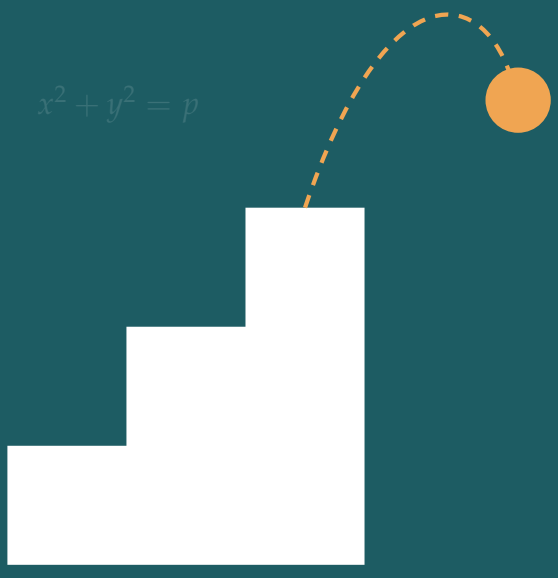
STEP 備試 · 大學數學

Beyond STEP

STEP I, II 及 III

2011 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

2011 年詳細解答

Keith McNulty

Beyond STEP：以歷屆 STEP 試題為門，走進大學數學

2011 年詳細解答

2011 年 STEP 試卷（STEP I, II 及 III）全部 39 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 隱函數微分法 純數學 . . . 3

第 2 題 代換積分法 純數學 . . . 7

第 3 題 三角恆等式 純數學 . . . 10

第 4 題 拋物線的切線 純數學 . . . 14

第 5 題 含絕對值的積分 純數學 . . . 18

第 6 題 二項式級數求和 純數學 . . . 21

第 7 題 可分離變量微分方程 純數學 . . . 24

第 8 題 丟番圖方程 純數學 . . . 28

第 9 題 拋體運動 力學 . . . 31

第 10 題 彈性碰撞 力學 . . . 35

第 11 題 剛體的平衡 力學 . . . 39

第 12 題 排隊與選票問題 概率與統計 . . . 43

第 13 題 連續隨機變量 概率與統計 . . . 47

STEP II

第 1 題 含平方根的方程 純數學 . . . 51

第 2 題 丟番圖方程 純數學 . . . 55

第 3 題 以單調性證明不等式 純數學 . . . 58

第 4 題 三角方程 純數學 . . . 62

第 5 題 向量與反射 純數學 . . . 66

第 6 題 分部積分法 純數學 . . . 69

第 7 題 數列與級數 純數學 . . . 73

第 8 題 圓的漸伸線 純數學 . . . 77

第 9 題 碰撞與恢復係數 力學 . . . 81

第 10 題 拋體與衝量 力學 85
第 11 題 三維空間中的平衡 力學 88
第 12 題 博弈中的等比級數 概率與統計 92
第 13 題 偏度的量度 概率與統計 96

STEP III

第 1 題 降階法 純數學 100
第 2 題 多項式的有理根 純數學 104
第 3 題 解三次方程 純數學 107
第 4 題 楊氏不等式 純數學 111
第 5 題 以積分求面積 純數學 115
第 6 題 積分代換 純數學 119
第 7 題 數學歸納法 純數學 122
第 8 題 複數軌跡 純數學 126
第 9 題 圓柱上的運動 力學 130
第 10 題 彈性繩與簡諧運動 力學 134
第 11 題 剛體轉動 力學 138
第 12 題 以概率生成函數處理複合分佈 概率與統計 142
第 13 題 離散分佈的眾數 概率與統計 145

詳細解答

STEP I 2011，第 1 題

純數學 • 隱函數微分法

> **STEP I 2011，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/11-s1-q1

關鍵思路

用隱函數求導（implicit differentiation），毋須把 y 解出來，就能得到每條曲線的斜率（gradient）。之後每一部分都只需寫出交點 (p,q) 須滿足的條件：它在直線上、它在曲線上，而兩者的斜率之間有某種關係（相切時相等，法線（normal）時乘積為 -1）。斜率條件給出 p 與 q 之間一個簡單的關係；把這關係代入另外兩條方程，便可用兩種不同方法確定 p，而兩者必須一致。

解答

第 (i) 部分

把 y 視為 x 的函數（function），將 $\dfrac{a}{x}+\dfrac{b}{y}=1$ 對 x 求導，得

$$-\frac{a}{x^2}-\frac{b}{y^2}\frac{\mathrm{d}y}{\mathrm{d}x}=0.$$

由於 $b\neq 0$，可移項得

$$\frac{\mathrm{d}y}{\mathrm{d}x}=-\frac{ay^2}{bx^2},$$

此即所求。（曲線上每一點的 x 和 y 均不為零，因為方程中含有 a/x 和 b/y。）

斜率相等。直線 $ax+by=1$ 的斜率為 $-\frac{a}{b}$。曲線在 (p,q) 的斜率為 $-\dfrac{aq^2}{bp^2}$，因此

$$-\frac{aq^2}{bp^2}=-\frac{a}{b}\quad\Longrightarrow\quad\frac{q^2}{p^2}=1\quad\Longrightarrow\quad q^2=p^2,$$

其中我們約去了 a/b，這是容許的，因為 $a\neq 0$。故 $p=\pm q$。

兩種情況。點 (p,q) 亦滿足

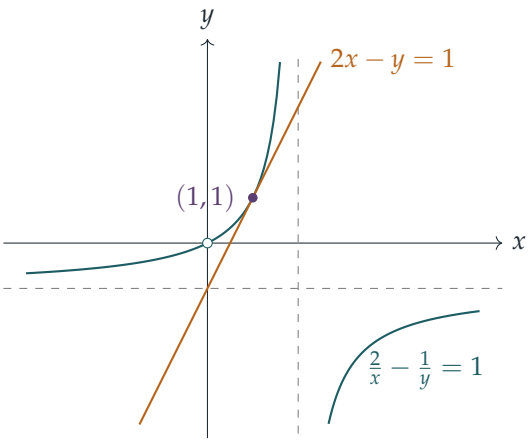
$$ap+bq=1 \qquad 及 \qquad \frac{a}{p}+\frac{b}{q}=1.$$

- 若 $q=p$：由直線得 $p(a+b)=1$，所以 $p=\dfrac{1}{a+b}$（特別地，$a+b\neq 0$）。由曲線得 $\dfrac{a+b}{p}=1$，所以 $p=a+b$。比較兩者，得 $a+b=\dfrac{1}{a+b}$，即 $(a+b)^2=1$。
- 若 $q=-p$：由直線得 $p(a-b)=1$，所以 $p=\dfrac{1}{a-b}$。由曲線得 $\dfrac{a-b}{p}=1$，所以 $p=a-b$。比較兩者，得 $(a-b)^2=1$。

所以 $(a-b)^2=1$ 或 $(a+b)^2=1$，此即所求。

> **想一想** 這樣的點真的存在嗎？還是我們只找到了一個必要條件（necessary condition）？

它們確實存在。取 $a=2$、$b=-1$，則 $a+b=1$。點 $(1,1)$ 既在直線 $2x-y=1$ 上，亦在曲線 $\frac{2}{x}-\frac{1}{y}=1$ 上，而兩者在該點的斜率都是 2。這裏曲線為 $y=\dfrac{x}{2-x}$，即去掉原點的一條直角雙曲線（rectangular hyperbola）。在公共點上斜率相等，表示該直線是切線（tangent）：



空心圓圈標示原點：它在雙曲線 $y=\frac{x}{2-x}$ 上，但不在原曲線上。

第 (ii) 部分

將 $\dfrac{a}{x}-\dfrac{b}{y}=1$ 隱函數求導，

$$-\frac{a}{x^2}+\frac{b}{y^2}\frac{\mathrm{d}y}{\mathrm{d}x}=0 \quad\Longrightarrow\quad \frac{\mathrm{d}y}{\mathrm{d}x}=\frac{ay^2}{bx^2}.$$

假設直線在 (p,q) 是曲線的法線。那麼 $p \neq 0$、$q \neq 0$，而且

$$ap + bq = 1 \qquad 及 \qquad \frac{a}{p} - \frac{b}{q} = 1.$$

切線斜率 $\dfrac{aq^2}{bp^2}$ 與直線斜率 $-\dfrac{a}{b}$ 都是有限且非零的，因為 a、b、p、q 全不為零。所以「法線」即表示兩者乘積為 -1：

$$\frac{aq^2}{bp^2} \cdot \left(-\frac{a}{b}\right) = -1 \quad \Longrightarrow \quad a^2q^2 = b^2p^2 \quad \Longrightarrow \quad aq = \pm bp.$$

想一想 哪一個正負號是可能的？

把每個符號代入曲線方程檢驗。

- 若 $aq = bp$，則 $\dfrac{b}{q} = \dfrac{a}{p}$，所以 $\dfrac{a}{p} - \dfrac{b}{q} = 0 \neq 1$。這個符號不可能。
- 若 $aq = -bp$，則 $\dfrac{b}{q} = -\dfrac{a}{p}$，於是曲線方程變為 $\dfrac{2a}{p} = 1$，得 $p = 2a$，從而 $q = -\dfrac{bp}{a} = -2b$。

把 $p = 2a$、$q = -2b$ 代入直線方程，得

$$2a^2 - 2b^2 = 1, \qquad 即 \qquad a^2 - b^2 = \tfrac{1}{2},$$

此即所求。法線的垂足為點 $(2a, -2b)$。

驗算

- 對於第 (i) 部分，上面 $a = 2$、$b = -1$ 的例子滿足 $(a+b)^2 = 1$，而直線在 $(1,1) = (p,p)$ 相切。同樣，$a = 2$、$b = 1$ 滿足 $(a-b)^2 = 1$，而直線 $2x + y = 1$ 在 $(1,-1)$ 與 $\frac{2}{x} + \frac{1}{y} = 1$ 相切。
- 對於第 (ii) 部分，取 $a = \frac{\sqrt{3}}{2}$、$b = \frac{1}{2}$，使 $a^2 - b^2 = \frac{1}{2}$，並取點 $(2a, -2b) = (\sqrt{3}, -1)$。直線給出 $\frac{3}{2} - \frac{1}{2} = 1$，曲線給出 $\frac{1}{2} + \frac{1}{2} = 1$。該點的切線斜率為 $\dfrac{aq^2}{bp^2} = \frac{1}{\sqrt{3}} \approx 0.577$，直線斜率為 $-\sqrt{3} \approx -1.732$，乘積為 -1。
- 數值上，在數以百計的隨機點，斜率公式都與有限差分（finite difference）所得的斜率吻合；而對 (a,b) 作隨機搜尋，只有當 $a^2 - b^2$ 在捨入誤差範圍內等於 0.5 時才會找到法線。

常見錯誤

- 在第 (i) 部分遺漏 $q = -p$ 的情況：正是這個情況給出 $(a-b)^2 = 1$。
- 在第 (ii) 部分停在 $aq = \pm bp$，沒有檢查曲線容許哪一個符號。

STEP 之外

隱函數求導讓我們初窺隱函數定理（implicit function theorem）：在任何滿足 $\partial F/\partial y \neq 0$ 的點附近，方程 $F(x,y) = 0$ 把 y 定義為 x 的可微函數，且 $\mathrm{d}y/\mathrm{d}x = -F_x/F_y$。這裏的相切論證亦有一個代數上的「孿生」版本：一條直線與一條圓錐曲線（conic）相切，當且僅當求兩者交點的二次方程有重根（repeated root）；你可以用這方法重新推導第 (i) 部分。若想了解這一切背後的分析（analysis），可讀 Abbott 的 *Understanding Analysis*（《理解分析》）。

詳細解答

STEP I 2011，第 2 題

純數學 • 代換積分法

> **STEP I 2011，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/11-s1-q2

關鍵思路

積分 E 沒有初等閉式（elementary closed form），所以我們根本不打算求它的值，而是把其餘每個積分變形，直到 E 本身出現為止。對於頭兩個積分，做法是用代數除法（algebraic division）把分式拆開，從而把麻煩的 $\frac{1}{1+x}$ 分離出來。對於 (i) 和 (ii)，做法是選取一個代換（substitution），使被積函數（integrand）在 $[0,1]$ 上變回 $\frac{\mathrm{e}^u}{1+u}$：觀察指數部分，想想怎樣才能把它變成新的變量。

解答

頭兩個積分

把 x 寫成 $x=(1+x)-1$，

$$\frac{x\,\mathrm{e}^x}{1+x}=\mathrm{e}^x-\frac{\mathrm{e}^x}{1+x},$$

因此

$$\int_0^1\frac{x\,\mathrm{e}^x}{1+x}\,\mathrm{d}x=\int_0^1\mathrm{e}^x\,\mathrm{d}x-E=\mathrm{e}-1-E,$$

正如所求。

至於第二個積分，把 x^2 除以 $1+x$：由於 $x^2=(x-1)(1+x)+1$，

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

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

$$\int_0^1 x\,\mathrm{e}^x\,\mathrm{d}x=\left[x\,\mathrm{e}^x\right]_0^1-\int_0^1\mathrm{e}^x\,\mathrm{d}x=\mathrm{e}-(\mathrm{e}-1)=1.$$

因此

$$\int_0^1 \frac{x^2 \mathrm{e}^x}{1+x}\,\mathrm{d}x = 1 - (\mathrm{e}-1) + E = 2 - \mathrm{e} + E.$$

（另一做法是寫成 $\frac{x^2}{1+x} = x - \frac{x}{1+x}$，然後直接利用第一個結果：$1-(\mathrm{e}-1-E)$ 得出相同答案。）

第 (i) 部分

指數是最棘手的部分，所以我們把它設為新的變量：

$$u = \frac{1-x}{1+x}.$$

解出 x：$u(1+x) = 1-x$，所以 $x = \dfrac{1-u}{1+u}$。於是

$$1+x = \frac{2}{1+u}, \qquad \frac{\mathrm{d}x}{\mathrm{d}u} = \frac{-(1+u)-(1-u)}{(1+u)^2} = -\frac{2}{(1+u)^2}.$$

當 $x=0$ 時，$u=1$；當 $x=1$ 時，$u=0$。因此

$$\begin{aligned}\int_0^1 \frac{\mathrm{e}^{\frac{1-x}{1+x}}}{1+x}\,\mathrm{d}x &= \int_1^0 \mathrm{e}^u \cdot \frac{1+u}{2} \cdot \left(-\frac{2}{(1+u)^2}\right)\mathrm{d}u \\ &= \int_0^1 \frac{\mathrm{e}^u}{1+u}\,\mathrm{d}u = E.\end{aligned}$$

$\mathrm{d}x/\mathrm{d}u$ 帶來的負號，用來把積分上下限調回遞增次序。

> **想一想** 為甚麼我們應該預期答案恰好是 E，而完全沒有 e？

映射 $x \mapsto \frac{1-x}{1+x}$ 把 $[0,1]$ 映到自身（它把兩個端點互換），而因子 $\frac{1}{1+x}$ 與 $\mathrm{d}x$ 結合後恰好得到 $\frac{\mathrm{d}u}{1+u}$。事實上 $\frac{\mathrm{d}x}{1+x} = -\frac{\mathrm{d}u}{1+u}$，這是記住整個計算的一個巧妙方法。

第 (ii) 部分

現在目標區間 $[0,1]$ 必須來自 $x \in [1,\sqrt{2}]$，而在這區間上 x^2 由 1 變到 2。所以我們試

$$u = x^2 - 1, \qquad \mathrm{d}u = 2x\,\mathrm{d}x,$$

當 $x=1$ 時 $u=0$，當 $x=\sqrt{2}$ 時 $u=1$。於是

$$\frac{\mathrm{d}x}{x} = \frac{x\,\mathrm{d}x}{x^2} = \frac{\mathrm{d}u}{2(1+u)}, \qquad \mathrm{e}^{x^2} = \mathrm{e}^{1+u} = \mathrm{e}\cdot\mathrm{e}^u,$$

因此

$$\int_1^{\sqrt{2}} \frac{\mathrm{e}^{x^2}}{x}\,\mathrm{d}x = \frac{\mathrm{e}}{2}\int_0^1 \frac{\mathrm{e}^u}{1+u}\,\mathrm{d}u = \frac{\mathrm{e}E}{2}.$$

想一想 為甚麼不用代換 $u = x^2$？

這樣也行得通，會得到 $\frac{1}{2}\int_1^2 \frac{\mathrm{e}^u}{u}\,\mathrm{d}u$，但區間是 $[1,2]$ 而不是 $[0,1]$。再平移一個單位，即 $u \mapsto u-1$，便得到相同的答案；$u = x^2 - 1$ 只不過是一步做完這兩步。

驗算

辛普森法則(Simpson's rule)給出 $E \approx 1.125386$。以數值計算，第一個積分是 $0.592896 = \mathrm{e} - 1 - E$，第二個是 $0.407104 = 2 - \mathrm{e} + E$，(i) 中的積分是 $1.125386 = E$，而 (ii) 中的積分是 $1.529558 = \frac{1}{2}\mathrm{e}E$。不用電腦也可以快速檢查是否合理：頭兩個被積函數都是正的，而的確 $\mathrm{e} - 1 - E \approx 0.59 > 0$ 及 $2 - \mathrm{e} + E \approx 0.41 > 0$；此外第二個比第一個小，這也理所當然，因為在 $[0,1]$ 上 $x^2 \leq x$。

常見錯誤

- 在 (i) 中忘記把上下限倒轉（或漏掉負號），結果得到 $-E$；對於正的被積函數，這是不可能的答案。
- 在 (ii) 中漏掉 $\mathrm{e}^{x^2} = \mathrm{e}^{1+u}$ 帶來的因子 e，或漏掉 $\mathrm{d}u = 2x\,\mathrm{d}x$ 帶來的因子 $\frac{1}{2}$。

STEP 之外

像 E 這樣的積分與指數積分（exponential integral）$\mathrm{Ei}(x) = \int_{-\infty}^{x} \frac{\mathrm{e}^t}{t}\,\mathrm{d}t$ 密切相關：代入 $t = 1 + x$ 可得 $E = \mathrm{e}^{-1}\big(\mathrm{Ei}(2) - \mathrm{Ei}(1)\big)$。劉維爾（Liouville）的理論證明了 $\int \frac{\mathrm{e}^x}{x}\,\mathrm{d}x$ 沒有初等原函數（elementary antiderivative），這正是題目把 E 當作一個具名常數、並要求我們把一切化歸為它的原因。Siklos 的 *Advanced Problems in Mathematics*（《數學進階問題》）收錄了更多同類型的題目。

詳細解答

STEP I 2011，第 3 題

純數學 • 三角恆等式

> **STEP I 2011，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/11-s1-q3

關鍵思路

在 $\frac{1}{3}\pi \mp \theta$ 處的兩個正弦相乘，可以用平方差（difference of squares）的技巧乾淨地化簡，於是一行就能證明 $(*)$。之後，餘切（cotangent）的和式來自對乘積求導：三個因子乘積的導數除以該乘積本身，就是三個「對數導數」（logarithmic derivative）之和，而對正弦而言，這些對數導數正是餘切。對餘弦的版本做同樣的事就得到正切，而 $\tan x + \cot x$ 是一個餘割（cosecant）。

解答

恆等式（identity）$(*)$

記 $a = \frac{1}{3}\pi$。由複角公式（compound-angle formulae），

$$\begin{aligned}\sin(a-\theta)\sin(a+\theta) &= (\sin a\cos\theta - \cos a\sin\theta)(\sin a\cos\theta + \cos a\sin\theta)\\ &= \sin^2 a\cos^2\theta - \cos^2 a\sin^2\theta\\ &= \sin^2 a\,(1-\sin^2\theta) - (1-\sin^2 a)\sin^2\theta = \sin^2 a - \sin^2\theta.\end{aligned}$$

由於 $\sin^2 a = \frac{3}{4}$，

$$4\sin\theta\sin(\tfrac{1}{3}\pi-\theta)\sin(\tfrac{1}{3}\pi+\theta) = 4\sin\theta\left(\tfrac{3}{4}-\sin^2\theta\right) = 3\sin\theta - 4\sin^3\theta = \sin 3\theta,$$

最後一步用了三倍角公式（triple-angle formula）。這就對所有 θ 證明了 $(*)$。

第 (i) 部分

由於 $(*)$ 對所有 θ 成立，我們可以對兩邊求導。由乘積法則（product rule），

$$4\Big[\cos\theta\sin(a-\theta)\sin(a+\theta)-\sin\theta\cos(a-\theta)\sin(a+\theta)\\ +\sin\theta\sin(a-\theta)\cos(a+\theta)\Big]=3\cos 3\theta,$$

其中中間一項的負號來自對 $a-\theta$ 應用鏈式法則（chain rule）。

想一想 怎樣把它變成餘切？

兩邊除以 $\sin 3\theta$，而由 $(*)$，它等於 $4\sin\theta\sin(a-\theta)\sin(a+\theta)$。左邊每一項中，除了被求導的那個因子外，其餘的正弦都會約去。只要 $\sin 3\theta\neq 0$，這樣做便是容許的，而此時三個正弦亦全不為零。我們得到

$$\cot\theta-\cot(\tfrac{1}{3}\pi-\theta)+\cot(\tfrac{1}{3}\pi+\theta)=3\cot 3\theta\qquad(\sin 3\theta\neq 0).\tag{A}$$

現在代入 $\theta=\frac{1}{9}\pi$，此時 $\sin 3\theta=\sin\frac{1}{3}\pi\neq 0$。由於 $\frac{1}{3}\pi-\frac{1}{9}\pi=\frac{2}{9}\pi$ 及 $\frac{1}{3}\pi+\frac{1}{9}\pi=\frac{4}{9}\pi$，

$$\cot\tfrac{1}{9}\pi-\cot\tfrac{2}{9}\pi+\cot\tfrac{4}{9}\pi=3\cot\tfrac{1}{3}\pi=\frac{3}{\sqrt{3}}=\sqrt{3},$$

即為所求。

第 (ii) 部分

在 $(*)$ 中代入 $\theta=\frac{1}{6}\pi-\phi$。四個正弦變成

$$\begin{aligned}\sin\theta=\sin(\tfrac{1}{6}\pi-\phi)&=\cos(\tfrac{1}{3}\pi+\phi),\\ \sin(\tfrac{1}{3}\pi-\theta)=\sin(\tfrac{1}{6}\pi+\phi)&=\cos(\tfrac{1}{3}\pi-\phi),\\ \sin(\tfrac{1}{3}\pi+\theta)=\sin(\tfrac{1}{2}\pi-\phi)&=\cos\phi,\\ \sin 3\theta=\sin(\tfrac{1}{2}\pi-3\phi)&=\cos 3\phi,\end{aligned}$$

每次都用了 $\sin(\frac{1}{2}\pi-x)=\cos x$。於是，把 ϕ 改記為 θ，

$$4\cos\theta\cos(\tfrac{1}{3}\pi-\theta)\cos(\tfrac{1}{3}\pi+\theta)=\cos 3\theta\qquad\text{對所有}\theta.\tag{**}$$

把 $(*)$ 除以 $(**)$，因子 4 互相抵消，而每個正弦除以餘弦都是正切；取倒數，

$$\cot\theta\cot(\tfrac{1}{3}\pi-\theta)\cot(\tfrac{1}{3}\pi+\theta)=\cot 3\theta,$$

此式在兩邊皆有定義之處成立，即 $\sin 3\theta\neq 0$ 及 $\cos 3\theta\neq 0$ 之處（每個條件都令相應的三個因子不為零）。

至於餘割的和，我們如第 (i) 部分一樣對 $(**)$ 求導：

$$4\Big[-\sin\theta\cos(a-\theta)\cos(a+\theta)+\cos\theta\sin(a-\theta)\cos(a+\theta)\\ -\cos\theta\cos(a-\theta)\sin(a+\theta)\Big]=-3\sin 3\theta.$$

除以 $\cos 3\theta = 4\cos\theta\cos(a-\theta)\cos(a+\theta)$（當它不為零時），再乘以 -1，

$$\tan\theta-\tan(\tfrac{1}{3}\pi-\theta)+\tan(\tfrac{1}{3}\pi+\theta)=3\tan 3\theta. \qquad \text{(B)}$$

(A) 與 (B) 的正負號一致，所以可以把兩式相加。

想一想 為甚麼把正切與餘切相加會有幫助？

因為 $\tan x+\cot x$ 是單一個餘割：

$$\tan x+\cot x=\frac{\sin^2 x+\cos^2 x}{\sin x\cos x}=\frac{2}{\sin 2x}=2\operatorname{cosec}2x.$$

所以只要 $\sin 6\theta\neq 0$，把 (A) 與 (B) 相加便得

$$2\operatorname{cosec}2\theta-2\operatorname{cosec}(\tfrac{2}{3}\pi-2\theta)+2\operatorname{cosec}(\tfrac{2}{3}\pi+2\theta)=6\operatorname{cosec}6\theta.$$

現在取 $\theta=\frac{1}{18}\pi$，使 $2\theta=\frac{1}{9}\pi$ 及 $6\theta=\frac{1}{3}\pi$（且 $\sin 6\theta\neq 0$）。由於 $\frac{2}{3}\pi-\frac{1}{9}\pi=\frac{5}{9}\pi$ 及 $\frac{2}{3}\pi+\frac{1}{9}\pi=\frac{7}{9}\pi$，除以 2 得

$$\operatorname{cosec}\tfrac{1}{9}\pi-\operatorname{cosec}\tfrac{5}{9}\pi+\operatorname{cosec}\tfrac{7}{9}\pi=3\operatorname{cosec}\tfrac{1}{3}\pi=3\cdot\frac{2}{\sqrt{3}}=2\sqrt{3},$$

即為所求。

驗算

- 在 $\theta=\frac{1}{6}\pi$ 時，$(*)$ 變成 $4\cdot\frac{1}{2}\cdot\frac{1}{2}\cdot 1=1=\sin\frac{1}{2}\pi$；而 $(**)$ 在 $\theta=0$ 時變成 $4\cdot 1\cdot\frac{1}{2}\cdot\frac{1}{2}=1=\cos 0$。
- 數值上，$\cot 20^\circ-\cot 40^\circ+\cot 80^\circ\approx 2.7475-1.1918+0.1763=1.7321\approx\sqrt{3}$，而 $\operatorname{cosec}20^\circ-\operatorname{cosec}100^\circ+\operatorname{cosec}140^\circ\approx 2.9238-1.0154+1.5557=3.4641\approx 2\sqrt{3}$。
- 恆等式 (A)、(B) 以及餘切乘積式都可以用計算機在任意隨機的 θ 測試；我們在數千個隨機值上驗證過。

常見錯誤

- 對 $\sin(\frac{1}{3}\pi-\theta)$ 求導時忘記鏈式法則帶來的負號，以致破壞了 (A) 中正負交替的符號。

- 除以 $\sin 3\theta$ 時沒有指出它在所代入的值上必須不為零。
- 試圖從餘切的乘積得出餘割的和：真正能夠合併的是和式 (A) 與 (B)。

STEP 之外

把導數除以函數本身的技巧就是對數微分法（logarithmic differentiation）：$\frac{\mathrm{d}}{\mathrm{d}\theta}\ln|f| = f'/f$ 把乘積變成和。再推進一步，sin 本身可以寫成遍及其零點的無窮乘積（infinite product），而對數微分法隨即給出著名的部分分式展開（partial-fraction expansion）$\pi\cot\pi x = \sum_n \frac{1}{x-n}$；恆等式 $(*)$ 正是它的一個有限的影子。這些想法在複分析（complex analysis）中有所發展；Needham 的 *Visual Complex Analysis*（《直觀複分析》）是一個美妙的起點。

詳細解答

STEP I 2011，第 4 題

純數學 • 拋物線的切線

> **STEP I 2011，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/11-s1-q4

關鍵思路

把拋物線（parabola）上一般點的切線（tangent）寫成不含分數的形式，兩條切線的交點只需相減便可得出。至於角度，則用 $\overrightarrow{FT}$ 與 $\overrightarrow{FP}$ 的純量積（scalar product）。各段長度其實比表面看來容易處理：FP 是焦半徑（focal distance），而 FT^2 可以因式分解。求得 $\cos\phi$ 之後，留意答案對 p 和 q 是對稱的。這種對稱性就是證明角平分的全部關鍵。

解答

點 T

對 $y^2 = 4ax$ 作隱函數求導（implicit differentiation），得 $2y\,\dfrac{\mathrm{d}y}{\mathrm{d}x} = 4a$。因此在 P 點，即 $y = 2ap \neq 0$ 之處（因為 $p \neq 0$），斜率為 $\dfrac{4a}{4ap} = \dfrac{1}{p}$。所以 P 點的切線為

$$y - 2ap = \frac{x - ap^2}{p} \quad \Longleftrightarrow \quad py = x + ap^2,$$

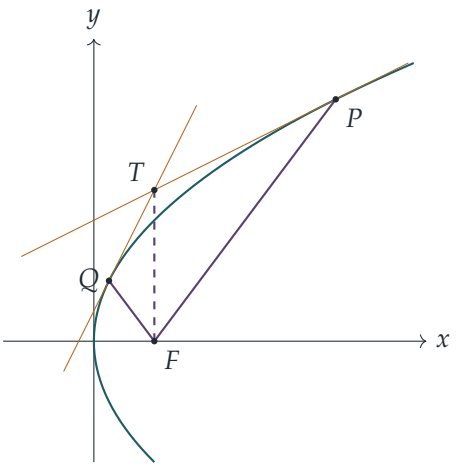
同理，Q 點的切線為 $qy = x + aq^2$。兩式相減，得

$$(p - q)\,y = a(p^2 - q^2) = a(p - q)(p + q).$$

由於 P 和 Q 是相異的點，$p \neq q$，故可以除以 $p - q$，得 $y = a(p + q)$。於是

$$x = py - ap^2 = ap^2 + apq - ap^2 = apq,$$

所以 T 是點 $\big(apq,\, a(p + q)\big)$，與所求相符。



圖中取 $a=1$、$p=2$、$q=\frac{1}{2}$，此時 FT 恰好是鉛垂的。

角 ϕ

首先注意 $T\neq F$：否則需要 $p+q=0$ 及 $pq=1$，從而 $p^2=-1$，這是不可能的。所以角 TFP 是有意義的。設 $F=(a,0)$，則

$$\overrightarrow{FP}=a\begin{pmatrix}p^2-1\\2p\end{pmatrix},\qquad \overrightarrow{FT}=a\begin{pmatrix}pq-1\\p+q\end{pmatrix}.$$

兩者的長度為

$$FP^2=a^2\left[(p^2-1)^2+4p^2\right]=a^2(p^2+1)^2,$$
$$FT^2=a^2\left[(pq-1)^2+(p+q)^2\right]=a^2\left[p^2q^2+p^2+q^2+1\right]=a^2(p^2+1)(q^2+1),$$

因為交叉項 $-2pq$ 與 $+2pq$ 互相抵消。所以 $FP=|a|(p^2+1)$，而 $FT=|a|\sqrt{(p^2+1)(q^2+1)}$。

? **想一想** 我們怎樣看出純量積可以因式分解？

我們預期 (p^2+1) 會與 FP 約去，所以刻意尋找這個因式。展開得

$$\begin{aligned}\overrightarrow{FP}\cdot\overrightarrow{FT}&=a^2\left[(p^2-1)(pq-1)+2p(p+q)\right]\\&=a^2\left[p^3q-p^2-pq+1+2p^2+2pq\right]\\&=a^2\left[p^3q+pq+p^2+1\right]\\&=a^2(p^2+1)(pq+1).\end{aligned}$$

因此，利用 $|a|^2=a^2$，

$$\cos\phi=\frac{\overrightarrow{FP}\cdot\overrightarrow{FT}}{FP\cdot FT}=\frac{a^2(p^2+1)(pq+1)}{a^2(p^2+1)\sqrt{(p^2+1)(q^2+1)}}=\frac{pq+1}{\sqrt{(p^2+1)(q^2+1)}},$$

這正是所求的結果。

角平分

T 的坐標對 p 和 q 是對稱的，所以在上面的計算中把 P 和 Q 的角色對調，便知角 TFQ 滿足

$$\cos\angle TFQ = \frac{qp+1}{\sqrt{(q^2+1)(p^2+1)}} = \cos\phi.$$

兩個角都在 $[0,\pi]$ 之內，而餘弦在此區間上是一一對應（one-to-one）的，所以 $\angle TFP = \angle TFQ = \phi$。

因此，射線（ray）FP 和 FQ 各自與射線 FT 成 ϕ 角。在平面上，從 F 出發而與 FT 成 ϕ 角的射線只有兩條，而且它們以直線 FT 為鏡互為鏡像。射線 FP 和 FQ 並不是同一條射線：一條穿過焦點（focus）的直線與拋物線相交於 F 兩側的兩點（或只相交一次），所以從 F 出發的一條射線不可能同時經過 P 和 Q。因此 FP 和 FQ 互為對直線 FT 的反射，而這正好表示直線 FT 平分角 PFQ。

有一個細節可以解釋為何題目說的是直線 FT。若 $pq+1>0$，則 $\phi<\frac{\pi}{2}$，射線 FT 本身位於角 $PFQ = 2\phi$ 之內。若 $pq+1<0$，則 $\phi>\frac{\pi}{2}$，角 PFQ 為 $2\pi-2\phi$，平分它的是 TF 越過 F 的延長線。若 $pq=-1$，則 PQ 是焦弦（focal chord），$\phi=\frac{\pi}{2}$，而 FT 與它垂直，平分這個平角。在所有情況下，直線 FT 都平分角 PFQ。

驗算

取圖中的 $a=1$、$p=2$、$q=\frac{1}{2}$：$T=(1,\frac{5}{2})$，$\overrightarrow{FP}=(3,4)$，$\overrightarrow{FQ}=(-\frac{3}{4},1)$，$\overrightarrow{FT}=(0,\frac{5}{2})$。於是 $\cos\angle TFP=\frac{4}{5}$，而 $\cos\angle TFQ=\frac{1}{5/4}=\frac{4}{5}$，與 $\dfrac{2}{\sqrt{5\cdot\frac{5}{4}}}=\dfrac{4}{5}$ 吻合。用電腦測試 10^5 組隨機選取的 a（正負皆有）、p 和 q，在 10^{-10} 的誤差範圍內證實了 T 的坐標、$\cos\phi$ 的公式以及角平分的結論。測試包括許多 $pq<-1$ 的情況，此時是 FT 向後的延長線起平分作用。

常見錯誤

- 除以 $p-q$ 時沒有說明它為何不等於零（因為兩點相異）。
- 寫成 $FP=a(p^2+1)$，這在 $a<0$ 時是錯的。$\cos\phi$ 中只出現 a^2，所以最終 a 的正負並無影響。
- 只說「餘弦相等，所以 FT 平分該角」，卻沒有解釋為何 FP 和 FQ 位於直線 FT 的兩側。

STEP 之外

這其實是拋物線的反射性質（reflective property）換了個樣子出現：從任何外部點引出的兩條切線段，在焦點處所張的角相等。橢圓（ellipse）和雙曲線（hyperbola）也有同樣的性質，三者都源於圓錐曲線（conic）的焦點－準線（focus–directrix）定義，大學一年級的幾何及射影幾何（projective geometry）課程會對此作更深入的發展。如想做更多同類風格的題目，Siklos 的 *Advanced Problems in Mathematics*（《高等數學難題》）可免費取得，非常值得逐題鑽研。

詳細解答

STEP I 2011，第 5 題

純數學 • 含絕對值的積分

STEP I 2011，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/11-s1-q5

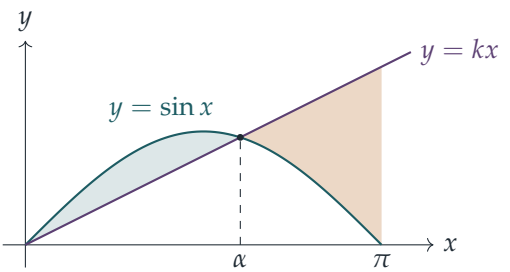
關鍵思路

只要知道 $\sin x - kx$ 在哪裏變號，絕對值符號（modulus）就不再麻煩，而草圖正告訴我們：只變號一次，就在 $x = \alpha$ 處。在該處把積分拆開，絕對值符號便會消失。最後一步是不再把 k 看作變量：由於 $k = \sin\alpha/\alpha$，所有東西都可以用 α 表示，而 I 的導數（derivative）便能整齊地因式分解。

解答

唯一的根

設 $g(x) = \sin x - kx$。直線 $y = kx$ 經過原點，斜率為 $k < 1$；而 $y = \sin x$ 離開原點時的斜率為 $\cos 0 = 1$。因此在 0 的右邊附近，正弦曲線位於直線之上。在 $x = \pi$ 處，正弦曲線回到 0，但直線的高度為 $k\pi > 0$，所以在那裏直線位於上方。由於 $y = \sin x$ 在 $(0, \pi)$ 上向下彎曲，而直線是直的，兩者在其間恰好相交一次。



要令「恰好一次」無懈可擊：$g(0) = 0$、$g'(x) = \cos x - k$，而且在 $(0, \pi)$ 上 $g''(x) = -\sin x < 0$。所以 g' 嚴格遞減（strictly decreasing），由 0 處的 $1 - k > 0$ 減至 π 處的

$-1-k<0$，並恰好在一點 β 等於零。因此 g 在 $(0,\beta]$ 上嚴格遞增，故在該處為正（因為它由 0 開始），並在 $[\beta,\pi)$ 上嚴格遞減，由正值減至 $g(\pi)=-k\pi<0$。所以它在 $(0,\pi)$ 中恰有一個零點 α，而且

$$\sin x > kx \text{ 當} 0<x<\alpha, \qquad \sin x < kx \text{ 當} \alpha<x<\pi.$$

計算 I

利用這些正負號（即草圖中的藍綠色和琥珀色區域），

$$I=\int_0^{\alpha}(\sin x-kx)\,\mathrm{d}x-\int_{\alpha}^{\pi}(\sin x-kx)\,\mathrm{d}x.$$

$\sin x-kx$ 的一個原函數（antiderivative）是 $F(x)=-\cos x-\frac{1}{2}kx^2$，所以

$$\begin{aligned} I &= \big(F(\alpha)-F(0)\big)-\big(F(\pi)-F(\alpha)\big)=2F(\alpha)-F(0)-F(\pi)\\ &= \big(-2\cos\alpha-k\alpha^2\big)+1-\big(1-\tfrac{1}{2}k\pi^2\big)\\ &= \tfrac{1}{2}k\pi^2-2\cos\alpha-k\alpha^2. \end{aligned}$$

由於 α 滿足 $(*)$，$k=\dfrac{\sin\alpha}{\alpha}$，所以 $k\alpha^2=\alpha\sin\alpha$，而且

$$I=\frac{\pi^2\sin\alpha}{2\alpha}-2\cos\alpha-\alpha\sin\alpha,$$

即為所求。

駐值（stationary value）

想一想 α 可以取甚麼值？

每個 $k\in(0,1)$ 都給出一個 $\alpha\in(0,\pi)$。反過來，每個 $\alpha\in(0,\pi)$ 都來自 $k=\sin\alpha/\alpha$，而它位於 $(0,1)$ 之中，因為在該範圍內 $0<\sin\alpha<\alpha$。因此我們必須研究 $I(\alpha)$ 在開區間 $(0,\pi)$ 內所有 α 的情況。

對第一項使用商法則（quotient rule）求導，得

$$\begin{aligned} \frac{\mathrm{d}I}{\mathrm{d}\alpha} &= \frac{\pi^2(\alpha\cos\alpha-\sin\alpha)}{2\alpha^2}+2\sin\alpha-(\sin\alpha+\alpha\cos\alpha)\\ &= \frac{\pi^2(\alpha\cos\alpha-\sin\alpha)}{2\alpha^2}-(\alpha\cos\alpha-\sin\alpha)\\ &= (\alpha\cos\alpha-\sin\alpha)\left(\frac{\pi^2}{2\alpha^2}-1\right). \end{aligned}$$

第一個因子在 $(0,\pi)$ 上永不為零。設 $h(\alpha)=\alpha\cos\alpha-\sin\alpha$。那麼 $h(0)=0$，而且在 $(0,\pi)$ 上 $h'(\alpha)=\cos\alpha-\alpha\sin\alpha-\cos\alpha=-\alpha\sin\alpha<0$，所以在整個區間上 $h(\alpha)<0$。（等價地，在 $(0,\frac{1}{2}\pi)$ 上 $\tan\alpha>\alpha$，而在此之後 $\cos\alpha\le 0$。）

所以 $\mathrm{d}I/\mathrm{d}\alpha = 0$ 當且僅當 $\pi^2 = 2\alpha^2$，即在 $\alpha = \pi/\sqrt{2}$ 處，而它確實位於 $(0, \pi)$ 之中。這是唯一的駐值。由於 $h < 0$，$\mathrm{d}I/\mathrm{d}\alpha$ 的正負號與 $\frac{\pi^2}{2\alpha^2} - 1$ 的正負號相反：

- 當 $0 < \alpha < \pi/\sqrt{2}$ 時，$\frac{\pi^2}{2\alpha^2} > 1$，所以 $\mathrm{d}I/\mathrm{d}\alpha < 0$；
- 當 $\pi/\sqrt{2} < \alpha < \pi$ 時，$\frac{\pi^2}{2\alpha^2} < 1$，所以 $\mathrm{d}I/\mathrm{d}\alpha > 0$。

所以 I 先減後增：該駐值是極小值，而由於它是區間上唯一的駐點，它就是 I 的最小值。

當 $\alpha = \pi/\sqrt{2}$ 時，我們有 $\dfrac{\pi^2}{2\alpha} = \dfrac{\pi^2}{\sqrt{2}\,\pi} = \dfrac{\pi}{\sqrt{2}} = \alpha$，所以 I 的第一項與第三項互相抵消：

$$I_{\min} = \alpha \sin\alpha - 2\cos\alpha - \alpha\sin\alpha = -2\cos\frac{\pi}{\sqrt{2}}.$$

這個值是正數，理應如此，因為 $\pi/\sqrt{2} \approx 2.22$ 位於 $\frac{1}{2}\pi$ 與 π 之間，而餘弦在該範圍內為負。

驗算

- 當 $k \to 0$ 時，$\alpha \to \pi$，公式給出 $I \to 2$，這正是 $\int_0^\pi \sin x\,\mathrm{d}x$。當 $k \to 1$ 時，$\alpha \to 0$，而 $I \to \frac{1}{2}\pi^2 - 2$，這正是 $\int_0^\pi (x - \sin x)\,\mathrm{d}x$。
- 數值上，$-2\cos(\pi/\sqrt{2}) \approx 1.2114$，比兩個端值 2 和 $\frac{1}{2}\pi^2 - 2 \approx 2.935$ 都小。直接以中點法數值積分（midpoint quadrature）計算 $\int_0^\pi |\sin x - kx|\,\mathrm{d}x$，取 $k = \sin\alpha/\alpha \approx 0.3582$，同樣得到 1.2114。

常見錯誤

- 對 α 求導時把 k 當作常數：k 依賴於 α，這正是我們要先代入 $k = \sin\alpha/\alpha$ 的原因。
- 只證明了 $\mathrm{d}I/\mathrm{d}\alpha = 0$ 僅在 $\alpha = \pi/\sqrt{2}$ 處成立，卻忘記證明 $\alpha\cos\alpha - \sin\alpha$ 在 $(0, \pi)$ 中沒有零點。

STEP 之外

唯一根的論證用到介值定理（intermediate value theorem）以及嚴格單調性，而求最小值時則默默用到一個事實：$\alpha \mapsto k$ 是 $(0, \pi)$ 與 $(0, 1)$ 之間的雙射（bijection），因此對 α 求最小值等同於對 k 求最小值。要把這類重新參數化（reparametrisation）和存在性論證變得嚴謹，正是實分析（real analysis）入門課程的工作；Abbott 的 *Understanding Analysis*（《理解分析》）是極佳的起點。

詳細解答

STEP I 2011，第 6 題

純數學 • 二項式級數求和

> **STEP I 2011，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/11-s1-q6

關鍵思路

只要知道 $(1-x)^{-3}$ 的係數（coefficients），再乘上一個多項式（polynomial），其實只是把幾個相鄰的係數混合起來，所以較複雜的有理函數（rational function）的係數幾乎可以不費吹灰之力得出。把這個過程倒過來看，一個各項為 (多項式，自變數為r) $\times\, x^r$ 的級數（series），只要看出是哪一個有理函數生成它，再代入適當的 x 值（這裏是 $x=\frac{1}{2}$），便可求和。

解答

開首的結果

當 $|x|<1$ 時，指數（index）為 -3 的二項式展開（binomial expansion）給出

$$(1-x)^{-3}=\sum_{r=0}^{\infty}\frac{(-3)(-4)\cdots(-3-r+1)}{r!}(-x)^r.$$

分子共有 r 個因子，每個都是負數，所以它們的積是 $(-1)^r\,\dfrac{(r+2)!}{2!}$，而這個 $(-1)^r$ 與 $(-x)^r$ 中的 $(-1)^r$ 互相抵消。因此 x^r 的係數是

$$a_r=\frac{(r+2)!}{2!\,r!}=\tfrac{1}{2}(r+1)(r+2),$$

正如所求。（所以 $a_0=1$、$a_1=3$、$a_2=6$、$a_3=10$、...）

第 (i) 部分

寫成

$$\frac{1-x+2x^2}{(1-x)^3} = (1-x+2x^2)\sum_{k\geq 0} a_k x^k.$$

x^r 項來自 $1\cdot a_r x^r$、$-x\cdot a_{r-1}x^{r-1}$ 及 $2x^2\cdot a_{r-2}x^{r-2}$，所以當 $r\geq 2$ 時，x^r 的係數是

$$\begin{aligned} a_r - a_{r-1} + 2a_{r-2} &= \tfrac{1}{2}(r+1)(r+2) - \tfrac{1}{2}r(r+1) + (r-1)r \\ &= (r+1) + r^2 - r \\ &= r^2 + 1. \end{aligned}$$

想一想 當 $r=0$ 及 $r=1$ 時，a_{r-1} 或 a_{r-2} 並不存在，這條公式是否仍然成立？

成立。若設 $a_{-1} = a_{-2} = 0$，公式 $\frac{1}{2}(r+1)(r+2)$ 在 $r=-1$ 及 $r=-2$ 時亦給出 0，所以上面的計算對所有 $r\geq 0$ 都成立。直接驗證：常數項是 $a_0 = 1 = 0^2+1$，而 x 的係數是 $a_1 - a_0 = 2 = 1^2+1$。所以對所有 $r\geq 0$，x^r 的係數都是 r^2+1，正如所求。

至於該級數，分子 $1,2,5,10,17,26,37,50,\ldots$ 是 $r=0,1,2,\ldots$ 時的 r^2+1，而分母是 2^r。所以該級數是

$$\sum_{r=0}^{\infty}(r^2+1)\left(\tfrac{1}{2}\right)^r,$$

即上述展開式在 $x=\frac{1}{2}$ 時的值。這樣做是合理的，因為 $(1-x)^{-3}$ 的二項式展開在 $|x|<1$ 時收斂（converges）。因此

$$1+\frac{2}{2}+\frac{5}{4}+\frac{10}{8}+\cdots = \frac{1-\frac{1}{2}+2\cdot\frac{1}{4}}{\left(\frac{1}{2}\right)^3} = \frac{1}{1/8} = 8.$$

第 (ii) 部分

首要任務是找出規律。把各項改寫成分母為 $1,2,4,8,16,32,64$：

$$1, \frac{4}{2}, \frac{9}{4}, \frac{16}{8}, \frac{25}{16}, \frac{36}{32}, \frac{49}{64}, \cdots$$

所以該級數是 $\displaystyle\sum_{r=0}^{\infty}\frac{(r+1)^2}{2^r}$。

想一想 $(1-x)^3$ 上面放甚麼分子，才會得出係數 $(r+1)^2$？

第 (i) 部分提示我們試用分子 $1+x$（同樣是一個簡短的多項式）。$(1+x)(1-x)^{-3}$ 中 x^r 的係數是

$$a_r + a_{r-1} = \tfrac{1}{2}(r+1)(r+2) + \tfrac{1}{2}r(r+1) = \tfrac{1}{2}(r+1)(2r+2) = (r+1)^2,$$

對所有 $r \geq 0$ 均成立（與之前一樣取 $a_{-1} = 0$）。代入 $x = \frac{1}{2}$，

$$1 + 2 + \frac{9}{4} + 2 + \frac{25}{16} + \cdots = \frac{1 + \frac{1}{2}}{\left(\frac{1}{2}\right)^3} = \frac{3}{2} \cdot 8 = 12.$$

另一方法：利用第 (i) 部分。以 $s = r + 1$ 平移下標，使得 $2^{-r} = 2 \cdot 2^{-s}$：

$$\sum_{r \geq 0} \frac{(r+1)^2}{2^r} = 2 \sum_{s \geq 1} \frac{s^2}{2^s} = 2 \left(\sum_{s \geq 0} \frac{s^2 + 1}{2^s} - \sum_{s \geq 0} \frac{1}{2^s} \right) = 2(8 - 2) = 12,$$

這裏用了第 (i) 部分的答案及幾何級數（geometric series）$\sum 2^{-s} = 2$。

驗算

- 以人手把 $(1 - x)^{-3}$ 展開為 $(1 + x + x^2 + \cdots)^3$，或把頭幾項乘開，得到 $1 + 3x + 6x^2 + 10x^3 + \cdots$，與 $\frac{1}{2}(r+1)(r+2)$ 吻合。
- 第 (i) 部分級數的部分和（partial sums）趨近 8：取 200 項後，和與 8 在機器精度內一致。第 (ii) 部分同樣如此，極限為 12。
- 第 (ii) 部分的兩個方法得出相同答案 12。

常見錯誤

- 二項式係數的正負號弄錯：$(-3)(-4)\cdots$ 中有 r 個負號，$(-x)^r$ 中亦有 r 個，兩者互相抵消。
- 在第 (ii) 部分，因為部分分數已約至最簡而看錯規律：$2 = \frac{4}{2} = \frac{16}{8}$，以及 $\frac{9}{8} = \frac{36}{32}$。
- 代入 $x = \frac{1}{2}$ 時沒有指出展開式在該處收斂（$|x| < 1$）。

STEP 之外

我們一直在做的，其實是運用生成函數（*generating functions*）：把數列 $c_0, c_1, c_2, \ldots$ 編碼成冪級數（power series）$\sum c_r x^r$，於是對數列的運算便變成對函數的代數運算。事實上，任何「r 的多項式乘以 x^r」之和，都是一個分母為 $(1 - x)$ 的冪的有理函數，這就是為何這類和總是得出有理數。要證明代入某個 x 值是合理的，需要依靠冪級數及其收斂半徑（radius of convergence）的理論，可參閱 Abbott 的 *Understanding Analysis*（《理解分析》）；至於把生成函數作為解題工具，可參閱 Siklos 的 *Advanced Problems in Mathematics*（《數學進階難題》）。

詳細解答

STEP I 2011，第 7 題

純數學 • 可分離變量微分方程

STEP I 2011，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/11-s1-q7

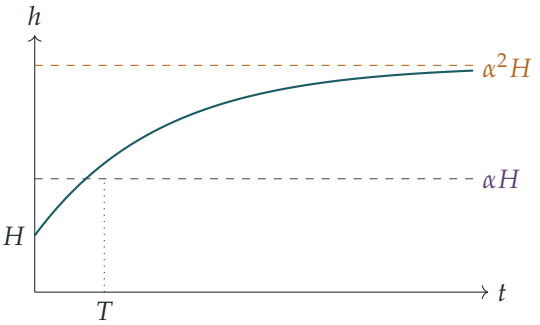
關鍵思路

體積等於（固定的截面面積）× 高度，所以高度的變化率就是（流入率 − 流出率）除以面積。題目中「高度保持不變」一句是確定流入率的線索：在那個高度，流入與流出恰好平衡。之後每一部分都是一條可分離變量（separable）的微分方程（differential equation）；而在第 (ii) 部分，代換（substitution）$u = \sqrt{h}$ 會把積分化為常規的計算。

解答

設固定的截面面積（cross-sectional area）為 A，固定的流入率為 q（單位時間內的體積）。水的體積為 Ah，所以

$$A\frac{\mathrm{d}h}{\mathrm{d}t} = q - (\text{流出率}).$$



高度從 H 開始上升，趨向流入與流出平衡的水平 $\alpha^2 H$，愈來愈接近卻永遠不會到達；問題是要多久才升到 αH，由於 $1 < \alpha < \alpha^2$，這個高度位於兩者之間。（圖中所示為第 (i) 部分的情況。）

第 (i) 部分

流出率為 λh，其中 $\lambda > 0$ 為某常數，所以

$$A\frac{\mathrm{d}h}{\mathrm{d}t} = q - \lambda h.$$

當 $h = \alpha^2 H$ 時高度保持不變，所以在該處 $\mathrm{d}h/\mathrm{d}t = 0$，得 $q = \lambda\alpha^2 H$。代入後得

$$\frac{\mathrm{d}h}{\mathrm{d}t} = \frac{\lambda}{A}\left(\alpha^2 H - h\right) = k\left(\alpha^2 H - h\right), \qquad k = \frac{\lambda}{A} > 0,$$

此即所求。

由於在整個上升過程中 $H \le h < \alpha^2 H$，所以 $\alpha^2 H - h > 0$，我們可以分離變量：

$$\begin{aligned} kT &= \int_H^{\alpha H} \frac{\mathrm{d}h}{\alpha^2 H - h} = \Big[-\ln\left(\alpha^2 H - h\right) \Big]_H^{\alpha H} = \ln\frac{\alpha^2 H - H}{\alpha^2 H - \alpha H} \\ &= \ln\frac{(\alpha-1)(\alpha+1)}{\alpha(\alpha-1)} = \ln\frac{\alpha+1}{\alpha} = \ln\left(1 + \frac{1}{\alpha}\right), \end{aligned}$$

其中因為 $\alpha > 1$，所以可約去 $\alpha - 1 \neq 0$。這就是所求的結果。

當 α 很大時，$x = 1/\alpha$ 很小，所以

$$kT \approx \frac{1}{\alpha} - \frac{1}{2\alpha^2} \approx \frac{1}{\alpha},$$

因為當 α 很大時，第二項相對於第一項可以忽略。

第 (ii) 部分

現在流出率為 $\mu\sqrt{h}$，其中 $\mu > 0$ 為某常數，所以 $A\,\mathrm{d}h/\mathrm{d}t = q - \mu\sqrt{h}$。高度在 $h = \alpha^2 H$ 時保持不變，所以 $q = \mu\sqrt{\alpha^2 H} = \mu\alpha\sqrt{H}$（利用了 $\alpha > 0$）。因此

$$\frac{\mathrm{d}h}{\mathrm{d}t} = c\left(\alpha\sqrt{H} - \sqrt{h}\right), \qquad c = \frac{\mu}{A} > 0.$$

同樣，當 $h < \alpha^2 H$ 時括號內的值為正，所以

$$cT' = \int_H^{\alpha H} \frac{\mathrm{d}h}{\alpha\sqrt{H} - \sqrt{h}}.$$

? **想一想** 分母中的 $\sqrt{h}$ 該如何處理？

代換 $u = \sqrt{h}$，於是 $h = u^2$ 及 $\mathrm{d}h = 2u\,\mathrm{d}u$；積分上下限變為 $u = \sqrt{H}$ 和 $u = \sqrt{\alpha H}$。為簡便起見記 $a = \alpha\sqrt{H}$，被積函數便是一個有理函數（rational function），作除法後得

$$\frac{2u}{a-u} = -2 + \frac{2a}{a-u}.$$

因此

$$\begin{aligned} cT' &= \int_{\sqrt{H}}^{\sqrt{\alpha H}} \left(-2 + \frac{2a}{a-u}\right) \mathrm{d}u = \Big[-2u - 2a\ln(a-u)\Big]_{\sqrt{H}}^{\sqrt{\alpha H}} \\ &= -2\left(\sqrt{\alpha H} - \sqrt{H}\right) + 2a\sqrt{H}\,\ln\frac{\alpha\sqrt{H} - \sqrt{H}}{\alpha\sqrt{H} - \sqrt{\alpha}\sqrt{H}} \\ &= 2\sqrt{H}\left(1 - \sqrt{\alpha} + \alpha\ln\frac{\alpha-1}{\alpha - \sqrt{\alpha}}\right). \end{aligned}$$

由於 $u \le \sqrt{\alpha H} < \alpha\sqrt{H} = a$（因為當 $\alpha > 1$ 時 $\sqrt{\alpha} < \alpha$），對數內的量都是正數。最後，因式分解 $\alpha - 1 = (\sqrt{\alpha}-1)(\sqrt{\alpha}+1)$ 及 $\alpha - \sqrt{\alpha} = \sqrt{\alpha}(\sqrt{\alpha}-1)$，得

$$\frac{\alpha-1}{\alpha-\sqrt{\alpha}} = \frac{\sqrt{\alpha}+1}{\sqrt{\alpha}} = 1 + \frac{1}{\sqrt{\alpha}},$$

所以

$$cT' = 2\sqrt{H}\left(1 - \sqrt{\alpha} + \alpha\ln\left(1 + \frac{1}{\sqrt{\alpha}}\right)\right),$$

此即所求。

當 α 很大時，$1/\sqrt{\alpha}$ 很小，所以

$$\alpha\ln\left(1 + \frac{1}{\sqrt{\alpha}}\right) \approx \alpha\left(\frac{1}{\sqrt{\alpha}} - \frac{1}{2\alpha}\right) = \sqrt{\alpha} - \frac{1}{2}.$$

兩個大項 $\pm\sqrt{\alpha}$ 互相抵消，剩下

$$cT' \approx 2\sqrt{H}\left(1 - \sqrt{\alpha} + \sqrt{\alpha} - \tfrac{1}{2}\right) = \sqrt{H}.$$

留意這裏我們需要用到對數展開中的 x^2 項：若只用 $\ln(1+x) \approx x$，括號內的值會是 1，於是得到 $2\sqrt{H}$，這是錯的。這正是題目給出兩項近似式的原因。

驗算

- 當 $\alpha \to 1^+$ 時，(i) 給出 $kT \to \ln 2$，而不是 0。這是對的：與平衡點（equilibrium）的差距起初為 $(\alpha^2-1)H \approx 2(\alpha-1)H$，而須縮小至 $(\alpha^2-\alpha)H \approx (\alpha-1)H$，即差距減半；而以速率 k 進行的指數衰減（exponential decay）減半所需時間為 $(\ln 2)/k$。在 (ii) 中，$cT' \to 2\sqrt{H}\ln 2$，這與在 $h = H$ 附近把 $\sqrt{h}$ 線性化（linearising）所得的等效速率 $c/(2\sqrt{H})$ 下同樣的減半相吻合。
- 取 $\alpha = 4$ 及 $H = 1$：$kT = \ln 1.25 \approx 0.2231$，$cT' = 2(-1 + 4\ln 1.5) \approx 1.2437$。對多組隨機的 H, α, k, c 用數值方法（龍格－庫塔法（Runge–Kutta））逐步求解兩條微分方程，結果與兩條公式吻合至 10^{-9}。
- 取 $\alpha = 10^4$、$H = 2$：$\alpha\ln(1 + 1/\alpha) \approx 0.99995$，$cT'/\sqrt{H} \approx 1.0066$，與極限值一致（下一項給出 $cT' \approx \sqrt{H}\,(1 + \frac{2}{3\sqrt{\alpha}})$）。

常見錯誤

- 忘記利用平衡條件消去流入率，結果剩下太多常數。
- 在第 (ii) 部分只用 $\ln(1+x) \approx x$，因而丟失了關鍵的 $-\frac{1}{2}$。
- 把 $\sqrt{\alpha^2 H}$ 寫成 $\alpha^2\sqrt{H}$。

STEP 之外

兩個模型都是一階自治微分方程（first-order autonomous differential equation）$\dot{h} = f(h)$，並在 $h = \alpha^2 H$ 處有一個穩定平衡點；線性的情況其實就是換了包裝的牛頓冷卻定律（Newton's law of cooling），而 $\sqrt{h}$ 的情況則是描述液體從小孔流出的托里拆利定律（Torricelli's law）。不必明確求解而去研究平衡點及其穩定性，正是常微分方程（ODE）定性理論（qualitative theory）的起點。Boyce 與 DiPrima 合著的 *Elementary Differential Equations*（《初等微分方程》）在前幾章便討論了自治方程和水箱混合模型。

詳細解答

STEP I 2011，第 8 題

純數學 • 丟番圖方程

> **STEP I 2011，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/11-s1-q8

關鍵思路

立方函數是嚴格遞增（strictly increasing）的，所以比較 m^3 與 n^3 及 $(n+1)^3$，等同於比較 m 與 n 及 $n+1$。若 m 嚴格位於兩個連續整數（consecutive integers）之間，它就不可能是整數。這樣便只剩下少數幾個 n 值需要逐一驗證。第 (ii) 部分則把同樣的夾逼（squeeze）方法用於一條新的方程。

解答

全題均利用 $x \mapsto x^3$ 在整個 $\mathbb{R}$ 上嚴格遞增這一事實，故對實數 a, b，

$$a < b \iff a^3 < b^3.$$

第 (i)(a) 部分

由 $(*)$ 得

$$m^3 - n^3 = n^2 + 1 > 0,$$

因為 $n^2 \geq 0$。所以 $m^3 > n^3$，因此 $m > n$。

其次，$m < n+1$ 當且僅當 $m^3 < (n+1)^3$。利用 $(*)$ 並展開，

$$\begin{aligned} m^3 < (n+1)^3 &\iff n^3 + n^2 + 1 < n^3 + 3n^2 + 3n + 1 \\ &\iff 0 < 2n^2 + 3n, \end{aligned}$$

這正是所要求的等價關係。每一步都是可逆的，所以「當且僅當」成立。

現在 $2n^2 + 3n = n(2n+3)$。這是一個二次式（quadratic），其根為 $n = -\frac{3}{2}$ 及 $n = 0$，且首項係數（leading coefficient）為正，所以它恰好在 $n < -\frac{3}{2}$ 或 $n > 0$ 時為正。對任

何這樣的 n，我們同時有 $m > n$ 及 $m < n+1$。因此

$$n < m < n+1 \quad \text{除非} \quad -\tfrac{3}{2} \le n \le 0.$$

第 (i)(b) 部分

設 m 和 n 是滿足 $(*)$ 的整數。若 $n < -\frac{3}{2}$ 或 $n > 0$，則由 (a) 部分得 $n < m < n+1$。連續整數 n 與 $n+1$ 之間嚴格來說沒有任何整數，所以這是不可能的。

想一想 (a) 部分對所有實數 n 都成立。那麼「是整數」這個條件在哪裏用上？

它用上了兩次。一是如上所述，排除區間以外的所有 n；二是把區間 $-\frac{3}{2} \le n \le 0$ 縮減為 $n \in \{-1, 0\}$。我們逐一驗證這兩個值：

- $n = 0$：$m^3 = 0+0+1 = 1$，所以 $m = 1$。
- $n = -1$：$m^3 = -1+1+1 = 1$，所以 $m = 1$。

兩者都給出整數 m。所以唯一的整數解是 $(m, n) = (1, 0)$ 及 $(1, -1)$，得證。

第 (ii) 部分

我們仿照第 (i) 部分的做法。首先比較 p 與 q：

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

只要 $q^2 > \frac{1}{2}$，即對每個整數 $q \neq 0$，此式皆為正。對這些 q，我們有 $p > q$。

其次比較 p 與 $q+1$：

$$\begin{aligned} p^3 < (q+1)^3 &\iff q^3 + 2q^2 - 1 < q^3 + 3q^2 + 3q + 1 \\ &\iff 0 < q^2 + 3q + 2 = (q+1)(q+2). \end{aligned}$$

此式恰好在 $q > -1$ 或 $q < -2$ 時成立。

把兩者結合起來。若 q 是整數，且 $q \neq 0$、$q \neq -1$ 及 $q \neq -2$，則 $q < p < q+1$，所以 p 不可能是整數。（若 $q \ge 1$ 或 $q \le -3$，兩個條件均成立。）剩下 $q \in \{-2, -1, 0\}$：

- $q = 0$：$p^3 = -1$，所以 $p = -1$。
- $q = -1$：$p^3 = -1+2-1 = 0$，所以 $p = 0$。
- $q = -2$：$p^3 = -8+8-1 = -1$，所以 $p = -1$。

三者都給出整數 p，所以整數解為

$$(p, q) = (-1, 0), \quad (0, -1), \quad (-1, -2).$$

驗算

把每個答案代回原式。例如，$(-1)^3 = -1 = (-2)^3 + 2(-2)^2 - 1 = -8 + 8 - 1$。用電腦搜尋所有 $|n|, |q| < 10^5$ 的整數，恰好找到 $(*)$ 的兩個解及第 (ii) 部分的三個解，再無其他。你也可以用數值方法測試夾逼的結果。當 $n = 5$ 時，$m = \sqrt[3]{151} \approx 5.32$，嚴格位於 5 與 6 之間。

常見錯誤

- 在第 (ii) 部分遺漏 $q = 0$。此時 $p^3 - q^3 = -1 < 0$，所以下界 $p > q$ 不成立，這個情況需要另行驗證。
- 寫「$m^3 > n^3$，所以 $m > n$」而不給理由。這是成立的，因為立方運算是嚴格遞增的（而平方運算對負數則不然），這一點值得寫明。
- 只寫到 $n < m < n + 1$ 便停下，沒有說明為何這意味着 m 不是整數。

STEP 之外

像 $y^3 = f(x)$ 這類求整數點的方程稱為*丟番圖方程*（Diophantine equations）。把未知數夾在相鄰的冪之間，是處理這類方程的標準 first 步。較難的版本，例如 $y^2 = x^3 + k$，會引出橢圓曲線（elliptic curves）以及 Baker 的對數線性型（linear forms in logarithms）理論。這些都是深奧的領域，但仍有不少情況可以用這種初等的估界方法解決。Davenport 的 *The Higher Arithmetic*（《高等算術》）是初等數論（number theory）的一本好入門書。

詳細解答

STEP I 2011，第 9 題

力學 • 拋體運動

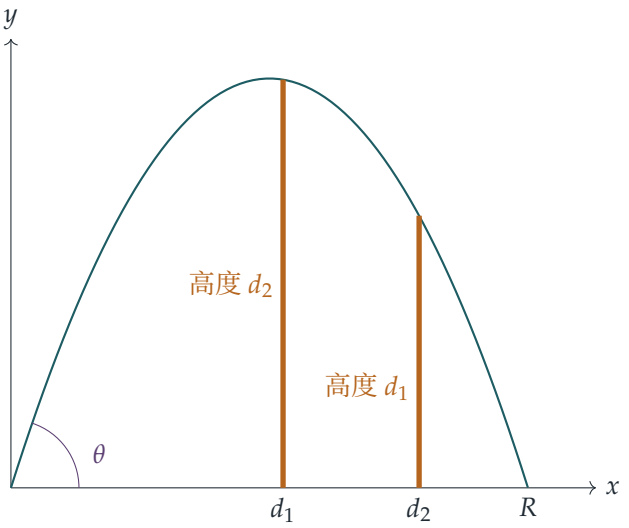
> **STEP I 2011，第 9 題**
>
> 繼續閱讀之前，請先看題目（原題為英文）。題目載於 STEP Support Programme 的試題資料庫：
>
> step.maths.org/questions/11-s1-q9

關鍵思路

把軌跡（trajectory）寫成單一方程 $y = x\tan\theta - ax^2$，其中常數 a 把 g、速率和角度合併在一起。「剛好越過」一幅牆，意思是路徑經過牆頂的角位，所以每幅牆都給出一條關於兩個未知數 $\tan\theta$ 和 a 的線性方程（linear equation）。解這對方程便得出角度，而射程（range）就只是 $y = 0$ 的另一個根（root）。

解答

我們把質點模型化為只受重力作用而運動的一點（不計空氣阻力（air resistance）），從原點 O 以速率 u、仰角 θ 拋出，x 為水平方向，y 為豎直向上方向。兩幅牆的高度不同，因此我們假設 $d_1 \neq d_2$。



（圖中採用 $d_1 < d_2$；但代數推導並不依賴哪一個較大。）

軌跡

在時間 t 時，$x = ut\cos\theta$ 及 $y = ut\sin\theta - \frac{1}{2}gt^2$。消去 $t = x/(u\cos\theta)$，得

$$y = x\tan\theta - ax^2, \qquad \text{其中} a = \frac{g}{2u^2\cos^2\theta} > 0.$$

設 $T = \tan\theta$。質點剛好越過每幅牆，所以它經過牆頂 (d_1, d_2) 和 (d_2, d_1)：

$$d_2 = d_1 T - ad_1^2, \tag{1}$$
$$d_1 = d_2 T - ad_2^2. \tag{2}$$

求 $\tan\theta$

為了消去 a，把 (1) 乘以 d_2^2，把 (2) 乘以 d_1^2，使兩條式中的 a 項都是 $ad_1^2d_2^2$：

$$d_2^3 = d_1d_2^2\,T - ad_1^2d_2^2,$$
$$d_1^3 = d_1^2d_2\,T - ad_1^2d_2^2.$$

相減，得

$$d_2^3 - d_1^3 = d_1d_2(d_2 - d_1)\,T.$$

想一想 我們可以除以 $d_2 - d_1$ 嗎？

可以：兩幅牆的高度 d_2 和 d_1 不同，所以 $d_1 \neq d_2$。（若兩者相等，就只有一幅牆，而單一個點無法同時確定 θ 和 u。）由於 $d_2^3 - d_1^3 = (d_2 - d_1)(d_2^2 + d_1d_2 + d_1^2)$，除以 $d_1d_2(d_2 -$

d_1) 便得

$$\tan\theta = \frac{d_1^2 + d_1 d_2 + d_2^2}{d_1 d_2},$$

此即所求。

射程

質點落地之處滿足 $y = 0$ 且 $x \neq 0$，即 $x = T/a$。因此我們也需要求出 a。把 (1) 除以 d_1，(2) 除以 d_2，得

$$\frac{d_2}{d_1} = T - ad_1, \qquad \frac{d_1}{d_2} = T - ad_2.$$

以第二式減第一式，消去 T：

$$a(d_2 - d_1) = \frac{d_2}{d_1} - \frac{d_1}{d_2} = \frac{(d_2 - d_1)(d_2 + d_1)}{d_1 d_2}, \qquad \text{所以} \qquad a = \frac{d_1 + d_2}{d_1 d_2},$$

這裏同樣用到 $d_1 \neq d_2$。因此射程為

$$R = \frac{T}{a} = \frac{d_1^2 + d_1 d_2 + d_2^2}{d_1 d_2} \cdot \frac{d_1 d_2}{d_1 + d_2} = \frac{d_1^2 + d_1 d_2 + d_2^2}{d_1 + d_2}.$$

作為合理性檢查，$R - d_2 = \dfrac{d_1^2}{d_1 + d_2} > 0$，同理 $R - d_1 = \dfrac{d_2^2}{d_1 + d_2} > 0$，所以質點落在兩幅牆之外，這是必然的。

> ? **想一想** g 和 u 到哪裏去了？

它們始終只透過單一組合 a 出現，而兩幅牆完全確定了 a：$u^2 = g/(2a\cos^2\theta)$。所以兩幅牆決定了整條軌跡，而射程只取決於 d_1 和 d_2。

驗算

- 對稱性（symmetry）：互換 d_1 和 d_2 只是把兩幅牆對調，而兩個答案關於 d_1, d_2 都是對稱的。
- 代入數值：取 $d_1 = 2$、$d_2 = 3$，得 $\tan\theta = \frac{19}{6}$、$a = \frac{5}{6}$ 及 $R = \frac{19}{5}$。於是 $y(2) = \frac{19}{3} - \frac{10}{3} = 3$，$y(3) = \frac{19}{2} - \frac{15}{2} = 2$，符合要求。以 $g = 9.8$ 對運動作數值逐步模擬，同樣落在 $x = 3.8$。
- 由於 $d_1^2 + d_2^2 \geq 2d_1 d_2$，得 $\tan\theta \geq 3$，所以拋射角總是很陡（至少約 71.6°）。這是合理的：要先越過較高的牆，然後這麼快就降到較矮的牆，路徑必須是一道窄拱。

常見錯誤

- 混淆哪幅牆有哪個高度：距離為 d_1 的牆，高度是 d_2。

- 把 a 寫成含 $\cos\theta$ 而非 $\cos^2\theta$，或沒有把 a 當作單一未知數處理，結果留下一個關於 u 和 θ 的混亂方程組。
- 除以 $d_2 - d_1$ 時沒有說明它為何不等於零。

STEP 之外

這裏的技巧在於：一個帶有兩個自由參數（parameter）（此處為 $\tan\theta$ 和 a）的曲線族，可由它必須經過的兩個點確定下來，這是插值（interpolation）的一個小例子。選取適當的未知數組合這個想法貫穿整個力學，例如求所有以固定拋射速率發射的軌跡的「包絡線」（envelope）（即安全拋物線（parabola of safety））。如想學習更多大學一年級程度的拋射體及質點力學，可參閱 Kleppner 和 Kolenkow 的 *An Introduction to Mechanics*（《力學導論》）。

詳細解答

STEP I 2011，第 10 題

力學 • 彈性碰撞

STEP I 2011，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/11-s1-q10

關鍵思路

兩個質點（particle）都由同一點從靜止開始運動，而在釋放與碰撞之間發生的一切（自由下落、一次完全彈性（perfectly elastic）的反彈）都不會損失任何能量。因此在任一高度，每個質點的速率都相同，而它們會在同一高度相遇。之後，碰撞就是一道標準的兩條方程問題（動量（momentum）守恆，以及 $e = 1$ 的牛頓碰撞定律（Newton's law of restitution）），而 B 的最大高度再次由能量求得。

解答

我們把 A 和 B 視為質點，忽略空氣阻力（air resistance），並全程取向上為正方向。

為何兩者速率相等

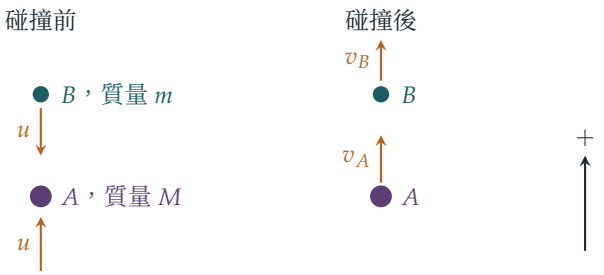
每個質點都在高度 h 處由靜止開始運動。自由飛行期間只有重力作用，所以機械能守恆（conservation of mechanical energy）；A 在平面上的反彈是完全彈性的，因此它只會令 A 的速度反向而不改變其速率，那裏也沒有損失能量。故此，在第一次碰撞之前，每個質點的單位質量總能量都是 gh，而處於高度 y 的質點的速率 v 滿足

$$\tfrac{1}{2}v^2 + gy = gh, \qquad \text{所以} \qquad v = \sqrt{2g(h-y)}.$$

速率只取決於高度，與質量或質點過去的運動經歷無關。在第一次碰撞的一刻，兩個質點處於同一高度 y，所以它們的速率相等。此外，A 尚未回到 P，所以 $y < h$，而這個共同速率 $u = \sqrt{2g(h-y)}$ 是正數。由於 A 已經反彈，它正向上運動；而 B（從未反彈）則正向下運動。

碰撞

碰撞前一刻，A 的速度為 $+u$，B 的速度為 $-u$。設碰撞後一刻兩者的速度分別為 v_A 和 v_B。



碰撞中的衝力（impulsive force）是這對質點之間的內力，而在撞擊的一瞬間重力不產生任何衝量（impulse），所以動量守恆：

$$Mu - mu = Mv_A + mv_B. \tag{1}$$

碰撞是完全彈性的（$e = 1$），所以分離速率等於接近速率：

$$v_B - v_A = u - (-u) = 2u. \tag{2}$$

把由 (2) 得到的 $v_B = v_A + 2u$ 代入 (1)：

$$(M-m)u = (M+m)v_A + 2mu \quad \Longrightarrow \quad v_A = \frac{(M-3m)u}{M+m},$$

然後

$$v_B = v_A + 2u = \frac{(M-3m)u + 2(M+m)u}{M+m} = \frac{(3M-m)u}{M+m}.$$

由於 $M > 3m$，我們有 $M - 3m > 0$，所以 $v_A > 0$；而 $v_B = v_A + 2u > 0$ 亦成立。因此第一次碰撞後兩個質點都向上運動，正如題目所要求。

想一想 我們怎樣知道 B 在下一次與 A 相遇之前，確實能到達其飛行的最高點？

碰撞後兩個質點的加速度都是 $-g$，所以兩者都在飛行時，相對速度一直保持 $v_B - v_A = 2u$：B 以穩定的速率離 A 而去，因此在 A 再次撞到平面之前，兩者不可能相遇。若 A 在 B 仍在上升時反彈，就比較兩者的單位質量能量：B 的是 $\frac{1}{2}v_B^2 + gy$，A 的是 $\frac{1}{2}v_A^2 + gy$；由於 $v_B > v_A > 0$，後者較小（而 A 的反彈不改變其能量）。所以每當兩者處於同一高度時，A 的速率都小於 B 的速率。要在 B 仍在上升時從下方追上 B，A 在相遇一刻向上的速率必須至少與 B 一樣快，而我們剛剛看到這是不可能的。因此第二次碰撞只會在 B 經過其最高點之後才發生，而這個最高點正是題目所求的最大高度。

B 的最大高度

碰撞發生在高度 y 處，其中 $u^2 = 2g(h-y)$，即

$$y = h - \frac{u^2}{2g}.$$

碰撞後 B 以初速率 v_B 只在重力作用下上升，所以它會再上升 $v_B^2/(2g)$（由 $v^2 = v_B^2 - 2gs$ 並取 $v = 0$ 得出）。因此它的最大高度是

$$\begin{aligned} y + \frac{v_B^2}{2g} &= h - \frac{u^2}{2g} + \frac{(3M-m)^2u^2}{2g(M+m)^2} \\ &= h + \frac{u^2}{2g(M+m)^2}\left[(3M-m)^2 - (M+m)^2\right] \\ &= h + \frac{u^2}{2g(M+m)^2}(2M-2m)(4M) \\ &= h + \frac{4M(M-m)u^2}{(M+m)^2 g}, \end{aligned}$$

其中第三行用了平方差（difference of two squares）公式。這正是所求的結果。

驗算

- 能量：以上述的 v_A 和 v_B，有 $Mv_A^2 + mv_B^2 = (M+m)u^2$，正如彈性碰撞所要求。
- 極限情況（limiting case）：若 $m \to 0$，則 $v_A \to u$ 而 $v_B \to 3u$（一個極輕的球撞上一個重球，就像撞上一道移動的牆一樣反彈，相對於它以速度 $2u$ 離開），而高度變為 $h + 4u^2/g$，與 $y + 9u^2/(2g)$ 吻合。
- 由於 $M > m$，額外高度是正數：B 最終到達 P 之上，這是理所當然的，因為它從 A 獲得了能量。
- 數值方面，我們在隨機時間釋放 B，取隨機的 h 及滿足 $M > 3m$ 的質量，逐步模擬運動（包括 A 其後的反彈）直至第二次碰撞，發現 B 的最大高度與公式吻合，相對誤差約為 10^{-10}。

常見錯誤

- 正負號錯誤：把碰撞前兩個速度都取作 $+u$，就忘記了 A 正在上升時 B 正在下落。
- 把額外上升的高度 $v_B^2/(2g)$ 從平面或從 P 起計，而不是從碰撞高度 $h - u^2/(2g)$ 起計。

STEP 之外

「速率只取決於高度」這個論證，其實就是保守力（conservative force）下的能量守恆：某一點的速率只取決於該點的勢能（potential energy），與所走的路徑無關。碰撞本身通常在質心參考系（centre-of-mass frame）中處理最為簡單：在該參考系中，彈性碰撞只是把每個質點的速度反向；再變換回原參考系，一行便得到 v_A 和 v_B。Kleppner 與 Kolenkow 的 *An Introduction to Mechanics*（《力學導論》）仔細地發展了這兩個概念。

詳細解答

STEP I 2011，第 11 題

力學 • 剛體的平衡

STEP I 2011，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：

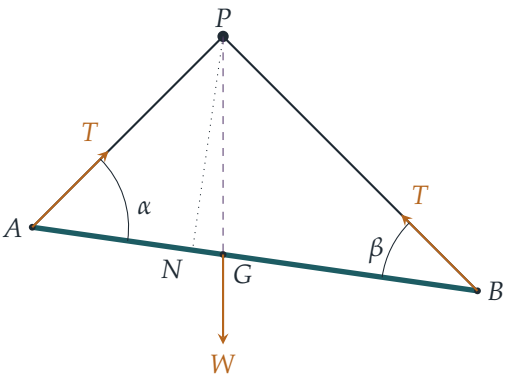
step.maths.org/questions/11-s1-q11

關鍵思路

桿上只受三個力：繩在 A 和 B 兩點的張力（tension），以及作用於 G 的重量。由於釘子光滑（smooth），兩個張力相等，因此對 G 取力矩（moments）（重量對 G 沒有力矩），便立即得出兩個角之間的關係。第二部分還需要一個事實：處於平衡（equilibrium）而又不全部平行的三個力，必定通過同一點。這一點就是釘子，所以 PG 是鉛垂的。

解答

我們把桿視為剛體桿（rigid rod），繩為輕（light）而不可伸長（inextensible）的，釘子為光滑的。設桿的重量為 W。



證明 $3\sin\alpha = 4\sin\beta$

繩是輕的，並繞過光滑的釘子，所以整條繩的張力 T 都相同。因此桿在 A 受到沿 AP 方向的力 T，在 B 受到沿 BP 方向的力 T，另外還有在 G 鉛垂向下作用的 W。

對 G 取力矩。重量沒有力矩。由 G 到直線 AP 的垂直距離是 $AG\sin\alpha = 3d\sin\alpha$，由 G 到直線 BP 的垂直距離是 $GB\sin\beta = 4d\sin\beta$，因為 $GB = 7d - 3d = 4d$。A 點的張力使桿繞 G 向一個方向轉動，B 點的張力則使桿向另一個方向轉動，所以平衡時

$$T\cdot 3d\sin\alpha = T\cdot 4d\sin\beta \quad\Longrightarrow\quad 3\sin\alpha = 4\sin\beta,$$

（兩邊除以 $Td \neq 0$）即為所求。

繩的長度

由 $\cos\beta = \frac{4}{5}$ 得 $\sin\beta = \frac{3}{5}$（β 是三角形的一個角，所以 $\sin\beta > 0$）。於是

$$\sin\alpha = \tfrac{4}{3}\sin\beta = \tfrac{4}{5}, \qquad \cos\alpha = \tfrac{3}{5},$$

這裏取正根，因為 α 是銳角（acute）。接着

$$\cos(\alpha+\beta) = \cos\alpha\cos\beta - \sin\alpha\sin\beta = \tfrac{3}{5}\cdot\tfrac{4}{5} - \tfrac{4}{5}\cdot\tfrac{3}{5} = 0,$$

而 $0 < \alpha+\beta < \pi$，所以 $\alpha+\beta = \frac{\pi}{2}$。因此三角形 PAB 在 P 處是直角，斜邊(hypotenuse)為 $AB = 7d$。故此

$$PA = 7d\cos\alpha = \tfrac{21}{5}d, \qquad PB = 7d\cos\beta = \tfrac{28}{5}d,$$

而繩的長度是

$$PA + PB = \tfrac{49}{5}d.$$

桿的傾角

想一想 我們知道 P 相對於桿的位置，但「鉛垂」方向在哪裏？

兩個張力都沿通過 P 的直線作用，所以它們的合力（resultant）也通過 P。平衡時這個合力必須與重量抵消，因此它必須與 W 大小相等、方向相反，並且在同一直線上。（處於平衡的三個不平行的力必定共點（concurrent）。）所以通過 G 的鉛垂線經過 P：PG 是鉛垂的。

設 N 為由 P 到 AB 的垂線的垂足。在直角三角形 PAN 中，

$$AN = PA\cos\alpha = \tfrac{21}{5}d\cdot\tfrac{3}{5} = \tfrac{63}{25}d, \qquad PN = PA\sin\alpha = \tfrac{21}{5}d\cdot\tfrac{4}{5} = \tfrac{84}{25}d.$$

由於 $AN = \frac{63}{25}d < 3d = AG$，點 N 位於 A 與 G 之間，且

$$NG = 3d - \tfrac{63}{25}d = \tfrac{12}{25}d.$$

在直角三角形 PNG 中，

$$\tan\angle PGN = \frac{PN}{NG} = \frac{84}{12} = 7.$$

角 PGN 是桿與鉛垂線 GP 之間、在 A 一側量度的角。它是銳角，所以端點 A 在通過 G 的水平線之上，而 B 在其之下，與題目給出的位置相符。桿與水平線之間的角 θ 等於 $\frac{\pi}{2} - \angle PGN$，所以

$$\tan\theta = \frac{1}{\tan\angle PGN} = \frac{1}{7}, \qquad \text{即} \qquad \theta = \arctan\tfrac{1}{7},$$

即為所求。

驗算

- 繩長 $\frac{49}{5}d = 9.8d$ 比桿長 $7d$ 長，符合三角形 PAB 中三角不等式（triangle inequality）的要求。
- 兩個相等張力的合力平分它們之間的夾角，所以 PG 平分角 APB。根據角平分線定理（angle bisector theorem），$PA : PB$ 應等於 $AG : GB = 3 : 4$，而 $\frac{21}{5} : \frac{28}{5} = 3 : 4$，確實如此。
- 作為數值驗算，我們把 P 放在以 A 和 B 為焦點（foci）的橢圓（ellipse）上（即在這個繩長下釘子所有可能的位置），然後找出使 G 在 P 之下的深度取駐定值（stationary）的傾角。結果是 $8.13°$，與 $\arctan\frac{1}{7} \approx 8.130°$ 吻合。

常見錯誤

- 給繩的兩段設定不同的張力。光滑的釘子不能改變輕繩中的張力。
- 把兩段距離對調，得出 $4\sin\alpha = 3\sin\beta$。較長的一段 $GB = 4d$ 應配 B 處的角。
- 忘記 α 是銳角，而容許 $\cos\alpha = -\frac{3}{5}$。這會得出另一個三角形，而在其中 $\alpha + \beta$ 並不是直角。

STEP 之外

在繩長固定的情況下，釘子必定位於以 A 和 B 為焦點的橢圓上。於是，PG 平分角 APB 這個事實正是橢圓的反射性質（reflection property）：在 P 點的法線（normal）平分兩條

焦半徑（focal radii）之間的夾角。此外，平衡位置恰好就是勢能（potential energy）取駐定值的位置，這正是大學一年級課程中虛功原理（principle of virtual work）和拉格朗日力學（Lagrangian mechanics）的起點。Kleppner 和 Kolenkow 所著的 *An Introduction to Mechanics*（《力學導論》）是學習平衡、力矩和能量方法的好讀物。

詳細解答

STEP I 2011，第 12 題

概率與統計 • 排隊與選票問題

STEP I 2011，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/11-s1-q12

本題在 *Beyond STEP* 第 9.10 節總能找贖中有完整的引導式講解：https://beyondstep.keithmcnulty.org/zh-hant/ch09/ballot.html

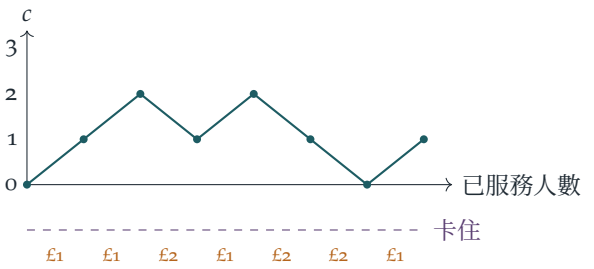
關鍵思路

只需追蹤一個數字：我手上有多少枚 £1 硬幣。每位付 £1 的顧客令這個數字加一；每位付 £2 的顧客令它減一（我收下他們的 £2 硬幣，但這枚硬幣不能用來找零）。我會卡住，當且僅當一位付 £2 的顧客到來時，我手上一枚 £1 硬幣也沒有。由於付 £2 的顧客只有寥寥幾位，危險全都集中在隊伍的前端，所以我們只需看隊伍最前面的幾個人，逐一分析每種開頭會帶來甚麼結果。

解答

重要的只是隊伍中硬幣的排列模式，而不是那些人是誰。由於所有人的排列都是等可能的，任何一個位置上的人，等可能是尚未排入隊伍的任何一人，因此我們可以從隊頭開始逐個建構隊伍：若尚餘 a 位持 £1 硬幣的人和 b 位持 £2 硬幣的人，則下一位持 £1 硬幣的概率（probability）為 $\frac{a}{a+b}$，持 £2 硬幣的概率為 $\frac{b}{a+b}$。

設 c 為我手上 £1 硬幣的數目。我能服務所有人，當且僅當 c 永遠不需要跌至 0 以下，即在隊伍中的每一點，已服務的 £1 顧客人數至少與已服務的 £2 顧客人數一樣多。



上圖顯示一條成功的隊伍：c 的路徑從未觸及虛線。

第 (i) 部分

若只有一位付 £2 的顧客，唯一會卡住的情況是此人排在隊首，那時我手上沒有零錢。若隊首是其他人，那麼到這位 £2 顧客到來時，我手上至少有一枚 £1 硬幣。這位 £2 顧客等可能排在 $m+1$ 個位置中的任何一個，因此

$$\mathrm{P}(\text{賣給所有人}) = 1 - \frac{1}{m+1} = \frac{m}{m+1}.$$

第 (ii) 部分

以 1 表示付 £1 的顧客，以 2 表示付 £2 的顧客，並列出所有開頭。

- 若第一人是 2，我立即卡住。
- 若隊伍以 $1,2,2$ 開頭，則 c 依次為 $1,0$，我在第三人時卡住。
- 若隊伍以 $1,1$ 開頭，則兩人之後 $c=2$。付 £2 的顧客總共只有兩位，而 c 只會在服務其中一人時下跌，因此 c 永遠不會跌至 $2-2=0$ 以下。成功。
- 若隊伍以 $1,2,1$ 開頭，則 c 依次為 $1,0,1$，而只剩一位付 £2 的顧客，因此 c 永遠不會跌至 0 以下。成功。

這些情況涵蓋了所有可能，因此我能服務所有人，當且僅當隊伍以 $1,1$ 或 $1,2,1$ 開頭。從隊頭開始建構隊伍，

$$\mathrm{P}(1,1) = \frac{m}{m+2}\cdot\frac{m-1}{m+1}, \quad \mathrm{P}(1,2,1) = \frac{m}{m+2}\cdot\frac{2}{m+1}\cdot\frac{m-1}{m} = \frac{2(m-1)}{(m+2)(m+1)}.$$

這兩個事件互斥（disjoint），因此

$$\mathrm{P}(\text{賣給所有人}) = \frac{m(m-1)+2(m-1)}{(m+2)(m+1)} = \frac{(m-1)(m+2)}{(m+2)(m+1)} = \frac{m-1}{m+1},$$

即為所求。（我們需要 $m \geq 2$，使最後一個因子中尚餘 $m-1 \geq 1$ 位 £1 顧客。）

第 (iii) 部分

現在有三位付 £2 的顧客。同樣地，看隊伍最前面的三個人。

- 以 2 開頭，或以 1, 2, 2 開頭：卡住，與 (ii) 完全相同。
- 以 1, 1, 1 開頭：此時 $c = 3$，三位 £2 顧客最多只能把 c 減至 0。成功。
- 以 1, 1, 2 或 1, 2, 1 開頭：兩種情況下，三人之後都有 $c = 1$，而餘下的是 $m - 2$ 位持 £1 硬幣的人和 2 位持 £2 硬幣的人，以隨機次序排列。

想一想 在最後一種情況中，隊伍餘下的部分何時會出問題？

以 $c = 1$ 開始處理隊伍餘下的部分，就好比在一條有兩位 £2 顧客的隊伍前面，多了一位免費的 £1 顧客，這正是改頭換面的第 (ii) 部分。在 $c = 1$ 且只剩兩位 £2 顧客的情況下，我卡住當且僅當接下來兩人都是 £2 顧客：c 先變成 0，然後我便無法服務。若接下來兩人中有任何一人持 £1 硬幣，則 c 永遠不會跌至 0 以下（仿照 (ii) 檢查 1, ... 和 2, 1, ... 即可）。在首三人之後餘下的 m 人中，

$$\begin{aligned} \text{P(接下來兩人都是2)} &= \frac{2}{m} \cdot \frac{1}{m-1}, \\ \text{故}\quad \text{P(其餘部分成功)} &= 1 - \frac{2}{m(m-1)} = \frac{(m-2)(m+1)}{m(m-1)}, \end{aligned}$$

這裏用了 $m^2 - m - 2 = (m-2)(m+1)$。

設 $D = (m+3)(m+2)(m+1)$。各個有用開頭的概率為

$$\begin{aligned} \text{P}(1,1,1) &= \frac{m(m-1)(m-2)}{D}, \\ \text{P}(1,1,2) &= \frac{m \cdot (m-1) \cdot 3}{D}, \qquad \text{P}(1,2,1) = \frac{m \cdot 3 \cdot (m-1)}{D}. \end{aligned}$$

由於各個開頭互斥，而且無論開頭如何，隊伍餘下的部分同樣是隨機次序，

$$\begin{aligned} \text{P(成功)} &= \frac{m(m-1)(m-2)}{D} + \frac{6m(m-1)}{D} \cdot \frac{(m-2)(m+1)}{m(m-1)} \\ &= \frac{(m-2)\big[m(m-1) + 6(m+1)\big]}{D} = \frac{(m-2)(m^2+5m+6)}{D} \\ &= \frac{(m-2)(m+2)(m+3)}{(m+3)(m+2)(m+1)} = \frac{m-2}{m+1}, \end{aligned}$$

即為所求。（這裏 $m \geq 3$ 確保 $m - 1 \geq 2$，使條件概率（conditional probability）有意義。）

驗算

- 最小的情況：當 $m = n$ 時，對 $n = 1, 2, 3$，公式分別給出 $\frac{1}{2}$、$\frac{1}{3}$、$\frac{1}{4}$。當 $m = n = 2$ 時，由 1 和 2 組成的六種排列模式等可能，而只有 1122 和 1212 成功：$\frac{2}{6} = \frac{1}{3}$。✓
- m 很大時：三個概率都趨向 1，這在 £1 硬幣充裕時理應如此。
- 透過完整列舉所有隊伍次序，當 $m = 5$ 時，對 $n = 1, 2, 3$ 我得到 $\frac{5}{6}$、$\frac{2}{3}$、$\frac{1}{2}$，與 $\frac{m}{m+1}$、$\frac{m-1}{m+1}$、$\frac{m-2}{m+1}$ 吻合；這些公式亦已對直至 14 的每個 m 作精確驗證。

常見錯誤

- 忘記我收到的 £2 硬幣不能用來找零：只有 £1 硬幣才算數。
- 把每個人當作以固定概率 $\frac{m}{m+n}$ 獨立抽出。隊伍是以不放回抽樣（without replacement）的方式產生的，所以概率會隨着人逐一排入而改變。
- 在 (iii) 中，以為三人之後 $c = 1$ 必定安全，因而漏掉了失敗情況 1, 1, 2, 2, 2 或 1, 2, 1, 2, 2。

STEP 之外

對每個 $n \leq m$，規律都是 $\frac{m+1-n}{m+1}$。這就是*選票問題*（ballot problem），最簡潔的證明運用了*反射原理*（reflection principle）：一條觸及卡住水平的路徑可以經反射得到一條終點不同的路徑，從而把壞隊伍的數目化為單一個二項式係數（binomial coefficient）。當 $m = n$ 時，好隊伍的數目是一個卡特蘭數（Catalan number），它在組合數學中隨處可見。關於隨機遊走(random walk)和反射原理，可參閱 Grimmett 與 Welsh 的 *Probability: An Introduction*（《概率導論》）。

詳細解答

STEP I 2011，第 13 題

概率與統計 • 連續隨機變量

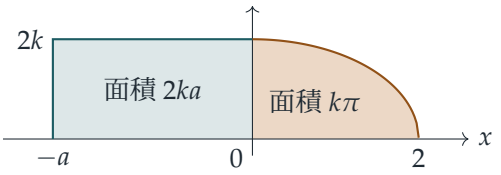
STEP I 2011，第 13 題

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/11-s1-q13

關鍵思路

先畫出密度函數（density function）的圖像。它由兩部分拼成：在 $[-a, 0)$ 上是一個高為 $2k$ 的長方形，在 $[0, 2]$ 上則是一個半徑為 2 的四分之一圓（垂直方向按比例 k 縮放）。圖像下方的面積就是概率，因此單靠長方形和四分之一圓，大部分工作已經完成，毋須做任何積分。總面積必須等於 1，由此便可定出 k。之後所有關於 d 的問題，都歸結為同一件事：某一點右方有多少面積？



解答

求 k

在 $[0, 2]$ 上，$\sqrt{4 - x^2}$ 的圖像是圓 $x^2 + y^2 = 4$ 的四分之一，所以

$$\int_0^2 \sqrt{4 - x^2}\,\mathrm{d}x = \tfrac{1}{4}\pi \cdot 2^2 = \pi.$$

（用題目給出的原函數（antiderivative）計算，結果一致：$2\arcsin 1 + 0 - 0 = \pi$。）因此總概率為 $2ka + k\pi = 1$，即

$$k = \frac{1}{2a + \pi}.$$

第 (i) 部分

把 $\mathrm{E}(X)$ 的積分在 0 處分拆：

$$\int_{-a}^{0} 2kx\,\mathrm{d}x = k\left[x^2\right]_{-a}^{0} = -ka^2.$$

至於第二部分，$(4-x^2)^{3/2}$ 的導數是 $-3x(4-x^2)^{1/2}$，所以

$$\int_{0}^{2} kx\sqrt{4-x^2}\,\mathrm{d}x = k\left[-\tfrac{1}{3}(4-x^2)^{3/2}\right]_{0}^{2} = k\cdot\tfrac{1}{3}\cdot 8 = \tfrac{8}{3}k.$$

兩部分相加，得

$$\mathrm{E}(X) = k\left(\frac{8}{3} - a^2\right) = \frac{8-3a^2}{3(2a+\pi)}.$$

第 (ii) 部分

0 右方的概率就是四分之一圓的面積：

$$\mathrm{P}(X>0) = k\pi = \frac{\pi}{2a+\pi}.$$

? **想一想** 知道了 $\mathrm{P}(X>0)$，為何就能判斷 d 的正負？

由於密度函數在 $(-a,2)$ 上為正，函數 $t \mapsto \mathrm{P}(X>t)$ 在 $-a \le t \le 2$ 上連續（continuous）且嚴格遞減（strictly decreasing）。它由 1 降至 0，所以恰好有一個 d 滿足 $\mathrm{P}(X>d) = \frac{1}{10}$。由於函數遞減，$d<0$ 當且僅當 $\mathrm{P}(X>0) < \frac{1}{10}$。換句話說，當 0 右方的概率不足十分之一時，$d$ 便是負數。

現在假設 $2a > 9\pi$，則 $2a+\pi > 10\pi$，所以

$$\mathrm{P}(X>0) = \frac{\pi}{2a+\pi} < \frac{1}{10},$$

因此 $d<0$，得證。（這個論證反過來亦成立：只有當 $2a>9\pi$ 時才有 $d<0$。）

在這情況下 $-a \le d < 0$，所以 d 右方的面積，就是長方形由 d 到 0、寬度為 $-d$ 的那一部分，再加上整個四分之一圓：

$$\mathrm{P}(X>d) = 2k(-d) + k\pi = \frac{\pi-2d}{2a+\pi}.$$

令它等於 $\frac{1}{10}$，得 $10\pi - 20d = 2a+\pi$，所以

$$d = \frac{9\pi-2a}{20}.$$

這個值確實位於 $[-a,0)$ 之內：因為 $2a>9\pi$，所以它是負數；而 $d \ge -a$ 等價於 $9\pi + 18a \ge 0$，這顯然成立。

第 (iii) 部分

現在 $d = \sqrt{2} > 0$，所以 d 落在圓弧部分之下，而

$$\mathrm{P}(X > \sqrt{2}) = k \int_{\sqrt{2}}^{2} \sqrt{4 - x^2}\,\mathrm{d}x.$$

利用題目給出的結果，

$$\begin{aligned}\int_{\sqrt{2}}^{2} \sqrt{4 - x^2}\,\mathrm{d}x &= \left[2\arcsin\left(\tfrac{1}{2}x\right) + \tfrac{1}{2}x\sqrt{4 - x^2}\right]_{\sqrt{2}}^{2} \\ &= \left(2 \cdot \tfrac{\pi}{2} + 0\right) - \left(2 \cdot \tfrac{\pi}{4} + \tfrac{\sqrt{2}}{2} \cdot \sqrt{2}\right) \\ &= \frac{\pi}{2} - 1.\end{aligned}$$

（從幾何角度看，這是被 $x = \sqrt{2}$ 處的弦（chord）割出的弓形（circular segment），它等於圓面積的八分之一，即 $\frac{\pi}{2}$，減去一個面積為 1 的直角三角形。）所以

$$\frac{\frac{\pi}{2} - 1}{2a + \pi} = \frac{1}{10} \quad \Longrightarrow \quad 5\pi - 10 = 2a + \pi \quad \Longrightarrow \quad a = 2\pi - 5.$$

這是正數（$a \approx 1.283$），並且滿足 $2a \leq 9\pi$；由於 $d > 0$，這與第 (ii) 部分的結論吻合。

驗算

- 在第 (i) 部分，取 $a = 1$，均值為 $\frac{5}{3(2+\pi)} \approx 0.3242$。以 2×10^5 個樣本進行蒙地卡羅模擬（Monte Carlo simulation），得出 0.3236。
- 在臨界情況 $2a = 9\pi$ 下，第 (ii) 部分的公式給出 $d = 0$，與條件一致。取 $a = 20$，公式給出 $d = \frac{9\pi - 40}{20} \approx -0.586$，而以數值方法解 $\mathrm{P}(X > d) = \frac{1}{10}$ 亦得出相同的值。
- 在第 (iii) 部分，取 $a = 2\pi - 5$，以辛普森法則（Simpson's rule）計算得 $\mathrm{P}(X > \sqrt{2}) = 0.1000$。

常見錯誤

- 沒有求出 k。歸一化（normalisation）條件 $2ka + k\pi = 1$ 是隱藏的第一步，而所有答案都只應以 a 表示。
- 在第 (ii) 部分漏掉四分之一圓，只寫 $\mathrm{P}(X > d) = -2kd$。其實 $[0, 2]$ 上的全部概率，同樣位於負數 d 的右方。
- 把 $\int_{-a}^{0} 2kx\,\mathrm{d}x$ 的正負號弄錯。由於整個區間上 $x < 0$，這個積分是負數。

STEP 之外

d 這個值就是 X 的上 10% 點，即 0.9 分位數（quantile）。要求出它，就要求累積分佈函數（cumulative distribution function）F 的反函數；而這個反函數，正正是用均勻分佈的隨機變量去模擬另一個隨機變量時所用的工具（逆變換抽樣（inverse transform sampling））。這個密度函數形狀混合，一邊平坦、一邊彎曲，說明了為何分位數往往要分段求出。Grimmett 和 Welsh 所著的 *Probability: An Introduction*（《概率導論》）對分佈函數和分位數有仔細的論述。

詳細解答

STEP II 2011，第 1 題

純數學 • 含平方根的方程

STEP II 2011，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/11-s2-q1

關鍵思路

兩部分都涉及平方根之和。只要找出定義域(domain)、端點的值和對稱性(symmetry)，這些函數的圖像便很容易理解。在第 (i) 部分，圖像給出一個不等式（inequality）：曲線永遠不會低於某個高度，而直線只能在一點達到該高度，所以根本不用解方程。在第 (ii) 部分，圖像告訴我們應該預期有多少個根，從而讓我們剔除平方時引入的多餘根。

解答

第 (i) 部分

設 $f(x)=\sqrt{1-x}+\sqrt{3+x}$。兩個平方根只有在 $1-x\geq 0$ 及 $3+x\geq 0$ 時才是實數，所以定義域為 $-3\leq x\leq 1$。在兩個端點，$f(-3)=2+0=2$ 及 $f(1)=0+2=2$。

令 $x=-1+t$，得 $f=\sqrt{2-t}+\sqrt{2+t}$。以 $-t$ 代替 t 時此式不變，所以曲線關於 $x=-1$ 對稱。求導數（derivative），

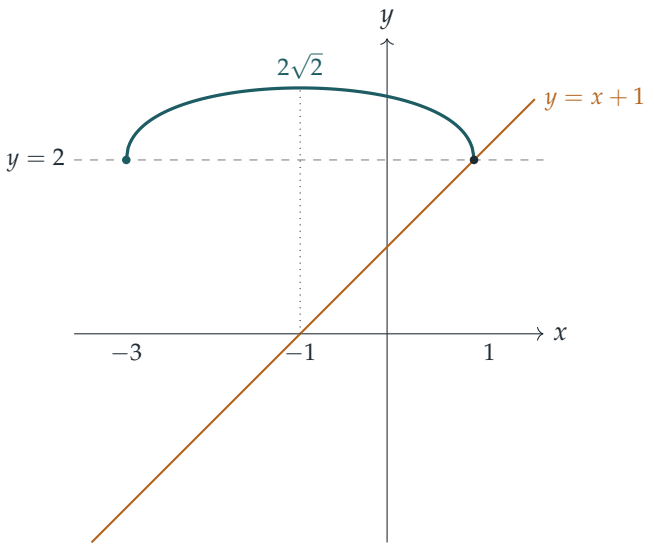
$$f'(x)=-\frac{1}{2\sqrt{1-x}}+\frac{1}{2\sqrt{3+x}},$$

它只在 $1-x=3+x$，即 $x=-1$ 時為零，此時 $f(-1)=2\sqrt{2}$。當 $x<-1$ 時，$3+x<1-x$，所以 $f'(x)>0$；當 $x>-1$ 時，$f'(x)<0$。因此 f 由 2 上升至極大值(maximum) $2\sqrt{2}$，再回落至 2。在 $x=-3$ 附近，f' 的第二項趨於無限大，所以曲線以垂直方向離開 $(-3,2)$；同理，它亦以垂直方向到達 $(1,2)$。f 的每一項都是凹函數（concave），所以整條曲線是一頂平滑的「帽子」。

要看出最小值，更簡潔的方法是平方：

$$f(x)^2=(1-x)+(3+x)+2\sqrt{(1-x)(3+x)}=4+2\sqrt{(1-x)(3+x)}\ \geq\ 4,$$

所以在整個定義域上 $f(x) \geq 2$，等號只在 $x = -3$ 及 $x = 1$ 時成立。



現在引入直線 $y = x + 1$。在定義域上 $x \leq 1$，所以 $x + 1 \leq 2$，等號只在 $x = 1$ 時成立。另一方面 $f(x) \geq 2$。因此 $f(x) = x + 1$ 迫使兩邊都等於 2，而這只在 $x = 1$ 時發生。驗證：$\sqrt{0} + \sqrt{4} = 2 = 1 + 1$。所以恰好有一個實數值滿足方程：

$$x = 1.$$

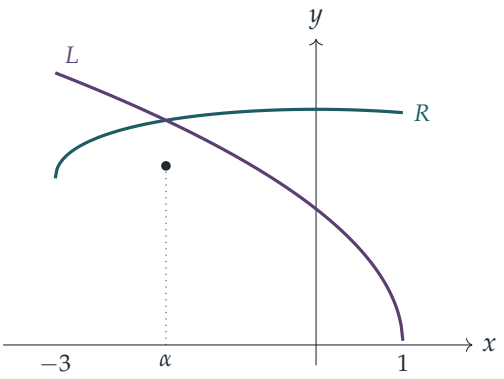
（在 $[-3, 1]$ 以外，左邊沒有定義，所以沒有其他候選值。）

第 (ii) 部分

同樣，我們需要 $x \leq 1$、$x \geq -3$ 及 $x \leq 3$，所以定義域為 $[-3, 1]$。比較

$$L(x) = 2\sqrt{1 - x} \qquad \text{和} \qquad R(x) = \sqrt{3 + x} + \sqrt{3 - x}.$$

L 嚴格遞減（strictly decreasing），由 $L(-3) = 4$ 降至 $L(1) = 0$。R 的形狀與第 (i) 部分的曲線相同：它關於 $x = 0$ 對稱（互換 x 與 $-x$），$R(\pm 3) = \sqrt{6}$，極大值為 $R(0) = 2\sqrt{3}$，在 $[-3, 0]$ 上遞增，在 $[0, 3]$ 上遞減。在我們的定義域上，只看到由 $x = -3$ 至 $x = 1$ 的一段，其中 $R(1) = 2 + \sqrt{2}$。



在 $x=-3$，$L=4>\sqrt{6}=R$；在 $x=0$，$L=2<2\sqrt{3}=R$。在 $[-3,0]$ 上，L 嚴格遞減而 R 遞增，所以 $L-R$ 嚴格遞減，並恰好變號一次。在 $[0,1]$ 上，$L\le 2$ 而 $R\ge 2+\sqrt{2}$，所以兩者不相交。因此方程**恰好有一個**實數解，且它位於 $-3<x<0$。

要求出這個解，就要平方。兩邊都是非負的，所以平方不會損失任何東西：

$$\begin{aligned} & 4(1-x)=(3+x)+(3-x)+2\sqrt{9-x^2} \\ \iff\quad & -1-2x=\sqrt{9-x^2}. \end{aligned}$$

想一想 再次平方時，新方程的哪些解才是真正的解？

右邊是平方根，所以它 ≥ 0；真正的解必須滿足 $-1-2x\ge 0$，即 $x\le -\frac{1}{2}$。在這個條件下，平方是可逆的：

$$1+4x+4x^2=9-x^2 \iff 5x^2+4x-8=0 \iff x=\frac{-2\pm 2\sqrt{11}}{5}.$$

根 $\frac{-2+2\sqrt{11}}{5}\approx 0.93$ 違反 $x\le -\frac{1}{2}$，所以我們捨去它。另一個根 $\frac{-2-2\sqrt{11}}{5}\approx -1.73$ 滿足條件，並位於 $[-3,0]$ 內，正如圖像所預測。所以唯一解是

$$x=-\frac{2+2\sqrt{11}}{5}.$$

驗算

- 在 $x\approx -1.72665$：$2\sqrt{1-x}\approx 3.3025$，而 $\sqrt{3+x}+\sqrt{3-x}\approx 1.1284+2.1741\approx 3.3025$。兩邊相符。
- 被捨去的根 $x\approx 0.92665$ 給出 $L\approx 0.542$，但 $R\approx 3.42$。它其實是 $\sqrt{3+x}-\sqrt{3-x}=2\sqrt{1-x}$ 的解，而這條方程平方後得出同一結果。
- 在第 (i) 部分，$f(-1)=2\sqrt{2}\approx 2.83>0=(-1)+1$，與直線除了在 $(1,2)$ 外處處位於曲線下方一致。

常見錯誤

- 在第 (i) 部分用平方和代數方法硬解。這既繁複，又可能引入增根（extraneous root），而界 $f \geq 2 \geq x+1$ 卻能立即解決問題。
- 保留 $5x^2+4x-8=0$ 的兩個根。每一步平方都需要附帶其符號條件，這裏是 $-1-2x \geq 0$。
- 忘記定義域：對於 $x>1$，兩條方程都沒有意義。

STEP 之外

第 (i) 部分是分析（analysis）中一個常見技巧的小例子：與其解 $f(x)=g(x)$，不如證明 $f \geq c \geq g$，再看等號何時成立。第 (ii) 部分在求出任何根之前，就結合介值定理（intermediate value theorem）與單調性（monotonicity）來數算根的個數，這正是大學第一門分析課程中證明存在性與唯一性（existence and uniqueness）的方法。Abbott 的 *Understanding Analysis*（《理解分析》）對連續性（continuity）與介值定理有細緻的論述。

詳細解答

STEP II 2011，第 2 題

純數學 • 丟番圖方程

> **STEP II 2011，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/11-s2-q2

關鍵思路

每當見到 x^3+y^3，就應把它因式分解（factorise）：$x^3+y^3=(x+y)(x^2-xy+y^2)$。一旦 $x+y$ 被確定（在 (i) 中為 k，在 (ii) 中為 z^2），因式 $x+y$ 便可與 kz^3 的一部分約去，剩下一個二次條件。對這個二次式配方（completing the square），可把它化成「某式 $=3\times$ 完全平方數（perfect square）」的形式，再利用上下界（bounds），便只剩下少數幾個立方數需要逐一嘗試。這正是題目先要求列出不超過 10^3 的立方數的原因。

以供參考：$1, 8, 27, 64, 125, 216, 343, 512, 729, 1000$。

解答

第 (i) 部分

由於 $x+y=k$，

$$kz^3=x^3+y^3=(x+y)(x^2-xy+y^2)=k(x^2-xy+y^2),$$

而 $k>0$，故 $z^3=x^2-xy+y^2$。代入 $y=k-x$：

$$\begin{aligned} z^3 &= x^2-x(k-x)+(k-x)^2 \\ &= x^2-kx+x^2+k^2-2kx+x^2 \\ &= k^2-3kx+3x^2, \end{aligned}$$

得證。

完全平方數。兩邊乘以 4 並配方：

$$4z^3-k^2=12x^2-12kx+3k^2=3\left(4x^2-4kx+k^2\right)=3(k-2x)^2.$$

所以 $\dfrac{4z^3-k^2}{3}=(k-2x)^2$，即整數 $k-2x$ 的平方。

上下界。由上式，$4z^3=k^2+3(k-2x)^2\geq k^2$，故 $\frac{1}{4}k^2\leq z^3$。至於上界，注意 $3kx-3x^2=3x(k-x)=3xy$，所以

$$z^3=k^2-3xy<k^2,$$

因為 x 和 y 都是正數。因此 $\frac{1}{4}k^2\leq z^3<k^2$。

$k=20$ 的情況。我們尋找滿足 $x+y=20$ 的解。上下界給出 $100\leq z^3<400$，從立方數列表可知 $z^3\in\{125,216,343\}$。接着檢驗 $(4z^3-400)/3$ 是否為完全平方數：

- $z=5$：$(500-400)/3=100/3$，不是整數；
- $z=6$：$(864-400)/3=464/3$，不是整數；
- $z=7$：$(1372-400)/3=324=18^2$。✓

所以在 $k=20$ 時 $(k-2x)^2=18^2$。由於 $x<y$ 意味着 $2x<x+y=k$，故 $k-2x>0$，因此 $k-2x=18$，得 $x=1$、$y=19$。驗算：$1^3+19^3=1+6859=6860=20\times343=20\times7^3$。所以

$$x=1,\quad y=19,\quad z=7$$

是 $(*)$ 在 $k=20$ 時的一個解。

第 (ii) 部分

現設 $x+y=z^2$。則

$$kz^3=(x+y)(x^2-xy+y^2)=z^2(x^2-xy+y^2),$$

兩邊除以 z^2（由於 $z>0$，這是容許的），得 $x^2-xy+y^2=kz$。

> ? **想一想** 怎樣才能令同樣的「完全平方」技巧在這裏奏效？

利用恆等式 $4(x^2-xy+y^2)=(x+y)^2+3(x-y)^2$，這其實就是 (i) 中所得結果的另一種面貌（展開右邊即可驗證）。代入 $x+y=z^2$ 和 $k=19$，得

$$76z=z^4+3(y-x)^2,\qquad \text{即}\qquad 3(y-x)^2=z\,(76-z^3).$$

由於 $y>x$，左邊為正，故 $z^3<76$，因此 $z\in\{1,2,3,4\}$。我們需要 $z(76-z^3)/3$ 為完全平方數：

- $z=1$：$75/3=25=5^2$，所以 $y-x=5$；但 $x+y=1$，對正整數而言不可能。

- $z=2$：$2\times 68/3=136/3$，不是整數。
- $z=3$：$3\times 49/3=49=7^2$，所以 $y-x=7$ 且 $x+y=9$，得 $x=1$、$y=8$。
- $z=4$：$4\times 12/3=16=4^2$，所以 $y-x=4$ 且 $x+y=16$，得 $x=6$、$y=10$。

驗算：$1^3+8^3=513=19\times 27=19\times 3^3$，以及 $6^3+10^3=1216=19\times 64=19\times 4^3$。所以 $(*)$ 在 $k=19$ 時的兩個解為

$$(x,y,z)=(1,8,3)\quad 及\quad (x,y,z)=(6,10,4).$$

（事實上，上述論證顯示在 $x+y=z^2$ 的條件下，只有這兩個解。）

驗算

- 用一個簡單例子檢驗 (i) 的恆等式：$k=5$、$x=2$、$y=3$ 給出 $x^3+y^3=35=5\times 7$，而 $k^2-3kx+3x^2=25-30+12=7$。✓
- 如上文所做，把每個答案代回 $(*)$：$6860=20\times 343$、$513=19\times 27$、$1216=19\times 64$。
- 用電腦搜尋所有滿足 $x+y=20$ 的 $x<y$，以及在 $k=19$ 時所有滿足 $x+y=z^2$ 的 $x<y$，結果正好找到這些解，再無其他。

常見錯誤

- 忘記正負號：$(k-2x)^2=18^2$ 也容許 $k-2x=-18$，但這會給出 $x=19>y$，被條件 $x<y$ 排除。
- 在 (ii) 中把 $z=1$ 保留為解：$y-x=5$ 本身沒有問題，但它不能與 $x+y=1$ 同時成立。

STEP 之外

像 $x^3+y^3=kz^3$ 這類方程，問的是哪些整數 k 可以寫成兩個有理數（rational number）立方之和。這個問題直接通向橢圓曲線（elliptic curves），至今仍未完全解決；著名的的士數（taxicab number）$1729=1^3+12^3=9^3+10^3$ 也屬於同一領域。這裏用到的技巧（因式分解、配方，再利用上下界把搜尋範圍收窄至有限）是初等數論（elementary number theory）的基本功。Davenport 的 *The Higher Arithmetic*（《高等算術》）是很好的下一步讀物。

詳細解答

STEP II 2011，第 3 題

純數學 • 以單調性證明不等式

> **STEP II 2011，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/11-s2-q3

關鍵思路

這裏每一條不等式（inequality）都源自同一個原則：如果一個函數（function）從零出發，而其導數（derivative）非負，那麼它會一直保持非負。最後兩個結果再加了一點變化。一旦知道某函數是遞增（increasing）的，我們便可以代入兩個早已知道大小次序的輸入值（例如 $x \leq \arcsin x$），從而讀出輸出值之間的不等式。看出要比較哪兩個輸入值，就是整個技巧所在。

解答

第 (i)(a) 部分

求導，並對 $x\cos x$ 使用乘積法則（product rule），

$$f'(x) = \cos x - (\cos x - x\sin x) = x\sin x.$$

當 $0 \leq x \leq \frac{1}{2}\pi$ 時，$x \geq 0$ 及 $\sin x \geq 0$ 均成立，所以 $f'(x) \geq 0$，因此 f 在此區間上遞增。

故此，當 $0 \leq x \leq \frac{1}{2}\pi$ 時，我們有 $x \geq 0$，所以 $f(x) \geq f(0) = \sin 0 - 0 = 0$，得證。

第 (i)(b) 部分

設 $h(x) = \arcsin x - x$，其中 $0 \leq x < 1$。利用題目給出的事實，

$$h'(x) = \frac{\mathrm{d}}{\mathrm{d}x}(\arcsin x) - 1 \geq 1 - 1 = 0,$$

所以 h 遞增，且 $h(x) \geq h(0) = 0$。即是說，當 $0 \leq x < 1$ 時，$\arcsin x \geq x$。

（題目給出的事實其實就是 $\dfrac{1}{\sqrt{1-x^2}} \geq 1$，這成立是因為 $0 < \sqrt{1-x^2} \leq 1$。）

第 (i)(c) 部分

記 $g(x) = \dfrac{x}{\sin x}$。當 $0 < x < \frac{1}{2}\pi$ 時，由商法則（quotient rule）得

$$g'(x) = \frac{\sin x - x\cos x}{\sin^2 x} = \frac{f(x)}{\sin^2 x}.$$

由 (a) 部分，此處 $f(x) \geq 0$，而 $\sin^2 x > 0$，所以 $g'(x) \geq 0$，因此 g 在 $(0, \frac{1}{2}\pi)$ 上遞增。

想一想 知道 g 是遞增的，對處理一條涉及 $\arcsin x$ 的不等式有甚麼幫助？

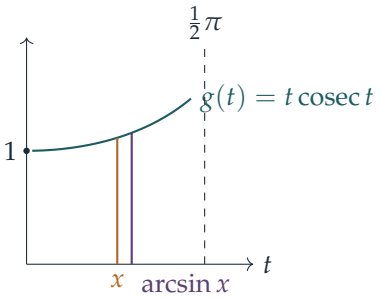
看看目標：$\dfrac{\arcsin x}{x} \geq \dfrac{x}{\sin x}$。右方是 $g(x)$。左方同樣是 g 的一個值。若 $y = \arcsin x$，則 $\sin y = x$，所以

$$g(y) = \frac{y}{\sin y} = \frac{\arcsin x}{x}.$$

因此我們需要證明 $g(\arcsin x) \geq g(x)$；由於 g 遞增，這可由 (b) 部分的 $\arcsin x \geq x$ 推出。

最後，取 $0 < x < 1$，並設 $y = \arcsin x$。於是 $0 < y < \frac{1}{2}\pi$。另外 $0 < x < 1 < \frac{1}{2}\pi$，所以 x 和 y 都位於 g 遞增的區間內。由 (b) 部分，$y \geq x$，所以 $g(y) \geq g(x)$：

$$(\arcsin x)\, x^{-1} = \frac{y}{\sin y} \geq \frac{x}{\sin x} = x \operatorname{cosec} x \qquad (0 < x < 1).$$



上圖展示 $x = 0.8$ 時的論證：輸入值 $\arcsin x$ 位於 x 的右方，所以遞增曲線在那裏較高。

第 (ii) 部分

我們照搬 (i) 的模式。首先是 (b) 的類比。設 $m(x) = x - \arctan x$，其中 $x \geq 0$。由題目給出的事實，$m'(x) = 1 - \dfrac{\mathrm{d}}{\mathrm{d}x}(\arctan x) \geq 0$，所以 $m(x) \geq m(0) = 0$。因此

$$\arctan x \leq x \qquad (x \geq 0).$$

接着是 (c) 的類比。設 $k(x) = x^{-1}\tan x$，其中 $0 < x < \frac{1}{2}\pi$。則

$$k'(x) = \frac{x\sec^2 x - \tan x}{x^2} = \frac{x - \sin x\cos x}{x^2\cos^2 x} = \frac{2x - \sin 2x}{2x^2\cos^2 x},$$

這裏先將分子分母同乘以 $\cos^2 x$（它是正數），再同乘以 2。

想一想 為甚麼 $2x - \sin 2x \geq 0$？

這就是熟悉的「當 $t \geq 0$ 時 $\sin t \leq t$」，取 $t = 2x$ 即可。它可以用與其他部分相同的方式證明：$p(t) = t - \sin t$ 滿足 $p'(t) = 1 - \cos t \geq 0$ 及 $p(0) = 0$，所以當 $t \geq 0$ 時 $p(t) \geq 0$。

所以 $k'(x)$ 的分子非負，分母為正，從而 $k'(x) \geq 0$。因此 k 在 $(0, \frac{1}{2}\pi)$ 上遞增。

現在取 $0 < x < \frac{1}{2}\pi$，並設 $y = \arctan x$。於是 $0 < y < \frac{1}{2}\pi$，所以 x 和 y 都位於 k 遞增的區間內，且 $\tan y = x$。由於 $y \leq x$，我們得到 $k(y) \leq k(x)$：

$$\frac{x}{\arctan x} = \frac{\tan y}{y} \leq \frac{\tan x}{x}.$$

x 和 $\arctan x$ 都是正數，所以兩邊同乘以 $x\arctan x$ 不會改變不等號的方向：

$$(\tan x)(\arctan x) \geq x^2 \qquad (0 < x < \tfrac{1}{2}\pi),$$

得證。

驗算

- 在 (c) 中取 $x = \frac{1}{2}$：$\arcsin(\frac{1}{2})/\frac{1}{2} = \pi/3 \approx 1.0472$，而 $\frac{1}{2}\operatorname{cosec}\frac{1}{2} \approx 1.0429$。不等式成立，而且只是勉強成立，這正符合預期：當 $x \to 0$ 時兩邊都趨向 1。
- 在 (ii) 中取 $x = 1$：$\tan 1 \cdot \arctan 1 = \frac{1}{4}\pi\tan 1 \approx 1.2232 \geq 1$。
- 當 x 很小時，$\tan x \approx x + \frac{1}{3}x^3$ 及 $\arctan x \approx x - \frac{1}{3}x^3$，所以兩者之積為 $x^2 + O(x^6)$：(ii) 中的不等式在 0 附近極之緊。我們在 20,000 個隨機點作數值驗算，確認了上述每一條不等式和每一條導數公式。

常見錯誤

- 在 (c) 中比較 g 在錯誤的一對點上的值，或忘記檢查 $\arcsin x$ 是否位於已知 g 遞增的區間 $(0, \frac{1}{2}\pi)$ 之內。
- 在 (ii) 中，輸入值之間的不等式方向相反（$\arctan x \leq x$），所以輸出值之間的不等式方向也相反。若把方向弄反，就會得出錯誤的不等式。
- 將不等式兩邊同乘以某數時，沒有說明該乘數（$x\arctan x$ 或 $\cos^2 x$）是正數。

STEP 之外

貫穿全題的原則——非負導數迫使函數遞增——是均值定理（mean value theorem）的推論，任何分析（analysis）入門課程都會證明這條定理。比較 $g(x)$ 與 $g(\arcsin x)$ 的技巧（其中 sin 與 arcsin 互為反函數），與你日後在反函數（inverse functions）和凸性（convexity）中會遇到的想法一脈相承。Abbott 的 *Understanding Analysis*（《理解分析》）對均值定理及其推論給出了嚴謹而易讀的證明。

詳細解答

STEP II 2011，第 4 題

純數學 • 三角方程

> **STEP II 2011，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/11-s2-q4

關鍵思路

把 $\cos\theta$ 寫成某個相關角的正弦（或餘弦），這樣「$\cos\theta = \sin 4\theta$」就變成兩個餘弦相等的方程，可以精確求解。將 $\sin 4\theta$ 展開成 $\sin\theta$ 的冪，會得到一個三次方程（cubic），其根正是這些角的正弦值，像 $\sin 18^\circ$ 這樣的根式（surd）就是這樣出現的。第 (ii) 部分是一個關於 $\sin^2 x$ 的二次方程（quadratic）；而第 (iii) 部分只要找到一個正弦平方等於 $\frac{1}{4}$ 的角，就會發現它其實是喬裝了的第 (ii) 部分。

解答

第 (i) 部分

為了把同類的東西放在一起比較，寫 $\sin 4\theta = \cos(90^\circ - 4\theta)$。兩個角的餘弦相等，當且僅當它們在相差 360° 的倍數之下相等或互為相反數，因此

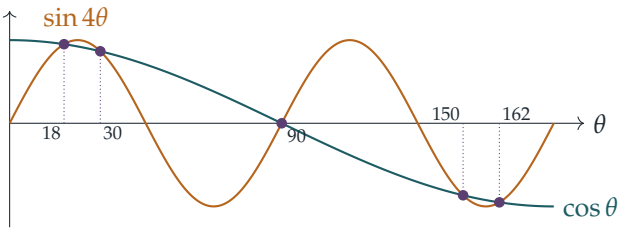
$$\cos\theta = \cos(90^\circ - 4\theta) \iff \theta = \pm(90^\circ - 4\theta) + 360^\circ k \quad \text{對某整數} k.$$

- 取 $+$ 號：$5\theta = 90^\circ + 360^\circ k$，所以 $\theta = 18^\circ + 72^\circ k$。在 $0^\circ < \theta < 180^\circ$ 內得 $\theta = 18^\circ, 90^\circ, 162^\circ$。
- 取 $-$ 號：$3\theta = 90^\circ - 360^\circ k$，所以 $\theta = 30^\circ - 120^\circ k$，得 $\theta = 30^\circ, 150^\circ$。

因此解為

$$\theta = 18^\circ, 30^\circ, 90^\circ, 150^\circ, 162^\circ.$$

下圖證實兩條曲線恰好相交五次。



接下來處理「由此」的部分。設 $s = \sin\theta$。兩次運用倍角公式（double-angle formulae），

$$\sin 4\theta = 2\sin 2\theta\cos 2\theta = 4\sin\theta\cos\theta\,(1 - 2\sin^2\theta),$$

於是方程 $\cos\theta = \sin 4\theta$ 變成

$$\cos\theta\big(1 - 4s(1 - 2s^2)\big) = 0.$$

想一想 為甚麼可以約去 $\cos\theta$？

在五個解之中，只有 $\theta = 90°$ 使 $\cos\theta = 0$。其餘四個解 $18°, 30°, 150°, 162°$ 都有 $\cos\theta \neq 0$，所以對它們可以除以 $\cos\theta$，得

$$8s^3 - 4s + 1 = 0.$$

我們知道 $s = \sin 30° = \frac{1}{2}$ 是其中一個根，因式分解（factorising）得

$$8s^3 - 4s + 1 = (2s - 1)(4s^2 + 2s - 1).$$

$\sin 18°$ 是這個三次方程的根，而由於正弦在 $[0°, 90°]$ 上遞增，$0 < \sin 18° < \sin 30° = \frac{1}{2}$。所以 $\sin 18° \neq \frac{1}{2}$，它必定是 $4s^2 + 2s - 1 = 0$ 的根，而這個方程的根為

$$s = \frac{-2 \pm \sqrt{4 + 16}}{8} = \frac{-1 \pm \sqrt{5}}{4}.$$

只有取 $+$ 號的根是正數，所以

$$\sin 18° = \tfrac{1}{4}(\sqrt{5} - 1),$$

得證。

第 (ii) 部分

設 $s = \sin x$。由於 $\sin^2 2x = 4\sin^2 x\cos^2 x = 4s^2(1 - s^2)$，方程只取決於 s：

$$4s^2 + 1 = 16s^2(1 - s^2) \iff 16s^4 - 12s^2 + 1 = 0.$$

這是關於 s^2 的二次方程：

$$s^2 = \frac{12 \pm \sqrt{144 - 64}}{32} = \frac{12 \pm 4\sqrt{5}}{32} = \frac{3 \pm \sqrt{5}}{8}.$$

要開平方根，留意

$$\left(\frac{1+\sqrt{5}}{4}\right)^2 = \frac{6+2\sqrt{5}}{16} = \frac{3+\sqrt{5}}{8}, \qquad \left(\frac{\sqrt{5}-1}{4}\right)^2 = \frac{6-2\sqrt{5}}{16} = \frac{3-\sqrt{5}}{8}.$$

s^2 的兩個值都介乎 0 與 1 之間，所以四個平方根每一個都確實是 $\sin x$ 可取的值；又因為以上每一步都可逆，它們都滿足原方程。可能的值為

$$\sin x = \tfrac{1}{4}+\tfrac{1}{4}\sqrt{5}, \qquad -\tfrac{1}{4}-\tfrac{1}{4}\sqrt{5}, \qquad -\tfrac{1}{4}+\tfrac{1}{4}\sqrt{5}, \qquad \tfrac{1}{4}-\tfrac{1}{4}\sqrt{5}.$$

由第 (i) 部分，$\frac{1}{4}(\sqrt{5}-1) = \sin 18°$；而 $\frac{1}{4}(1+\sqrt{5}) = \sin 54°$（它等於 $\cos 36° = 1 - 2\sin^2 18°$，你可以驗證兩者結果相同）。

第 (iii) 部分

把第 (ii) 部分的方程除以 4：

$$\sin^2 x + \tfrac{1}{4} = \sin^2 2x.$$

將它與 $\sin^2 3\alpha + \sin^2 5\alpha = \sin^2 6\alpha$ 比較。若設 $x = 3\alpha$，則自然有 $2x = 6\alpha$，而只要

$$\sin^2 5\alpha = \tfrac{1}{4} \qquad \text{及} \qquad \sin 3\alpha \text{ 是第 (ii) 部分四個值之一，}$$

兩條方程便吻合。

想一想 $\sin 3\alpha$ 是那些值之一就足夠了嗎？

足夠：在第 (ii) 部分我們看到，方程 $\sin^2 x + \frac{1}{4} = \sin^2 2x$ 只透過 $s = \sin x$ 依賴於 x，而那四個 s 值都滿足它。所以任何正弦取這些值之一的 x 都可行。

取 $\alpha = 6°$。則 $5\alpha = 30°$，所以 $\sin^2 5\alpha = \frac{1}{4}$；而 $3\alpha = 18°$，所以 $\sin 3\alpha = \frac{1}{4}(\sqrt{5}-1)$，正是我們的值之一。因此

$$\sin^2 18° + \sin^2 30° = \sin^2 36°,$$

$\alpha = 6°$ 是一個解。

取 $\alpha = 42°$。則 $5\alpha = 210°$，所以 $\sin 5\alpha = -\frac{1}{2}$，$\sin^2 5\alpha = \frac{1}{4}$；而 $3\alpha = 126°$，所以 $\sin 3\alpha = \sin 54° = \frac{1}{4}(1+\sqrt{5})$，同樣是我們的值之一。所以 $\alpha = 42°$ 是一個解。

因此兩個合適的值是 $\alpha = 6°$ 和 $\alpha = 42°$。

同樣的方法亦給出 $\alpha = 66°$（$5\alpha = 330°$，$\sin 3\alpha = \sin 198° = -\sin 18°$）及 $\alpha = 78°$（$5\alpha = 390°$，$\sin 3\alpha = \sin 234° = -\sin 54°$）。範圍內另一個滿足 $\sin^2 5\alpha = \frac{1}{4}$ 的角是 $\alpha = 30°$，但它不成立，因為 $\sin 90° = 1$ 不是我們的值之一。（事實上 $1 + \frac{1}{4} \neq \sin^2 180° = 0$。）

驗算

▶ $\sin 18° \approx 0.30902$，而 $\frac{1}{4}(\sqrt{5}-1) \approx 0.30902$。

- 對 $\alpha = 6°$：$\sin^2 18° + \sin^2 30° \approx 0.09549 + 0.25 = 0.34549$，而 $\sin^2 36° \approx 0.34549$。
- 在 $(0°, 90°)$ 上以數值方法掃描 $\sin^2 3\alpha + \sin^2 5\alpha - \sin^2 6\alpha$，找到的零點恰好是 $6°, 42°, 45°, 66°, 78°$。多出來的 $45°$ 並非來自第 (ii) 部分：在該處 $\frac{1}{2} + \frac{1}{2} = 1$。

常見錯誤

- 忘記 $\cos A = \cos B$ 的第二組解（即 $A = -B + 360°k$），因而在第 (i) 部分漏掉 $30°$ 和 $150°$。
- 不加說明就約去 $\cos\theta$。這裏這樣做沒問題，但只是因為使 $\cos\theta = 0$ 的解是 $90°$，而不是 $18°$。
- 在第 (ii) 部分只給出 $\sin x$ 的正值。四個值全部都是可能的。

STEP 之外

$\sin 18°$ 之所以有精確值，是因為 $18°$ 是直角的五分之一，因此它與正五邊形（regular pentagon）及黃金比例（golden ratio）息息相關：$\frac{1}{4}(1 + \sqrt{5}) = \frac{1}{2}\varphi$。更一般地，$\cos(360°/n)$ 能只用平方根表示，當且僅當正 n 邊形可以用直尺和圓規作出（尺規作圖，ruler and compasses）；高斯（Gauss）證明了這取決於整除 n 的費馬質數（Fermat primes）。這是域論（field theory）和伽羅瓦理論（Galois theory）的經典入門應用。Allenby 的 *Rings, Fields and Groups*（《環、域與羣》）逐步建立所需的代數，而 Needham 的 *Visual Complex Analysis*（《直觀複分析》）則說明為甚麼單位根（roots of unity）與多邊形密不可分。

詳細解答

STEP II 2011，第 5 題

純數學 • 向量與反射

STEP II 2011，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/11-s2-q5

關鍵思路

要把一點關於一條經過原點的直線作反射（reflection），可把它的位置向量（position vector）分解為沿該直線的部分和與該直線垂直的部分：反射保留前者，並把後者反向。$\mathbf{b}$ 沿 $\mathbf{a}$ 方向的分量（component）就是我們熟悉的投影（projection）$\frac{\mathbf{a}\cdot\mathbf{b}}{\mathbf{a}\cdot\mathbf{a}}\mathbf{a}$，其餘的都只是點積（dot product）的運算整理。至於共線（collinear）問題，記住：一點位於直線 AB 上，當且僅當它的位置向量可寫成 $s\mathbf{a}+t\mathbf{b}$，其中 $s+t=1$。

解答

整個解答中，$\mathbf{a}$ 和 $\mathbf{b}$ 均非零且互不平行，因為 O、A、B 不共線。特別地，$\mathbf{a}\cdot\mathbf{a}\neq 0$ 且 $\mathbf{b}\cdot\mathbf{b}\neq 0$，所以可以除以它們。記 $\theta=\angle AOB$。

反射點 C

設 P 為從 B 到直線 OA 的垂足（foot of the perpendicular）。它的位置向量是某個倍數 $k\mathbf{a}$，而 $\mathbf{b}-k\mathbf{a}$ 與 $\mathbf{a}$ 垂直：

$$(\mathbf{b}-k\mathbf{a})\cdot\mathbf{a}=0 \quad\Longrightarrow\quad k=\frac{\mathbf{a}\cdot\mathbf{b}}{\mathbf{a}\cdot\mathbf{a}}.$$

關於直線 OA 的反射保持 P 不動，並把 B 映到另一側的點 C，使 P 為 BC 的中點。所以 $\mathbf{c}=2\overrightarrow{OP}-\mathbf{b}$，即

$$\mathbf{c}=\frac{2\,\mathbf{a}\cdot\mathbf{b}}{\mathbf{a}\cdot\mathbf{a}}\mathbf{a}-\mathbf{b}=\lambda\mathbf{a}-\mathbf{b},\qquad \lambda=\frac{2\,\mathbf{a}\cdot\mathbf{b}}{\mathbf{a}\cdot\mathbf{a}},$$

即為所求。

反射點 D

以直線 OB 代替 OA，同樣的論證給出

$$\mathbf{d} = \frac{2\,\mathbf{b}\cdot\mathbf{c}}{\mathbf{b}\cdot\mathbf{b}}\,\mathbf{b} - \mathbf{c}.$$

而 $\mathbf{b}\cdot\mathbf{c} = \lambda\,\mathbf{a}\cdot\mathbf{b} - \mathbf{b}\cdot\mathbf{b}$，所以

$$\begin{aligned}\mathbf{d} &= \left(\frac{2\lambda\,\mathbf{a}\cdot\mathbf{b}}{\mathbf{b}\cdot\mathbf{b}} - 2\right)\mathbf{b} - (\lambda\mathbf{a} - \mathbf{b}) \\ &= \left(\frac{2\lambda\,\mathbf{a}\cdot\mathbf{b}}{\mathbf{b}\cdot\mathbf{b}} - 1\right)\mathbf{b} - \lambda\mathbf{a}.\end{aligned}$$

這就是 $\mathbf{d} = \mu\mathbf{b} - \lambda\mathbf{a}$，其中

$$\mu = \frac{2\lambda\,\mathbf{a}\cdot\mathbf{b}}{\mathbf{b}\cdot\mathbf{b}} - 1 = \frac{4(\mathbf{a}\cdot\mathbf{b})^2}{(\mathbf{a}\cdot\mathbf{a})(\mathbf{b}\cdot\mathbf{b})} - 1 = 4\cos^2\theta - 1,$$

最後一步用了 $\mathbf{a}\cdot\mathbf{b} = |\mathbf{a}||\mathbf{b}|\cos\theta$。下面需要用到的正是 $4\cos^2\theta - 1$ 這個形式。

A、B、D 共線

點 D 位於直線 AB 上，當且僅當對某個純量（scalar）t 有 $\mathbf{d} = \mathbf{a} + t(\mathbf{b} - \mathbf{a}) = (1-t)\mathbf{a} + t\mathbf{b}$。另一方面，我們亦有 $\mathbf{d} = -\lambda\mathbf{a} + \mu\mathbf{b}$。

> ? **想一想** 為甚麼可以比較 $\mathbf{a}$ 和 $\mathbf{b}$ 的係數（coefficient）？

因為 $\mathbf{a}$ 和 $\mathbf{b}$ 互不平行。若 $s\mathbf{a} + t\mathbf{b} = s'\mathbf{a} + t'\mathbf{b}$，則 $(s - s')\mathbf{a} = (t' - t)\mathbf{b}$；除非兩個括號都等於零，否則這會使 $\mathbf{a}$ 和 $\mathbf{b}$ 平行。這正是用到 O、A、B 不共線的地方。

所以 $-\lambda = 1 - t$ 且 $\mu = t$。消去 t，可知 A、B、D 共線當且僅當

$$\mu = \lambda + 1.$$

$\lambda = -\frac{1}{2}$ 的情況

此時共線條件給出 $\mu = \frac{1}{2}$，所以 $4\cos^2\theta - 1 = \frac{1}{2}$，即 $\cos^2\theta = \frac{3}{8}$。由於 $\lambda < 0$，我們有 $\mathbf{a}\cdot\mathbf{b} < 0$，所以 $\cos\theta < 0$，故

$$\cos\angle AOB = -\sqrt{\tfrac{3}{8}} = -\frac{\sqrt{6}}{4}.$$

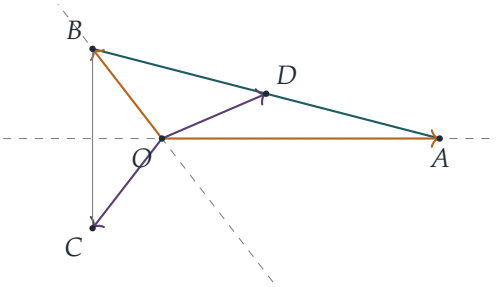
於是

$$\mathbf{d} = \tfrac{1}{2}\mathbf{b} + \tfrac{1}{2}\mathbf{a} = \tfrac{1}{2}(\mathbf{a} + \mathbf{b}),$$

所以 D **是** AB **的中點**。關於圖形的形狀，我們還可以說得更多。由 $\lambda = \dfrac{2|\mathbf{b}|\cos\theta}{|\mathbf{a}|} = -\frac{1}{2}$，

$$\frac{|\mathbf{a}|}{|\mathbf{b}|} = -4\cos\theta = \sqrt{6},$$

所以 $OA = \sqrt{6}\,OB$，而角 AOB 是鈍角（obtuse angle）（約 127.8°）。此外，反射保持與 O 的距離不變，所以 $OD = OC = OB$：AB 的中點到 O 的距離恰好與 B 相同。



驗算

- 先關於 OA、再關於 OB 反射兩次，等同於繞 O 旋轉（rotation）2θ。所以 D 應是 B 旋轉 2θ 後的位置，且 $\cos\angle BOD = \cos 2\theta = 2\cos^2\theta - 1$。在最後的情況中，這等於 $2\cdot\frac{3}{8} - 1 = -\frac{1}{4}$。直接計算，$\mathbf{b}\cdot\frac{1}{2}(\mathbf{a}+\mathbf{b}) = \frac{1}{2}(-\frac{3}{2}+1)|\mathbf{b}|^2 = -\frac{1}{4}|\mathbf{b}|^2$，其中用了 $\mathbf{a}\cdot\mathbf{b} = -\frac{1}{4}\,\mathbf{a}\cdot\mathbf{a} = -\frac{3}{2}\,\mathbf{b}\cdot\mathbf{b}$，結果吻合。
- 若 $\theta = 90°$，則 $\lambda = 0$，C 是 $-\mathbf{b}$，而 D 亦是 $-\mathbf{b}$。公式給出 $\mu = 4\cdot 0 - 1 = -1$，與預期相符。
- 數值上，取 $\mathbf{a} = (\sqrt{6}, 0)$ 及 $\mathbf{b} = (-0.6124,\ 0.7906)$，反射兩次得 $\mathbf{d} = (0.9186,\ 0.3953)$，這正是 AB 的中點，而且 $|\mathbf{d}| = 1 = |\mathbf{b}|$。

常見錯誤

- 把反射寫成 $\mathbf{b} - 2\times$（投影）。這其實是關於與 OA 垂直的直線作反射。可用 $\mathbf{b} = \mathbf{a}$ 來測試：它應保持不動。
- 取 $\cos\theta = +\sqrt{6}/4$。λ 的正負決定了 $\mathbf{a}\cdot\mathbf{b}$ 的正負。
- 比較係數時沒有說明 $\mathbf{a}$ 和 $\mathbf{b}$ 線性無關（linearly independent）。

STEP 之外

映射 $\mathbf{x} \mapsto \frac{2\,\mathbf{a}\cdot\mathbf{x}}{\mathbf{a}\cdot\mathbf{a}}\,\mathbf{a} - \mathbf{x}$ 是一個線性映射（linear map），它的矩陣是大學一年級課程中行列式（determinant）為 -1 的正交矩陣（orthogonal matrix）的標準例子。兩次反射的複合是一個旋轉，旋轉角為兩面鏡子之間夾角的兩倍。這是研究等距變換（isometry）以及由反射生成的羣（group）——例如正多邊形的對稱羣——的起點。Strang 的 *Introduction to Linear Algebra*（《線性代數導論》）對投影和反射矩陣有很好的講解。

詳細解答

STEP II 2011，第 6 題

純數學 • 分部積分法

STEP II 2011，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/11-s2-q6

關鍵思路

被積函數（integrand）$[f'(x)]^2[f(x)]^n$ 只差一個 f' 的因子，就會變成一眼便能積出來的形式，因為 $f'(x)[f(x)]^n$ 正是 $[f(x)]^{n+1}/(n+1)$ 的導數（derivative）。所以我們用分部積分（integration by parts），把這一部分拆出來。剩下的積分含有 f''，而條件 $f''=kff'$ 又把它變回同樣容易處理的形式 $f'f^m$。至於各個應用，關鍵就在於選取 f 和 n，使所給的被積函數等於 $[f']^2f^n$，再求出 k。

解答

一般結果

以 f 代表 $f(x)$，其餘類推。用分部積分，取 $u=f'$ 及 $\dfrac{\mathrm{d}v}{\mathrm{d}x}=f'f^n$，於是 $\dfrac{\mathrm{d}u}{\mathrm{d}x}=f''$ 及 $v=\dfrac{f^{n+1}}{n+1}$（這沒有問題，因為 $n\geq 1$ 表示 $n+1\neq 0$）：

$$\begin{aligned} I=\int f'\cdot f'f^n\,\mathrm{d}x &= \frac{f'f^{n+1}}{n+1}-\frac{1}{n+1}\int f''f^{n+1}\,\mathrm{d}x \\ &= \frac{f'f^{n+1}}{n+1}-\frac{k}{n+1}\int f'f^{n+2}\,\mathrm{d}x \qquad \text{利用} f''=kff' \\ &= \frac{f'f^{n+1}}{n+1}-\frac{kf^{n+3}}{(n+1)(n+3)}+C. \end{aligned}$$

因此

$$\boxed{I=\frac{f'(x)\,[f(x)]^{n+1}}{n+1}-\frac{k\,[f(x)]^{n+3}}{(n+1)(n+3)}+C}$$

這正是題目要求的、以 $f(x)$、$f'(x)$、k 和 n 表示的式子。

第 (i) 部分

對於 $f(x) = \tan x$，我們有 $f'(x) = \sec^2 x$ 及

$$f''(x) = 2\sec x \cdot \sec x \tan x = 2\tan x \sec^2 x = 2f(x)f'(x),$$

所以條件成立，且 $k = 2$。代入公式得

$$\int \sec^4 x \tan^n x \,\mathrm{d}x = \frac{\sec^2 x \tan^{n+1} x}{n+1} - \frac{2\tan^{n+3} x}{(n+1)(n+3)} + C. \tag{1}$$

驗證。直接計算：利用 $\sec^2 x = 1 + \tan^2 x$ 及代換（substitution）$t = \tan x$，$\mathrm{d}t = \sec^2 x \,\mathrm{d}x$：

$$\int \sec^4 x \tan^n x \,\mathrm{d}x = \int (1+t^2)t^n \,\mathrm{d}t = \frac{t^{n+1}}{n+1} + \frac{t^{n+3}}{n+3} + C.$$

用 $\sec^2 x = 1 + t^2$ 改寫 (1) 的右方，得

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

答案相同。所以對 $f = \tan$ 而言，結果得到驗證。

所求積分。由於

$$\frac{\sin^4 x}{\cos^8 x} = \frac{\sin^4 x}{\cos^4 x} \cdot \frac{1}{\cos^4 x} = \tan^4 x \sec^4 x = [f'(x)]^2[f(x)]^4,$$

這是 (1) 中 $n = 4$ 的情況。利用剛才得到的化簡形式，

$$\int \frac{\sin^4 x}{\cos^8 x}\,\mathrm{d}x = \frac{\tan^5 x}{5} + \frac{\tan^7 x}{7} + C.$$

第 (ii) 部分

想一想 應該取哪一個 f？那個 $\sec^2 x$ 又從何而來？

括號 $(\sec x + \tan x)$ 顯然是 f 的候選，所以試取 $f(x) = \sec x + \tan x$。它的導數是

$$f'(x) = \sec x \tan x + \sec^2 x = \sec x(\tan x + \sec x) = \sec x\, f(x).$$

這就是關鍵事實：對 f 求導，只不過是把它乘以 $\sec x$。因此

$$\begin{aligned} f''(x) &= \sec x \tan x\, f(x) + \sec x\, f'(x) = \sec x \tan x\, f + \sec^2 x\, f \\ &= \sec x\, f\,(\tan x + \sec x) = \sec x\, f^2, \end{aligned}$$

又因為 $f\,f' = f \cdot \sec x\, f = \sec x\, f^2$ 亦成立，所以 $f'' = ff'$，即 $k = 1$。

此外 $[f'(x)]^2 = \sec^2 x\,[f(x)]^2$，所以

$$\sec^2 x\,(\sec x + \tan x)^6 = \sec^2 x\, f^2 \cdot f^4 = [f'(x)]^2[f(x)]^4,$$

這是一般結果中 $n = 4$、$k = 1$ 的情況：

$$\begin{aligned}\int \sec^2 x\,(\sec x + \tan x)^6\,\mathrm{d}x &= \frac{f'f^5}{5} - \frac{f^7}{35} + C = \frac{\sec x\, f^6}{5} - \frac{f^7}{35} + C \\ &= \frac{f^6}{35}\big(7\sec x - (\sec x + \tan x)\big) + C \\ &= \frac{(\sec x + \tan x)^6\,(6\sec x - \tan x)}{35} + C.\end{aligned}$$

驗算

- 對 (i) 的答案求導：$\sec^2 x \tan^4 x + \sec^2 x \tan^6 x = \sec^2 x \tan^4 x(1 + \tan^2 x) = \tan^4 x \sec^4 x$。
- 以數值計算，辛普森法則（Simpson's rule）給出 $\int_{-0.7}^{1.1} \sin^4 x / \cos^8 x\,\mathrm{d}x \approx 22.12954$，與 $\left[\frac{1}{5}\tan^5 x + \frac{1}{7}\tan^7 x\right]_{-0.7}^{1.1}$ 吻合；而 $\int_{-1}^{0.9} \sec^2 x(\sec x + \tan x)^6\,\mathrm{d}x \approx 133.68566$，與 (ii) 的答案吻合。
- 對於 $f = \tan$ 及 1 至 6 的每一個 n，一般公式都與數值積分（quadrature）的結果一致。

常見錯誤

- 在分部積分中選取 $u = f^n$：這樣會產生 $f^{n-1}[f']^3$，情況只會更糟。應把一個 f' 與 f^n 配成一對，使它能恰好積出來。
- 對 $\tan x$ 取 $k = 1$：其實 $f'' = 2\sec^2 x \tan x$，所以 $k = 2$。
- 在 (ii) 中，忘記 $\sec^2 x$ 必須以 $[f']^2/f^2$ 的形式吸收進去，這正是指數由 6 降至 $n = 4$ 的原因。

STEP 之外

條件 $f'' = kff'$ 本身就是一條微分方程（differential equation）：積分一次得 $f' = \frac{1}{2}kf^2 + c$，這是一條一階可分離變數方程（separable equation），其解包括 tan（$k = 2$，$c = 1$）及 $\sec x + \tan x$（$k = 1$，$c = \frac{1}{2}$）。認識到某個被積函數之所以容易處理，是因為該函數滿足某條常微分方程（ODE），這個主題在大學一年級有關特殊函數（special functions）和遞推公式（reduction formulae）的課程中會反覆出現。Boyce 和 DiPrima

的 *Elementary Differential Equations*（《初等微分方程》）講解了像這樣的可分離方程和自治方程（autonomous equations）。

詳細解答

STEP II 2011，第 7 題

純數學 • 數列與級數

> **STEP II 2011，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/11-s2-q7

關鍵思路

切勿逐項處理這兩個數列（sequence）：把每個和拆成 λ 的冪之和與 μ 的冪之和，再分別套用幾何級數（geometric series）公式。這兩個數滿足 $\lambda+\mu=2$ 及 $\lambda\mu=-1$，所以它們是 $x^2=2x+1$ 的根。單憑這條二次方程（quadratic equation），所有棘手的分母都會化簡掉。另一個工具是平方：b_k^2 和 a_k^2 都會化成 a_{2k} 加一個常數。

解答

首先列出一些全程都會用到的事實：

$$\lambda-1=\sqrt{2}, \qquad \mu-1=-\sqrt{2}, \qquad \lambda\mu=1-2=-1, \qquad \lambda-\mu=2\sqrt{2}.$$

此外 $a_0=2$ 及 $b_0=0$。

第 (i) 部分

由於 $\lambda\neq 1$ 且 $\mu\neq 1$，我們可以對每個和套用幾何級數公式：

$$\sum_{r=0}^{n}\lambda^r=\frac{\lambda^{n+1}-1}{\lambda-1}=\frac{\lambda^{n+1}-1}{\sqrt{2}}, \qquad \sum_{r=0}^{n}\mu^r=\frac{\mu^{n+1}-1}{\mu-1}=-\frac{\mu^{n+1}-1}{\sqrt{2}}.$$

把第一個和減去第二個和，

$$\sum_{r=0}^{n}b_r=\frac{\lambda^{n+1}-1}{\sqrt{2}}+\frac{\mu^{n+1}-1}{\sqrt{2}}=\frac{a_{n+1}-2}{\sqrt{2}}=-\sqrt{2}+\frac{1}{\sqrt{2}}\,a_{n+1},$$

正是所求。若改為把兩個和相加，兩個 -1 便互相抵消：

$$\boxed{\sum_{r=0}^{n} a_r = \frac{\lambda^{n+1}-1}{\sqrt{2}} - \frac{\mu^{n+1}-1}{\sqrt{2}} = \frac{1}{\sqrt{2}}\, b_{n+1}.}$$

（快速檢驗：當 $n=0$ 時，這式表示 $a_0 = b_1/\sqrt{2} = 2\sqrt{2}/\sqrt{2} = 2$，是正確的。）

第 (ii) 部分

由第 (i) 部分，內層的和是 $b_{m+1}/\sqrt{2}$。因此

$$\sum_{m=0}^{2n}\left(\sum_{r=0}^{m} a_r\right) = \frac{1}{\sqrt{2}}\sum_{m=0}^{2n} b_{m+1} = \frac{1}{\sqrt{2}}\sum_{s=1}^{2n+1} b_s = \frac{1}{\sqrt{2}}\sum_{s=0}^{2n+1} b_s,$$

最後一步成立，是因為 $b_0 = 0$。現在以 $2n+1$ 代替 n 應用第 (i) 部分，得

$$\sum_{m=0}^{2n}\left(\sum_{r=0}^{m} a_r\right) = \frac{1}{\sqrt{2}}\left(-\sqrt{2} + \frac{a_{2n+2}}{\sqrt{2}}\right) = \tfrac{1}{2}a_{2n+2} - 1. \qquad (*)$$

想一想 怎樣把 a_{2n+2} 與 b_{n+1}^2 聯繫起來？

我們取平方，並利用 $\lambda\mu = -1$ 處理交叉項：

$$b_{n+1}^2 = \lambda^{2n+2} + \mu^{2n+2} - 2(\lambda\mu)^{n+1} = a_{2n+2} - 2(-1)^{n+1}, \qquad a_{n+1}^2 = a_{2n+2} + 2(-1)^{n+1}.$$

若 n 是奇數，則 $n+1$ 是偶數，所以 $b_{n+1}^2 = a_{2n+2} - 2$。於是 $(*)$ 等於 $\frac{1}{2}(a_{2n+2}-2) = \frac{1}{2}b_{n+1}^2$，正是所求。

若 n 是偶數，則 $(-1)^{n+1} = -1$，所以這時等於 $a_{2n+2} - 2$ 的是 a_{n+1}^2。因此

$$\boxed{n \text{ 為偶數：} \quad \sum_{m=0}^{2n}\left(\sum_{r=0}^{m} a_r\right) = \tfrac{1}{2}a_{n+1}^2}$$

或等價地寫成 $\frac{1}{2}b_{n+1}^2 - 2$。

第 (iii) 部分

由第 (i) 部分，再用第 (ii) 部分的平方恆等式，

$$\left(\sum_{r=0}^{n} a_r\right)^2 = \tfrac{1}{2}b_{n+1}^2 = \tfrac{1}{2}a_{2n+2} - (-1)^{n+1}.$$

對於第二個和，我們同樣把它拆成兩個幾何級數，這次公比分別是 λ^2 和 μ^2：

$$\sum_{r=0}^{n} \lambda^{2r+1} = \lambda\,\frac{\lambda^{2n+2}-1}{\lambda^2-1}.$$

想一想 λ^2-1 可以化簡嗎？

可以。由於 $\lambda^2=2\lambda+1$，得 $\lambda^2-1=2\lambda$，於是因子 λ 被約去，和為 $\frac{1}{2}(\lambda^{2n+2}-1)$。由於 μ 滿足同一條二次方程，$\mu^2-1=2\mu$，而 μ 的和為 $\frac{1}{2}(\mu^{2n+2}-1)$（注意 $\mu^2\neq 1$，所以公式適用）。兩者相加，

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

因此

$$\left(\sum_{r=0}^{n} a_r\right)^2 - \sum_{r=0}^{n} a_{2r+1} = \left(\tfrac{1}{2}a_{2n+2}-(-1)^{n+1}\right) - \left(\tfrac{1}{2}a_{2n+2}-1\right) = 1+(-1)^n.$$

若 n 是偶數，此式等於 2，正是所求。若 n 是奇數，此式等於 0，所以

$$\boxed{n \text{ 為奇數：} \quad \left(\sum_{r=0}^{n} a_r\right)^2 = \sum_{r=0}^{n} a_{2r+1}.}$$

驗算

各 a_n 都是整數，並滿足 $a_{n+1}=2a_n+a_{n-1}$：$2,2,6,14,34,82,198,478,\ldots$。在第 (ii) 部分取 $n=3$，部分和 $2,4,10,24,58,140,338$ 加起來是 576。由於 $b_4=24\sqrt{2}$，得 $\frac{1}{2}b_4^2=576$，兩者吻合。在第 (iii) 部分取 $n=2$：$(2+2+6)^2-(2+14+82)=100-98=2$。取 $n=1$：$(2+2)^2-(2+14)=0$。我們用電腦驗證了全部三個部分，包括奇偶兩種情況（parity），直至 n 等於 24 的每一個值。

常見錯誤

- 忘記 $\mu-1=-\sqrt{2}$ 是負數，以致在第 (i) 部分把 a 和 b 的角色對調。
- 寫成 $(\lambda\mu)^{n+1}=1$ 而不是 $(-1)^{n+1}$；這個正負號正是要按奇偶分情況的關鍵所在。
- 在第 (ii) 部分移動 b_{m+1} 之和的下標時，沒有留意到正是 $b_0=0$ 才容許和由 0 開始。

STEP 之外

兩個數列都滿足線性遞推關係（linear recurrence）$x_{n+1}=2x_n+x_{n-1}$，其特徵方程（characteristic equation）為 $x^2=2x+1$。所以 λ 和 μ 是它的根，而 a_n 和 b_n 是特解。由於 $\lambda^n=\frac{1}{2}(a_n+b_n)$，恆等式 $\lambda^n\mu^n=(-1)^n$ 便化為 $a_n^2-b_n^2=4(-1)^n$，這是環（ring）$\mathbb{Z}[\sqrt{2}]$

中的一條佩爾型方程（Pell-type equation）。Davenport 的 *The Higher Arithmetic*（《高等算術》）闡述了這類單位（units）與 $\sqrt{2}$ 的連分數（continued fractions）之間如何互相配合。

詳細解答

STEP II 2011，第 8 題

純數學 • 圓的漸伸線

> **STEP II 2011，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/11-s2-q8

關鍵思路

在任何時刻，繩子都由兩段組成：一段是圓上的弧，其長度可以由角度直接讀出；另一段是沿切線（tangent）離開圓的直線段。由於總長度固定，直線段的長度就是剩下的部分；而切線與半徑垂直，所以用向量（vector）就能輕易定出 B 的位置。至於最後的面積，留意整個被掃過的區域就是「外支下方的面積，減去內支下方的面積，再減去半圓」。

解答

參數方程

設 $P = (\cos t, \sin t)$ 為繩子離開圓的點，因此 t 是由正 x 軸到 OP 的角度。開始時 $P = A = (-1, 0)$，所以 $t = \pi$；當繩子從圓的頂部繞過去時，P 沿順時針方向移動，t 遞減至 0。

與圓接觸的那段繩子是由角度 π 到角度 t 的弧。半徑為 1，所以這段弧長為 $\pi - t$。繩子長度為 π，所以直線段 PB 的長度為

$$\pi - (\pi - t) = t.$$

由於繩子拉緊，PB 是圓在 P 點的切線，因此它垂直於 $OP = (\cos t, \sin t)$。與 OP 垂直的兩個單位向量是 $\pm(\sin t, -\cos t)$。開始時（$t = \pi$）繩子由 A 垂直向上，方向為 $(0, 1) = (\sin \pi, -\cos \pi)$，而根據連續性（continuity），符號在整個過程中保持不變。因此

$$\overrightarrow{OB} = \overrightarrow{OP} + t\,(\sin t, -\cos t),$$

即

$$x = \cos t + t \sin t, \qquad y = \sin t - t \cos t,$$

正如所求。（驗算：$t=\pi$ 給出 $(-1,\pi)$，即 B 的起始位置；$t=0$ 給出 $(1,0)$，即 B 碰到圓的位置。）

x 的最大值

求導後，$\pm\sin t$ 與 $\pm\cos t$ 各項互相抵消：

$$\frac{\mathrm{d}x}{\mathrm{d}t}=-\sin t+\sin t+t\cos t=t\cos t,\qquad \frac{\mathrm{d}y}{\mathrm{d}t}=\cos t-\cos t+t\sin t=t\sin t.$$

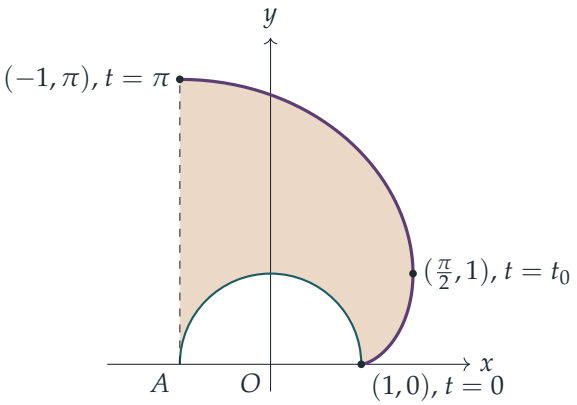
當 $0<t<\pi$ 時，$\frac{\mathrm{d}x}{\mathrm{d}t}=t\cos t$ 在 $t<\frac{\pi}{2}$ 時為正，在 $t>\frac{\pi}{2}$ 時為負。所以當 t 取遍 $[0,\pi]$ 時，x 先遞增後遞減，最大值出現在

$$t_0=\frac{\pi}{2},\qquad \text{此時}(x,y)=\left(\frac{\pi}{2},1\right).$$

草圖

有兩個事實有幫助。第一，當 $0<t\le\pi$ 時 $y>0$：若 $t\le\frac{\pi}{2}$，則由於 $\tan t>t$，$y=\cos t\,(\tan t-t)>0$（當 $t=\frac{\pi}{2}$ 時直接有 $y=1$）；若 $t>\frac{\pi}{2}$，則 $\cos t<0$，所以 y 的兩項都非負，且 $\sin t-t\cos t>0$。第二，$x^2+y^2=1+t^2$，所以 B 永遠在圓外，而且離圓的距離隨 t 增大。此外 $\frac{\mathrm{d}y}{\mathrm{d}t}=t\sin t\ge 0$，所以 y 隨 t 穩定遞增。

因此曲線由 $(1,0)$ 出發（該處 $\frac{\mathrm{d}y}{\mathrm{d}x}=\tan t=0$，所以它水平地離開 x 軸），上升至最右點 $(\frac{\pi}{2},1)$，該處切線垂直，然後掃過圓的頂部回到 $(-1,\pi)$，該處再次水平。這條曲線是圓的漸伸線（involute）。



陰影區域介乎曲線、繩子的初始位置（虛線）與半圓之間，就是繩子掃過的面積。

$\pi \geq t \geq t_0$ 部分的面積

在曲線的這一部分，x 由 $\frac{\pi}{2}$（在 $t = \frac{\pi}{2}$）遞減至 -1（在 $t = \pi$），且 $y > 0$。所以這段弧與 x 軸之間的面積為

$$\int_{-1}^{\pi/2} y \,\mathrm{d}x = \int_{\pi}^{\pi/2} y \frac{\mathrm{d}x}{\mathrm{d}t} \,\mathrm{d}t = \int_{\pi/2}^{\pi} \left(t^2 \cos^2 t - t \sin t \cos t\right) \mathrm{d}t,$$

其中我們對調了上下限以吸收負號。我們需要兩個積分，都是先用倍角公式（double-angle formula），再用分部積分（integration by parts）。寫 $\sin t \cos t = \frac{1}{2} \sin 2t$ 及 $\cos^2 t = \frac{1}{2}(1 + \cos 2t)$。用分部積分，

$$\int t \sin 2t \,\mathrm{d}t = -\tfrac{1}{2} t \cos 2t + \tfrac{1}{4} \sin 2t,$$
$$\int t^2 \cos 2t \,\mathrm{d}t = \tfrac{1}{2} t^2 \sin 2t - \int t \sin 2t \,\mathrm{d}t = \tfrac{1}{2} t^2 \sin 2t + \tfrac{1}{2} t \cos 2t - \tfrac{1}{4} \sin 2t.$$

在 $t = \pi$ 時，$\sin 2t = 0$，$\cos 2t = 1$；在 $t = \frac{\pi}{2}$ 時，$\sin 2t = 0$，$\cos 2t = -1$。因此

$$\int_{\pi/2}^{\pi} t \sin t \cos t \,\mathrm{d}t = \tfrac{1}{2}\left[-\tfrac{1}{2} t \cos 2t\right]_{\pi/2}^{\pi} = \tfrac{1}{2}\left(-\frac{\pi}{2} - \frac{\pi}{4}\right) = -\frac{3\pi}{8},$$
$$\int_{\pi/2}^{\pi} t^2 \cos^2 t \,\mathrm{d}t = \tfrac{1}{2}\left[\tfrac{1}{3} t^3\right]_{\pi/2}^{\pi} + \tfrac{1}{2}\left[\tfrac{1}{2} t \cos 2t\right]_{\pi/2}^{\pi} = \frac{7\pi^3}{48} + \tfrac{1}{2}\left(\frac{\pi}{2} + \frac{\pi}{4}\right) = \frac{7\pi^3}{48} + \frac{3\pi}{8}.$$

所以所求面積為

$$\frac{7\pi^3}{48} + \frac{3\pi}{8} - \left(-\frac{3\pi}{8}\right) = \frac{7\pi^3}{48} + \frac{3\pi}{4}.$$

繩子掃過的面積

想一想 $t \geq t_0$ 那段弧下方的區域比掃過的區域多。必須除去甚麼？

剛求得的面積，是外弧（$t_0 \leq t \leq \pi$）與 x 軸之間、$-1 \leq x \leq \frac{\pi}{2}$ 範圍內的全部區域。掃過的區域等於這個面積減去位於其下方的兩塊：上半圓（semicircle），面積為 $\frac{\pi}{2}$；以及內弧（$0 \leq t \leq t_0$）與 x 軸之間、$1 \leq x \leq \frac{\pi}{2}$ 範圍內的狹長小塊。內弧位於圓外（因為當 $t > 0$ 時 $x^2 + y^2 = 1 + t^2 > 1$），並位於外弧下方，所以這兩塊互不重疊，相減是合理的。

在內弧上 x 隨 t 遞增，所以狹長小塊的面積為

$$\int_{0}^{\pi/2} y \frac{\mathrm{d}x}{\mathrm{d}t} \,\mathrm{d}t = \int_{0}^{\pi/2} \left(t \sin t \cos t - t^2 \cos^2 t\right) \mathrm{d}t.$$

使用同樣的原函數（antiderivative），現在以 $t = 0$ 為下限（該處 $\sin 2t = 0$，$\cos 2t = 1$），

$$\int_{0}^{\pi/2} t \sin t \cos t \,\mathrm{d}t = \tfrac{1}{2}\left[-\tfrac{1}{2} t \cos 2t\right]_{0}^{\pi/2} = \frac{\pi}{8},$$
$$\int_{0}^{\pi/2} t^2 \cos^2 t \,\mathrm{d}t = \tfrac{1}{2} \cdot \frac{\pi^3}{24} + \tfrac{1}{2}\left[\tfrac{1}{2} t \cos 2t\right]_{0}^{\pi/2} = \frac{\pi^3}{48} - \frac{\pi}{8},$$

所以狹長小塊的面積為 $\frac{\pi}{4} - \frac{\pi^3}{48}$（約 0.14，又小又正，令人放心）。因此繩子掃過的面積為

$$\left(\frac{7\pi^3}{48} + \frac{3\pi}{4}\right) - \left(\frac{\pi}{4} - \frac{\pi^3}{48}\right) - \frac{\pi}{2} = \frac{8\pi^3}{48} = \frac{\pi^3}{6}.$$

驗算

- 最後的答案有更簡潔的求法。當 t 減少 δt 時，繩子的直線段（長度為 t）繞着一個基本上固定的點轉過角度 δt，掃出面積約為 $\frac{1}{2}t^2\,\delta t$ 的細小扇形（sector）。把這些加起來得 $\int_0^\pi \frac{1}{2}t^2\,\mathrm{d}t = \frac{\pi^3}{6}$，結果一致。
- 數值上：用辛普森法則（Simpson's rule）求外弧下方的面積得 6.87794，與 $\frac{7\pi^3}{48} + \frac{3\pi}{4}$ 吻合；對陰影區域的邊界使用鞋帶公式（shoelace formula）得 5.16771 $\approx \frac{\pi^3}{6}$。

常見錯誤

- 以為直線段的長度是 $\pi - t$（那是弧長），或者在兩個切線方向中選錯了一個；檢查 $t = \pi$ 的情況可以同時發現這兩種錯誤。
- 把積分上下限保留為 t_0 到 π，卻沒有留意 x 沿該段弧遞減，結果得出負的「面積」。
- 忘記內弧下方的小區域，結果得出掃過面積為 $\frac{7\pi^3}{48} + \frac{\pi}{4}$。

STEP 之外

繩子從一條曲線上解開時，其末端描出的曲線稱為漸伸線，而繩子所解開的那條曲線則稱為它的漸屈線（evolute）：漸伸線在 B 點的曲率中心（centre of curvature）正是繩子離開圓的點 P。大多數齒輪的齒形都採用圓的漸伸線，因為兩個這樣的齒互相滾動時，速度比保持不變。「驗算」一節中的「細小扇形」論證，其實就是極坐標面積公式（polar-area formula）$\frac{1}{2}\int r^2\,\mathrm{d}\theta$ 的化身。想進一步了解曲率及這類曲線，可參閱 Siklos 的 *Advanced Problems in Mathematics*（《數學進階難題》）；至於背後嚴謹的微積分，可參閱 Abbott 的 *Understanding Analysis*（《理解分析》）。

詳細解答

STEP II 2011，第 9 題

力學 • 碰撞與恢復係數

STEP II 2011，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：

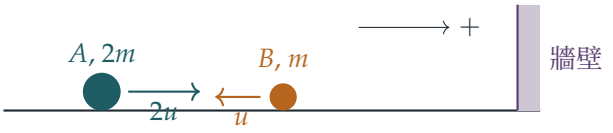
step.maths.org/questions/11-s2-q9

關鍵思路

這裏每一次碰撞都是正碰（direct collision），所以每次都用同樣兩條方程解決：動量守恆（conservation of momentum）和牛頓（Newton）恢復定律（law of restitution）。真正的危險只在於正負號。一開始就固定一個正方向（向着牆壁），把每個速度都寫成該方向上帶正負號的數，代數運算便會自然理順。最後的論斷歸結為證明某個表達式恆為正；由於它是 f 的線性（linear）函數，我們只需考察 f 取值範圍的兩個端點。

解答

模型。 我們把 A 和 B 視為光滑（smooth）平面上的質點（particle），所以除碰撞時刻外沒有水平力作用；牆壁亦是光滑的，因此它只會令 B 的速度反向（並按 f 縮放）。第一次碰撞後 B 繼續前進並撞向牆壁，所以牆壁位於 B 那一側，即沿 A 原本運動方向、在 B 之外。我們取該方向為正方向。於是沿直線的次序始終是 A、B、牆壁。



第一次碰撞

碰撞前 A 的速度為 $2u$，B 的速度為 $-u$。設碰撞後兩者的速度分別為 v_A 和 v_B。由動量守恆：

$$2m(2u) + m(-u) = 2m\,v_A + m\,v_B \quad \Longrightarrow \quad 2v_A + v_B = 3u.$$

接近速率為 $2u-(-u)=3u$。由於 B 在 A 的前方，分離速率為 v_B-v_A，牛頓恢復定律給出

$$v_B-v_A=3eu.$$

第一條方程減去第二條方程，得 $3v_A=3u-3eu$，所以

$$v_A=(1-e)u,\qquad v_B=(1+2e)u.$$

兩者皆 ≥ 0，所以碰撞後 A 以速率 $(1-e)u$ 向牆壁移動（若 $e=1$ 則靜止），而 B 以速率 $(1+2e)u$ 向牆壁移動。這與 B 隨後撞到牆壁相符。

牆壁

牆壁光滑而碰撞為正碰，所以 B 沿直線以速率 $f(1+2e)u$ 反彈，其速度變為 $-f(1+2e)u$。與此同時，A 仍以速度 $(1-e)u\geq 0$ 運動，因為沒有任何力作用於它。B 現正向 A 移動，所以兩者必定再次相遇。

第二次碰撞

第二次碰撞前一刻，速度為

$$A:\ (1-e)u,\qquad B:\ -f(1+2e)u.$$

設碰撞後的速度為 w_A 和 w_B。由動量守恆得

$$2w_A+w_B=2(1-e)u-f(1+2e)u.$$

想一想 這一次恢復方程應該怎樣寫，哪個減哪個？

B 仍然是較接近牆壁的質點，所以分離速度仍是 w_B-w_A。接近速率是 A 的速度減去 B 的速度，即 $(1-e)u+f(1+2e)u$。因此

$$w_B-w_A=e\big[(1-e)u+f(1+2e)u\big].$$

為求出 w_B，把這條方程的兩倍加到動量方程上：

$$\begin{aligned}3w_B&=2(1-e)u-f(1+2e)u+2e(1-e)u+2ef(1+2e)u\\&=2(1-e)(1+e)u+f(1+2e)(2e-1)u\\&=2(1-e^2)u-(1-4e^2)fu.\end{aligned}$$

所以

$$w_B=\tfrac{2}{3}(1-e^2)u-\tfrac{1}{3}(1-4e^2)fu,$$

此速度以正方向（即向着牆壁）量度。這就是所求的結果。

B 總是向牆壁移動

我們須證明對所有 $0 < e \le 1$ 及 $0 < f \le 1$，都有 $w_B > 0$。由於 $u > 0$，這等同於證明

$$g(f) = 2(1-e^2) + (4e^2-1)f > 0.$$

$1 - 4e^2$ 一項可正可負，所以不能直接把它略去。改為固定 e，注意 g 是 f 的線性函數，而且

$$g(0) = 2(1-e^2) \ge 0, \qquad g(1) = 2 - 2e^2 + 4e^2 - 1 = 1 + 2e^2 > 0.$$

$[0,1]$ 上的線性函數是其兩端值的加權平均：$g(f) = (1-f)\,g(0) + f\,g(1)$。當 $0 < f \le 1$ 時，第一項 ≥ 0，而第二項嚴格為正（因為 $f > 0$ 且 $g(1) > 0$），所以 $g(f) > 0$。因此 $w_B > 0$，無論 e 和 f 取何值，B 在第二次碰撞後都向牆壁移動。

驗算

- 取 $e = f = 1$（全程完全彈性），A 在第一次碰撞時完全停下，B 以 $3u$ 離開，再以 $3u$ 返回，撞向一個質量為它兩倍的靜止質點。公式給出 $w_B = 0 + u = u$，與彈性碰撞的結果 $\frac{m-2m}{3m}(-3u) = u$ 吻合。
- 取 $e = f = \frac{1}{2}$，公式給出 $w_B = \frac{2}{3}\cdot\frac{3}{4}u = \frac{1}{2}u$。逐步模擬位置和碰撞，亦得到相同的值 $0.5u$。
- 對正負號而言，最壞情況是 $e = 1$ 而 f 很小。此時 $w_B = fu$，雖然很小，但仍然為正。

常見錯誤

- 在第二次碰撞中用速率代替帶正負號的速度，因而把 B 的反彈速度符號弄錯。
- 因為 B「從牆壁那邊過來」而把恢復方程寫成 $w_A - w_B$。質點沿直線的次序並沒有改變，所以分離速度仍是 $w_B - w_A$。
- 假設 $1 - 4e^2 > 0$ 來斷言 $w_B > 0$。當 $e > \frac{1}{2}$ 時此假設不成立，這正是論證必須用到 $f \le 1$ 的原因。

STEP 之外

恢復定律是一個經驗模型（empirical model）。在大學一年級的課程中，你會同時接觸一個更簡潔的概念——質心參考系（centre-of-mass frame）：在該參考系中總動量為零，而恢復係數為 e 的碰撞只是把兩個速度都乘以 $-e$。這令第一次碰撞之類的結果幾乎一

目了然。與牆壁的反覆碰撞亦會在撞球問題（billiard problems）和計數論證中再次出現，例如那個著名的問題：兩個木塊與一面牆之間的碰撞次數，竟然給出 π 的各位數字。Kleppner 和 Kolenkow 的 *An Introduction to Mechanics*（《力學導論》）正是以這種精神講解動量與碰撞。

詳細解答

STEP II 2011，第 10 題

力學 • 拋體與衝量

STEP II 2011，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/11-s2-q10

關鍵思路

由於 R 始終位於 P 的正下方，兩顆質點在整個過程中水平速度相同，因此弦線保持鉛垂，只需考慮鉛垂方向的運動。當 P 恰好在平面上方 $\frac{1}{2}H$ 時，弦線繃緊（taut）；此時的急拉是一個鉛垂方向的衝量（impulse），把 P 向上的動量（momentum）分給 R。此後兩顆質點速度完全相同，因此作為同一個拋射體（projectile）一起飛行，直至 R 落地。

解答

最大高度

把質點視為只受重力作用的點。速度的鉛垂分量為 $u\sin\theta - gt$，在最高點時為零。在鉛垂方向使用 $v^2 = u^2 + 2as$，末鉛垂速率為 0，

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

值得記下以備後用：由此得 $u\sin\theta = \sqrt{2gH}$。

弦線何時繃緊

以 O 為原點，x 為水平方向，y 為鉛垂向上方向。質點 P 的水平速度始終為 $u\cos\theta$。要令 R 保持在 P 的正下方，R 必須以速率 $u\cos\theta$ 射出；由於平面光滑（smooth），R 便以這個恆定速率滑行。因此弦線始終鉛垂，而距離 PR 就是 P 的高度：

$$y_P = u\sin\theta\, t - \tfrac{1}{2}gt^2 = \sqrt{2gH}\, t - \tfrac{1}{2}gt^2.$$

當 y_P 首次達到 $\frac{1}{2}H$ 時，弦線繃緊（這必定會發生，因為 P 上升至 $H > \frac{1}{2}H$）：

$$\tfrac{1}{2}gt^2 - \sqrt{2gH}\,t + \tfrac{1}{2}H = 0 \quad\Longrightarrow\quad t = \frac{\sqrt{2gH} \pm \sqrt{2gH - gH}}{g} = (\sqrt{2} \pm 1)\sqrt{\frac{H}{g}}.$$

較小的根是首次到達的時刻，因此弦線在以下時間後繃緊：

$$t_1 = (\sqrt{2} - 1)\sqrt{H/g},$$

正如所求。（較大的根是 P 下降時再次經過高度 $\frac{1}{2}H$ 的時刻。）

急拉

在 t_1 時，P 的鉛垂速度為

$$u\sin\theta - gt_1 = \sqrt{2gH} - (\sqrt{2} - 1)\sqrt{gH} = \sqrt{gH},$$

方向向上，而 R 沒有鉛垂速度。兩者的水平速度均為 $u\cos\theta$。

? **想一想** 在弦線急拉的一瞬間，有哪些衝量作用，方向如何？

弦線鉛垂，因此張力（tension）衝量 J 是鉛垂的：對 P 向下，對 R 向上。平面光滑，只能向上推 R；由於 R 被向上拉離平面，平面不施加任何衝量。所以水平速度不變，而在鉛垂方向上兩顆質點的總動量守恆。弦線不可伸長（inextensible）且保持繃緊，因此緊接急拉之後，P 和 R 有相同的鉛垂速度 w。兩者質量同為 m，

$$m\sqrt{gH} + m\cdot 0 = 2mw \qquad\Longrightarrow\qquad w = \tfrac{1}{2}\sqrt{gH}.$$

距離 D

急拉之後，若張力為零，P 和 R 速度相同，加速度亦同為向下的 g。因此它們如同兩個相同的拋射體，相隔固定距離 $\frac{1}{2}H$ 運動：弦線恰好保持繃緊而張力為零，這是自洽的。R 從高度 0 以向上速度 w 出發，因此在以下時間後回到平面：

$$t_2 = \frac{2w}{g} = \sqrt{\frac{H}{g}}.$$

所以 R 運動的總時間為

$$t_1 + t_2 = \sqrt{2}\sqrt{\frac{H}{g}} = \sqrt{\frac{2H}{g}} = \sqrt{\frac{u^2\sin^2\theta}{g^2}} = \frac{u\sin\theta}{g},$$

而在整段時間內 R 以 $u\cos\theta$ 水平移動。因此

$$D = u\cos\theta\cdot\frac{u\sin\theta}{g} = \frac{u^2\sin\theta\cos\theta}{g} = \frac{u^2\sin 2\theta}{2g}.$$

這恰好是 P 單獨運動時射程（range）的一半：R 落地的時刻，正是沒有弦線時 P 到達其軌跡最高點的時刻。

$D = H$ 的情況

設

$$\frac{u^2 \sin\theta\cos\theta}{g} = \frac{u^2 \sin^2\theta}{2g}$$

並除以 $u^2 \sin\theta/g$（它不為零，因為質點要上升，須有 $0 < \theta < \frac{1}{2}\pi$），得 $\cos\theta = \frac{1}{2}\sin\theta$，所以

$$\tan\theta = 2.$$

驗算

- t 的兩個根之積為 H/g，之和為 $2\sqrt{2H/g}$，即 P 到達最高點再回到平面所需的時間；根據對稱性，理應如此。
- 以隨機的 u 和 θ 進行數值模擬（逐步推進 P 的飛行直至其高度為 $\frac{1}{2}H$，施加急拉，再逐步推進直至 R 落地），結果與 t_1 和 D 吻合；例如 $u = 18.11$、$\theta = 0.711$、$g = 9.8$，兩種方法均得 $D \approx 16.543$。
- 取 $\tan\theta = 2$、$u = 10$、$g = 9.8$：$D = H \approx 4.082$。

常見錯誤

- 取了較大的根 $(\sqrt{2}+1)\sqrt{H/g}$，那是 P 下降時再次經過高度 $\frac{1}{2}H$ 的時刻。
- 忘記 R 在急拉前滑行的時間 t_1，只用急拉後的飛行來計算 D。
- 在急拉時使用能量守恆：不可伸長的弦線急拉繃緊時會損失動能；只有動量守恆。

STEP 之外

急拉之後，這對質點的行為與位於其質心（centre of mass）的單一質點完全相同：內力（這裏是張力）永遠不會改變系統質心的運動。這是大學一年級課程中質點系及剛體（rigid body）力學的起點，而像這題一樣的衝擊約束亦以同樣方法處理。Kleppner 和 Kolenkow 的 *An Introduction to Mechanics*（《力學導論》）對這些概念有仔細的闡述。

詳細解答

STEP II 2011，第 11 題

力學 • 三維空間中的平衡

STEP II 2011，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/11-s2-q11

關鍵思路

一切都建基於為每條繩寫出一個沿繩方向的單位向量（unit vector）。每個單位向量都可分拆為鉛垂部分和水平部分：鉛垂部分由繩與鉛垂線所成的角決定；水平部分則由 O 指向繩在天花板上的端點，並由該方向與 OA 所成的角決定。一旦知道三個單位向量，平衡（equilibrium）條件便只是三條純量方程，$\mathbf{i}$、$\mathbf{j}$、$\mathbf{k}$ 各一條。

解答

我們把 P 視為質點（particle），並把繩視為輕而不可伸長（light and inextensible），因此每條繩都以沿繩方向的力把 P 拉向其固定端。取 O 為原點，$\mathbf{j}$ 沿 OA，$\mathbf{k}$ 鉛垂向上；$\mathbf{i}$ 為水平方向，並垂直於 OA。

先處理三角學。由於 $\tan\theta = \sqrt{2}$，以 $1, \sqrt{2}, \sqrt{3}$ 為邊的直角三角形給出

$$\sin\theta = \frac{\sqrt{2}}{\sqrt{3}}, \qquad \cos\theta = \frac{1}{\sqrt{3}}.$$

由於 $\tan\phi = \frac{1}{4}\sqrt{2}$，兩直角邊為 $4, \sqrt{2}$，斜邊為 $\sqrt{18} = 3\sqrt{2}$，所以

$$\sin\phi = \frac{1}{3}, \qquad \cos\phi = \frac{2\sqrt{2}}{3}.$$

第 (i) 部分

與鉛垂線成角 α 的繩，其單位向量（由 P 指向天花板）的鉛垂分量為 $\cos\alpha$，水平部分的長度為 $\sin\alpha$。

對於 PB，$\alpha = 90^\circ - \theta$，所以鉛垂分量為 $\cos(90^\circ - \theta) = \sin\theta = \frac{\sqrt{2}}{\sqrt{3}}$，而水平部分的長度為 $\sin(90^\circ - \theta) = \cos\theta = \frac{1}{\sqrt{3}}$。

水平部分沿 OB 方向（因為 O 在 P 的正上方），與 $\mathbf{j}$ 成角 $90^\circ+\theta$。與 $\mathbf{j}$ 成角 $90^\circ+\theta$ 的水平單位向量為 $\pm\sin(90^\circ+\theta)\,\mathbf{i}+\cos(90^\circ+\theta)\,\mathbf{j}=\pm\cos\theta\,\mathbf{i}-\sin\theta\,\mathbf{j}$。我們選取 $\mathbf{i}$ 的指向，使 B 位於 OA 的負 $\mathbf{i}$ 一側；題目所要求的形式決定了這個選擇。於是沿 PB 的單位向量的水平部分為

$$\tfrac{1}{\sqrt{3}}(-\cos\theta\,\mathbf{i}-\sin\theta\,\mathbf{j})=\tfrac{1}{\sqrt{3}}\left(-\tfrac{1}{\sqrt{3}}\,\mathbf{i}-\tfrac{\sqrt{2}}{\sqrt{3}}\,\mathbf{j}\right)=-\tfrac{1}{3}\,\mathbf{i}-\tfrac{\sqrt{2}}{3}\,\mathbf{j}.$$

加上鉛垂分量，沿 PB 方向的單位向量為

$$-\tfrac{1}{3}\,\mathbf{i}-\tfrac{\sqrt{2}}{3}\,\mathbf{j}+\tfrac{\sqrt{2}}{\sqrt{3}}\,\mathbf{k},$$

正如所求。（驗算：$\frac{1}{9}+\frac{2}{9}+\frac{2}{3}=1$。）

第 (ii) 部分

用同樣的方法可得另外兩條繩的單位向量。

繩 PA：與鉛垂線成 30°，水平部分沿 $\mathbf{j}$：

$$\mathbf{e}_A=\tfrac{1}{2}\,\mathbf{j}+\tfrac{\sqrt{3}}{2}\,\mathbf{k}.$$

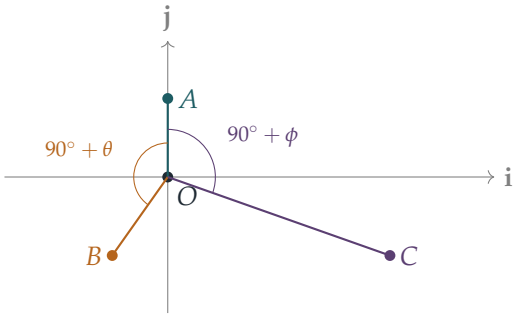
繩 PC：與鉛垂線成 60°，所以鉛垂分量為 $\frac{1}{2}$，水平長度為 $\frac{\sqrt{3}}{2}$。方向 OC 與 $\mathbf{j}$ 成角 $90^\circ+\phi$，所以其水平單位向量為 $\pm\cos\phi\,\mathbf{i}-\sin\phi\,\mathbf{j}$。

> ? **想一想** C 是在 OA 與 B 相同的一側，還是在相反的一側？

圖中顯示的是相反一側，而且平衡條件本身亦要求如此。PA 中的張力（tension）沒有 $\mathbf{i}$ 分量，而 PB 中的張力則有負分量。若 PC 也有負的 $\mathbf{i}$ 分量，便沒有任何力能與它們平衡。所以符號為 $+$，且

$$\mathbf{e}_C=\tfrac{\sqrt{3}}{2}\left(\tfrac{2\sqrt{2}}{3}\,\mathbf{i}-\tfrac{1}{3}\,\mathbf{j}\right)+\tfrac{1}{2}\,\mathbf{k}=\tfrac{\sqrt{6}}{3}\,\mathbf{i}-\tfrac{\sqrt{3}}{6}\,\mathbf{j}+\tfrac{1}{2}\,\mathbf{k}.$$

下面的天花板平面圖（取 $OP=1$，故 $OA=\tan30^\circ$、$OB=\cot\theta$、$OC=\tan60^\circ$）顯示了各點的布局。



因此作用於 P 的力為

$$\begin{aligned} \text{張力，繩}PA: &\quad T\left(\tfrac{1}{2}\,\mathbf{j}+\tfrac{\sqrt{3}}{2}\,\mathbf{k}\right),\\ \text{張力，繩}PB: &\quad U\left(-\tfrac{1}{3}\,\mathbf{i}-\tfrac{\sqrt{2}}{3}\,\mathbf{j}+\tfrac{\sqrt{2}}{\sqrt{3}}\,\mathbf{k}\right),\\ \text{張力，繩}PC: &\quad V\left(\tfrac{\sqrt{6}}{3}\,\mathbf{i}-\tfrac{\sqrt{3}}{6}\,\mathbf{j}+\tfrac{1}{2}\,\mathbf{k}\right),\\ \text{重量}: &\quad -W\,\mathbf{k}. \end{aligned}$$

第 (iii) 部分

P 處於平衡狀態，所以四個力之和為零。我們逐一取分量。

i：

$$-\tfrac{1}{3}U+\tfrac{\sqrt{6}}{3}V=0 \quad\Longrightarrow\quad U=\sqrt{6}\,V,$$

正如所求。

j：

$$\tfrac{1}{2}T-\tfrac{\sqrt{2}}{3}U-\tfrac{\sqrt{3}}{6}V=0.$$

代入 $U=\sqrt{6}V$，並利用 $\sqrt{2}\sqrt{6}=2\sqrt{3}$：

$$\tfrac{1}{2}T=\tfrac{2\sqrt{3}}{3}V+\tfrac{\sqrt{3}}{6}V=\tfrac{5\sqrt{3}}{6}V \quad\Longrightarrow\quad T=\tfrac{5\sqrt{3}}{3}V.$$

k：

$$\tfrac{\sqrt{3}}{2}T+\tfrac{\sqrt{2}}{\sqrt{3}}U+\tfrac{1}{2}V=W.$$

代入以上兩個結果，

$$\tfrac{\sqrt{3}}{2}\cdot\tfrac{5\sqrt{3}}{3}V+\tfrac{\sqrt{2}\sqrt{6}}{\sqrt{3}}V+\tfrac{1}{2}V=\tfrac{5}{2}V+2V+\tfrac{1}{2}V=5V=W.$$

因此

$$V=\frac{W}{5},\qquad U=\frac{\sqrt{6}\,W}{5},\qquad T=\frac{5\sqrt{3}}{3}\cdot\frac{W}{5}=\frac{W}{\sqrt{3}}=\frac{\sqrt{3}\,W}{3}.$$

三者皆為正，所以每條繩都是拉緊的（taut），與模型一致。

驗算

- 每個單位向量的長度都是 1：對於 $\mathbf{e}_C$，$\frac{6}{9}+\frac{3}{36}+\frac{1}{4}=1$。
- $\mathbf{e}_A$ 與 $\mathbf{e}_C$ 的水平部分之間的夾角，其餘弦為 $-\frac{1}{3}=-\sin\phi$，所以夾角為 $90^\circ+\phi\approx 109.47^\circ$，理應如此。
- 取 $W=1$：$T\approx 0.5774$、$U\approx 0.4899$、$V=0.2$。於是鉛垂分量 $0.5774\times 0.8660+0.4899\times 0.8165+0.2\times 0.5=0.5+0.4+0.1=1$ 與重量平衡。留意這個整齊的分配：PA、PB、PC 分別承受重量的 $\frac{1}{2}$、$\frac{2}{5}$、$\frac{1}{10}$。

常見錯誤

- 處理繩 PB 時混淆 sin 與 cos：它與鉛垂線所成的角是 $90^\circ - \theta$，所以其鉛垂分量為 $\sin\theta$，而不是 $\cos\theta$。
- 把 C 放在 OA 與 B 相同的一側。這樣 **i** 方程便會迫使張力為負（即 $V = -\frac{1}{3}W$），而繩是不可能有負張力的。
- 忘記 $\cos(90^\circ + \phi) = -\sin\phi$，以致 $\mathbf{e}_B$ 與 $\mathbf{e}_C$ 的 **j** 分量符號出錯。

STEP 之外

三條沿三個獨立方向的繩使系統成為靜定（statically determinate）：平衡方程組成一個 3×3 線性方程組，並有唯一解。若再加第四條繩，單靠靜力學便不能確定各張力，工程師必須引入繩的彈性。這正是結構分析（structural analysis）以及有關秩（rank）與零空間（null space）的線性代數的起點。Kleppner 與 Kolenkow 的 *An Introduction to Mechanics*（《力學導論》）就是以這種方式把力當作向量處理，而 Strang 的 *Introduction to Linear Algebra*（《線性代數導論》）則解釋這類方程組何時有唯一解。

詳細解答

STEP II 2011，第 12 題

概率與統計 • 博弈中的等比級數

STEP II 2011，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/11-s2-q12

關鍵思路

每局最多只有三分，所以我們可以列出一局所有可能的進程（只有八個序列），並把它們分為「Xavier 勝」、「Younis 勝」和「無人勝出，再打一局」三類。由於每一局都重新開始，整場比賽是一個幾何過程（geometric process）：Younis 贏得整場比賽的概率（probability），等於 Younis 贏得一局的概率除以一局分出勝負的概率。

解答

把每一局寫成一串得分者，例如 XYY 表示 Xavier 贏第一分，而 Younis 贏第二和第三分。第一分以概率 p 歸 X。此後，每一分以概率 p 歸於與上一分相同的球員，並以概率 $1-p$ 轉換到另一位球員。

只要同一位球員連贏兩分，該局便分出勝負，因此可能的局為：

序列	概率	結果
XX	$p \cdot p$	Xavier 勝
YXX	$(1-p)(1-p)p$	Xavier 勝
YY	$(1-p)p$	Younis 勝
XYY	$p(1-p)p$	Younis 勝
XYX	$p(1-p)(1-p)$	再打一局
YXY	$(1-p)(1-p)(1-p)$	再打一局

（一局一旦出現 XX 或 YY，便不需要第三分；我們設想第三分有沒有打，都沒有分別。）未分勝負的局，正好是得分交替出現的局，即後兩分都「轉換」。

第 (i) 部分

設 A 為 Younis 贏得某一局的事件（event），C 為該局未分勝負的事件。由上表，

$$\mathrm{P}(A) = (1-p)p + p^2(1-p) = p(1-p)(1+p) = p(1-p^2),$$
$$\mathrm{P}(C) = p(1-p)^2 + (1-p)^3 = (1-p)^2\big(p + (1-p)\big) = (1-p)^2.$$

各局在相同規則下獨立（independent）進行，所以 Younis 贏得整場比賽，當且僅當對某個 $n \geq 1$，首 $n-1$ 局都未分勝負，而 Younis 贏第 n 局。因此

$$w = \sum_{n=1}^{\infty} \mathrm{P}(C)^{n-1}\,\mathrm{P}(A) = \frac{\mathrm{P}(A)}{1-\mathrm{P}(C)},$$

其中由於當 $p \neq 0$ 時 $\mathrm{P}(C) = (1-p)^2 < 1$，這個幾何級數（geometric series）收斂（converges）。（等價地：若第一局未分勝負，我們便回到起點，所以 $w = \mathrm{P}(A) + \mathrm{P}(C)\,w$。）而 $1-(1-p)^2 = 2p - p^2 = p(2-p)$，所以

$$w = \frac{p(1-p^2)}{p(2-p)} = \frac{1-p^2}{2-p},$$

其中約去 p 用到了 $p \neq 0$。這就是所求的結果。

與 $\frac{1}{2}$ 比較。我們有

$$w - \tfrac{1}{2} = \frac{2(1-p^2) - (2-p)}{2(2-p)} = \frac{p - 2p^2}{2(2-p)} = \frac{p(1-2p)}{2(2-p)}.$$

當 $0 < p \leq 1$ 時，因子 p 和 $2-p$ 都是正數，所以 $w - \frac{1}{2}$ 與 $1-2p$ 同號。因此當 $0 < p < \frac{1}{2}$ 時 $w > \frac{1}{2}$，當 $p > \frac{1}{2}$ 時 $w < \frac{1}{2}$（而在 $p = \frac{1}{2}$ 時 $w = \frac{1}{2}$，正如對稱性所要求）。

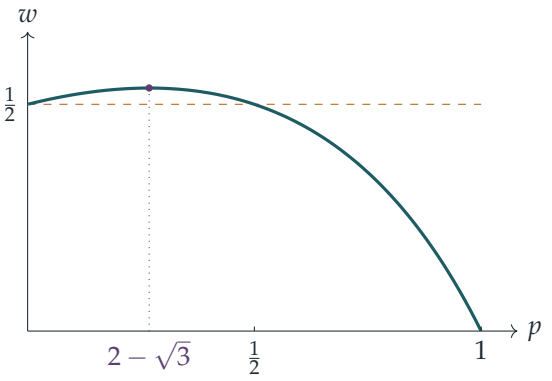
? **想一想** 降低 p 會令 Younis 更有可能贏得第一分。這是否必定對 Younis 有利？

w 是否隨 p 遞減？不是。用商法則（quotient rule）求導，

$$\frac{\mathrm{d}w}{\mathrm{d}p} = \frac{-2p(2-p) + (1-p^2)}{(2-p)^2} = \frac{p^2 - 4p + 1}{(2-p)^2}.$$

分子的根為 $p = 2 \pm \sqrt{3}$，其中只有 $2-\sqrt{3} \approx 0.268$ 位於 $[0,1]$ 之內。所以當 $0 < p < 2-\sqrt{3}$ 時 $\mathrm{d}w/\mathrm{d}p > 0$，當 $2-\sqrt{3} < p \leq 1$ 時 $\mathrm{d}w/\mathrm{d}p < 0$。在區間 $0 < p < 2-\sqrt{3}$ 上，減小 p 實際上會減小 w。例如 $w(0.2) = \frac{8}{15} \approx 0.533$，但 $w(0.1) = \frac{99}{190} \approx 0.521$。

原因是 p 同時扮演兩個角色。較小的 p 讓 Younis 取得第一分，但同時也令連勝不太可能出現，因為贏得一分的球員只會以概率 p 繼續得分。當 p 極小時，得分傾向交替出現，第一分的優勢便被抵銷。最大值為 $w(2-\sqrt{3}) = 4 - 2\sqrt{3} \approx 0.536$，而當 $p \to 0$ 時 $w \to \frac{1}{2}$。



第 (ii) 部分

當 $p = \frac{2}{3}$ 時，

$$w = \frac{1 - \frac{4}{9}}{2 - \frac{2}{3}} = \frac{5/9}{4/3} = \frac{5}{12}.$$

由於 $\mathrm{P}(C) = \frac{1}{9} < 1$，比賽以概率 1 結束，所以 Xavier 贏得比賽的概率為 $1 - \frac{5}{12} = \frac{7}{12}$。Xavier 的期望收益（expected gain）為

$$\tfrac{7}{12} \times 1 - \tfrac{5}{12} \times k,$$

而當此值為零時，比賽是公平的（fair），即

$$k = \frac{7}{5}.$$

所以當 Younis 勝出時，Xavier 應付給 Younis £1.40。

第 (iii) 部分

當 $p = 0$ 時，Younis 必定贏得每一局的第一分，而之後每一分都必定由輸掉上一分的球員贏得。所以每一局都以概率 1 按 YXY 進行。沒有人能連贏兩分，比賽永不結束：兩位球員都不會勝出。

這就是第 (i) 部分排除 $p = 0$ 的原因。此時 $\mathrm{P}(A) = 0$ 且 $1 - \mathrm{P}(C) = 0$，所以公式 $w = \mathrm{P}(A)/(1 - \mathrm{P}(C))$ 變成 $0/0$。當 $p \to 0$ 時 $\frac{1-p^2}{2-p} \to \frac{1}{2}$ 這個事實，描述的是那些鄰近而最終會結束的比賽；它並不是 $p = 0$ 這個情況本身的概率。

驗算

- 當 $p = \frac{1}{2}$ 時，每一分都等於擲一枚公平的硬幣，所以 w 應為 $\frac{1}{2}$。公式給出 $\frac{3/4}{3/2} = \frac{1}{2}$。
- 當 $p = 1$ 時，Xavier 必定贏得首兩分，所以 $w = 0$。公式給出 0。

- 表中六局的概率之和為 1。
- 對全部八個得分序列作精確枚舉（p 由 0.01 至 0.99），重現了 w；而在 $p = \frac{2}{3}$ 下模擬 200 000 場比賽，Younis 勝出的比例為 41.6%，對比 $\frac{5}{12} \approx 41.7\%$。

常見錯誤

- 把轉換規則弄反。概率 p 屬於贏得上一分的球員，而不是屬於 Xavier。
- 把錯誤的乘積設為相等，得出 $k = \frac{5}{7}$。公平的意思是 $\mathrm{P}(\text{X 勝}) \times 1 = \mathrm{P}(\text{Y 勝}) \times k$，所以較不可能勝出的一方得到較大的賠付。
- 在第 (iii) 部分把 $p = 0$ 代入公式，並聲稱每位球員的概率都是 $\frac{1}{2}$，但事實上比賽永不結束。

STEP 之外

這裏的得分構成一條馬可夫鏈（Markov chain）：每一分只取決於前一分。把比賽分拆為相同而獨立的局，是更新（renewal）論證的一個例子；而方程 $w = \mathrm{P}(A) + \mathrm{P}(C)w$ 是「首步分析」（first-step analysis）最簡單的例子，這是計算馬可夫鏈中吸收概率（absorption probabilities）的標準方法。第 (iii) 部分展示了吸收失敗時的情況：鏈永遠循環下去。Grimmett 與 Welsh 的 *Probability: An Introduction*（《概率導論》）以大學一年級的程度介紹馬可夫鏈及這些方法。

詳細解答

STEP II 2011，第 13 題

概率與統計 • 偏度的量度

STEP II 2011，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/11-s2-q13

關鍵思路

偏度（skewness）描述的是分佈的「不對稱」程度。這裏的三種量度從三個不同角度把它找出來：透過三階中心矩（third central moment）、透過中位數（median）在兩個十分位數（decile）之間的位置，以及透過平均值與中位數之間的差距。對於 $[0,1]$ 上的密度 $2x$，一切都可以明確算出：各階矩都只是一行就能完成的積分，而累積分佈函數（cumulative distribution function）是 x^2，所以它的反函數就是平方根。真正要花功夫的只是小心處理根式（surd）的運算，以及在最後找一個乾淨利落的方法比較它們的大小。

解答

偏度量度甚麼。偏度量度的是分佈相對其中心的*不對稱性*。對稱分佈的偏度為零。若較長的尾部在右邊，偏度為正；若較長的尾部在左邊，偏度為負。

第 (i) 部分

把立方展開，並利用期望值的線性性質（linearity of expectation），記住 μ 是常數：

$$\begin{aligned}\mathrm{E}\big((X-\mu)^3\big) &= \mathrm{E}(X^3) - 3\mu\,\mathrm{E}(X^2) + 3\mu^2\,\mathrm{E}(X) - \mu^3 \\ &= \mathrm{E}(X^3) - 3\mu(\sigma^2+\mu^2) + 3\mu^3 - \mu^3 \\ &= \mathrm{E}(X^3) - 3\mu\sigma^2 - \mu^3,\end{aligned}$$

這裏用了 $\mathrm{E}(X)=\mu$ 和 $\mathrm{E}(X^2)=\sigma^2+\mu^2$（即方差（variance）的常用公式移項後的形式）。除以 σ^3，便得到所求的 γ 表達式。

密度為 $2x$ 的分佈。對於 $k \geq 0$，

$$\mathrm{E}(X^k) = \int_0^1 x^k \cdot 2x \,\mathrm{d}x = \frac{2}{k+2}.$$

所以 $\mu = \frac{2}{3}$、$\mathrm{E}(X^2) = \frac{1}{2}$ 及 $\mathrm{E}(X^3) = \frac{2}{5}$。因此

$$\sigma^2 = \tfrac{1}{2} - \tfrac{4}{9} = \tfrac{1}{18}, \qquad \sigma = \frac{1}{3\sqrt{2}}, \qquad \sigma^3 = \frac{1}{27 \cdot 2\sqrt{2}} = \frac{1}{54\sqrt{2}}.$$

分子為

$$\mathrm{E}(X^3) - 3\mu\sigma^2 - \mu^3 = \frac{2}{5} - 3 \cdot \frac{2}{3} \cdot \frac{1}{18} - \frac{8}{27} = \frac{2}{5} - \frac{1}{9} - \frac{8}{27} = \frac{54 - 15 - 40}{135} = -\frac{1}{135}.$$

因此

$$\gamma = -\frac{1}{135} \cdot 54\sqrt{2} = -\frac{54\sqrt{2}}{135} = -\frac{2\sqrt{2}}{5},$$

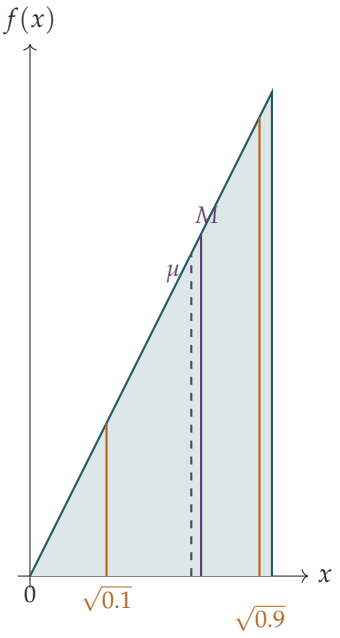
正如所求。負號是合理的：密度集中在 $x = 1$ 附近，在左邊留下一條伸向 0 的長尾。

第 (ii) 部分

對於 $0 \leq x \leq 1$，累積分佈函數為

$$\mathrm{F}(x) = \int_0^x 2t \,\mathrm{d}t = x^2,$$

它在 $[0,1]$ 上遞增，所以對於 $0 \leq p \leq 1$，其反函數為 $\mathrm{F}^{-1}(p) = \sqrt{p}$。



十分位數偏度。把 D 的分子和分母同乘以 $\sqrt{10}$，便可消去所有根式，只剩下一個：

$$D = \frac{\sqrt{\frac{9}{10}} - 2\sqrt{\frac{1}{2}} + \sqrt{\frac{1}{10}}}{\sqrt{\frac{9}{10}} - \sqrt{\frac{1}{10}}} = \frac{3 - 2\sqrt{5} + 1}{3 - 1} = \frac{4 - 2\sqrt{5}}{2} = 2 - \sqrt{5},$$

因為 $\sqrt{10} \cdot \sqrt{\frac{1}{2}} = \sqrt{5}$。這正是要證明的結果。

皮爾遜偏度。中位數滿足 $\mathrm{F}(M) = \frac{1}{2}$，所以 $M = \frac{1}{\sqrt{2}}$。利用 $\frac{1}{\sigma} = 3\sqrt{2}$，

$$P = 3\Big(\frac{2}{3} - \frac{1}{\sqrt{2}}\Big) \cdot 3\sqrt{2} = 9\sqrt{2} \cdot \frac{2}{3} - 9 = 6\sqrt{2} - 9.$$

數值上，$D \approx -0.236$、$P \approx -0.515$ 及 $\gamma \approx -0.566$，所以大小次序看來是對的；現在我們加以證明。

> ? **想一想** 不用計算機，怎樣比較 $2 - \sqrt{5}$ 和 $6\sqrt{2} - 9$ 這類數的大小？

先移項使兩邊都是正數，然後平方：對於正數，$a < b$ 當且僅當 $a^2 < b^2$。以下每一步都是可逆的，所以最後的陳述成立，便證明了最初的陳述。

$D > P$。我們需要證明 $2 - \sqrt{5} > 6\sqrt{2} - 9$，即 $\sqrt{5} + 6\sqrt{2} < 11$。兩邊都是正數，所以這等價於

$$\begin{aligned}(\sqrt{5} + 6\sqrt{2})^2 = 77 + 12\sqrt{10} < 121 &\iff 12\sqrt{10} < 44 \\ &\iff 3\sqrt{10} < 11 \\ &\iff 90 < 121,\end{aligned}$$

其中最後一步再次把兩個正數平方。由於 $90 < 121$ 成立，所以 $D > P$。

$P > \gamma$。我們需要證明 $6\sqrt{2} - 9 > -\frac{2\sqrt{2}}{5}$。乘以 5 並移項，即 $32\sqrt{2} > 45$。兩邊都是正數，所以這等價於 $32^2 \cdot 2 > 45^2$，即 $2048 > 2025$，這是成立的。所以 $P > \gamma$。

因此 $D > P > \gamma$。三種量度都是負數，一致認為這個分佈向左偏，但對偏的程度卻各有不同的看法。

驗算

- (i) 中的恆等式對任何對稱分佈都給出 0：例如，若 X 以相等概率取 ± 1，則 $\mu = 0$、$\sigma = 1$ 及 $\mathrm{E}(X^3) = 0$。
- 數值上，用中點法則計算 $\mathrm{E}\big((X-\mu)^3\big)/\sigma^3$，得到 -0.56569，與 $-2\sqrt{2}/5$ 吻合。
- 以 U 為 $[0,1]$ 上的均勻分佈，模擬 $X = \sqrt{U}$（它滿足 $\mathrm{F}(x) = x^2$），得到的樣本中位數接近 0.707，十分位數接近 0.316 和 0.949，與草圖一致。

- 平均值 $\frac{2}{3}$ 位於中位數 $\frac{1}{\sqrt{2}} \approx 0.707$ 的左邊，這是左偏分佈的典型特徵，也是 $P < 0$ 的原因。

常見錯誤

- 把 $\sigma^3 = (\frac{1}{18})^3$ 寫成這樣，而不是 $(\frac{1}{18})^{3/2}$：要立方的是標準差（standard deviation），不是方差。
- 在其中一邊可能是負數的情況下把不等式平方。應先移項使兩邊都是正數。
- 在 γ 的公式或 P 的分母中，該用 $\sigma^2 = \frac{1}{18}$ 的地方卻用了 $\mathrm{E}(X^2) = \frac{1}{2}$。

STEP 之外

$\mathrm{E}\big((X-\mu)^3\big)$ 這個量是三階中心矩，而 (i) 中的展開正是矩與累積量（cumulant）之間關係的起點；對於獨立隨機變量之和，累積量可以簡單地相加。像 D 這類以分位數（quantile）為基礎的量度在統計學中備受重視，因為它們是穩健的（robust）：單一個極端的觀測值幾乎不會移動十分位數，卻可以大幅改變三階矩。Grimmett 和 Welsh 的 *Probability: An Introduction*（《概率導論》）以合適的程度講解矩與矩母函數（moment generating function）。

詳細解答

STEP III 2011，第 1 題

純數學 • 降階法

STEP III 2011，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/11-s3-q1

關鍵思路

第 (i) 部分是一條可分離方程（separable equation），而它的出現是有原因的：把 $y = ze^{-x}$ 代入 $(*)$ 後，得出的正正就是這條方程。含 z 本身的項會互相抵消，只剩下一條只涉及 z' 和 z'' 的方程，而對 $u = z'$ 來說它是一階的。這就是*降階法*（reduction of order）：只要知道一個解（e^{-x}），便可以求出另一個解。第 (iii) 部分則在右方不為零的情況下作同樣的代換，結果變成一條線性（linear）一階方程，可用積分因子（integrating factor）求解。

解答

全題我們都在 $x \neq -1$ 的區間上討論，因此可以除以 $x+1$。

第 (i) 部分

由於

$$\frac{x+2}{x+1} = 1 + \frac{1}{x+1},$$

分離變數（對 $u \neq 0$）得

$$\int \frac{\mathrm{d}u}{u} = \int \left(1 + \frac{1}{x+1}\right) \mathrm{d}x \quad \Longrightarrow \quad \ln|u| = x + \ln|x+1| + c.$$

取指數，得 $|u| = e^c\,|x+1|\,e^x$。正負號與因子 e^c 可合併成一個任意常數，而 $u = 0$ 亦是一個解，故通解（general solution）為

$$u = C(x+1)e^x, \qquad C \text{ 為任意常數。}$$

第 (ii) 部分

設 $y = z\mathrm{e}^{-x}$，由乘積法則（product rule）得

$$y' = (z' - z)\mathrm{e}^{-x},$$
$$y'' = (z'' - z')\mathrm{e}^{-x} - (z' - z)\mathrm{e}^{-x} = (z'' - 2z' + z)\mathrm{e}^{-x}.$$

代入 $(*)$，並消去非零因子 e^{-x}：

$$\begin{aligned} 0 &= (x+1)(z'' - 2z' + z) + x(z' - z) - z \\ &= (x+1)z'' + \big(-2(x+1) + x\big)z' + \big((x+1) - x - 1\big)z \\ &= (x+1)z'' - (x+2)z'. \end{aligned}$$

所有 z 項都完全抵消了。記 $u = \dfrac{\mathrm{d}z}{\mathrm{d}x}$，再除以 $x+1$，得

$$\frac{\mathrm{d}u}{\mathrm{d}x} - \left(\frac{x+2}{x+1}\right)u = 0,$$

這是一條關於 $\dfrac{\mathrm{d}z}{\mathrm{d}x}$ 的一階方程，正如題目所要求。它恰好就是第 (i) 部分的方程。

想一想 為甚麼 z 項會消失？我們能否事先預料到？

z 的係數就是代入 $z = 1$ 所得的結果，即在 $(*)$ 中檢驗 $y = \mathrm{e}^{-x}$。由於 $(x+1)\mathrm{e}^{-x} - x\mathrm{e}^{-x} - \mathrm{e}^{-x} = 0$，$y = \mathrm{e}^{-x}$ 本身已經是一個解，所以常數 z 必定滿足方程，因此不可能剩下單獨的 z 項。這正是選用這個代換的原因。

由第 (i) 部分，$z' = C(x+1)\mathrm{e}^{x}$。用分部積分（integration by parts），

$$\int (x+1)\mathrm{e}^{x}\,\mathrm{d}x = (x+1)\mathrm{e}^{x} - \int \mathrm{e}^{x}\,\mathrm{d}x = x\mathrm{e}^{x} + \text{常數},$$

所以

$$z = Cx\mathrm{e}^{x} + B.$$

因此 $y = z\mathrm{e}^{-x} = Cx + B\mathrm{e}^{-x}$，把 C 改名為 A，得

$$y = Ax + B\mathrm{e}^{-x},$$

其中 A 和 B 為任意常數。$(*)$ 的每一個解都可以這樣得到，因為任何 y 都可以寫成 $z\mathrm{e}^{-x}$，其中 $z = y\mathrm{e}^{x}$；所以這就是通解，正如題目所要求。

第 (iii) 部分

同樣的代換只會改變右方，右方變成 $(x+1)^2$ 除以 e^{-x}：

$$(x+1)z'' - (x+2)z' = (x+1)^2\mathrm{e}^{x}.$$

設 $u = z'$，再除以 $x+1$，得

$$\frac{\mathrm{d}u}{\mathrm{d}x} - \left(\frac{x+2}{x+1}\right)u = (x+1)\mathrm{e}^x.$$

這是線性方程。由第 (i) 部分，$(x+1)\mathrm{e}^x$ 是齊次（homogeneous）方程的解，所以它的倒數 $\dfrac{\mathrm{e}^{-x}}{x+1}$ 是一個積分因子。（它在相差正負號的意義下等於 $\exp\left(-\int \frac{x+2}{x+1}\,\mathrm{d}x\right)$。）兩邊同乘，得

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{u\,\mathrm{e}^{-x}}{x+1}\right) = 1 \quad\Longrightarrow\quad u = (x+1)\mathrm{e}^x(x+k)$$

其中 k 為常數。所以

$$z' = (x^2+x)\mathrm{e}^x + k(x+1)\mathrm{e}^x.$$

第二項積分後得 $kx\mathrm{e}^x$，與第 (ii) 部分相同。至於第一項，用兩次分部積分，

$$\begin{aligned}\int (x^2+x)\mathrm{e}^x\,\mathrm{d}x &= (x^2+x)\mathrm{e}^x - \int (2x+1)\mathrm{e}^x\,\mathrm{d}x \\ &= (x^2+x)\mathrm{e}^x - (2x+1)\mathrm{e}^x + 2\mathrm{e}^x + \text{常數} = (x^2-x+1)\mathrm{e}^x + \text{常數}.\end{aligned}$$

因此 $z = (x^2-x+1)\mathrm{e}^x + kx\mathrm{e}^x + B$，而

$$y = z\mathrm{e}^{-x} = x^2 - x + 1 + kx + B\mathrm{e}^{-x}.$$

$-x$ 可併入任意的 kx 中，所以記 $A = k-1$，通解為

$$y = x^2 + 1 + Ax + B\mathrm{e}^{-x},$$

其中 A 和 B 為任意常數。

另一方法：由於右方是二次式，我們可以嘗試特解（particular integral）$y = px^2 + qx + r$。代入得 $(x+1)2p + x(2px+q) - (px^2+qx+r) = px^2 + 2px + 2p - r$，與 x^2+2x+1 比較係數，得 $p = 1$、$r = 1$，而 q 可任意取值（因為 x 是一個餘函數（complementary function））。再加上第 (ii) 部分的餘函數，便得出相同的答案。

驗算

直接代入 $y = x^2+1$：$y' = 2x$，$y'' = 2$，所以左方為 $2(x+1) + 2x^2 - x^2 - 1 = x^2 + 2x + 1 = (x+1)^2$，符合要求。數值上，隨機選取 A、B 及 $x \in (-0.9, 3)$，以有限差分（finite difference）計算 $y = x^2 + 1 + Ax + B\mathrm{e}^{-x}$ 的導數，結果滿足方程的誤差約在 3×10^{-6} 以內；同樣的測試亦驗證了第 (i) 和第 (ii) 部分。

常見錯誤

- 把 y'' 算錯：它是 $(z'' - 2z' + z)\mathrm{e}^{-x}$，$z$ 前面是*加*號；這裏若出現正負號錯誤，便會留下一個多餘的 z 項。

- 在第 (iii) 部分消去因子 e^{-x} 時，忘記把右方乘以 e^{x}。
- 漏掉常數：最終答案需要兩個任意常數，每次積分各產生一個。

STEP 之外

降階法是一種通用技巧：若 y_1 是某條線性二階方程的解，代入 $y = y_1 v$ 必定會消去 v 項，剩下一條關於 v' 的一階方程。這亦是*參數變易法*（variation of parameters）背後的思路，該方法對任何右方都能求出特解。兩者在 Boyce 和 DiPrima 的 *Elementary Differential Equations*（《初等微分方程》）中都有清晰的講解。

詳細解答

STEP III 2011，第 2 題

純數學 • 多項式的有理根

> **STEP III 2011，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/11-s3-q2

關鍵思路

將 $\mathrm{f}(p/q) = 0$ 乘以 q^{n-1}，除了一個分母外，其餘分母全部消去：首項變成 p^n/q，而其餘各項都是整數。因此 p^n/q 必須是整數；由於 p 與 q 沒有公因數，這就迫使 $q = 1$。此後，每一部分都變成關於整數根的問題，而這些根可以從大小、整除性，或者更巧妙地從奇偶性（parity）加以排除。

解答

開首的結果

設 $\mathrm{f}(p/q) = 0$，其中 p、q 為互質（coprime）整數，且 $q > 0$。那麼

$$0 = q^{n-1}\,\mathrm{f}\left(\frac{p}{q}\right) = \frac{p^n}{q} + a_{n-1}p^{n-1} + a_{n-2}p^{n-2}q + \cdots + a_1 p\,q^{n-2} + a_0 q^{n-1},$$

這是因為項 $a_k(p/q)^k$ 變成 $a_k p^k q^{n-1-k}$，而對每個 $k \le n-1$ 都有 $n-1-k \ge 0$。首項以外的每一項都是整數，所以

$$\frac{p^n}{q} = -\left(a_{n-1}p^{n-1} + \cdots + a_0 q^{n-1}\right)$$

是整數；即 q 整除 p^n。

> ? **想一想** 為何 $q \mid p^n$ 會迫使 $q = 1$？

若 $q > 1$，則它有某個質因數 r。於是 r 整除 p^n；由於 r 是質數，它必定整除 p 本身。但這樣 r 便是 p 與 q 一個大於 1 的公因數，這是不可能的。所以 q 沒有任何質因數，又因

$q > 0$，得

$$q = 1.$$

因此 $\mathrm{f}(x) = 0$ 的每個有理根 p/q 都等於 p，即一個整數。（留意這裏正是用上了 f 是首一（monic）多項式這一點：若首項係數為 a_n，就會剩下 $a_n p^n/q$，而我們只能推斷 $q \mid a_n$。）

以下兩個簡單事實會有幫助。若 m 是一個整數根，則

$$m\left(m^{n-1} + a_{n-1}m^{n-2} + \cdots + a_1\right) = -a_0,$$

所以 m 整除 a_0；又因 $a_0 \neq 0$，故 $m \neq 0$。

第 (i) 部分

數 $2^{1/n}$ 是 $x^n - 2 = 0$ 的根，而這個多項式是首一的、係數為整數，且 $a_0 = -2 \neq 0$。因此若 $2^{1/n}$ 是有理數，它便會是一個滿足 $m^n = 2$ 的整數 m。但

$$1^n = 1 < 2 < 4 \leq 2^n,$$

所以正實數 $2^{1/n}$ 嚴格地位於相鄰整數 1 與 2 之間，不可能是整數。因此對每個 $n \geq 2$，$2^{1/n}$ 都是無理數。

第 (ii) 部分

這裏 $n = 3$，$a_0 = 1$。任何有理根都是整除 1 的整數 m，所以 $m = \pm 1$。但

$$1^3 - 1 + 1 = 1 \neq 0, \qquad (-1)^3 - (-1) + 1 = 1 \neq 0.$$

所以 $x^3 - x + 1 = 0$ 沒有有理根。

（另一方法：$m^3 - m = (m-1)m(m+1)$ 是一個包含兩個相鄰整數的乘積，所以是偶數，於是 $m^3 - m + 1$ 是奇數，永不為零。這個奇偶性的想法正是第 (iii) 部分所需要的。）

第 (iii) 部分

多項式 $x^n - 5x + 7$ 是首一的、係數為整數，常數項 $7 \neq 0$，所以任何有理根都是整數 m。我們證明 $m^n - 5m + 7$ 恆為奇數。

- 若 m 為偶數，則 m^n 與 $5m$ 都是偶數，所以 $m^n - 5m + 7$ 是偶數 − 偶數 + 奇數，即奇數。
- 若 m 為奇數，則 m^n 與 $5m$ 都是奇數，所以 $m^n - 5m$ 是偶數，而 $m^n - 5m + 7$ 是奇數。

無論哪種情況，$m^n - 5m + 7$ 都是奇數，特別地它不等於 0。因此對任何 $n \geq 2$，$x^n - 5x + 7 = 0$ 都沒有整數根，從而沒有有理根。

（若改為逐一檢查 7 的因數，便要證明對 $m = \pm 1, \pm 7$ 及每一個 n 都有 $m^n - 5m + 7 \neq 0$；這樣做也行得通，但較為繁瑣；奇偶性則一次過處理所有 n。）

驗算

- 試一個*確實*有有理根的多項式，例如 $x^2 - x - 6 = (x-3)(x+2)$：根 3 與 -2 正如所料都是整數，而且都整除 $a_0 = -6$。
- 非首一的多項式可以有非整數的有理根，例如 $2x - 1$；這就是首項係數為 1 之所以重要的原因。
- 數值上，$2^{1/2} \approx 1.414$ 與 $2^{1/3} \approx 1.260$ 都位於 1 與 2 之間；而 $x^3 - x + 1$ 在 $x = \pm 1$ 時取值 1。以電腦搜尋 3000 個隨機的首一整係數多項式，所找到的有理根全是整數；並確認了對 $|m| \leq 200$、$2 \leq n \leq 14$，$m^n - 5m + 7$ 都是奇數。

常見錯誤

- 乘以 q^n 而非 q^{n-1}：你會得到 $p^n = -q(\cdots)$，這樣也行得通，但之後必須說明 $q \mid p^n$；若乘以 q^{n-1}，分數 p^n/q 便令這一點一目了然。
- 斷言「$q \mid p^n$，所以 $q \mid p$」而不說明原因；論證須經由 q 的一個質因數。
- 在 (iii) 中只檢查了幾個 n 值。論證必須同時對所有 $n \geq 2$ 成立。

STEP 之外

這是首一形式的*有理根定理*（rational root theorem）：若一個有理數是某個首一整係數多項式的根，則它是整數。用代數數論（algebraic number theory）的語言來說，整數是*整閉*（integrally closed）的，而首一整係數多項式的根稱為代數整數（algebraic integers）。(iii) 中的奇偶性技巧，是將多項式模一個質數（modulo a prime）約化的最簡單情況；這是證明多項式沒有根或不可約（irreducible）的有力工具。Davenport 的 *The Higher Arithmetic*（《高等算術》）是很好的下一步讀物，而 Allenby 的 *Rings, Fields and Groups*（《環、域與羣》）則發展了代數方面的內容。

詳細解答

STEP III 2011，第 3 題

純數學 • 解三次方程

STEP III 2011，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/11-s3-q3

關鍵思路

兩個立方 $(x-\alpha)^3$ 和 $(x-\beta)^3$ 為我們提供兩個可自由選取的常數 a 和 b，而一個沒有 x^2 項的首一三次多項式（monic cubic）要符合四個條件。因此我們用其中兩個條件求出 a 和 b，其餘兩個條件就必須恰好是二次方程所給出、關於 $\alpha+\beta$ 和 $\alpha\beta$ 的事實。一旦把三次式寫成兩個立方之和（sum of two cubes），令它等於零便化為「某個比的立方等於一個常數」，只要取立方根（cube root）便可解出，而且三個立方根都要取。

解答

把三次式寫成兩個立方之和

由於 $p\neq 0$，二次方程 $pt^2-qt+p^2=0$ 滿足

$$\alpha+\beta=\frac{q}{p},\qquad \alpha\beta=\frac{p^2}{p}=p,$$

而其判別式（discriminant）為 q^2-4p^3。因此條件 $q^2\neq 4p^3$ 正好表示 $\alpha\neq\beta$。此外 $\alpha\beta=p\neq 0$，故兩個根都不為零。

展開得

$$a(x-\alpha)^3+b(x-\beta)^3=(a+b)x^3-3(a\alpha+b\beta)x^2+3(a\alpha^2+b\beta^2)x-(a\alpha^3+b\beta^3).$$

要與 $x^3-3px+q$ 的係數相符，我們需要

$$a+b=1,\qquad a\alpha+b\beta=0,\qquad a\alpha^2+b\beta^2=-p,\qquad a\alpha^3+b\beta^3=-q.$$

頭兩條是 a 和 b 的線性方程。把 $b=1-a$ 代入第二條，得 $a(\alpha-\beta)=-\beta$；由於 $\alpha\neq\beta$，

可以兩邊相除：

$$a = \frac{\beta}{\beta - \alpha}, \qquad b = \frac{\alpha}{\alpha - \beta}.$$

現在驗證其餘兩個條件會自動成立：

$$a\alpha^2 + b\beta^2 = \frac{\beta\alpha^2 - \alpha\beta^2}{\beta - \alpha} = \frac{\alpha\beta(\alpha - \beta)}{\beta - \alpha} = -\alpha\beta = -p,$$
$$a\alpha^3 + b\beta^3 = \frac{\beta\alpha^3 - \alpha\beta^3}{\beta - \alpha} = \frac{\alpha\beta(\alpha - \beta)(\alpha + \beta)}{\beta - \alpha} = -\alpha\beta(\alpha + \beta) = -p \cdot \frac{q}{p} = -q.$$

所以，取這兩個 a 和 b 的值，

$$x^3 - 3px + q = \frac{\beta}{\beta - \alpha}(x - \alpha)^3 + \frac{\alpha}{\alpha - \beta}(x - \beta)^3,$$

此即所求。

解 $x^3 - 24x + 48 = 0$

這裏 $p = 8$、$q = 48$，而 $q^2 = 2304 \neq 2048 = 4p^3$，所以上述結果適用。對應的二次方程是 $8t^2 - 48t + 64 = 0$，即 $t^2 - 6t + 8 = 0$，其根為 2 和 4。取 $\alpha = 2$、$\beta = 4$；則 $a = 2$、$b = -1$，於是

$$x^3 - 24x + 48 = 2(x - 2)^3 - (x - 4)^3.$$

（快速展開即可確認：$2x^3 - 12x^2 + 24x - 16 - x^3 + 12x^2 - 48x + 64$。）

? **想一想** 我們怎樣把 $2(x - 2)^3 = (x - 4)^3$ 變成可以求解的形式？

兩邊除以 $(x - 4)^3$。這樣做是容許的，因為 $x = 4$ 不是根（$64 - 96 + 48 = 16$）。方程變成

$$\left(\frac{x - 2}{x - 4}\right)^3 = \frac{1}{2}.$$

記 $c = 2^{1/3}$。立方等於 $\frac{1}{2}$ 的複數為 $k = c^{-1}$、$c^{-1}\omega$ 和 $c^{-1}\omega^2$。對每一個值，解 $\dfrac{x - 2}{x - 4} = k$：

$$x - 2 = kx - 4k \quad \Longrightarrow \quad x = \frac{2 - 4k}{1 - k},$$

由於 $k \neq 1$，這一步是成立的。映射 $k \mapsto x$ 是一一對應（one-to-one）的，所以 k 的三個值給出三個不同的根，它們必然就是該三次方程的全部根。

當 $k = c^{-1}$ 時，分子分母同乘以 c：

$$x = \frac{2c - 4}{c - 1} = \frac{2(2 - 2^{1/3})}{1 - 2^{1/3}},$$

這正是題目所給的解。當 $k = c^{-1}\omega$ 時，分子分母同乘以 $c\,\omega^{-1} = c\,\omega^2$（利用 $\omega^3 = 1$）：

$$x = \frac{2c\omega^2 - 4}{c\omega^2 - 1} = \frac{2(2 - 2^{1/3}\omega^2)}{1 - 2^{1/3}\omega^2},$$

同理，$k = c^{-1}\omega^2$ 給出

$$x = \frac{2(2 - 2^{1/3}\omega)}{1 - 2^{1/3}\omega}.$$

所以三個解為 $x = \dfrac{2(2 - 2^{1/3}\omega^j)}{1 - 2^{1/3}\omega^j}$，其中 $j = 0, 1, 2$。

$p = r^2$、$q = 2r^3$ 的情況

此時 $q^2 = 4r^6 = 4p^3$，正是上述方法所排除的情況：二次式 $r^2t^2 - 2r^3t + r^4 = r^2(t - r)^2$ 有重根（repeated root）$t = r$，我們無法求出 a 和 b。不過，二次方程有重根，暗示三次方程也有重根。事實上，代入 $x = r$ 得 $r^3 - 3r^3 + 2r^3 = 0$，而導數（derivative）$3x^2 - 3r^2$ 在 $x = r$ 處亦為零，所以 $(x - r)^2$ 是一個因式。比較 x^3 項和常數項，得

$$x^3 - 3r^2x + 2r^3 = (x - r)^2(x + 2r).$$

根為 $x = r$（二重根）和 $x = -2r$。

> **想一想** 我們能否從先前的公式得出這些根？

不能。當 $\alpha = \beta$ 時，兩個立方合而為一，所以 $a(x - \alpha)^3 + b(x - \alpha)^3$ 只可能是 $(x - \alpha)^3$，它沒有 x 項可與 $-3px$ 相配。這就是我們需要另一種方法的原因。把三次式寫成 $(x - r)$ 的冪，便可看出實際情況：$x^3 - 3r^2x + 2r^3 = (x - r)^3 + 3r(x - r)^2$，所以二重根正好位於二次方程的重根 $t = r$ 處。

驗算

- 實根為 $2(2 - 1.259921)/(1 - 1.259921) \approx -5.694644$，代入後得 $x^3 - 24x + 48 \approx 0$，誤差約為 10^{-13}。另外兩個根為 $2.847322 \pm 0.567212\,\mathrm{i}$，殘差同樣為零。
- 三個根之和為 0（因為沒有 x^2 項），積為 -48：$-5.694644 \times (2.847322^2 + 0.567212^2) \approx -48.0$。
- 一個實根加兩個複根，與理論相符：$x^3 - 3px + q$ 只有在 $q^2 < 4p^3$ 時才有三個實根，而這裏 $q^2 > 4p^3$。
- 該恆等式（identity）可以用數值檢驗：我們對 2000 組隨機的 (p, q)，在隨機複數 x 處作了檢驗，並對多個 r 值檢驗了因式分解 $(x - r)^2(x + 2r)$。

常見錯誤

- 只取 $\frac{1}{2}$ 的實立方根，因而只找到一個解。三次方程有三個根，它們分別來自三個立方根。
- 搞錯 a、b 各自對應哪個根：a（即 $(x-\alpha)^3$ 的係數）的分子是 β。
- 在消去分母時把 ω 和 ω^2 對調。由於這兩個答案是一對共軛（conjugate pair），無論哪種做法，最終得到的根集合都相同。

STEP 之外

這是十六世紀解三次方程方法的一個版本，通常歸功於卡爾達諾（Cardano）和塔爾塔利亞（Tartaglia）：先化為沒有 x^2 項的「缺項」三次方程（depressed cubic），再解一個輔助二次方程（即預解式（resolvent））。伽羅瓦理論（Galois theory）解釋了為何三次和四次方程有這樣的預解式，而一般的五次方程卻沒有。Allenby 的 *Rings, Fields and Groups*（《環、域與羣》）為通往這些概念提供了一條易讀的路徑。

詳細解答

STEP III 2011，第 4 題

純數學 • 楊氏不等式

> **STEP III 2011，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/11-s3-q4

本題在 *Beyond STEP* 第 3.4 節以面積證明不等式中有完整的引導式講解：https://beyondstep.keithmcnulty.org/zh-hant/ch03/areas-inequalities.html

關鍵思路

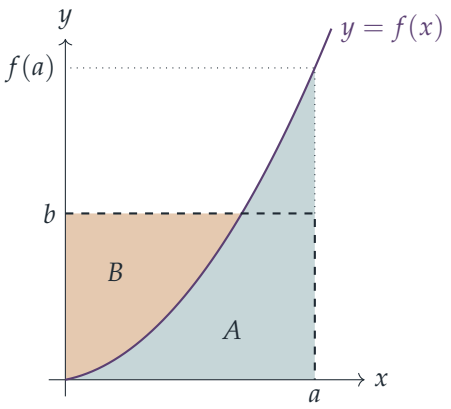
$(*)$ 中的兩個積分（integral）都是同一幅圖上的面積：$\int_0^a f(x)\,\mathrm{d}x$ 是曲線與 x 軸之間的面積，而 $\int_0^b f^{-1}(y)\,\mathrm{d}y$ 是曲線與 y 軸之間的面積。兩者合起來總能覆蓋以 $(0,0)$ 和 (a,b) 為頂角的矩形，而且除非頂角 (a,b) 剛好落在曲線上，否則還會有剩餘。這樣一來，第 (ii) 和第 (iii) 部分只不過是選取適當的 f，再計算兩個積分而已。

解答

第 (i) 部分

由於 f 連續（continuous）且斜率（gradient）為正，它是嚴格遞增（strictly increasing）的，因此反函數（inverse function）f^{-1} 存在，而且同樣遞增；又 $f(0) = 0$ 表示曲線 $y = f(x)$ 從原點出發。

設 A 為 $0 \le x \le a$ 範圍內曲線與 x 軸之間的區域，其面積為 $\int_0^a f(x)\,\mathrm{d}x$。設 B 為 $0 \le y \le b$ 範圍內曲線與 y 軸之間的區域；對 y 積分，其面積為 $\int_0^b f^{-1}(y)\,\mathrm{d}y$。這兩個區域位於曲線的兩側，所以不會重疊。



先假設 $b \le f(a)$，如圖所示。矩形 $0 \le x \le a$、$0 \le y \le b$ 中的每一點，要麼位於曲線上或其下方（因而屬於 A），要麼位於曲線上或其左方（因而屬於 B）。所以 A 和 B 合起來覆蓋了面積為 ab 的矩形，$(*)$ 成立。多出的部分是 A 中位於直線 $y = b$ 以上的部分，即 $f^{-1}(b) \le x \le a$ 範圍內夾在該直線與曲線之間的狹長區域。

若改為 $b \ge f(a)$，只需把兩者的角色對調，同樣的論證仍然成立：矩形被覆蓋，而這次多出的部分是 B 中位於直線 $x = a$ 右方的部分。

因此 $(*)$ 在所有情況下都成立，而等號成立需要多出部分的面積為零。當 $a, b > 0$ 且 $b \ne f(a)$ 時，多出的部分由曲線和一條直線圍成，兩者只在一端相交，所以其面積為正。因此，對於 $a > 0$ 及 $b > 0$，

$$\text{等號成立當且僅當} b = f(a),$$

即矩形的頂角 (a, b) 落在曲線上的時候。

第 (ii) 部分

當 $p > 1$ 時，$f(x) = x^{p-1}$ 在 $x \ge 0$ 上連續，$f(0) = 0$，並在該範圍內嚴格遞增，所以 $(*)$ 適用。（其斜率 $(p-1)x^{p-2}$ 在 $x > 0$ 時為正；若 $p < 2$，斜率在 0 處趨於無窮大，但這不影響面積的論證。）其反函數為 $f^{-1}(y) = y^{1/(p-1)}$。

條件 $\frac{1}{p} + \frac{1}{q} = 1$ 給出 $\frac{1}{q} = \frac{p-1}{p}$，所以

$$q = \frac{p}{p-1}, \qquad \frac{1}{p-1} + 1 = \frac{p}{p-1} = q.$$

於是 $f^{-1}(y) = y^{q-1}$，而兩個積分分別為

$$\int_0^a x^{p-1}\,\mathrm{d}x = \frac{a^p}{p}, \qquad \int_0^b y^{q-1}\,\mathrm{d}y = \frac{b^q}{q}.$$

代入 $(*)$ 得

$$ab \le \frac{a^p}{p} + \frac{b^q}{q},$$

即為所求。

想一想 等號條件 $b = f(a)$ 在這裏代表甚麼？又應如何化簡 b^q？

條件是 $b = a^{p-1}$。由於 $(p-1)q = p$，故 $b^q = a^{(p-1)q} = a^p$。所以

$$\frac{a^p}{p} + \frac{b^q}{q} = a^p\left(\frac{1}{p} + \frac{1}{q}\right) = a^p = a \cdot a^{p-1} = ab,$$

等號確實成立。

第 (iii) 部分

取 $0 \le x \le \frac{1}{2}\pi$ 上的 $f(x) = \sin x$。它是連續的，$f(0) = 0$，而其斜率 $\cos x$ 在 $0 \le x < \frac{1}{2}\pi$ 時為正，所以它嚴格遞增（斜率只在端點處為零，這不影響論證）。它把 $[0, \frac{1}{2}\pi]$ 映射到 $[0,1]$ 上，反函數為 arcsin，而這正好是 a 和 b 容許的取值範圍。

第一個積分是

$$\int_0^a \sin x \,\mathrm{d}x = 1 - \cos a.$$

至於第二個積分，取 $u = \arcsin y$ 及 $\mathrm{d}v = \mathrm{d}y$，用分部積分法（integration by parts）：

$$\begin{aligned}\int_0^b \arcsin y \,\mathrm{d}y &= \Big[y \arcsin y\Big]_0^b - \int_0^b \frac{y}{\sqrt{1-y^2}}\,\mathrm{d}y \\ &= b \arcsin b + \Big[\sqrt{1-y^2}\Big]_0^b \\ &= b \arcsin b + \sqrt{1-b^2} - 1.\end{aligned}$$

兩者相加，兩個 1 互相抵消，$(*)$ 變成

$$ab \le b \arcsin b + \sqrt{1-b^2} - \cos a,$$

這就是所求的不等式（inequality），等號在 $b = \sin a$ 時成立。

想一想 a 可以自由選取。選哪一個值才能得到我們想要的不等式？

對於 $t \ge 1$，令 $b = t^{-1}$，它按要求落在 $(0,1]$ 內。把不等式乘以 $t > 0$，並利用 $t\sqrt{1-t^{-2}} = \sqrt{t^2-1}$（因為 $t > 0$，此式成立），得

$$a \le \arcsin(t^{-1}) + \sqrt{t^2-1} - t\cos a \quad \text{對所有} a \in [0, \tfrac{1}{2}\pi].$$

目標不等式中沒有 a，而右方有一個 t，這正是 $t\cos a$ 在 $a = 0$ 時的值。取 $a = 0$：

$$0 \le \arcsin(t^{-1}) + \sqrt{t^2-1} - t, \qquad \text{即} \qquad \arcsin(t^{-1}) \ge t - \sqrt{t^2-1} \quad (t \ge 1).$$

驗算

- 當 $t = 1$ 時，最後的不等式變成 $\frac{1}{2}\pi \geq 1$；當 $t = 2$ 時，則變成 $\frac{1}{6}\pi \approx 0.5236 \geq 2 - \sqrt{3} \approx 0.2679$。
- 當 $t \to \infty$ 時，兩邊都是 $\frac{1}{t}$ 的量級：$\arcsin(t^{-1}) \approx t^{-1} + \frac{1}{6}t^{-3}$，而 $t - \sqrt{t^2 - 1} = \frac{1}{t + \sqrt{t^2 - 1}} \approx \frac{1}{2t}$。所以這個下界約為真實值的一半。
- 我們在 2×10^5 個隨機點 (a, b) 上測試了第 (iii) 部分的不等式；它從未失效，而兩邊的差距只在曲線 $b = \sin a$ 附近才縮小至零。我們亦以隨機的 p、a 和 b 驗證了楊氏不等式（Young's inequality），並用辛普森法則（Simpson's rule）核對了兩個積分。

常見錯誤

- 在第 (i) 部分只畫出 $b \leq f(a)$ 的情況。$b \geq f(a)$ 的情況是對稱的，但你應該明確指出這一點。
- 寫出 $f^{-1}(y) = y^{1/(p-1)}$ 之後，沒有察覺 $1/(p-1) = q - 1$，以致得不出積分 b^q/q。
- 在最後一步中，試圖尋找使界限最佳的 a。那個 a 是 $\arcsin(t^{-1})$，但它只會給出一個平凡的恆等式。最簡單的選擇 $a = 0$ 才是行得通的那一個。

STEP 之外

第 (ii) 部分就是楊氏不等式，它是推導赫爾德(Hölder)不等式 $\sum |x_i y_i| \leq \left(\sum |x_i|^p\right)^{1/p} \left(\sum |y_i|^q\right)^{1/q}$ 的標準途徑，而赫爾德不等式又是分析學中 L^p 空間（L^p spaces）的基礎。把 $\int f$ 與 $\int f^{-1}$ 配對的做法，正是勒讓德變換（Legendre transform）和凸對偶（convex duality）的雛形，這些概念在力學（由拉格朗日量（Lagrangian）過渡到哈密頓量（Hamiltonian））和最佳化（optimisation）中都會再次出現。有關分析學的背景知識，可參閱 Bartle 與 Sherbert 的 *Introduction to Real Analysis*（《實分析導論》）。

詳細解答

STEP III 2011，第 5 題

純數學 • 以積分求面積

STEP III 2011，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/11-s3-q5

本題在 *Beyond STEP* 第 4.9 節 *Holditch* 定理中有完整的引導式講解：https://beyondstep.keithmcnulty.org/zh-hant/ch04/holditch.html

關鍵思路

直角坐標（cartesian）公式 $(*)$ 之所以有用，是因為它是雙線性（bilinear）的：若 A 是 P 平移一個只依賴 t 的向量所得，那麼 OA 掃過的面積可拆成三部分：OP 掃過的面積、單由平移產生的面積，以及一個交叉項（cross term）。這根桿正好給出這類平移，即 $\mp$ 單位向量（unit vector）$(\cos t, \sin t)$ 的某個倍數。交叉項 f 看似無從計算，但我們根本不需要算它：A 和 B 都沿同一條曲線 $\mathcal{C}$ 走一圈，所以 $[A] = [B]$，而這條等式就確定了 f。

解答

面積公式

設 $x = r\cos\theta$、$y = r\sin\theta$，其中 r 和 θ 都是 t 的函數（function），由乘積法則（product rule）得

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \dot{r}\cos\theta - r\sin\theta\,\dot{\theta}, \qquad \frac{\mathrm{d}y}{\mathrm{d}t} = \dot{r}\sin\theta + r\cos\theta\,\dot{\theta},$$

其中上方的點表示 $\mathrm{d}/\mathrm{d}t$。於是

$$\begin{aligned} x\frac{\mathrm{d}y}{\mathrm{d}t} - y\frac{\mathrm{d}x}{\mathrm{d}t} &= r\cos\theta\,(\dot{r}\sin\theta + r\cos\theta\,\dot{\theta}) - r\sin\theta\,(\dot{r}\cos\theta - r\sin\theta\,\dot{\theta}) \\ &= r^2(\cos^2\theta + \sin^2\theta)\,\dot{\theta} = r^2\frac{\mathrm{d}\theta}{\mathrm{d}t}, \end{aligned}$$

因為含 $r\dot{r}\sin\theta\cos\theta$ 的項互相抵消。把積分（integral）變數由 θ 換成 t，

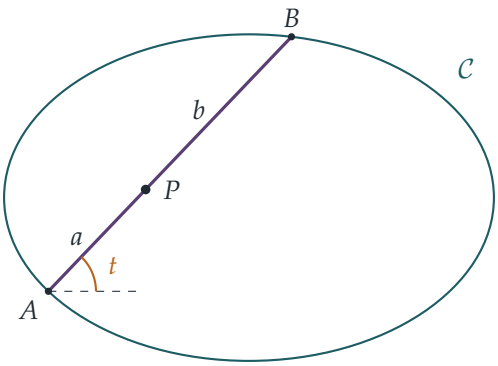
$$\tfrac{1}{2}\int r^2\,\mathrm{d}\theta = \tfrac{1}{2}\int r^2\frac{\mathrm{d}\theta}{\mathrm{d}t}\,\mathrm{d}t = \tfrac{1}{2}\int\left(x\frac{\mathrm{d}y}{\mathrm{d}t} - y\frac{\mathrm{d}x}{\mathrm{d}t}\right)\mathrm{d}t,$$

這正是 $(*)$。

A 和 B 的坐標

桿的方向是單位向量 $\mathbf{u} = (\cos t, \sin t)$，由 A 指向 B，而 P 位於兩者之間，且 $AP = a$、$PB = b$。所以 A 是 $P - a\mathbf{u}$，B 是 $P + b\mathbf{u}$：

$$A = (x - a\cos t,\ y - a\sin t), \qquad B = (x + b\cos t,\ y + b\sin t).$$



OA 掃過的面積

同樣以點表示 $\mathrm{d}/\mathrm{d}t$。對 A 有

$$x_A = x - a\cos t, \quad \dot{x}_A = \dot{x} + a\sin t, \qquad y_A = y - a\sin t, \quad \dot{y}_A = \dot{y} - a\cos t.$$

展開得

$$x_A\dot{y}_A = x\dot{y} - a\cos t\,(x + \dot{y}) + a^2\cos^2 t,$$
$$y_A\dot{x}_A = y\dot{x} + a\sin t\,(y - \dot{x}) - a^2\sin^2 t,$$

所以

$$x_A\dot{y}_A - y_A\dot{x}_A = (x\dot{y} - y\dot{x}) + a^2 - a\Big[(x + \dot{y})\cos t + (y - \dot{x})\sin t\Big].$$

現在以 t 由 0 至 2π 應用 $(*)$。第一個括號給出 $[P]$，常數 a^2 給出 $\tfrac{1}{2}\cdot 2\pi a^2 = \pi a^2$，而根據 f 的定義，最後一項給出 $-af$。因此

$$[A] = [P] + \pi a^2 - af,$$

即為所求。

對 B 而言，唯一的改變是處處把 $-a$ 換成 $+b$（即在上面的推導中以 $-b$ 代替 a）。所以

$$[B] = [P] + \pi b^2 + bf.$$

$\mathcal{C}$ 與 $\mathcal{D}$ 之間的面積

? **想一想** $[A]$、$[B]$ 和 $[P]$ 實際上等於甚麼？

當 t 由 0 變到 2π，桿轉了整整一圈。兩個端點都沿 $\mathcal{C}$ 逆時針走一圈，並回到起點：轉了半圈時，A 和 B 互換位置；轉了一整圈後，兩者都回到原位。原點位於 $\mathcal{D}$ 之內，因此也位於 $\mathcal{C}$ 之內。所以 OA 和 OB 各自恰好掃過 $\mathcal{C}$ 所圍的區域一次，而 OP 掃過 $\mathcal{D}$ 所圍的區域一次。以 $|\mathcal{C}|$ 和 $|\mathcal{D}|$ 表示這兩個面積，

$$[A] = [B] = |\mathcal{C}|, \qquad [P] = |\mathcal{D}|.$$

令 $[A] = [B]$，便可消去 $[P]$：

$$\pi a^2 - af = \pi b^2 + bf \quad \Longrightarrow \quad (a+b)f = \pi(a^2 - b^2) = \pi(a-b)(a+b).$$

由於 $a + b > 0$（它是桿的長度），可以兩邊相除，得 $f = \pi(a-b)$。因此

$$|\mathcal{C}| - |\mathcal{D}| = [A] - [P] = \pi a^2 - a\pi(a-b) = \pi ab.$$

所以兩曲線之間的面積是 πab。改用 $[B]$ 計算，同樣得到 $\pi b^2 + b\pi(a-b) = \pi ab$，這是一個令人安心的一致性檢查。

驗算

- 圓形。若 $\mathcal{C}$ 是半徑為 R 的圓，長度為 $a+b$ 的弦與圓心的距離恆為 $d = \sqrt{R^2 - \frac{1}{4}(a+b)^2}$。由畢氏定理（Pythagoras），$P$ 與圓心的距離是 $\sqrt{d^2 + \frac{1}{4}(a-b)^2}$，所以 $\mathcal{D}$ 是一個同心圓（concentric circle），其半徑的平方為 $R^2 - ab$。兩者之間的面積是 $\pi R^2 - \pi(R^2 - ab) = \pi ab$。
- 對稱性。答案關於 a 和 b 對稱，這是必然的，因為把桿倒轉便會互換兩者的角色。
- 數值驗證。我們讓一根 $a = 1$、$b = 1.5$ 的桿沿橢圓 $x^2/16 + y^2/9 = 1$ 移動，並用鞋帶公式（shoelace formula）計算所描出曲線的面積。兩曲線之間的面積算得 4.712390，與 $\pi ab = 1.5\pi \approx 4.712389$ 吻合。我們亦對隨機選取的曲線 $P(t)$，用數值積分（quadrature）驗證了 $[A]$ 和 $[B]$ 的恆等式。

常見錯誤

- 把 $\dot{x}_A = \dot{x} + a\sin t$ 的正負號弄錯。對 $-a\cos t$ 求導得 $+a\sin t$。
- 試圖直接計算 f。它取決於 $\mathcal{D}$ 的形狀，沒有更多資料便無法算出。關鍵在於 $[A] = [B]$ 就能確定它。

STEP 之外

這就是 Holditch 定理（Holditch's theorem，1858 年）：$\mathcal{C}$ 與 $\mathcal{D}$ 之間的面積只取決於 a 和 b，而與 $\mathcal{C}$ 的形狀無關。公式 $(*)$ 是格林定理（Green's theorem）的特例，該定理把區域上的面積分轉化為沿其邊界的線積分。它也是求積儀（planimeter）背後的原理；求積儀是一種沿邊界描畫來量度面積的機械裝置。格林定理是大學一年級向量微積分（vector calculus）的核心內容，而 Needham 的 *Visual Complex Analysis*（《直觀複分析》）對這類邊界積分給出了優美的幾何觀點。

詳細解答

STEP III 2011，第 6 題

純數學 • 積分代換

STEP III 2011，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/11-s3-q6

關鍵思路

題目不容許我們計算任何積分的值，所以可用的工具只有那些把一個積分變成另一個積分的方法：代換法（substitution）和分部積分法（integration by parts）。積分上下限就是線索。數 $\frac{1}{3}, \frac{1}{2}$ 與 $\ln 2, \ln 3$ 由 $v = e^{-u}$ 聯繫起來，而 $\frac{1}{2}\ln 2, \frac{1}{2}\ln 3$ 則由 $t = e^{-2x}$ 與 $\frac{1}{2}, \frac{1}{3}$ 聯繫起來。積分 T 和 V 的上下限相同，而 $\frac{\mathrm{d}}{\mathrm{d}t}\operatorname{artanh} t = \frac{1}{1-t^2}$ 提示我們可以在兩者之間用分部積分。三條聯繫已足以使四個積分全部相等。



解答

全文使用

$$\operatorname{artanh} t = \tfrac{1}{2}\ln\frac{1+t}{1-t} \quad (|t|<1), \qquad \frac{\mathrm{d}}{\mathrm{d}t}\operatorname{artanh} t = \frac{1}{1-t^2},$$

特別地，$\operatorname{artanh}\frac{1}{2} = \frac{1}{2}\ln 3$ 及 $\operatorname{artanh}\frac{1}{3} = \frac{1}{2}\ln 2$。所有被積函數（integrand）在其（閉且有界的）區間上都是連續的，因此每個積分都存在，下面的運算都是合理的。

$T = V$：分部積分

在 T 中，我們對 $\frac{1}{t}$ 積分、對 $\operatorname{artanh} t$ 求導，取 $u = \operatorname{artanh} t$ 及 $\mathrm{d}w = \frac{\mathrm{d}t}{t}$，於是 $w = \ln t$（由於在區間上 $t > 0$，這沒有問題）：

$$\begin{aligned} T &= \Big[\ln t \, \operatorname{artanh} t\Big]_{1/3}^{1/2} - \int_{1/3}^{1/2} \frac{\ln t}{1-t^2}\,\mathrm{d}t \\ &= \Big[\ln t \, \operatorname{artanh} t\Big]_{1/3}^{1/2} + V, \end{aligned}$$

因為餘下的積分正正就是 V（啞變量（dummy variable）用甚麼名稱並不重要）。

> **想一想** 邊界項（boundary term）不會破壞等式嗎？

它等於零，而這正是本題巧妙之處。在上限，$\ln\frac{1}{2}\cdot\operatorname{artanh}\frac{1}{2} = (-\ln 2)(\frac{1}{2}\ln 3)$；在下限，$\ln\frac{1}{3}\cdot\operatorname{artanh}\frac{1}{3} = (-\ln 3)(\frac{1}{2}\ln 2)$。兩者都等於 $-\frac{1}{2}\ln 2\ln 3$，所以

$$\Big[\ln t \, \operatorname{artanh} t\Big]_{1/3}^{1/2} = -\tfrac{1}{2}\ln 2\ln 3 + \tfrac{1}{2}\ln 2\ln 3 = 0,$$

因此 $T = V$。

$V = U$：代換 $v = e^{-u}$

設 $v = e^{-u}$，則 $\ln v = -u$ 及 $\mathrm{d}v = -e^{-u}\,\mathrm{d}u$。當 $v = \frac{1}{3}$ 時，$u = \ln 3$；當 $v = \frac{1}{2}$ 時，$u = \ln 2$。於是

$$\begin{aligned} V &= -\int_{\ln 3}^{\ln 2} \frac{-u}{1-e^{-2u}}\left(-e^{-u}\right)\mathrm{d}u = -\int_{\ln 3}^{\ln 2} \frac{u\,e^{-u}}{1-e^{-2u}}\,\mathrm{d}u \\ &= \int_{\ln 2}^{\ln 3} \frac{u\,e^{-u}}{1-e^{-2u}}\,\mathrm{d}u \qquad \text{（交換上下限）} \\ &= \int_{\ln 2}^{\ln 3} \frac{u}{e^{u}-e^{-u}}\,\mathrm{d}u \qquad \text{（分子分母同乘以 } e^{u}\text{）} \\ &= \int_{\ln 2}^{\ln 3} \frac{u}{2\sinh u}\,\mathrm{d}u = U. \end{aligned}$$

$X = T$：代換 $t = e^{-2x}$

首先改寫 X 的被積函數。對 $x > 0$，

$$\coth x = \frac{e^{x}+e^{-x}}{e^{x}-e^{-x}} = \frac{1+e^{-2x}}{1-e^{-2x}}, \qquad \text{所以} \qquad \ln(\coth x) = \ln\frac{1+e^{-2x}}{1-e^{-2x}} = 2\operatorname{artanh}\left(e^{-2x}\right),$$

由於 $0 < e^{-2x} < 1$，此式成立。現設 $t = e^{-2x}$，則 $\mathrm{d}t = -2t\,\mathrm{d}x$，即 $\mathrm{d}x = -\frac{\mathrm{d}t}{2t}$。當 $x = \frac{1}{2}\ln 2$ 時，$t = \frac{1}{2}$；當 $x = \frac{1}{2}\ln 3$ 時，$t = \frac{1}{3}$。因此

$$X = \int_{1/2}^{1/3} 2\operatorname{artanh} t\left(-\frac{1}{2t}\right)\mathrm{d}t = \int_{1/3}^{1/2} \frac{\operatorname{artanh} t}{t}\,\mathrm{d}t = T.$$

結論

把三條聯繫合起來，得 $X = T = V = U$，所以四個積分全部相等，得證。我們始終無須求出任何一個積分的值。

驗算

對每個積分分別使用辛普森法則（Simpson's rule），得 $T \approx U \approx V \approx X \approx 0.177704$，大約在小數點後 12 位都吻合。用計算機快速驗證邊界項為零也令人安心：$\ln \frac{1}{2} \cdot \operatorname{artanh} \frac{1}{2}$ 和 $\ln \frac{1}{3} \cdot \operatorname{artanh} \frac{1}{3}$ 的結果都是 $-0.3807\ldots$。

常見錯誤

- 在 $V = U$ 中漏掉負號：共有三個負號（V 前面的一個、$\ln v = -u$ 中的一個及 $\mathrm{d}v = -e^{-u}\,\mathrm{d}u$ 中的一個），而交換上下限又帶來第四個。
- 因為邊界項看似不為零，便放棄分部積分，卻沒有在兩端把它算出來。
- 寫成 $\ln\coth x = \operatorname{artanh}(e^{-2x})$，漏掉了因子 2；這個因子稍後會被 $\mathrm{d}x = -\frac{\mathrm{d}t}{2t}$ 中的 $\frac{1}{2}$ 抵消。

STEP 之外

把 $\operatorname{artanh} t/t$ 展開成冪級數（power series），可見這個共同值等於 $\chi_2(\frac{1}{2}) - \chi_2(\frac{1}{3})$，其中 $\chi_2(z) = \sum_{k\geq 0} z^{2k+1}/(2k+1)^2$ 是勒讓德（Legendre）chi 函數，與雙對數函數（dilogarithm）$\mathrm{Li}_2(z) = \sum_{n\geq 1} z^n/n^2$ 關係密切。這些函數的許多恆等式，正是來自這類變量代換。不計算任何一個積分而證明兩個積分相等，是數學分析（analysis）中的標準技巧；而支持代換法和分部積分法的定理（連續性（continuity）、微積分基本定理（fundamental theorem of calculus））會在第一門分析課程中得到嚴格的表述。Abbott 的 *Understanding Analysis*（《理解分析》）是看這些定理被仔細證明的好去處。

詳細解答

STEP III 2011，第 7 題

純數學 • 數學歸納法

STEP III 2011，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/11-s3-q7

關鍵思路

把 $\sqrt{a+1}+\sqrt{a}$ 平方，就能消去兩個獨立的平方根，剩下形如 $p+q\sqrt{a(a+1)}$ 的數；因此當 n 為偶數時，我們可以透過乘以 T_2，由 T_n 推進到 T_{n+2}。接着會發現，量 $A^2-a(a+1)B^2$ 在這個乘法下保持不變，而這正是歸納法（induction）所需要的。再多乘一個因子 $\sqrt{a+1}+\sqrt{a}$ 便能處理 n 為奇數的情況；而 (i) 和 (ii) 中的恆等式（identity）說明，兩項分別是兩個相鄰整數的平方根。

解答

全文記 $N=a(a+1)$，於是 $\sqrt{a+1}\,\sqrt{a}=\sqrt{N}$。

第 (i) 部分

基本情況（*base case*），$n=2$。展開得

$$T_2=(a+1)+2\sqrt{a+1}\,\sqrt{a}+a=(2a+1)+2\sqrt{N},$$

所以 $A_2=2a+1$ 和 $B_2=2$ 都是整數，而且

$$A_2^2-NB_2^2=4a^2+4a+1-4a(a+1)=1.$$

歸納步驟（*inductive step*）。假設對某個偶數 $n\geq 2$，有 $T_n=A_n+B_n\sqrt{N}$，其中整數 A_n, B_n 滿足 $A_n^2-NB_n^2=1$。為簡便起見，記 $p=2a+1$，並注意 $p^2=4a^2+4a+1=$

$4N+1$。由於 $T_{n+2}=T_nT_2$，

$$\begin{aligned}T_{n+2}&=\big(A_n+B_n\sqrt{N}\big)\big(p+2\sqrt{N}\big)\\&=\big(pA_n+2NB_n\big)+\big(2A_n+pB_n\big)\sqrt{N}.\end{aligned}$$

因此可取

$$A_{n+2}=pA_n+2NB_n,\qquad B_{n+2}=2A_n+pB_n,$$

由於 p、N、A_n 和 B_n 都是整數，它們也是整數。至於恆等式，交叉項 $\pm4pNA_nB_n$ 互相抵消：

$$\begin{aligned}A_{n+2}^2-NB_{n+2}^2&=\big(p^2A_n^2+4pNA_nB_n+4N^2B_n^2\big)-N\big(4A_n^2+4pA_nB_n+p^2B_n^2\big)\\&=(p^2-4N)A_n^2-N(p^2-4N)B_n^2\\&=A_n^2-NB_n^2=1,\end{aligned}$$

這裏先用了 $p^2-4N=1$，再用歸納假設（inductive hypothesis）。所以命題對 $n+2$ 成立；由歸納法，命題對每個正偶數 n 都成立。

另外注意 $A_2,B_2>0$，而遞推關係（recurrence）的係數都是正數，所以對每個偶數 n，都有 $A_n>0$ 和 $B_n>0$。第 (iii) 部分會用到這一點。

想一想 為甚麼 A^2-NB^2 在乘以 T_2 之後仍然保持不變？

因為 $A^2-NB^2=(A+B\sqrt{N})(A-B\sqrt{N})$，而同樣的計算顯示，「共軛」（conjugate）$A-B\sqrt{N}$ 在每一步都被乘以 $p-2\sqrt{N}$。因此兩者之積每一步都被乘以 $(p+2\sqrt{N})(p-2\sqrt{N})=p^2-4N=1$。上面的代數運算只不過是把這個事實逐步寫出來而已。

第 (ii) 部分

$n=1$ 的情況。$T_1=1\cdot\sqrt{a+1}+1\cdot\sqrt{a}$，所以 $C_1=D_1=1$，且 $(a+1)C_1^2=a+1=aD_1^2+1$。

$n\geq3$ 爲奇數的情況。此時 $m=n-1$ 是正偶數，所以由第 (i) 部分，$T_m=A_m+B_m\sqrt{N}$，且 $A_m^2-NB_m^2=1$。利用 $\sqrt{N}\sqrt{a+1}=(a+1)\sqrt{a}$ 和 $\sqrt{N}\sqrt{a}=a\sqrt{a+1}$，

$$\begin{aligned}T_n=\big(\sqrt{a+1}+\sqrt{a}\big)\big(A_m+B_m\sqrt{N}\big)&=A_m\sqrt{a+1}+A_m\sqrt{a}+(a+1)B_m\sqrt{a}+aB_m\sqrt{a+1}\\&=\big(A_m+aB_m\big)\sqrt{a+1}+\big(A_m+(a+1)B_m\big)\sqrt{a}.\end{aligned}$$

所以 $C_n=A_m+aB_m$ 和 $D_n=A_m+(a+1)B_m$ 都是整數（而且是正數）。於是

$$\begin{aligned}(a+1)C_n^2-aD_n^2&=(a+1)\big(A_m^2+2aA_mB_m+a^2B_m^2\big)\\&\qquad-a\big(A_m^2+2(a+1)A_mB_m+(a+1)^2B_m^2\big)\\&=A_m^2+a(a+1)\big[a-(a+1)\big]B_m^2\\&=A_m^2-NB_m^2=1,\end{aligned}$$

因為含 A_mB_m 的項互相抵消。因此 $(a+1)C_n^2 = aD_n^2 + 1$，得證。（$n=1$ 的情況其實是同一個計算，只是取 $T_0 = 1$，即 $A_0 = 1$、$B_0 = 0$。）

第 (iii) 部分

n 爲偶數。由於 $A_n > 0$ 且 $B_n > 0$，

$$T_n = A_n + B_n\sqrt{N} = \sqrt{A_n^2} + \sqrt{NB_n^2},$$

而由第 (i) 部分，整數 NB_n^2 和 $A_n^2 = NB_n^2 + 1$ 是相鄰的。

> **想一想** 為甚麼我們需要知道 A_n 和 B_n 的正負號？

因為 $\sqrt{A^2} = |A|$，而不是 A。只有在知道 $A_n \geq 0$ 之後，寫 $A_n = \sqrt{A_n^2}$ 才成立，這就是我們在第 (i) 部分記下正性的原因。

n 爲奇數。由於 $C_n, D_n > 0$，

$$T_n = C_n\sqrt{a+1} + D_n\sqrt{a} = \sqrt{(a+1)C_n^2} + \sqrt{aD_n^2},$$

而由第 (ii) 部分，整數 aD_n^2 和 $(a+1)C_n^2 = aD_n^2 + 1$ 是相鄰的。

所以對每個正整數 n，都存在某個正整數 k 使 $T_n = \sqrt{k+1} + \sqrt{k}$：具體而言，$n$ 為偶數時 $k = a(a+1)B_n^2$，n 為奇數時 $k = aD_n^2$。

驗算

取 $a = 3$，則 $T_n = (2+\sqrt{3})^n$。於是 $T_2 = 7 + 4\sqrt{3} = 7 + 2\sqrt{12} = \sqrt{49} + \sqrt{48}$，且 $49 = 12 \cdot 4 + 1$。接着，由 $A_2 = 7$、$B_2 = 2$ 得 $C_3 = 13$、$D_3 = 15$，以及 $T_3 = 26 + 15\sqrt{3} = \sqrt{676} + \sqrt{675} \approx 51.9808$，與直接計算 $(2+\sqrt{3})^3$ 的結果一致。我們用電腦對 $1 \leq a \leq 39$ 和 $n \leq 23$ 運行了遞推關係，以整數運算準確驗證了兩條恆等式，並以數值方式驗證了 T_n 的展開式。

常見錯誤

- 忘記基本情況，或者在第 (i) 部分只涉及偶數 n 的情況下，卻從 $n = 1$ 開始歸納。
- 在第 (iii) 部分寫 $A_n = \sqrt{A_n^2}$，卻沒有指出 A_n（以及 B_n、C_n、D_n）是正數。

STEP 之外

方程 $x^2 - Ny^2 = 1$ 就是佩爾方程（Pell's equation），而第 (i) 部分說明，$(2a+1) + 2\sqrt{a(a+1)}$ 的各次冪為 $N = a(a+1)$ 生成無限多組解。關鍵事實是：$x + y\sqrt{N}$ 的「範數」（norm）$x^2 - Ny^2$ 具有積性（multiplicative），這正是二次數域（quadratic number field）及其單位（unit）理論的起點。達文波特（Davenport）的 *The Higher Arithmetic*（《高等算術》）對佩爾方程和連分數（continued fraction）有非常易讀的論述。

詳細解答

STEP III 2011，第 8 題

純數學 • 複數軌跡

> **STEP III 2011，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/11-s3-q8

關鍵思路

將分子和分母同乘以分母的共軛（conjugate），便得到 u 和 v；之後每一部分只需代入題目給出的直線，再認出一條熟悉的曲線。半角代換（half-angle substitution）把 $\frac{2x}{1+x^2}$ 和 $\frac{x^2-1}{x^2+1}$ 變成正弦和餘弦，這正是題目建議用它的原因。整個過程中，要時刻留意曲線上哪些點真正被取到：端點和缺失的點佔了答案的一半。

解答

設 $z = x + \mathrm{i}y$。則 $1 + \mathrm{i}z = (1-y) + \mathrm{i}x$ 及 $\mathrm{i} + z = x + \mathrm{i}(1+y)$，因此把分子和分母同乘以 $x - \mathrm{i}(1+y)$，得

$$\begin{aligned} w &= \frac{\big((1-y)+\mathrm{i}x\big)\big(x-\mathrm{i}(1+y)\big)}{x^2+(1+y)^2} \\ &= \frac{(1-y)x + x(1+y) + \mathrm{i}\big(x^2-(1-y)(1+y)\big)}{x^2+(1+y)^2}. \end{aligned}$$

因此

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

適用於 $z \neq -\mathrm{i}$（在該點 w 無定義）。

知道逆映射（inverse map）亦很有幫助。由 $w(\mathrm{i}+z) = 1 + \mathrm{i}z$ 得 $z(w-\mathrm{i}) = 1 - \mathrm{i}w$，所以

$$z = \frac{1-\mathrm{i}w}{w-\mathrm{i}}, \qquad w \neq \mathrm{i}.$$

因此每個 $w \neq \mathrm{i}$ 都恰好來自一個 z，而 $w = \mathrm{i}$ 永遠取不到。這就是我們會一再遇到的那個缺失點。

第 (i) 部分

當 $y = 0$ 時，

$$u = \frac{2x}{1+x^2}, \qquad v = \frac{x^2-1}{x^2+1}.$$

令 $x = \tan(\theta/2)$。當 x 取遍所有實數時，θ 取遍 $-\pi < \theta < \pi$，而每個 x 恰好對應一個這樣的 θ。由半角公式，

$$u = \frac{2\tan(\theta/2)}{1+\tan^2(\theta/2)} = \sin\theta, \qquad v = -\frac{1-\tan^2(\theta/2)}{1+\tan^2(\theta/2)} = -\cos\theta.$$

所以 $u^2+v^2 = 1$。此外，當 $-\pi < \theta < \pi$ 時，$(\sin\theta, -\cos\theta)$ 描出整個單位圓（unit circle），只缺 $\theta = \pm\pi$ 所對應的點，即 $(u,v) = (0,1)$。因此 w 的軌跡（locus）是圓 $u^2+v^2=1$ 去掉單一點 $w = \mathrm{i}$，正如所求。（缺失點是 $x \to \pm\infty$ 時 w 的極限。）

第 (ii) 部分

現在 $-1 < x < 1$，對應 $-\frac{\pi}{2} < \theta < \frac{\pi}{2}$。在此範圍內 $v = -\cos\theta < 0$，而 $u = \sin\theta$ 由 -1 遞增至 1（不包括兩端）。所以 w 的軌跡是

$$u^2+v^2 = 1, \quad v < 0,$$

即單位圓的下半部分，不包括其端點 $w = \pm 1$（它們來自 $x = \pm 1$）。

第 (iii) 部分

當 $x = 0$ 時，得 $u = 0$ 及

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

由於 $y > -1$，約去 $y+1$ 是沒有問題的。

想一想 是否每個負的 v 值都能取到，還是只取到其中一部分？

當 y 由 -1 增加至 1，$y+1$ 在 $(0,2)$ 中遞增，所以 $\frac{2}{y+1}$ 在 $(1,\infty)$ 中遞減，而 v 則在 $(-\infty,0)$ 中連續且嚴格遞增。當 $y \to -1^+$ 時，$v \to -\infty$；當 $y \to 1^-$ 時，$v \to 0^-$。所以每個負的 v 都恰好取到一次。w 的軌跡是負虛軸

$$u = 0, \quad v < 0,$$

不包括原點（它來自 $y = 1$）。

第 (iv) 部分

當 $y = 1$ 時，

$$u = \frac{2x}{x^2+4}, \qquad v = \frac{x^2}{x^2+4}.$$

於是

$$u^2 + v^2 = \frac{4x^2 + x^4}{(x^2+4)^2} = \frac{x^2(x^2+4)}{(x^2+4)^2} = \frac{x^2}{x^2+4} = v,$$

所以 w 位於 $u^2 + v^2 = v$ 上，即

$$u^2 + \left(v - \tfrac{1}{2}\right)^2 = \tfrac{1}{4},$$

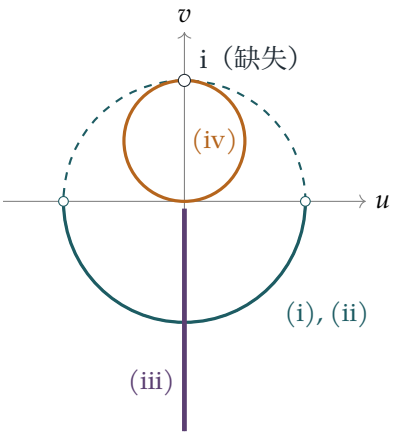
這是圓心為 $\frac{1}{2}\mathrm{i}$、半徑為 $\frac{1}{2}$ 的圓。

想一想 這個圓上哪些點真正被取到？

由於 $v = 1 - \frac{4}{x^2+4}$，我們有 $0 \le v < 1$，所以最高點 $w = \mathrm{i}$ 永遠取不到。其餘每一點都能取到：令 $x = 2\tan(\phi/2)$，其中 $-\pi < \phi < \pi$，得

$$u = \frac{\tan(\phi/2)}{1+\tan^2(\phi/2)} = \tfrac{1}{2}\sin\phi, \qquad v = \frac{\tan^2(\phi/2)}{1+\tan^2(\phi/2)} = \tfrac{1}{2}(1-\cos\phi),$$

它繞整個圓一周，只缺 $\phi = \pm\pi$，即點 $(0,1)$。（另一種方法是用逆映射，它說明每個 $w \ne \mathrm{i}$ 都恰好來自一個 z。）所以軌跡是圓 $u^2 + v^2 = v$ 去掉點 $w = \mathrm{i}$。它在 $w = 0$ 與實軸相切，該點是 $z = \mathrm{i}$ 的像。



由虛線和實線組成的青色圓是實軸的像，即第 (i) 部分；實線的下半部分是第 (ii) 部分。紫紅色射線是第 (iii) 部分，琥珀色圓是第 (iv) 部分。

驗算

- $z = 0$ 給出 $w = 1/\mathrm{i} = -\mathrm{i}$：它既在下半圓 (ii) 上，也在射線 (iii) 上，而 $z = 0$ 確實同時位於這兩段線上。
- $z = 1 + \mathrm{i}$ 給出 $w = \frac{\mathrm{i}}{1+2\mathrm{i}} = 0.4 + 0.2\mathrm{i}$，而 $0.4^2 + 0.2^2 = 0.2 = v$，所以它位於 (iv) 的圓上。
- 很大的實數 z：$w(10^8) \approx 2 \times 10^{-8} + \mathrm{i}$，趨近缺失點 i。
- 我們亦以數值方法對數以千計的隨機點作了驗證，既驗證了 u, v 的公式，也驗證了每個所聲稱的軌跡上的每一點，都在正確的直線或線段上有原像（preimage）。

常見錯誤

- 只寫出圓或半圓，卻沒有說明哪些點被去掉或不包括在內：這裏是 (i) 和 (iv) 中的 $w = \mathrm{i}$，以及 (ii) 中的端點 ± 1。
- 聲稱 (iii) 的軌跡是整條虛軸：對每個 $-1 < y < 1$，$v = \frac{y-1}{y+1}$ 都是負數。

STEP 之外

這個映射是一個默比烏斯變換（Möbius transformation）；事實上 $w = \mathrm{i}\,\dfrac{z - \mathrm{i}}{z + \mathrm{i}}$，所以 $|w| = |z - \mathrm{i}|/|z + \mathrm{i}|$，而它等於 1 當且僅當 z 與 $\pm\mathrm{i}$ 等距，亦即當它是實數時。默比烏斯映射把直線和圓映成直線和圓（直線被視為「經過 ∞ 的圓」），而缺失點 i 正是 ∞ 的像。這個映射（即凱萊變換（Cayley transform））把上半平面映成單位圓盤，是複分析（complex analysis）和雙曲幾何（hyperbolic geometry）中的重要工具。Needham 的 *Visual Complex Analysis*（《直觀複分析》）以圖像為主導，對默比烏斯變換有十分精彩的論述。

詳細解答

STEP III 2011，第 9 題

力學 • 圓柱上的運動

STEP III 2011，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：

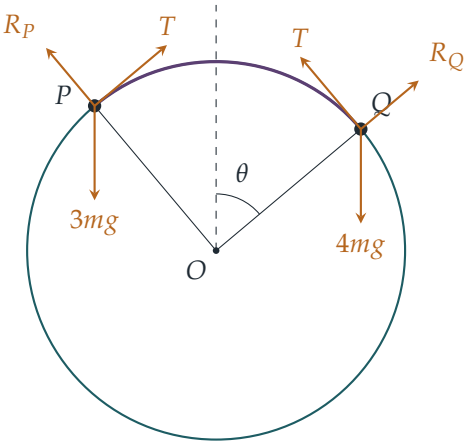
step.maths.org/questions/11-s3-q9

關鍵思路

長度為 $\frac{1}{2}\pi a$ 的繩子貼在圓柱面上，在 O 點所張的角是直角，因此 OP 與 OQ 始終互相垂直，整個系統只需一個坐標 θ 來描述。由能量守恆（conservation of energy）可直接得出 $\dot{\theta}^2$，再沿指向圓心的方向應用牛頓第二定律（Newton's second law），便可把每個法向反作用力（normal reaction）以 $\dot{\theta}^2$ 表示。質點離開表面，即表示法向反作用力降至零。至於張力（tension），則用切向（tangential）方程：把兩條方程相加可消去 T 並得出 $\ddot{\theta}$，之後任取其中一條便可求出 T。

解答

模型。P 和 Q 是質點（particle），繩子輕而不可伸長（inextensible），圓柱面光滑（smooth）。輕繩繞過光滑表面時，各處張力同為 T。兩質點之間的弧長為 $\frac{1}{2}\pi a$，所以 $\angle POQ = \frac{1}{2}\pi$。設 θ 為 OQ 與向上鉛垂線所成的角，Q 位於最高點的一側。那麼 OP 在另一側與向上鉛垂線所成的角為 $\frac{1}{2}\pi - \theta$。只要兩質點仍留在表面上，它們都在半徑為 a 的圓上以角速率（angular speed）$\dot{\theta}$ 運動，因此各自的速率均為 $a\dot{\theta}$。



平衡。在每個質點處沿表面方向分解力（法向反作用力沒有切向分量，而繩子沿切向拉向最高點）：對 Q 得 $T = 4mg\sin\theta$，對 P 得 $T = 3mg\sin(\frac{1}{2}\pi - \theta) = 3mg\cos\theta$。因此在平衡（equilibrium）位置時 $\tan\theta = \frac{3}{4}$，即 $\sin\theta = \frac{3}{5}$ 及 $\cos\theta = \frac{4}{5}$。

能量。取 O 所在高度為勢能（potential energy）零點。Q 的高度為 $a\cos\theta$，P 的高度為 $a\cos(\frac{1}{2}\pi - \theta) = a\sin\theta$。法向反作用力不做功（它們與運動方向垂直），而張力對整個系統不做淨功（繩子不可伸長），所以能量守恆。由於擾動輕微，系統由靜止開始運動，故

$$\tfrac{1}{2}(3m + 4m)a^2\dot{\theta}^2 + 4mga\cos\theta + 3mga\sin\theta = 4mga\cdot\tfrac{4}{5} + 3mga\cdot\tfrac{3}{5} = 5mga.$$

兩邊乘以 $2/(ma)$，得

$$7a\dot{\theta}^2 + 8g\cos\theta + 6g\sin\theta = 10g,$$

此即所求。當 Q 向下移動時，θ 由 $\tan^{-1}\frac{3}{4}$ 開始增大。

第 (i) 部分

Q 在半徑為 a 的圓上以速率 $a\dot{\theta}$ 運動，所以它有指向 O 的加速度 $a\dot{\theta}^2$。其重量指向 O 的分量為 $4mg\cos\theta$，而 R_Q 向外作用。繩子沿切向，因此在徑向上沒有貢獻。於是

$$4mg\cos\theta - R_Q = 4ma\dot{\theta}^2, \qquad \text{所以} \qquad R_Q = 4mg\cos\theta - 4ma\dot{\theta}^2.$$

當 R_Q 降至零，即 $a\dot{\theta}^2 = g\cos\theta$ 時，Q 離開表面。把此式代入 $\theta = \beta$ 時的能量方程：

$$7g\cos\beta + 8g\cos\beta + 6g\sin\beta = 10g \quad\Longrightarrow\quad 15\cos\beta + 6\sin\beta = 10,$$

此即所求。

想一想 為甚麼能量方程一直到 $\theta = \beta$ 仍然成立？

只要兩個質點都留在圓柱面上，能量方程便成立，而題目告訴我們 Q 先離開。你也可以自行驗證這一點。對 P 作同樣的推導，得 $R_P = 3mg\sin\theta - 3ma\dot{\theta}^2$，它在 $13\sin\theta + 8\cos\theta = 10$ 時為零。在 $\beta \approx 73.6°$ 時，左邊約為 14.7，所以 R_P 仍然是正數。

第 (ii) 部分

現在沿切向分解力，以運動方向為正方向。Q 以切向加速度 $a\ddot{\theta}$ 遠離最高點，而 P 以同樣的切向加速度 $a\ddot{\theta}$ 移向最高點（繩子使兩者速率保持相等）。對 Q 而言，重力有向前的分量 $4mg\sin\theta$，張力則向後拉：

$$4ma\ddot{\theta} = 4mg\sin\theta - T. \tag{1}$$

對 P 而言，張力把它向前拉，重力有向後的分量 $3mg\cos\theta$：

$$3ma\ddot{\theta} = T - 3mg\cos\theta. \tag{2}$$

把 (1) 和 (2) 相加可消去 T：

$$7ma\ddot{\theta} = 4mg\sin\theta - 3mg\cos\theta.$$

（這與把能量方程對時間求導再除以 $2\dot{\theta}$ 的結果一致。）代入 (1)：

$$\begin{aligned} T &= 4mg\sin\theta - 4ma\ddot{\theta} \\ &= 4mg\sin\theta - \tfrac{4}{7}m(4g\sin\theta - 3g\cos\theta) \\ &= \tfrac{12}{7}mg\sin\theta + \tfrac{12}{7}mg\cos\theta = \tfrac{12}{7}mg(\sin\theta + \cos\theta), \end{aligned}$$

此即所求。當 $0 < \theta < \tfrac{1}{2}\pi$ 時此值為正，所以繩子在整個過程中都保持拉緊。

驗算

- 在平衡位置時，$\sin\theta + \cos\theta = \tfrac{7}{5}$，所以 $T = \tfrac{12}{5}mg$。這與 $4mg\cdot\tfrac{3}{5}$ 及 $3mg\cdot\tfrac{4}{5}$ 相符。
- 寫成 $15\cos\beta + 6\sin\beta = \sqrt{261}\cos(\beta - \alpha)$，其中 $\tan\alpha = \tfrac{2}{5}$，可得 $\beta \approx 1.2839$ 弧度 $\approx 73.56°$。我們從剛越過平衡位置之處開始以數值方法逐步模擬運動，發現 R_Q 正是在這個角度首次降至零，而 R_P 則保持為正。在整條路徑上，能量關係式和張力公式在捨入誤差範圍內均成立。

常見錯誤

- 把 P 的高度當作 $a\cos\theta$ 而不是 $a\sin\theta$。繩長的設定，關鍵正在於直角 $\angle POQ$。
- 給兩個質點不同的張力，或把 T 放進徑向方程。繩子沿切向接觸質點，所以 T 只出現在切向方程中。

STEP 之外

由於一個坐標 θ 便足以描述一切，這是一個單自由度（one degree of freedom）系統。拉格朗日力學（Lagrangian mechanics）可以直接由動能和勢能得出運動方程，完全不需要考慮約束力（forces of constraint）。法向反作用力和張力之後會以約束條件的「拉格朗日乘數」（Lagrange multipliers）的形式出現。在接觸拉格朗日（Lagrange）的理論之前，Kleppner 和 Kolenkow 所著的 *An Introduction to Mechanics*（《力學導論》）是加深你對約束運動和能量方法理解的好起點。

詳細解答

STEP III 2011，第 10 題

力學 • 彈性繩與簡諧運動

STEP III 2011，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/11-s3-q10

關鍵思路

兩個質點組成的系統不受任何外來的水平力，所以它們的質心（centre of mass）以由衝量（impulse）決定的恆定速率運動。相對於質心，弦的伸長量（extension）遵從簡諧運動（simple harmonic motion），而弦恰好在半個振動週期內保持拉緊。把運動拆分為「質心」加「伸長量」兩部分，每個答案幾乎不費功夫便會自然得出。

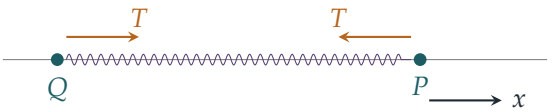
解答

模型與設定

兩質點在光滑水平面上運動，因此它們所受的水平力只有輕弦的張力（tension），而這兩個力大小相等、方向相反。取直線 QP 為 x 軸，方向由 Q 指向 P，並設 Q 起初位於 $x = 0$，P 起初位於 $x = a$。設 x_P 和 x_Q 為施加衝量後時間 t 時兩者的位置，並設

$$e = (x_P - x_Q) - a$$

為弦的伸長量。施加衝量後一刻，$\dot{x}_P = u$，$\dot{x}_Q = 0$ 及 $e = 0$，且 $\dot{e} = u > 0$，因此弦立即變得拉緊。



弦拉緊期間，由胡克定律（Hooke's law）得張力

$$T = \frac{\lambda e}{a} = \frac{\frac{1}{2}ma\omega^2 e}{a} = \frac{1}{2}m\omega^2 e.$$

張力把 P 往回拉向 Q，並把 Q 向前拉向 P：

$$m\ddot{x}_P = -\tfrac{1}{2}m\omega^2 e, \qquad m\ddot{x}_Q = +\tfrac{1}{2}m\omega^2 e.$$

質心

把兩條方程相加，得 $\ddot{x}_P + \ddot{x}_Q = 0$。因此位於 $\bar{x} = \frac{1}{2}(x_P + x_Q)$ 的質心 G 以恆定速度運動，該速度等於其初始值：

$$\dot{\bar{x}} = \tfrac{1}{2}(u + 0) = \tfrac{1}{2}u.$$

（這其實就是動量守恆（conservation of momentum）：總動量 mu 由質量 $2m$ 共同分擔。）

伸長量

把兩條方程相減，

$$\ddot{e} = \ddot{x}_P - \ddot{x}_Q = -\omega^2 e.$$

這是角頻率（angular frequency）為 ω 的簡諧運動。由 $e(0) = 0$ 及 $\dot{e}(0) = u$，得

$$e = \frac{u}{\omega}\sin\omega t.$$

想一想 為甚麼角頻率是 ω，而不是 $\omega/\sqrt{2}$ 或 $\sqrt{2}\,\omega$？

每個質點所受的力都是 $\frac{1}{2}m\omega^2 e$，但 e 的變化源於兩個質點都在運動：把方程相減時，兩者的加速度相加。換一個角度看，兩個質量同為 m 的質點，其約化質量（reduced mass）為 $\frac{1}{2}m$，而弦的勁度（stiffness）為 $\lambda/a = \frac{1}{2}m\omega^2$，所以 $\omega^2 = \frac{1}{2}m\omega^2 / \frac{1}{2}m$。題目特意選取這個彈性模量（modulus），令結果整齊漂亮。

弦的長度下一次等於 a，是在 $e = 0$ 再次成立之時，即 $t > 0$ 時首次出現 $\sin\omega t = 0$ 之時：

$$t_1 = \frac{\pi}{\omega}.$$

在整段 $0 < t < \pi/\omega$ 期間都有 $e > 0$，所以弦確實一直拉緊，上述方程亦一直成立。

所走的距離

把每個位置寫成質心位置加上或減去間距的一半：

$$x_P = \bar{x} + \tfrac{1}{2}(a + e), \qquad x_Q = \bar{x} - \tfrac{1}{2}(a + e).$$

在 t_1 時，$e = 0$ 再次成立，所以每個質點移動的量都與 G 完全相同，即

$$\tfrac{1}{2}u \cdot \frac{\pi}{\omega} = \frac{\pi u}{2\omega}.$$

為確定這個位移就是所走的距離，我們要驗證兩個質點都從不折返。求導得

$$\dot{x}_P = \tfrac{1}{2}u + \tfrac{1}{2}\dot{e} = \tfrac{1}{2}u(1+\cos\omega t), \qquad \dot{x}_Q = \tfrac{1}{2}u - \tfrac{1}{2}\dot{e} = \tfrac{1}{2}u(1-\cos\omega t),$$

而對所有 t，兩者皆 ≥ 0。所以每個質點都只向前移動，各自走了距離 $\dfrac{\pi u}{2\omega}$，正如所求。

速率

在 $t_1 = \pi/\omega$ 時，$\cos\omega t_1 = -1$，所以

$$\dot{x}_P = 0, \qquad \dot{x}_Q = u.$$

因此 P 瞬間靜止，而 Q 以速率 u 向 P 移動。兩個質點恰好交換了速度，就如兩個等質量物體發生完全彈性碰撞（perfectly elastic collision）一樣。這並非巧合：動量守恆，而儲存的彈性勢能已全數轉回動能。

到碰撞為止的時間

在 t_1 之後，Q 向 P 移動，所以間距小於 a。弦變得鬆弛，不施加任何力，兩個質點都不受任何水平力。因此 P 保持靜止，而 Q 由距離 P 為 a 之處開始，以恆定速率 u 移動。它再經過時間 a/u 便到達 P。由施加衝量到碰撞的總時間為

$$\frac{\pi}{\omega} + \frac{a}{u}.$$

驗算

- 動量：在 t_1 時，總動量為 $m\cdot 0 + mu = mu$，與開始時相同。能量：兩個時刻的動能都是 $\frac{1}{2}mu^2$，而兩次弦都處於自然長度。
- 量綱：u/ω 是長度，a/u 是時間，理應如此。
- 我們取 $m = a = \omega = u = 1$，以數值方法逐步計算運動方程。弦在 $t = 3.14159$ 時變得鬆弛，此時每個質點都移動了 $1.57080 = \pi/2$；兩者的速率分別為 0 和 1，而碰撞發生在 $t = 4.14159 = \pi + 1$。

常見錯誤

- 把 P 當作繫於固定點，這會得出角頻率 $\omega/\sqrt{2}$ 和錯誤的時間。弦的兩端都在移動。
- 在 t_1 之後仍沿用簡諧運動。弦一旦短於 a 便會鬆弛，運動變為勻速。
- 完全忘記最後一段：碰撞發生在弦回復自然長度之後再過 a/u。

STEP 之外

把二體問題（two-body problem）拆分為質心的運動和間距的運動，而後者由約化質量 $\mu = m_1 m_2/(m_1 + m_2)$ 支配，是大學一年級力學中最有用的技巧之一。它把行星軌道、振動中的雙原子分子以及碰撞的粒子都化為單體問題。在本題中，彈性弦就像一次「柔軟」的碰撞，歷時 π/ω，但結果與瞬間發生的彈性碰撞完全相同。Kleppner 與 Kolenkow 所著的 *An Introduction to Mechanics*（《力學導論》）對二體約化和約化質量有細緻的推導。

詳細解答

STEP III 2011，第 11 題

力學 • 剛體轉動

STEP III 2011，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/11-s3-q11

關鍵思路

扭轉之後，每條繩仍然由天花板上的固定點連到圓盤邊緣上已轉過 θ 的那一點。因此每條繩的水平部分是圓的一條弦（chord），所以幾何關係要從上方俯視來看。由對稱性，每條繩的情況都一樣，所以鉛直方向的平衡（equilibrium）決定張力（tension），而各張力的水平分量對軸的力矩（moment）之和就給出力偶（couple）。圓盤一經釋放，餘下的就交給能量守恆（energy conservation）。

解答

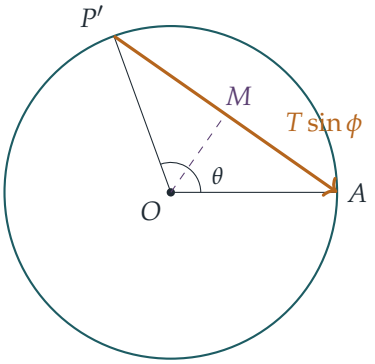
我們把圓盤模型化為剛性均勻薄片（lamina），把繩模型化為輕而不可伸長（inextensible）的繩。由對稱性，所有繩都是拉緊的，張力同為 T，傾角同為 ϕ，而圓盤保持水平，圓心在 z 軸上。當圓盤轉過 θ 時，它會上升至天花板下方深度 h 處。

繩的角度

經過 P 的繩繫於天花板上的 $A = (a, 0, 0)$。轉動後，繩的下端位於

$$P' = (a\cos\theta,\ a\sin\theta,\ -h).$$

從上方看，A 和 P' 都在以軸為圓心、半徑為 a 的圓上，兩者在圓心 O 所夾的角為 θ。所以繩的水平投影（projection）是一條長度為 $2a\sin\frac{1}{2}\theta$ 的弦。



從上方俯視的圖。M 是弦 $P'A$ 的中點，所以 OM 平分角 θ，與弦垂直，且長度為 $OM = a\cos\frac{1}{2}\theta$。繩本身長 b，與鉛直方向成角 ϕ，所以它的水平投影長度為 $b\sin\phi$。令兩者相等，得

$$b\sin\phi = 2a\sin\tfrac{1}{2}\theta,$$

此即所求。繩的鉛直投影則給出圓盤的新深度：

$$h = b\cos\phi = \sqrt{b^2 - 4a^2\sin^2\tfrac{1}{2}\theta}.$$

由於 $b > 2a$，這個值是實數。由於 $0 < \theta < \pi$，我們有 $0 < \sin\frac{1}{2}\theta < 1$，所以 $0 < \phi < \frac{1}{2}\pi$。

力偶

對圓盤作鉛直方向的分解，n 個相等的張力承托着重量：

$$nT\cos\phi = mg.$$

力偶 C 作用於圓盤所在的平面內，所以對這個平衡沒有影響。

每個張力都有一個鉛直分量，以及一個沿弦由 P' 指向 A 的水平分量 $T\sin\phi$。鉛直分量與軸平行，所以對軸沒有力矩。水平分量沿弦作用，而弦與 O 的垂直距離為 $OM = a\cos\frac{1}{2}\theta$。每條繩的水平拉力都傾向把圓盤轉回 $\theta = 0$，所以 n 個力矩的轉向全部相同。對軸取力矩，

$$\begin{aligned} C &= n\,T\sin\phi\cdot a\cos\tfrac{1}{2}\theta = mg\tan\phi\, a\cos\tfrac{1}{2}\theta \\ &= mg\,a\cos\tfrac{1}{2}\theta\cdot\frac{2a\sin\frac{1}{2}\theta}{\sqrt{b^2-4a^2\sin^2\frac{1}{2}\theta}} = \frac{mga^2\sin\theta}{\sqrt{b^2-4a^2\sin^2\frac{1}{2}\theta}}, \end{aligned}$$

其中先用了 $nT = mg/\cos\phi$，再用 $\tan\phi = b\sin\phi/(b\cos\phi)$，最後用 $2\sin\frac{1}{2}\theta\cos\frac{1}{2}\theta = \sin\theta$。這就是所求的大小。

由靜止釋放

移去力偶後，圓盤會反向扭回並下降。繩是不可伸長的，且一直保持拉緊，所以不做功：每條繩的張力都與其繫點相對於天花板固定點的速度垂直。重力是唯一做功的力，所以機械能守恆。

由上面的幾何關係，當圓盤由繩鉛直的位置轉過 α 時，其深度為

$$z(\alpha) = \sqrt{b^2 - 4a^2 \sin^2 \tfrac{1}{2}\alpha} = \sqrt{b^2 - 2a^2(1 - \cos\alpha)}.$$

圓盤的動能（kinetic energy）為 $\frac{1}{2}m\dot{z}^2 + \frac{1}{2}I\omega^2$，其中對軸的轉動慣量（moment of inertia）為 $I = \frac{1}{2}ma^2$。

> **想一想** 當繩處於鉛直時，圓盤的圓心是否仍有鉛直方向的速度？

沒有。求導得

$$\dot{z} = \frac{\mathrm{d}z}{\mathrm{d}\alpha}\,\dot{\alpha} = -\frac{a^2 \sin\alpha}{z}\,\dot{\alpha},$$

在 $\alpha = 0$ 時這等於零。那一刻繩是鉛直的，圓盤處於最低點，所以此時全部動能都是轉動動能。

圓盤由深度 $\sqrt{b^2 - 4a^2 \sin^2 \frac{1}{2}\theta}$ 下降至深度 b。能量守恆給出

$$\tfrac{1}{2} \cdot \tfrac{1}{2}ma^2\,\omega^2 = mg\left(b - \sqrt{b^2 - 4a^2 \sin^2 \tfrac{1}{2}\theta}\right),$$

即

$$\frac{a^2\omega^2}{4g} = b - \sqrt{b^2 - 4a^2 \sin^2 \tfrac{1}{2}\theta},$$

此即所求。

驗算

- *虛功（virtual work）*。逆着力偶把圓盤再轉 $\delta\theta$，圓盤會升高 $-\delta z$，所以 $C = mg\,|\mathrm{d}z/\mathrm{d}\theta| = mg\,a^2 \sin\theta / z$。這與我們取力矩所得的力偶完全一致。
- *小扭轉*。當 θ 很小時，$C \approx mga^2\theta/b$，這正是大家熟悉的雙線懸掛（bifilar suspension，這裏是 n 線懸掛）的回復力偶。
- *數值驗證*。我們在三維空間中明確放置 n 條繩，並對隨機的 a, b, θ, m, n 把各張力的力矩加起來，結果與公式吻合。我們亦取 $a = 1$、$b = 3$、$\theta = 2$、$g = 9.81$，以數值方法積分釋放後的運動，得出繩變成鉛直時 $a^2\omega^2/(4g) = 0.51651$，與 $3 - \sqrt{9 - 4\sin^2 1} = 0.51651$ 相符。

常見錯誤

- 把水平張力的力臂（lever arm）當作 a 而不是 $a\cos\frac{1}{2}\theta$。這個力沿弦作用，而不是沿切線作用。
- 忘記圓盤在扭轉時會上升。這個高度變化正是最後一部分中能量的來源。
- 把平移動能 $\frac{1}{2}m\dot{z}^2$ 計算在內，卻沒有留意到它在繩鉛直時等於零。

STEP 之外

這是一個運動受約束、只有一個自由度（degree of freedom）的系統。深度是扭轉角的固定函數（function），所以能量只取決於 α 和 $\dot{\alpha}$。拉格朗日力學（Lagrangian mechanics）把這個觀察轉化為運動方程。它也解釋了為何力偶恰好是 $mg\,\mathrm{d}z/\mathrm{d}\alpha$，這就是虛功原理。在 $\alpha = 0$ 附近線性化，便得到簡諧（simple harmonic）扭轉振動（torsional oscillations），而雙線懸掛正是藉此用來量度轉動慣量。Kleppner 和 Kolenkow 的 *An Introduction to Mechanics*（《力學導論》）以這個程度講解剛體轉動和能量方法。

詳細解答

STEP III 2011，第 12 題

概率與統計 • 以概率生成函數處理複合分佈

STEP III 2011，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/11-s3-q12

關鍵思路

關於一個分佈的平均值（mean）和方差（variance），我們所需要的一切都藏在它的概率母函數（probability generating function, pgf）在 $t = 1$ 處的首兩階導數之中。因此，我們用連鎖法則（chain rule）對複合函數 $G(H(t))$ 求導（兩次），代入 $t = 1$，並利用 $H(1) = 1$。至於擲硬幣的部分，N 服從幾何分佈，而每個 X_i 都是單次擲公平硬幣的指示變量（indicator），所以兩個概率母函數都很簡單，而 $G(H(t))$ 是一個有理函數（rational function），可以展開成幾何級數（geometric series）。

解答

一般結果

對於任何取非負整數值、概率母函數為 $P(t) = \mathrm{E}(t^Z)$ 的隨機變量 Z，我們有 $P(1) = 1$，

$$P'(1) = \mathrm{E}(Z), \qquad P''(1) = \mathrm{E}\big(Z(Z-1)\big),$$

因此

$$\mathrm{Var}(Z) = P''(1) + P'(1) - \big(P'(1)\big)^2.$$

記 $\mu = \mathrm{E}(X_i) = H'(1)$、$\sigma^2 = \mathrm{Var}(X_i)$，並以 $K(t) = G(H(t))$ 表示 Y 的概率母函數。由連鎖法則，$K'(t) = G'(H(t))\,H'(t)$。由於 $H(1) = 1$，

$$\mathrm{E}(Y) = K'(1) = G'(1)H'(1) = \mathrm{E}(N)\,\mathrm{E}(X_i),$$

這就是第一個結果。

再求導一次（先用乘積法則（product rule），再用連鎖法則），

$$K''(t) = G''(H(t))\big(H'(t)\big)^2 + G'(H(t))\,H''(t),$$

所以，再次利用 $H(1) = 1$，

$$K''(1) = G''(1)\mu^2 + \mathrm{E}(N)\,H''(1).$$

把上面的方差公式分別應用於 N 和 X_i，得

$$G''(1) = \mathrm{Var}(N) - \mathrm{E}(N) + \big(\mathrm{E}(N)\big)^2, \qquad H''(1) = \sigma^2 - \mu + \mu^2.$$

想一想 為甚麼不直接由 $K''(1)$ 計算 $\mathrm{Var}(Y)$，而要先改寫 $G''(1)$ 和 $H''(1)$？

因為我們想要的答案是以方差表示，而不是以階乘矩（factorial moments）表示；透過 $\mathrm{Var}(N)$ 和 σ^2 來表達 $G''(1)$ 和 $H''(1)$，可以讓多餘的項互相抵消。記 $m = \mathrm{E}(N)$，

$$\begin{aligned}\mathrm{Var}(Y) &= K''(1) + K'(1) - \big(K'(1)\big)^2 \\ &= \big(\mathrm{Var}(N) - m + m^2\big)\mu^2 + m\big(\sigma^2 - \mu + \mu^2\big) + m\mu - m^2\mu^2 \\ &= \mathrm{Var}(N)\mu^2 + m\sigma^2 + \big(-m\mu^2 + m\mu^2\big) + \big(m^2\mu^2 - m^2\mu^2\big) + \big(-m\mu + m\mu\big) \\ &= \mathrm{Var}(N)\big(\mathrm{E}(X_i)\big)^2 + \mathrm{E}(N)\,\mathrm{Var}(X_i),\end{aligned}$$

即為所求。

擲硬幣

*N 的概率母函數。*第一次擲出正面發生在第 n 次，當且僅當首 $n-1$ 次都是反面而第 n 次是正面，所以對 $n = 1, 2, 3, \ldots$，$\mathrm{P}(N = n) = \left(\frac{1}{2}\right)^n$（這是幾何分佈（geometric distribution））。因此，當 $|t| < 2$ 時，

$$G(t) = \sum_{n=1}^{\infty} \left(\frac{t}{2}\right)^n = \frac{t/2}{1 - t/2} = \frac{t}{2-t}.$$

*每個 X_i 的概率母函數。*若其後的第 i 次擲出正面，令 $X_i = 1$，否則令 $X_i = 0$。於是 $Y = X_1 + \cdots + X_N$，而 $H(t) = \frac{1}{2} + \frac{1}{2}t = \frac{1}{2}(1+t)$。其後各次擲硬幣彼此獨立，亦與決定 N 的各次擲硬幣獨立，所以一般結果的條件成立。

Y 的概率母函數。

$$G(H(t)) = \frac{\frac{1}{2}(1+t)}{2 - \frac{1}{2}(1+t)} = \frac{1+t}{3-t}.$$

*平均值與方差。*對於 $p = \frac{1}{2}$ 的幾何分佈，$\mathrm{E}(N) = 1/p = 2$，$\mathrm{Var}(N) = (1-p)/p^2 = 2$（或者對 G 求導：$G'(t) = 2/(2-t)^2$、$G''(t) = 4/(2-t)^3$，得 $G'(1) = 2$、$G''(1) = 4$，

以及 $\operatorname{Var}(N) = 4 + 2 - 4 = 2$）。每個 X_i 都服從伯努利（Bernoulli）分佈 $\left(\frac{1}{2}\right)$，所以 $\mathrm{E}(X_i) = \frac{1}{2}$，$\operatorname{Var}(X_i) = \frac{1}{4}$。因此

$$\mathrm{E}(Y) = 2 \cdot \tfrac{1}{2} = 1, \qquad \operatorname{Var}(Y) = 2 \cdot \tfrac{1}{4} + 2 \cdot \tfrac{1}{4} = 1.$$

Y 的分佈。當 $|t| < 3$ 時，我們展開

$$\frac{1+t}{3-t} = \frac{1+t}{3} \cdot \frac{1}{1 - t/3} = \frac{1+t}{3} \sum_{k=0}^{\infty} \frac{t^k}{3^k} = \sum_{k=0}^{\infty} \frac{t^k}{3^{k+1}} + \sum_{k=0}^{\infty} \frac{t^{k+1}}{3^{k+1}}.$$

t^0 的係數只來自第一個和。當 $r \geq 1$ 時，t^r 的係數是 $\dfrac{1}{3^{r+1}} + \dfrac{1}{3^r} = \dfrac{4}{3^{r+1}}$。所以

$$\mathrm{P}(Y = 0) = \frac{1}{3}, \qquad \mathrm{P}(Y = r) = \frac{4}{3^{r+1}} \quad (r = 1, 2, 3, \ldots).$$

驗算

- 各概率之和為 1：$\frac{1}{3} + 4\sum_{r\geq 1} 3^{-(r+1)} = \frac{1}{3} + 4 \cdot \frac{1}{6} = 1$。
- 直接計算 $\mathrm{P}(Y = 0)$：$\sum_{n\geq 1} \left(\frac{1}{2}\right)^n \left(\frac{1}{2}\right)^n = \sum_{n\geq 1} 4^{-n} = \frac{1}{3}$。同理，$\mathrm{P}(Y = 1) = \sum_{n\geq 1} n\, 4^{-n} = \frac{4}{9}$，與 $4/3^2$ 吻合。
- 對 $n \leq 200$ 精確求和 $\mathrm{P}(N = n)\binom{n}{r}2^{-n}$，可重現 $r \leq 7$ 時的 $\mathrm{P}(Y = r)$，而所得平均值和方差均為 1.000；以 200 000 次模擬，得到 0.996 和 0.994。

常見錯誤

- 把 N 當作從 0 開始（概率母函數為 $1/(2-t)$）：這裏 N 計算的是擲出正面的那一次，所以 $N \geq 1$。
- 忘記 $P''(1)$ 是 $\mathrm{E}(Z^2) - \mathrm{E}(Z)$ 而不是 $\mathrm{E}(Z^2)$，以致在方差公式中漏掉 $+P'(1)$。
- 對所有 $r \geq 0$ 都給出 $4/3^{r+1}$：$r = 0$ 的一項是不同的。

STEP 之外

Y 是一個隨機和（random sum）（即複合分佈（compound distribution）），而上述方差公式是全方差定律（law of total variance）$\operatorname{Var}(Y) = \mathrm{E}\big(\operatorname{Var}(Y \mid N)\big) + \operatorname{Var}\big(\mathrm{E}(Y \mid N)\big)$ 的一個特例。概率母函數的複合也是分支過程（branching process）的核心：在分支過程中，$G(G(\cdots G(t)))$ 給出第 n 代的規模，而滅絕的概率是 G 的一個不動點（fixed point）。Grimmett 和 Welsh 的 *Probability: An Introduction*（《概率導論》）正是在這個程度上講解母函數、隨機和及分支過程。

詳細解答

STEP III 2011，第 13 題

概率與統計 • 離散分佈的眾數

> **STEP III 2011，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/11-s3-q13

關鍵思路

要找出離散隨機變量（discrete random variable）最可能的取值，我們不用微分，而是比較相鄰的概率（probability）。比值 $\mathrm{P}(X = r+1)/\mathrm{P}(X = r)$ 是 r 的一個簡單有理函數（rational function），而問它何時大於 1，就變成一個關於 r 的線性不等式（linear inequality）。比值大於 1 時概率上升，一旦跌到 1 以下概率便下降，所以峯值就在轉折前的最後一個 r。同一方法無論放回抽樣（with replacement）還是不放回抽樣（without replacement）都適用，改變的只是那個比值。

解答

第 (i) 部分

在放回抽樣下，各次抽取互相獨立，每次抽到黑球的概率都是 $p = b/n$，所以 $X \sim \mathrm{B}(k, p)$：

$$\mathrm{P}(X = r) = \binom{k}{r} p^r (1-p)^{k-r}, \qquad r = 0, 1, \ldots, k.$$

先假設 $0 < b < n$，使得這些概率全部為正。對於 $0 \le r \le k-1$，

$$\frac{\mathrm{P}(X = r+1)}{\mathrm{P}(X = r)} = \frac{k-r}{r+1} \cdot \frac{p}{1-p} = \frac{(k-r)\,b}{(r+1)(n-b)}.$$

由於分母為正，這個比值大於 1 當且僅當

$$\begin{aligned} & (k-r)b > (r+1)(n-b) \\ \iff \quad & kb - rb > rn - rb + n - b \\ \iff \quad & (k+1)b > (r+1)n \\ \iff \quad & r+1 < \frac{(k+1)b}{n}. \end{aligned}$$

同理，比值等於 1 當且僅當 $r+1=m$，小於 1 當且僅當 $r+1>m$，其中

$$m = \frac{(k+1)b}{n}.$$

因此，隨着 r 增大，每一步若終點 $r+1<m$，$\mathrm{P}(X=r)$ 便嚴格遞增；若終點 $r+1>m$，便嚴格遞減。注意由於 $0<b<n$，有 $0<m<k+1$，所以轉折發生在取值範圍 $0,\ldots,k$ 之內。

想一想 如果 m 不是整數，最大的概率究竟在哪裏？

若 m 不是整數，記 $M=\lfloor m \rfloor$，則 $M<m<M+1$ 且 $0 \le M \le k$。概率嚴格遞增直至 $\mathrm{P}(X=M)$（每一步的終點都 $\le M < m$），之後嚴格遞減（之後每一步的終點都 $\ge M+1 > m$）。因此黑球數目唯一的最可能取值是

$$\left\lfloor \frac{(k+1)b}{n} \right\rfloor,$$

正如所求。

不唯一的情況。若 m 是整數，由 $r=m-1$ 到 $r=m$ 這一步的比值恰好是 1，所以 $\mathrm{P}(X=m-1)=\mathrm{P}(X=m)$，而這兩個相等的值比其餘所有值都大（概率嚴格上升至 $m-1$，並在 m 之後嚴格下降）。由於 $1 \le m \le k$，兩個取值都是可能的。所以答案不唯一，當且僅當 $(k+1)b/n$ 是整數（且 $0<b<n$），此時 $m-1$ 和 m 同樣是最可能的取值。例如 $n=3$、$b=1$、$k=5$ 給出 $m=2$，而確實 $\mathrm{P}(X=1)=\mathrm{P}(X=2)=\frac{80}{243}$。

平凡情況不在上述討論之內：若 $b=0$，則必然 $X=0$，與公式相符；若 $b=n$，則必然 $X=k$，而公式（會給出 $k+1$）並不適用，因為上面的比值涉及除以 $n-b$。

第 (ii) 部分

在不放回抽樣下，X 服從超幾何分佈（hypergeometric distribution）：把樣本視為 k 個球組成的無序集合，所有 $\binom{n}{k}$ 個集合機會均等，

$$\mathrm{P}(X=r) = \frac{\binom{b}{r}\binom{n-b}{k-r}}{\binom{n}{k}}.$$

由於 $k \le b$，最大可能取值是 $r = k$；最小的是 $r_0 = \max(0, k-(n-b))$，因為非黑球最多只能抽出 $n-b$ 個。對於 $r_0 \le r \le k$，概率為正；而對於 $r_0 \le r \le k-1$，利用 $\binom{a}{r+1} = \binom{a}{r}\frac{a-r}{r+1}$ 和 $\binom{a}{s-1} = \binom{a}{s}\frac{s}{a-s+1}$，

$$\frac{\mathrm{P}(X=r+1)}{\mathrm{P}(X=r)} = \frac{b-r}{r+1}\cdot\frac{k-r}{n-b-k+r+1}.$$

在此範圍內兩個分母都為正，所以比值大於 1 當且僅當

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

展開後，左邊是 $bk-r(b+k)+r^2$，右邊是 $(r+1)(n-b-k+1)+r^2+r$。r^2 項互相抵消，不等式變為

$$\begin{aligned} & bk-r(b+k) > r(n-b-k+1)+(n-b-k+1)+r \\ \iff\quad & bk+b+k+1 > r(n+2)+(n+2) \\ \iff\quad & r+1 < \frac{(b+1)(k+1)}{n+2}. \end{aligned}$$

想一想 轉折點是否真的落在可能取值的範圍 $r_0,\ldots,k$ 之內？

設 $m = \dfrac{(b+1)(k+1)}{n+2}$。由於 $b+1 < n+2$，有 $m < k+1$。至於下端，$m > 0$ 是顯然的；而若 $r_0 = k+b-n > 0$，則需要 $(b+1)(k+1) > (n+2)(k+b-n)$。記 $x = n-b$ 及 $y = n-k$，左邊減右邊等於

$$(n+1-x)(n+1-y) - \left[(n+1)^2 - 1 - (n+2)(x+y)\right] = 1+x+y+xy > 0.$$

所以在所有情況下都有 $r_0 < m < k+1$，第 (i) 部分的論證可以原封不動地套用：當 $r+1 < m$ 時概率嚴格上升，一旦 $r+1 > m$ 便嚴格下降。因此，在唯一的情況下，黑球數目最可能的取值是

$$\left\lfloor \frac{(b+1)(k+1)}{n+2} \right\rfloor.$$

它不唯一，當且僅當 $(b+1)(k+1)/(n+2)$ 是整數；此時 $m-1$ 和 m（兩者都是可能的取值，因為 $r_0 \le m-1$ 且 $m \le k$）機會相等，並且比其他任何取值都更可能。例如 $n=10$、$b=7$、$k=5$ 給出 $m=4$，而 $X=3$ 與 $X=4$ 打和，概率各為 $\frac{105}{252} = \frac{5}{12}$。

驗算

- 我們用電腦以分數精確計算了所有 $n \le 24$、$0 \le b \le n$ 及 $1 \le k \le 19$（第 (ii) 部分另要求 $k \le b$）的兩個分佈，以窮舉法找出眾數，並在所有情況下確認了兩條公式及兩個打和條件（第 (ii) 部分共 2565 個情況）。

- 對於 $n = 10$、$b = 6$、$k = 4$：$m = \frac{35}{12}$，公式給出 2，而 $X = 2$ 確實是唯一的眾數。
- 若 n 遠大於 k，取走幾個球幾乎不改變黑球所佔的比例，所以第 (ii) 部分應該與第 (i) 部分相若。事實正是如此：$(b+1)(k+1)/(n+2) \approx (k+1)b/n$。

常見錯誤

- 斷定眾數（mode）是 $\lfloor kb/n \rfloor$，即平均值（mean）的整數部分。眾數接近平均值，但不一定等於其整數部分。
- 忘記當比值恰好等於 1 時，眾數有兩個，而不是一個。
- 在第 (ii) 部分展開 $(r+1)(n-b-k+r+1)$ 時弄錯正負號；r^2 項互相抵消是一個有用的檢查。

STEP 之外

比較相鄰概率是找出任何離散分佈眾數的標準方法，而同一計算作為第 (i) 部分的極限情況，給出泊松分佈（Poisson distribution）的眾數 $\lfloor \lambda \rfloor$。超幾何分佈是抽樣理論的核心，也是統計學中費雪（Fisher）精確檢驗（Fisher's exact test）的基礎。概率像這樣先升後降的分佈稱為單峯分佈（unimodal）；對於二項分佈和超幾何分佈，這源自一個更強的性質——對數凹性（log-concavity）。如欲深入了解，可參閱 Grimmett 和 Welsh 的 *Probability: An Introduction*（《概率導論》），或 Blitzstein 和 Hwang 的 *Introduction to Probability*（《概率導論》）。

Beyond STEP

STEP 是大多數中學生一生中應考過最艱深的數學考試，也是最出色的考試之一：題目簡短、陌生而深刻，當中最精彩的，其實是喬裝了的大學數學小品。

本書精選最有趣的 STEP 試題，由 1987 年首次開考之前的樣本試卷（specimen papers），到 2025 年的試卷，並以每一道題目作為一扇門。每道題目都會仔細解答，沿途提出優秀學生會問的問題，然後帶你穿過這扇門：這個想法從何而來，到了大學會演變成甚麼，以及接下來該讀些甚麼。

本書為正在準備 STEP、有天分又有志投身數學的學生而寫。他們想要的不止是模擬試卷：他們想開始像數學系本科生那樣思考，並一探前方有甚麼正等着自己。

- 共 13 章，涵蓋純數學（pure mathematics）、力學（mechanics）、概率（probability）與統計（statistics）
- 136 道 STEP 歷屆試題，每題均附引導式解答
- 每道題目背後的大學數學
- 附答案的延伸練習，以及延伸閱讀

2011 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。