得獎者的泊松概率的最大值，然後再確保近似所帶來的誤差不會把答案推過 60%。

## 解答

### 第 (i) 部分

設 $E$ 為事件 $A > k$，$F$ 為事件 $B < 1-k$。那麼 $A > k$ 表示每個 $X_i$ 都大於 $k$，而 $B < 1-k$ 表示每個 $X_i$ 都小於 $1-k$。由獨立性（independence），

$$\mathrm{P}(E) = (1-k)^n, \qquad \mathrm{P}(F) = (1-k)^n, \qquad \mathrm{P}(E \cap F) = (1-2k)^n,$$

其中最後一式成立，是因為 $E \cap F$ 表示每個 $X_i$ 都落在 $(k, 1-k)$ 之內，而這個區間的長度為 $1-2k > 0$。

我們要求的事件是 $E \cup F$ 的補集，所以由容斥原理得

$$p_n = 1 - \mathrm{P}(E \cup F) = 1 - 2(1-k)^n + (1-2k)^n.$$

由於 $0 < 1-2k < 1-k < 1$，兩個冪都趨向 0，因此

$$p_n \to 1 \quad \text{當} n \to \infty.$$

評註。無論把 $k$ 取得多小，只要觀測值夠多，我們便幾乎可以肯定：有某個觀測值與 0 相距不超過 $k$，又有另一個觀測值與 1 相距不超過 $k$。換言之，樣本最小值收斂於 0，樣本最大值收斂於 1（均指依概率收斂），所以樣本全距（sample range）$B - A$ 趨向 1，即整個區間的長度。這正是為甚麼（舉例說）$A$ 和 $B$ 是未知均勻分佈（uniform distribution）兩個端點的自然估計量（estimator）。

### 第 (ii) 部分

設 $N = 4\,000\,000$，並設 $X$ 為得獎人數。各讀者是否得獎互相獨立，每人得獎的概率為 $p$，所以 $X \sim \mathrm{B}(N, p)$。記 $q = 1 - p$。你能保住工作，當且僅當 $X = 1$ 或 $X = 2$，其概率為

$$K(p) = Npq^{N-1} + \binom{N}{2} p^2 q^{N-2}.$$

*從泊松分佈的角度看。*由於 $N$ 極大而 $p$ 很小，$X$ 近似服從平均值為 $\lambda = Np$ 的泊松分佈，所以

$$K(p) \approx h(\lambda), \qquad \text{其中} h(\lambda) = \mathrm{e}^{-\lambda}\left(\lambda + \frac{\lambda^2}{2}\right).$$

求導得

$$h'(\lambda) = \mathrm{e}^{-\lambda}\left(1 + \lambda - \lambda - \frac{\lambda^2}{2}\right) = \mathrm{e}^{-\lambda}\left(1 - \frac{\lambda^2}{2}\right),$$

它在 $0 < \lambda < \sqrt{2}$ 時為正，在 $\lambda > \sqrt{2}$ 時為負。所以 $h$ 在 $\lambda = \sqrt{2}$ 處取得最大值：

$$h(\sqrt{2}) = \mathrm{e}^{-\sqrt{2}}(\sqrt{2} + 1) \approx 0.2431 \times 2.4142 \approx 0.5869.$$

你的最佳選擇是 $p = \sqrt{2}/N \approx 3.5 \times 10^{-7}$，但即使如此，你保住工作的機會也只有約 58.7%。這就是題目所要求的推理。

> ? **想一想** 泊松分佈只是一個近似。真正的二項概率會不會悄悄越過 0.6？

不會。我們可以直接為 $K(p)$ 找出上界。由於 $q \le 1$ 及 $\binom{N}{2} \le \frac{1}{2}N^2$，

$$K(p) = q^{N-2}\left(Npq + \binom{N}{2}p^2\right) \le q^{N-2}\left(\lambda + \frac{\lambda^2}{2}\right).$$

對所有 $p$ 都有 $1 - p \le \mathrm{e}^{-p}$（直線 $1 - p$ 是凸曲線 $\mathrm{e}^{-p}$ 在 $p = 0$ 處的切線），所以 $q^{N-2} \le \mathrm{e}^{-(N-2)p} = \mathrm{e}^{-\mu}$，其中 $\mu = (N-2)p$。設 $c = N/(N-2)$，則 $\lambda = c\mu$，且 $c > 1$。於是

$$K(p) \le \mathrm{e}^{-\mu}\left(c\mu + \frac{c^2\mu^2}{2}\right) \le c^2\,\mathrm{e}^{-\mu}\left(\mu + \frac{\mu^2}{2}\right) = c^2 h(\mu) \le c^2 h(\sqrt{2}).$$

這裏 $c^2 = \left(1 - \frac{2}{N}\right)^{-2} < 1.000002$，所以無論 $p$ 取甚麼值，

$$K(p) < 1.000002 \times 0.58694 < 0.587 < 0.6.$$

因此，無論你怎樣選擇 $p$，保住工作的機會都少於 60%，這正是所要證明的。

## 驗算

- 在第 (i) 部分，取 $n = 1$ 得 $p_1 = 1 - 2(1-k) + (1-2k) = 0$。這是對的：當 $k < \frac{1}{2}$ 時，單獨一個觀測值不可能既 $\le k$ 又 $\ge 1-k$。取 $n = 2$ 則得 $p_2 = 2k^2$，即兩點分別落在兩端的窄帶內（次序不拘）的概率。
- 以 $n = 10$、$k = 0.3$ 進行模擬，結果為 0.944，而公式給出 0.9436。
- 在 $\lambda = Np$ 的細密網格上計算精確的二項概率 $K(p)$，得出最大值為 0.58694，在 $\lambda \approx 1.414$ 處取得，與泊松值在小數點後五位吻合。

## 常見錯誤

- 寫成 $p_n = \mathrm{P}(A \le k)\,\mathrm{P}(B \ge 1-k)$。同一個樣本的最小值和最大值並不獨立。
- 直接對 $p$ 求二項概率的最大值。這樣做是可行的，但代數運算相當繁瑣；改用 $\lambda = Np$ 來處理，問題便化為一行就能完成的微積分題。
- 做到泊松近似便停筆，沒有留意到 0.587 與 0.6 相當接近，值得補上一些論證。

## STEP 之外

第 (i) 部分讓我們初嘗次序統計量（order statistics）和依概率收斂（convergence in probability）的滋味：樣本的最小值和最大值會收斂於支集（support）的兩端。第 (ii) 部分是泊松的「稀有事件定律」（law of rare events），而我們所用的上界，正是以下一般事實的一個簡單例子：當 $p$ 很小時，二項分佈與泊松分佈在全變差（total variation）的意義下非常接近。這兩個想法在 Grimmett 與 Welsh 的 *Probability: An Introduction*，以及 Blitzstein 與 Hwang 的 *Introduction to Probability* 中都有詳細闡述。

# 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

**1987 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。