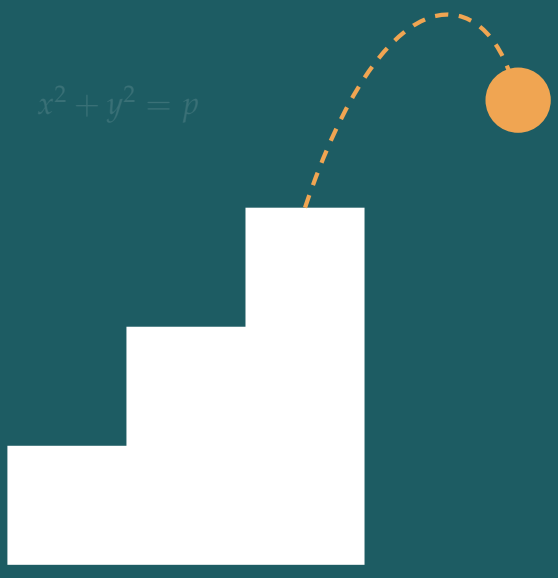
STEP 備試 · 大學數學

Beyond STEP

STEP I, II 及 III

1998 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

1998 年詳細解答

Keith McNulty

Beyond STEP：以歷屆 STEP 試題為門，走進大學數學

1998 年詳細解答

1998 年 STEP 試卷（STEP I, II 及 III）全部 42 題的詳細解答。自樣本試卷以來每一道 STEP 題目的詳細解答、*Beyond STEP* 全書以及各章的單行本，均可在 `beyondstep.keithmcnulty.org` 免費閱讀及下載。

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

由機器從英文翻譯而成；尚未經審閱。

2026 年 11 月首次於網上發表。

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

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

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

目錄

STEP I

第 1 題 數字模式的計數 純數學 . . . 3

第 2 題 代換積分法 純數學 . . . 6

第 3 題 真假命題的證明 純數學 . . . 9

第 4 題 約束最優化 純數學 . . . 13

第 5 題 複數與旋轉 純數學 . . . 17

第 6 題 三角函數遞推關係 純數學 . . . 20

第 7 題 等比級數 純數學 . . . 23

第 8 題 二階微分方程 純數學 . . . 26

第 9 題 擺的碰撞 力學 . . . 29

第 10 題 拋體運動 力學 . . . 32

第 11 題 繩過滑輪的連接質點 力學 . . . 35

第 12 題 離散數與獨立性 概率與統計 . . . 38

第 13 題 放回抽樣 概率與統計 . . . 42

第 14 題 幾何概率 概率與統計 . . . 45

STEP II

第 1 題 丟番圖方程與奇偶性 純數學 . . . 49

第 2 題 二項式級數近似 純數學 . . . 52

第 3 題 裂項相消級數 純數學 . . . 55

第 4 題 分部積分法 純數學 . . . 58

第 5 題 複數的三角不等式 純數學 . . . 61

第 6 題 擺線與法線 純數學 . . . 64

第 7 題 以導數證明不等式 純數學 . . . 68

第 8 題 向量幾何 純數學 . . . 72

第 9 題 桁架的靜力學 力學 . . . 76
第 10 題 移動楔上的質點 力學 . . . 80
第 11 題 拋體運動 力學 . . . 84
第 12 題 貝葉斯定理 概率與統計 . . . 88
第 13 題 指數分佈 概率與統計 . . . 91
第 14 題 項分佈與泊松分佈的期望值 概率與統計 . . . 94

STEP III

第 1 題 曲線描繪 純數學 . . . 97
第 2 題 Beta 積分的遞推關係 純數學 . . . 101
第 3 題 線性遞推關係 純數學 . . . 104
第 4 題 極坐標玫瑰線 純數學 . . . 108
第 5 題 矩陣指數 純數學 . . . 112
第 6 題 多面體的體積 純數學 . . . 116
第 7 題 曲線描繪與積分 純數學 . . . 120
第 8 題 球面的向量幾何 純數學 . . . 124
第 9 題 轉動慣量 力學 . . . 127
第 10 題 鏈碰撞 力學 . . . 131
第 11 題 振幅擺 力學 . . . 135
第 12 題 網絡可靠度 概率與統計 . . . 139
第 13 題 最大似然估計 概率與統計 . . . 142
第 14 題 無偏估計量 概率與統計 . . . 146

詳細解答

STEP I 1998，第 1 題

純數學 • 數字模式的計數

> **STEP I 1998，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/98-s1-q1

關鍵思路

不要嘗試把數逐個列出。應該先選定出現的是哪兩個數字，再點算同時用上這兩個數字的五位數字串（string）。唯一的麻煩是一個數不能以 0 開頭，所以我們按 0 是否兩個數字之一來分開點算。另外，別忘了檢查範圍的兩端。

解答

由 10 000 到 99 999 的每個整數都恰好有五位，而且首位數字不是零。範圍內餘下的唯一整數是 100 000，我們留到最後處理。

一個五位數恰好含有兩個不同的數字，即是說它的所有數字都來自某一對 $\{a, b\}$，而且 a 和 b 都確實出現。每個這樣的數都決定了它所屬的那一對，所以我們可以逐對點算，然後相加。

情況 1：兩個數字都不是 0

從 $1, 2, \ldots, 9$ 中選出兩個不同的數字，共有 $\binom{9}{2} = 36$ 種選法。對於固定的一對 $\{a, b\}$，五個位置各自可以是 a 或 b，得出 $2^5 = 32$ 個數字串。其中兩個只用了一個數字（$aaaaa$ 和 $bbbbb$），所以有 $32 - 2 = 30$ 個同時用上兩個數字。它們全都不以 0 開頭，所以這 30 個都是真正的五位數。這個情況得出

$$36 \times 30 = 1080.$$

> **想一想** 為甚麼不能用完全相同的方法處理 $\{0, a\}$ 這一對？

因為在 32 個由 0 和 a 組成的數字串之中，有一半以 0 開頭，而這些並不是五位數（例如

$0a0aa$ 其實是一個四位數）。所以我們必須先固定首位數字，然後才點算。

情況 2：其中一個數字是 0

另一個數字 a 可以是 $1,\ldots,9$ 中的任何一個，共 9 種選擇。首位數字不能是 0，所以必定是 a。餘下四個位置各自可以是 0 或 a，得出 $2^4 = 16$ 種可能。當中恰好有一個，即 $aaaa$，不含 0，這樣整個數便只有一個不同的數字。所以對每個 a 都有 $16 - 1 = 15$ 個數，這個情況得出

$$9 \times 15 = 135.$$

端點 100 000

這個範圍包括兩個端點，而 100 000 只用了 1 和 0 這兩個數字，所以它也要計算在內。

總數

$$1080 + 135 + 1 = 1216.$$

所以這樣的整數共有 **1216** 個。

驗算

- 用電腦檢查由 10 000 到 100 000 的每一個整數，點算其中數字所組成的集合大小為 2 的數，結果是 1216。這也印證了上面的分拆：1080 個不含 0，135 個含有它，再加上 100 000。
- 把同一方法用於一個可以用手逐一列出的較小情況。對於兩位數，這方法得出 $\binom{9}{2}(2^2 - 2) + 9(2^1 - 1) = 72 + 9 = 81$。這與直接點算的結果吻合：共有 90 個兩位數，減去 9 個各位數字相同的數（repdigit）$11, 22, \ldots, 99$。

常見錯誤

- 忘記了 100 000，得出 1215。遇到「包括首尾」的範圍，務必留意兩個端點。
- 計算成 $\binom{10}{2} \times 30 = 1350$，這把以 0 開頭的數也算了進去。
- 在情況 2 中減去 2 而不是 1：全是零的數字串在我們把首位數字固定為 a 時已經排除了。

STEP 之外

從所有由 $\{a,b\}$ 組成的數字串中減去只用一個數字的數字串，這個技巧是容斥原理（inclusion–exclusion principle）最簡單的情況。一般而言，長度為 n 而且用上全部 k 個給定符號的字串，數目是 $\sum_{j=0}^{k}(-1)^j\binom{k}{j}(k-j)^n$，也就是 $k!$ 乘以一個第二類斯特靈數（Stirling number of the second kind）。這類點算結果在組合數學（combinatorics）和概率（probability）中隨處可見，例如優惠券收集問題（coupon collector problem）。想多做這類點算題，可以試試 Grimmett and Welsh 的 *Probability: An Introduction*，或者 Siklos 的 *Advanced Problems in Mathematics* 中的組合數學題目。

詳細解答

STEP I 1998，第 2 題

純數學 • 代換積分法

STEP I 1998，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/98-s1-q2

關鍵思路

那個不好處理的式子 $(x^2+1)^{1/2}+x$ 位於 f 之內，所以最自然的做法就是把它整個設為 t。出人意表的是，這個關係很容易反過來解：先把根式單獨移到一邊再平方，x^2 項便會互相抵消，剩下的 x 只是 t 的一個簡單函數。第一條恆等式（identity）一旦到手，第二個積分不過是 $\mathrm{f}(t)=t^{-3}$ 的特殊情況。

解答

變數代換

設

$$t=(x^2+1)^{1/2}+x.$$

當 x 由 0 走到 ∞ 時，$(x^2+1)^{1/2}$ 和 x 都在遞增，所以 t 嚴格遞增（strictly increasing），由 $x=0$ 時的 $t=1$ 出發，並隨 $x\to\infty$ 趨向 ∞。因此這個代換（substitution）是由 $[0,\infty)$ 到 $[1,\infty)$ 的一對一（one-to-one）滿射，新的積分上下限是 1 和 ∞。

想一想 怎樣把 x 用 t 表示出來？

把平方根單獨移到一邊，然後平方：由 $t-x=(x^2+1)^{1/2}$ 得

$$t^2-2tx+x^2=x^2+1 \quad\Longrightarrow\quad x=\frac{t^2-1}{2t}=\frac{1}{2}\left(t-t^{-1}\right),$$

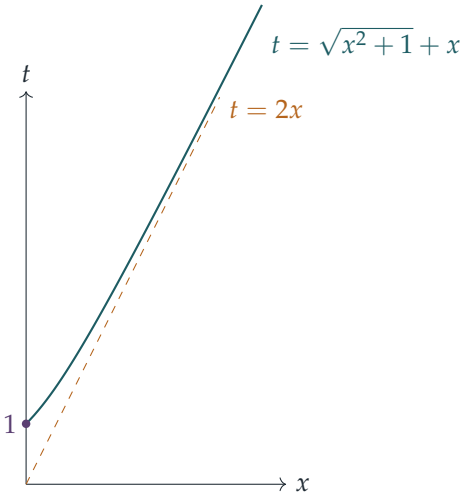
其中因為 $t\geq 1$，除以 t 並無問題。（平方可能會引入增根，但這裏 $t-x=(t+t^{-1})/2>0$，所以我們平方的那個根確實是正數。）求導得

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

因此

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

這正是所求的恆等式。



從草圖可見，t 由 1 開始穩定上升，所以值域中沒有任何部分被重複覆蓋。

最後部分

取 $\mathrm{f}(t) = t^{-3}$。恆等式的左邊正好就是我們要求的積分，所以

$$\begin{aligned}\int_0^\infty \big((x^2+1)^{1/2}+x\big)^{-3}\,\mathrm{d}x &= \frac{1}{2}\int_1^\infty \big(t^{-3}+t^{-5}\big)\,\mathrm{d}t \\ &= \frac{1}{2}\Big[-\tfrac{1}{2}t^{-2}-\tfrac{1}{4}t^{-4}\Big]_1^\infty \\ &= \frac{1}{2}\Big(\frac{1}{2}+\frac{1}{4}\Big) = \frac{3}{8},\end{aligned}$$

正如所求。兩個廣義積分（improper integral）都收斂（converge），因為 t^{-3} 和 t^{-5} 在 $[1,\infty)$ 上可積。

想一想 有沒有更簡潔的方法去理解這個代換？

有：設 $x = \sinh\theta$。這樣 $(x^2+1)^{1/2} = \cosh\theta$，所以 $t = \cosh\theta + \sinh\theta = \mathrm{e}^\theta$。我們的公式 $x = \frac{1}{2}(t - t^{-1})$ 其實就是 $\sinh\theta = \frac{1}{2}(\mathrm{e}^\theta - \mathrm{e}^{-\theta})$ 換了個樣子而已。

驗算

- 當 $x = 0$ 時，$t = 1$，而 $x = \frac{1}{2}(1-1) = 0$：反函數公式吻合。當 $x = \frac{3}{4}$ 時，$t = \frac{5}{4} + \frac{3}{4} = 2$，而 $\frac{1}{2}(2 - \frac{1}{2}) = \frac{3}{4}$。
- 當 x 很大時，$t \approx 2x$，所以最後部分的被積函數（integrand）表現得像 $(2x)^{-3}$，積分必定收斂。
- 數值上，對原來的積分使用辛普森法則（Simpson's rule）得 0.375000，與 $\frac{3}{8}$ 相符。第一部分的恆等式亦可用數值方法驗證：取 $\mathrm{f}(t) = \mathrm{e}^{-t}$（兩邊均為 0.258187）和 $\mathrm{f}(t) = 1/(1+t^4)$（兩邊均為 0.188387）時都成立。

常見錯誤

- 忘記更改積分上下限：下限變成 $t = 1$，而不是 $t = 0$。
- 把 $\mathrm{d}t$ 用 x 表示，然後苦苦設法消去 x；先把 x 用 t 表示，再對它求導，會容易得多。

STEP 之外

這條恆等式說的是在單調（monotonic）映射下的變數代換；在分析學中，對於廣義積分，這是透過取 $[0, X]$ 上積分的極限來證明其合理性的。這個代換本身就是 sinh 的反函數：$\ln\left(x + \sqrt{x^2+1}\right) = \operatorname{arsinh} x$，這正是 t 原來等於 $\mathrm{e}^{\operatorname{arsinh} x}$ 的原因。如欲參考有關廣義積分和代換的嚴謹論述，可參閱 Bartle 與 Sherbert 的 *Introduction to Real Analysis*（實分析入門）。

詳細解答

STEP I 1998，第 3 題

純數學 • 真假命題的證明

> **STEP I 1998，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/98-s1-q3

關鍵思路

每一部分所需的工具都不同，而首要工作是判斷該命題是真是假。答「真」需要一個涵蓋所有情況的證明；答「假」則只需一個反例（counterexample），或者一個說明符合要求的對象根本不可能存在的論證。取對數（logarithm）可解決 (i)，選一個合適的 θ 值可解決 (ii)，多項式（polynomial）在 θ 很大時的表現可解決 (iii)，而一個完全平方（perfect square）可解決 (iv)。

解答

第 (i) 部分：真

由於 $a > 0$，可以寫成 $a = \mathrm{e}^{\ln a}$。於是

$$a^{\ln b} = \left(\mathrm{e}^{\ln a}\right)^{\ln b} = \mathrm{e}^{\ln a \ln b},$$
$$b^{\ln a} = \left(\mathrm{e}^{\ln b}\right)^{\ln a} = \mathrm{e}^{\ln b \ln a}.$$

因為乘法符合交換律，兩個指數相等，所以對所有 $a, b > 0$ 都有 $a^{\ln b} = b^{\ln a}$。（等價的說法是：兩邊都是正數，而且對數相同，都是 $\ln a \ln b$。）

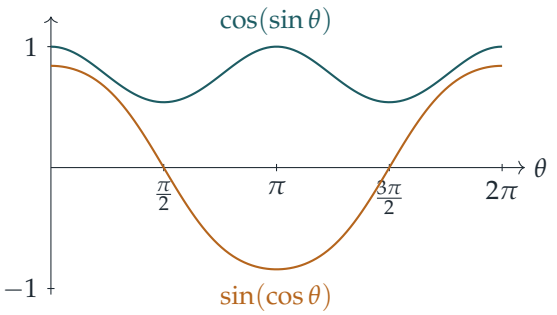
第 (ii) 部分：假

一個反例已經足夠。取 $\theta = 0$：

$$\cos(\sin 0) = \cos 0 = 1, \qquad \sin(\cos 0) = \sin 1.$$

由於 $0 < 1 < \frac{\pi}{2}$，可知 $0 < \sin 1 < 1$（事實上 $\sin 1 \approx 0.841$），所以兩邊並不相等，命題為假。

事實上兩邊永遠不會相等；從圖像可見，$\cos(\sin\theta)$ 總是在 $\sin(\cos\theta)$ 之上。



要證明這一點（本題並不要求），留意 $\sin(\cos\theta) \le \sin|\cos\theta| = \cos\left(\frac{\pi}{2} - |\cos\theta|\right)$，以及 $\cos(\sin\theta) = \cos|\sin\theta|$。$|\sin\theta|$ 和 $\frac{\pi}{2} - |\cos\theta|$ 都落在 $[0, \frac{\pi}{2}]$ 之內，而餘弦在這個區間上遞減；又因為 $|\sin\theta| + |\cos\theta| \le \sqrt{2} < \frac{\pi}{2}$，所以 $|\sin\theta| < \frac{\pi}{2} - |\cos\theta|$。因此對每一個 θ 都有 $\cos(\sin\theta) > \sin(\cos\theta)$。

第 (iii) 部分：假

設 P 為任意多項式。我們要證明存在某個實數 θ 使得 $|P(\theta) - \cos\theta| > 10^{-6}$，以下分兩種情況討論。

P 爲常數，設 $P(\theta) = c$。取 $\theta = 0$ 和 $\theta = \pi$，此時 $\cos\theta$ 分別是 1 和 -1，由三角不等式（triangle inequality）得

$$|c-1| + |c+1| \ge |(c+1) - (c-1)| = 2,$$

所以 $|c-1|$、$|c+1|$ 之中至少有一個不小於 1，遠遠大於 10^{-6}。

P 的次數（*degree*）爲 $n \ge 1$，設 $P(\theta) = a_n\theta^n + \cdots + a_0$，其中 $a_n \ne 0$。當 $\theta \ne 0$ 時，

$$P(\theta) = a_n\theta^n\left(1 + \frac{a_{n-1}}{a_n\theta} + \cdots + \frac{a_0}{a_n\theta^n}\right),$$

當 $\theta \to \infty$ 時，括號內的式子趨向 1，而 $|a_n\theta^n| \to \infty$。所以 $|P(\theta)| \to \infty$，我們可以選一個 θ 使得 $|P(\theta)| > 2$。對這個 θ，由於 $|\cos\theta| \le 1$，

$$|P(\theta) - \cos\theta| \ge |P(\theta)| - |\cos\theta| > 2 - 1 = 1 > 10^{-6}.$$

在兩種情況下，該不等式都會在某處不成立，所以這樣的多項式並不存在。

想一想 可是 $\cos\theta$ 的麥克勞林級數（Maclaurin series）是一個類似多項式的和，而且可以要多接近就有多接近。為甚麼把它截斷不行呢？

在任何固定的有界區間（bounded interval）上，只要取足夠多項，截斷後的麥克勞林級數與 $\cos\theta$ 的差距*確實*在 10^{-6} 之內。但命題要求的是同一個多項式對*所有*實數 θ 都適用，而每一個非常數多項式最終都會奔向 $\pm\infty$，$\cos\theta$ 卻始終困在 $[-1,1]$ 之內。正是「對所有實數 θ」這幾個字令命題為假。

第 (iv) 部分：真

對 $x>0$，

$$x^4+3+x^{-4}-5=x^4-2+x^{-4}=\left(x^2-x^{-2}\right)^2\geq 0,$$

因為實數的平方必為非負數。因此對所有 $x>0$ 都有 $x^4+3+x^{-4}\geq 5$，而等號恰好在 $x^2=x^{-2}$ 時成立，即 $x=1$ 時。（另一個做法：由算術平均—幾何平均不等式（AM–GM inequality）得 $x^4+x^{-4}\geq 2\sqrt{x^4\cdot x^{-4}}=2$。）

總結：(i) 真，(ii) 假，(iii) 假，(iv) 真。

驗算

- (i)：取 $a=2$、$b=3$，$2^{\ln 3}$ 和 $3^{\ln 2}$ 都約等於 2.141；隨機選取正數 a,b，兩邊相符，相對誤差約為 10^{-15}。
- (ii)：在一整個週期內對 θ 取樣，$\cos(\sin\theta)-\sin(\cos\theta)$ 從未低於約 0.107。
- (iii)：cos 的 20 次麥克勞林多項式在 $\theta=1$ 時很準確，但在 $\theta=10$ 時誤差約為 7.5，在 $\theta=30$ 時誤差約為 10^{11}。
- (iv)：在細密的格點上，x^4+3+x^{-4} 的最小值是 5，在 $x=1$ 時取得。

常見錯誤

- 只檢查幾個數值就想「證明」一個假命題，或者因為代數運算出錯而「否證」了一個真命題：動筆之前，先判斷每一部分是真是假。
- 在 (iii) 中忘記了常數多項式：「趨向無窮」的論證需要次數至少為 1。
- 在 (iii) 中引用泰勒級數（Taylor series），把它當作處處適用的多項式；它並不是。

STEP 之外

第 (iii) 部分與魏爾斯特拉斯逼近定理（Weierstrass approximation theorem）形成對比：在任何有界閉區間上，每一個連續函數（continuous function）都可以用多項式一致逼近（uniform approximation），精確度要多高就有多高。但在整條實數線上一致逼近是不可能的，原因正是上面的增長論證。Abbott 的 *Understanding Analysis*（一本實分析入門書）討論了魏爾斯特拉斯定理和一致收斂（uniform convergence）。

詳細解答

STEP I 1998，第 4 題

純數學 • 約束最優化

STEP I 1998，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/98-s1-q4

關鍵思路

內接（inscribed）於圓的長方形，其對角線就是直徑（diameter），所以它的邊長 a 和 b 由畢氏定理（Pythagoras' theorem）連繫起來：a^2+b^2 是固定的。於是整個問題就變成：當 a^2+b^2 保持不變時，$a+b$、a^2+b^2 和 a^3+b^3 會怎樣變化。$n=2$ 這個情況是轉捩點：低於此值時正方形勝出，恰在此值時全部打成平手，高於此值時正方形反而敬陪末座。

解答

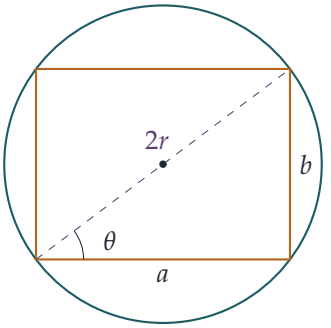
設圓的半徑為 r，長方形的邊長為 $a, b > 0$。長方形的每個角都是直角，所以它所對的弦是直徑（半圓上的圓周角）；因此每條對角線都是直徑，並且

$$a^2+b^2=(2r)^2=4r^2.$$

反過來，任何滿足 $a^2+b^2=4r^2$ 的 $a, b>0$ 都給出一個可放進圓內、且頂點都在圓上的長方形。所以不妨寫成

$$a=2r\cos\theta, \qquad b=2r\sin\theta, \qquad 0<\theta<\tfrac{\pi}{2},$$

其中 θ 是對角線與長度為 a 的邊之間的夾角。



周界

周界（perimeter）為 $P = 2(a+b)$。由於 $(a-b)^2 \geq 0$，我們有 $2ab \leq a^2 + b^2$，所以

$$(a+b)^2 = a^2 + b^2 + 2ab \leq 2(a^2+b^2) = 8r^2,$$

等號成立當且僅當 $a = b$。因為 $a + b > 0$，由此得 $a + b \leq 2\sqrt{2}\,r$，即

$$P \leq 4\sqrt{2}\,r,$$

而 $4\sqrt{2}\,r$ 這個值恰好在 $a = b = \sqrt{2}\,r$ 時取得。所以周界最大的長方形就是正方形，正如題目所要求證明的。

$n = 2$ 的情況

四條邊的平方和是

$$2(a^2+b^2) = 8r^2,$$

不論是哪一個長方形都一樣。所以每個內接長方形都給出相同的值：它們之間無分高下，特別是正方形並不比其他長方形優勝。

$n = 3$ 的情況

現在考慮 $S = 2(a^3+b^3)$。由於每條邊都比直徑短，$0 < a < 2r$ 且 $0 < b < 2r$，所以 $a^3 < 2r\,a^2$ 及 $b^3 < 2r\,b^2$。兩式相加，

$$S = 2(a^3+b^3) < 2 \cdot 2r(a^2+b^2) = 16r^3.$$

另一方面，當 $\theta \to 0$ 時（一個又長又窄的長方形，其中 $a \to 2r$ 而 $b \to 0$），$S \to 2(2r)^3 = 16r^3$。所以 S 可以任意接近 $16r^3$，卻永遠達不到它：*並不存在使 S 最大的長方形*。唯一能達到 $16r^3$ 的「形狀」是退化（degenerate）的情形，即把一條直徑來回描兩次，而那並不是長方形。

想一想 那麼正方形又怎樣了？

它由最好變成了最差。設 $f(\theta) = \cos^3\theta + \sin^3\theta$，則 $S = 16r^3 f(\theta)$，而

$$f'(\theta) = -3\cos^2\theta\sin\theta + 3\sin^2\theta\cos\theta = 3\sin\theta\cos\theta\,(\sin\theta - \cos\theta).$$

在 $0 < \theta < \frac{\pi}{2}$ 上，因子 $\sin\theta\cos\theta$ 是正的，而 $\sin\theta - \cos\theta$ 在 $\theta < \frac{\pi}{4}$ 時為負，在 $\theta > \frac{\pi}{4}$ 時為正。所以 f 先遞減後遞增，而正方形（$\theta = \frac{\pi}{4}$）給出的是最小值，

$$S_{\min} = 16r^3\left(2\cdot\frac{1}{2\sqrt{2}}\right) = 8\sqrt{2}\,r^3.$$

總括而言，對於 $n = 3$，S 的取值填滿區間 $[8\sqrt{2}\,r^3,\ 16r^3)$：正方形給出最小值，而沒有任何長方形給出最大值。

對每個 $n > 2$ 情況都一樣：在 $0 < \theta < \frac{\pi}{2}$ 上有 $\cos^n\theta + \sin^n\theta < \cos^2\theta + \sin^2\theta = 1$，因為 $\cos\theta$ 和 $\sin\theta$ 都嚴格介乎 0 與 1 之間，所以只有把長方形壓扁，才能趨近上確界 $2(2r)^n$。

驗算

- 取 $r = 1.7$，我們以數值方法讓 θ 掃過 $(0, \frac{\pi}{2})$：周界在 $\theta = \frac{\pi}{4}$ 處達到峯值，其值為 $4\sqrt{2}\,r \approx 9.617$；平方和恆等於 $8r^2 = 23.12$；而立方和則由接近 $16r^3 \approx 78.61$ 下降至正方形處的 $8\sqrt{2}\,r^3 \approx 55.58$，然後再度上升，始終達不到 78.61。
- 一個簡單的合理性檢查：單位圓內的 $\sqrt{2}\times\sqrt{2}$ 正方形周界為 $4\sqrt{2} \approx 5.66$，而一個 1.9×0.62 的長方形（幾乎就是一條直徑）周界約為 5.04。

常見錯誤

- 說 $n = 3$ 時的最大值是 $16r^3$，「在 $\theta = 0$ 處取得」。那是一條線段，不是長方形，所以最大值並不存在；你必須指出這一點，並說明為何不等式是嚴格的。
- 求出 $n = 3$ 時的駐點（stationary point）$\theta = \frac{\pi}{4}$ 後，沒有檢查其性質就把它當作極大點。

STEP 之外

$n = 3$ 的情況讓我們初次體會上確界（supremum）與最大值（maximum）的分別：開區間上的連續函數（continuous function）不一定取得它的界，這正是極值定理（extreme value theorem）堅持要求有界閉區間的原因。在 $n = 2$ 處出現的逆轉，是冪平均不等

式（power-mean inequalities）的一個例子，這類不等式比較不同 p 值下的 $\left(\frac{a^p+b^p}{2}\right)^{1/p}$。Abbott 的 *Understanding Analysis*（一本數學分析入門書）對上確界和極值定理有嚴謹細緻的論述。

詳細解答

STEP I 1998，第 5 題

純數學 • 複數與旋轉

> **STEP I 1998，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/98-s1-q5

關鍵思路

兩部分都是讓複數（complex number）的代數運算替我們完成工作。在第 (i) 部分，由正六邊形（regular hexagon）的一個頂點走到下一個頂點，就是繞中心旋轉（rotation）60°，而繞原點的旋轉不過是乘以一個模（modulus）為 1 的複數，所以我們相對於 Q 來計算。在第 (ii) 部分，一個數是實數，當且僅當它的虛部（imaginary part）為零，所以我們把每個分母實數化（realise），再令虛部等於零。

解答

第 (i) 部分

相對於中心來計算。由 Q 到 A 的向量為

$$A - Q = (10 + 2i) - (4 + 6i) = 6 - 4i.$$

正六邊形的六個頂點均勻地分佈在以 Q 為圓心的圓上，所以角 AQB 是 $360°/6 = 60°$，而 QB 與 QA 等長。順時針移動即是轉過 $-60°$，也就是乘以

$$\mathrm{e}^{-i\pi/3} = \cos\tfrac{\pi}{3} - i\sin\tfrac{\pi}{3} = \tfrac{1}{2} - \tfrac{\sqrt{3}}{2}\,i.$$

? **想一想** 為甚麼要乘 $A - Q$，而不是 A 本身？

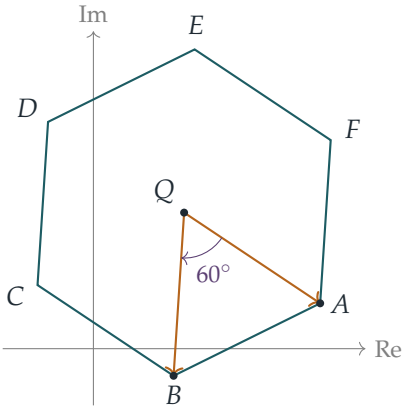
乘以 $\mathrm{e}^{-i\pi/3}$ 是繞原點旋轉。要繞 Q 旋轉，我們先把 Q 移到原點（減去 Q），旋轉，然後

移回去（加上 Q）。所以

$$\begin{aligned} B - Q &= (6-4i)\left(\tfrac{1}{2} - \tfrac{\sqrt{3}}{2}\,i\right) \\ &= 3 - 3\sqrt{3}\,i - 2i + 2\sqrt{3}\,i^2 \\ &= (3-2\sqrt{3}) - (2+3\sqrt{3})\,i, \end{aligned}$$

再加上 $Q = 4+6i$，得

$$\boxed{B = (7-2\sqrt{3}) + (4-3\sqrt{3})\,i} \approx 3.536 - 1.196\,i.$$



圖形印證了方向：A 位於 Q 的右下方，由它出發順時針走，下一個頂點幾乎在 Q 的正下方，正如我們求得的 B 那樣。

第 (ii) 部分

把分子和分母同乘以共軛複數（conjugate），將每個分母實數化，當中用到 $(1+ki)(1-ki) = 1+k^2$：

$$\frac{1}{1+i} = \frac{1-i}{2}, \qquad \frac{1}{1+2i} = \frac{1-2i}{5}, \qquad \frac{1}{1+3i} = \frac{1-3i}{10}.$$

由於 a、b、c 都是實數，整個算式的虛部為

$$-\frac{a}{2} - \frac{2b}{5} - \frac{3c}{10} = -\frac{5a+4b+3c}{10}.$$

這個算式是實數，當且僅當這個虛部為零，所以所求的條件是

$$\boxed{5a+4b+3c=0}.$$

（任何非零整數倍，例如 $-5a-4b-3c=0$，都是同一個條件。）事實上，由於我們求出的是虛部的準確值，這個條件既是充分條件（sufficient），也是必要條件（necessary）。

驗算

- 在第 (i) 部分，$|A-Q| = |6-4i| = \sqrt{52} = 2\sqrt{13}$。我們求得的 B 給出 $|B-Q|^2 = (3-2\sqrt{3})^2 + (2+3\sqrt{3})^2 = 21 - 12\sqrt{3} + 31 + 12\sqrt{3} = 52$，所以 B 確實在正確的圓上；而數值上亦有 $|B-A| \approx 7.211 = |A-Q|$，這正是等邊三角形 QAB 所必須滿足的。
- 在第 (ii) 部分，試取 $a=3$、$b=0$、$c=-5$（它們滿足 $5a+4b+3c=0$）：$\frac{3(1-i)}{2} - \frac{5(1-3i)}{10} = \frac{3}{2} - \frac{1}{2} + \left(-\frac{3}{2} + \frac{3}{2}\right) i = 1$，這是實數。
- 以 10 000 組滿足 $c = -(5a+4b)/3$ 的隨機三元組作數值檢驗，每一次虛部都是零。

常見錯誤

- 逆時針旋轉（乘以 $e^{i\pi/3}$）會得到 $(7+2\sqrt{3}) + (4+3\sqrt{3})i$，那是頂點 F，而不是 B。
- 把 A 繞原點旋轉，而不是繞 Q 旋轉。
- 在第 (ii) 部分，令實部等於零；或者忘記了 a、b、c 必須是實數，才可以這樣直接讀出虛部。

STEP 之外

乘以 $e^{i\theta}$ 是一個旋轉，而映射 $z \mapsto \alpha z + \beta$ 則是先旋轉和放大（enlargement），再作平移（translation）；這些都是莫比烏斯變換（Möbius transformations）最簡單的例子，而莫比烏斯變換是複分析（complex analysis）入門課程的研習內容。以原點為中心的正多邊形，其頂點就是 z_0 乘以各個單位根（roots of unity），這個事實是大量代數和幾何的起點。Needham 的 *Visual Complex Analysis*（《複分析：可視化方法》）把這種從幾何角度看複數乘法的觀點發揮得十分精彩。

詳細解答

STEP I 1998，第 6 題

純數學 • 三角函數遞推關係

STEP I 1998，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/98-s1-q6

關鍵思路

遞推關係（recurrence）的每一步都是把某個量與一個餘弦取平均，而半角恆等式（half-angle identity）$\frac{1}{2}(1+\cos\theta) = \cos^2\frac{\theta}{2}$ 會把這個平均變成一個完全平方。b_{n+1} 的公式中的平方根隨即又把這個平方消去。一旦留意到每一項都以前一項為因子，一般規律就不難猜出，而驗證它只需一步歸納法（induction）。

解答

全題中 $0 < x < \frac{\pi}{2}$，所以每個角 $x/2^k$（$k \geq 1$）都位於 $(0, \frac{\pi}{4})$ 內，其餘弦為正。在 b_{n+1} 的定義中，我們取正的平方根。

求 a_2 和 b_2

利用 $1+\cos\theta = 2\cos^2\frac{\theta}{2}$，

$$a_2 = \tfrac{1}{2}(\cos x + 1) = \cos^2\frac{x}{2}, \qquad b_2 = (a_2 b_1)^{1/2} = \left(\cos^2\frac{x}{2}\right)^{1/2} = \cos\frac{x}{2},$$

其中最後一步用到 $\cos\frac{x}{2} > 0$。

求 a_3 和 b_3

抽出公因子 $\cos\frac{x}{2}$，然後再用一次半角恆等式，這次取 $\theta = \frac{x}{2}$：

$$a_3 = \tfrac{1}{2}\left(\cos^2\frac{x}{2} + \cos\frac{x}{2}\right) = \cos\frac{x}{2}\cdot\tfrac{1}{2}\left(1+\cos\frac{x}{2}\right) = \cos\frac{x}{2}\cos^2\frac{x}{4}.$$

於是

$$b_3 = (a_3 b_2)^{1/2} = \left(\cos^2\frac{x}{2}\cos^2\frac{x}{4}\right)^{1/2} = \cos\frac{x}{2}\cos\frac{x}{4},$$

因為兩個餘弦都是正數。這就是題目要求的兩個表達式。

一般規律

比較目前得到的各項，

$$b_2 = \cos\frac{x}{2},\quad b_3 = \cos\frac{x}{2}\cos\frac{x}{4},\qquad a_2 = b_2\cos\frac{x}{2},\quad a_3 = b_3\cos\frac{x}{4}.$$

每個 b 都比前一個多乘一個餘弦，角度減半；而每個 a 則是對應的 b 再乘一次它最後的那個餘弦。因此我們猜想，對 $n \geq 2$，

$$b_n = \cos\frac{x}{2}\cos\frac{x}{4}\cdots\cos\frac{x}{2^{n-1}} = \prod_{k=1}^{n-1}\cos\frac{x}{2^k},$$
$$a_n = b_n\cos\frac{x}{2^{n-1}} = \cos\frac{x}{2}\cos\frac{x}{4}\cdots\cos\frac{x}{2^{n-2}}\,\cos^2\frac{x}{2^{n-1}}.$$

想一想 要驗證一個猜想滿足遞推關係，究竟需要檢查甚麼？

我們要檢查兩件事：猜想與初始值吻合；以及若公式對 a_n 和 b_n 成立，則遞推關係會給出 a_{n+1} 和 b_{n+1} 的公式。兩者合起來，便由歸納法證明公式對所有 $n \geq 2$ 成立。

初始值。當 $n = 2$ 時，猜想給出 $b_2 = \cos\frac{x}{2}$ 和 $a_2 = \cos^2\frac{x}{2}$，與上面求得的結果吻合。

歸納步驟。假設公式對某個 $n \geq 2$ 成立。設 $\theta = x/2^{n-1}$，則 $x/2^n = \theta/2$。於是

$$a_{n+1} = \tfrac{1}{2}(a_n + b_n) = b_n\cdot\tfrac{1}{2}\big(1+\cos\theta\big) = b_n\cos^2\frac{\theta}{2}.$$

對 b_{n+1} 的猜想是 $b_n\cos\frac{\theta}{2}$（多乘一個餘弦，角度為 $x/2^n$），所以上式即 $a_{n+1} = b_{n+1}\cos\frac{x}{2^n}$，正是所猜想的 a_{n+1} 的形式。接著，

$$b_{n+1} = \big(a_{n+1}b_n\big)^{1/2} = \left(b_n^2\cos^2\frac{\theta}{2}\right)^{1/2} = b_n\cos\frac{x}{2^n} = \prod_{k=1}^{n}\cos\frac{x}{2^k},$$

這裏取正根，而這樣做是合理的，因為 b_n 和 $\cos\frac{x}{2^n}$ 都是正數。這正是所猜想的 b_{n+1} 的形式，所以遞推關係的兩條等式都得到滿足，公式對所有 $n \geq 2$ 成立。

事實上，若約定空積（empty product）等於 1，公式亦給出 $b_1 = 1$ 和 $a_1 = \cos x$，所以公式由 $n = 1$ 起已經成立。

驗算

- 取 $x = \frac{\pi}{3}$。直接按遞推關係計算，得 $a_3 \approx 0.808013$ 和 $b_3 \approx 0.836516$，而 $\cos\frac{\pi}{6}\cos^2\frac{\pi}{12}$ 和 $\cos\frac{\pi}{6}\cos\frac{\pi}{12}$ 給出相同的數值。

- 在 $(0, \frac{\pi}{2})$ 內隨機抽取 2000 個 x，我們把遞推關係與乘積公式逐一比較，直至 $n = 24$：兩者相差不超過 10^{-15}。
- 當 $x \to 0$ 時，每個餘弦都趨向 1，a_1 和 b_1 亦然。所有項都趨向 1，與預期相符。

常見錯誤

- 寫下 $\sqrt{\cos^2\theta} = \cos\theta$ 而沒有說明為何 $\cos\theta > 0$。這裏等式之所以成立，是因為每個角都位於 $(0, \frac{\pi}{4})$ 內。
- 由 $n = 2$ 和 3 猜出規律後便停筆。題目要求我們驗證遞推關係，所以歸納步驟必不可少。

STEP 之外

把 b_n 乘以 $\sin\frac{x}{2^{n-1}}$，並反覆運用 $2\sin\phi\cos\phi = \sin 2\phi$，可得 $b_n = \dfrac{\sin x}{2^{n-1}\sin(x/2^{n-1})}$。當 $n \to \infty$ 時，它趨向 $\dfrac{\sin x}{x}$（取 $x = 1$，遞推關係在數十步之後給出 0.841471）。取 $x = \frac{\pi}{2}$，便得到關於 $2/\pi$ 的韋達（Viète）無窮乘積（infinite product）。這種「先取平均、再取幾何平均」的做法，與高斯（Gauss）的算術幾何平均（arithmetic–geometric mean）一脈相承，也是學習證明數列（sequence）收斂（converge）並求其極限（limit）的一個很好的入門例子。關於數列極限的嚴謹論述，可參閱 Abbott, *Understanding Analysis*（一本實分析入門教科書）。

詳細解答

STEP I 1998，第 7 題

純數學 • 等比級數

> **STEP I 1998，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/98-s1-q7

關鍵思路

不要逐年追蹤結餘，而應分開追蹤每一筆存款：一筆 £a 的存款在銀行裏賺取了 k 年利息後，會增長至 £ar^k，而總額就是一個等比級數（geometric series）。至於提款，把同一想法反過來用即可。每一筆提款都是由起始結餘中的某一部分支付的，而這部分在提取之前一直在賺取利息，所以我們比較的是同一時刻的價值。

解答

靜止期結束時的金額

利率為 $\rho\%$，即每年年終結餘乘以 $r = 1 + \frac{\rho}{100}$。

把存款的年份編號為 $1, 2, \ldots, m$。在第 k 年年初存入的款項，會在第 $k, k+1, \ldots, m$ 年的年終賺取利息，共 $m - k + 1$ 次。因此在第 m 年年終，它的價值為 ar^{m-k+1}。對 k 求和，第 m 年年終的結餘為

$$\begin{aligned} a\left(r^m + r^{m-1} + \cdots + r\right) &= ar\left(1 + r + \cdots + r^{m-1}\right) \\ &= ar\,\frac{r^m - 1}{r - 1}, \end{aligned}$$

這裏用了等比級數的求和公式（由於 $\rho > 0$，故有 $r \neq 1$）。

在其後的 l 年裏，既無存款亦無提款，所以這個金額再乘以 r 共 l 次。因此這段時期結束時的總額為

$$£a\,\frac{r^{l+1}(r^m - 1)}{r - 1},$$

正如題目所求。把這個金額記作 P。

提款

第一次提款在緊接靜止期之後那一年的年初進行，這正是結餘為 P 的時刻。設 B_k 為剛完成第 k 次提款後的結餘，則

$$B_1 = P - b, \qquad B_k = rB_{k-1} - b \quad (k = 2, \ldots, n),$$

因為在兩次提款之間，這筆錢會賺取一年的利息。把這個遞歸關係（recurrence）逐層展開，得

$$\begin{aligned} B_n &= r^{n-1}P - b\left(r^{n-1} + r^{n-2} + \cdots + r + 1\right) \\ &= r^{n-1}P - b\,\frac{r^n - 1}{r - 1}. \end{aligned}$$

想一想 為甚麼 P 所乘的因子是 r^{n-1}，而不是 r^n？

由第一次提款到第 n 次提款之間，只有 $n-1$ 個年終，所以這筆錢只賺取了 $n-1$ 次利息。最後一次提款之後所派的利息無關重要，因為那時戶口已經清空。同樣道理，第 j 次提款（由最後一次倒數，最後一次為 $j = 0$）在第 n 次提款之前會乘以 r^j。

第 n 次提款之後戶口清空，所以 $B_n = 0$。由此得

$$r^{n-1}P = b\,\frac{r^n - 1}{r - 1}.$$

代入 P：

$$a\,\frac{r^{l+n}(r^m - 1)}{r - 1} = b\,\frac{r^n - 1}{r - 1},$$

於是

$$\boxed{\frac{a}{b} = \frac{r^n - 1}{r^{l+n}\,(r^m - 1)}}$$

要理解這條等式，另一個等價的方法是計算「現值」（present value）。以第一次提款的時刻為準，第 k 次提款的價值為 b/r^{k-1}，而這些價值的總和必須等於 P：

$$P = b\left(1 + r^{-1} + \cdots + r^{-(n-1)}\right).$$

驗算

- 若 $m = n$，答案簡化為 $a/b = r^{-(l+n)}$，即 $b = ar^{l+n}$。這是合理的：把第 k 筆存款與第 k 次提款配對，兩者剛好相隔 $m + l = l + n$ 年，所以每筆存款恰好支付一次提款。
- 當 $r \to 1$（沒有利息）時，$\dfrac{r^n - 1}{r^m - 1} \to \dfrac{n}{m}$，所以 $a/b \to n/m$，即 $ma = nb$。存入的錢等於提出的錢。

- 取 $\rho = 5$、$m = n = 10$、$l = 5$ 及 $a = 1000$，逐年模擬得出 $P \approx 16855.58$。公式給出 $a/b = 1.05^{-15} \approx 0.48102$，所以 $b \approx 2078.93$；用這個 b 值，模擬的戶口最後剛好歸零。此外，用電腦以分數作精確運算，隨機選取 500 組 r、l、m 和 n 的值，兩條公式亦全部得到確認。

常見錯誤

- 在第一個和式中把 r^{m+1} 寫成 r^m。由於存款是在每年年初存入的，即使是最後一筆存款，也會賺取一年的利息。
- 把 $r^{n-1}P$ 寫成 r^nP。第 n 年年終的利息，是在戶口清空之後才派發的。

STEP 之外

這就是年金（annuity）的數學。把各筆款項折現（discounting）至同一日期（即現值）再作比較，是精算數學和金融數學的基本原則。遞歸關係 $B_k = rB_{k-1} - b$ 是一條一階線性差分方程（first-order linear difference equation）。它的解是一個特定的常數加上 r^k 的倍數，與我們求解一階線性微分方程（differential equation）的方法如出一轍；大學一年級有關差分方程和微分方程的課程會探討這種對應關係。想做更多這類題目，可參閱 Siklos 的 *Advanced Problems in Mathematics*。

詳細解答

STEP I 1998，第 8 題

純數學 • 二階微分方程

> **STEP I 1998，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/98-s1-q8

關鍵思路

這條題目雖然披着流體力學的外衣，骨子裏卻是一道純粹的微分方程（differential equation）題。左邊是「導數的導數」，所以只需積分兩次，並記住兩次積分之間除以 r 會令常數變成對數。接着，物理上的合理性（軸上的速度必須有限）確定其中一個常數，而管壁上的無滑移條件（no-slip condition）則確定另一個。

解答

通解

將方程乘以 r，得

$$\frac{\mathrm{d}}{\mathrm{d}r}\left(r\frac{\mathrm{d}v}{\mathrm{d}r}\right) = -kr.$$

對 r 積分一次，

$$r\frac{\mathrm{d}v}{\mathrm{d}r} = -\tfrac{1}{2}kr^2 + A,$$

其中 A 為常數。兩邊除以 r（這樣做沒有問題，因為在管內軸線以外的地方 $r > 0$）：

$$\frac{\mathrm{d}v}{\mathrm{d}r} = -\tfrac{1}{2}kr + \frac{A}{r}.$$

再積分一次，由於 $r > 0$，可寫 $\ln r$ 而不必寫 $\ln|r|$，

$$\boxed{v = -\tfrac{1}{4}kr^2 + A\ln r + B}$$

其中 A 和 B 為任意常數。二階方程的通解（general solution）含有兩個常數，正如所料。

軸線附近的情況

當 $r \to 0^+$ 時，$-\frac{1}{4}kr^2$ 和 B 兩項保持有界（分別趨向 0 和 B），而 $\ln r \to -\infty$。因此若 $A \neq 0$，則 $A \ln r$ 趨向 $-\infty$ 或 $+\infty$（視乎 A 的正負），於是 $|v| \to \infty$。要避免這種情況，唯一的方法是取 $A = 0$，此時 $v \to B$，是有限值。

> **想一想** 為甚麼我們要堅持 $A = 0$？

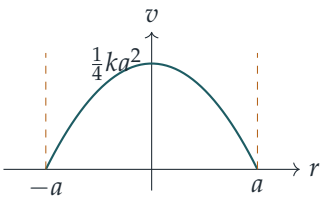
軸線 $r = 0$ 位於流體之內，所以該處的速率必須有限：現實中的流體不會沿着管道中心以無限大的速度流動。$\ln r$ 這個解在 $r > 0$ 時完全滿足方程，但在這裏從物理角度看卻不可接受。（若是兩個同軸圓柱之間的流動，軸線並不屬於流體的一部分，這個解便可以保留。）

速度分佈

取 $A = 0$，便有 $v = B - \frac{1}{4}kr^2$。由條件「在 $r = a$ 處 $v = 0$」得 $B = \frac{1}{4}ka^2$，所以

$$\boxed{v = \tfrac{1}{4}k\left(a^2 - r^2\right).}$$

速度分佈（velocity profile）是一條拋物線：速率在軸上最大，為 $\frac{1}{4}ka^2$，到管壁處則降至零。



（圖中所示為沿直徑切開的截面，虛線代表管壁。）

流量

代入題目給出的公式，

$$\begin{aligned} F &= 2\pi \int_0^a r \cdot \tfrac{1}{4}k\left(a^2 - r^2\right) \mathrm{d}r = \frac{\pi k}{2} \int_0^a \left(a^2 r - r^3\right) \mathrm{d}r \\ &= \frac{\pi k}{2}\left[\frac{a^2 r^2}{2} - \frac{r^4}{4}\right]_0^a = \frac{\pi k}{2} \cdot \frac{a^4}{4}, \end{aligned}$$

所以

$$\boxed{F = \frac{\pi k a^4}{8}.}$$

驗算

- 將 $v = \frac{1}{4}k(a^2 - r^2)$ 代回原方程：$r\,v' = -\frac{1}{2}kr^2$，其導數為 $-kr$，再除以 r 得 $-k$，符合要求。
- 平均速率為 $F/(\pi a^2) = \frac{1}{8}ka^2$，恰好是最大速率 $\frac{1}{4}ka^2$ 的一半，這正是拋物線形速度分佈廣為人知的性質。
- 數值上，取 $k = 2$ 和 $a = 1$，用辛普森法則（Simpson's rule）得 $F \approx 0.785398$，與 $\pi/4$ 吻合。

常見錯誤

- 第一次積分時漏掉常數，於是失去 $A \ln r$ 一項，令題目中「除非」（unless）的那一部分變得毫無意義。
- 將 $-kr$ 積分得 $-\frac{1}{2}kr^2$ 之後，忘記先除以 r 便再積分。
- 在 F 的積分中漏掉因子 r：通過半徑為 r 的薄圓環的流量與其圓周 $2\pi r$ 成正比，而不是與某個常數成正比。

STEP 之外

算子 $\frac{1}{r}\frac{\mathrm{d}}{\mathrm{d}r}\left(r\frac{\mathrm{d}}{\mathrm{d}r}\right)$ 正是拉普拉斯算子（Laplacian）∇^2 在柱坐標（cylindrical coordinates）下作用於只依賴 r 的函數時的形式，所以這條題目其實就是泊松方程（Poisson's equation）$\nabla^2 v = -k$；而 $\ln r$ 這個解，則是點電荷的 $1/r$ 勢在二維的對應。$F \propto a^4$ 這個結果稱為泊肅葉定律（Poiseuille's law）：管道半徑減半，流量便減至原來的十六分之一，這正是動脈收窄影響如此嚴重的原因。微分方程方面，可參閱 Boyce and DiPrima, *Elementary Differential Equations*（一本微分方程入門教科書）。

詳細解答

STEP I 1998，第 9 題

力學 • 擺的碰撞

> **STEP I 1998，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/98-s1-q9

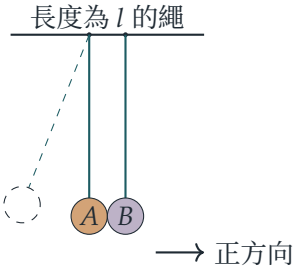
關鍵思路

由於位移非常小，每個球都像單擺（simple pendulum）一樣作簡諧運動（simple harmonic motion），而周期 $2\pi\sqrt{l/g}$ 與振幅和速率都無關。兩個單擺長度相同，所以在最低點發生任何一次碰撞之後，兩個球都會在同一時刻（即半個周期之後）回到最低點，而且速度方向相反。只要知道第二次碰撞在何處、何時發生，每次碰撞都只需寫兩行：動量守恆和牛頓實驗定律（Newton's experimental law）。

解答

模型

我們把 A 和 B 視為質量為 m 的質點，各自繫於長度同為 l 的不可伸長輕繩上，忽略空氣阻力，並假設碰撞在瞬間完成，衝量（impulse）沿水平的連心線方向。取第一次碰撞時 A 的運動方向（由 A 指向 B）為正方向，並設 $\omega = \sqrt{g/l}$。



為何每次碰撞都在最低點發生

在小幅振動下，不論振幅多大，每個球都作角頻率（angular frequency）為 ω 的簡諧運動。因此，一個以速率 v 離開最低點的球，其水平位移為 $\frac{v}{\omega}\sin\omega t$，並且在時間 π/ω（半個周期）之後回到最低點，此時速度為 $-v$——對任何 v 都是如此。

第一次碰撞在最低點發生：B 靜止地懸在該處，而 A 以速率 u 到達。假設某次碰撞在最低點發生，碰撞後 A 和 B 的速度分別為 $p \le q$（剛碰撞完，兩球不可能互相靠近）。從最低點量起，在其後時間 t，兩球之間的距離為

$$\frac{q}{\omega}\sin\omega t - \frac{p}{\omega}\sin\omega t = \frac{q-p}{\omega}\sin\omega t,$$

當 $0 < t < \pi/\omega$ 時此值為正，而在 $t = \pi/\omega$ 時為零。所以兩球不可能更早相遇，而且恰好在兩者都回到最低點時再次相遇。這正是題目的提示：兩個單擺周期相同，所以它們總是一同回到最低點。由歸納法（induction）可知，每次碰撞都在最低點發生，而且相鄰兩次碰撞相隔半個周期。

第一次碰撞

設第一次碰撞剛結束時 A 和 B 的速度分別為 v_A 和 v_B。由動量守恆（conservation of momentum）（在最低點時繩的張力沿鉛垂方向，所以不產生水平衝量）和牛頓實驗定律，得

$$mv_A + mv_B = mu,$$
$$v_B - v_A = eu.$$

兩式相加和相減，得

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

兩者都 ≥ 0，所以兩個球都擺向 B 的一方（若 $e = 1$，A 會完全停下，一直懸在最低點）。

想一想 第二次碰撞前一刻的速度是多少？

半個周期之後，每個球都回到最低點，速度方向相反：A 的速度為 $-\frac{1}{2}u(1-e)$，B 的速度為 $-\frac{1}{2}u(1+e)$。B 較快，所以從後面追上 A。它相對於 A 的接近速率為 $\frac{1}{2}u(1+e) - \frac{1}{2}u(1-e) = eu$。

第二次碰撞

設第二次碰撞剛結束時 A 和 B 的速度分別為 w_A 和 w_B。由動量守恆，得

$$mw_A + mw_B = -\tfrac{1}{2}mu(1-e) - \tfrac{1}{2}mu(1+e) = -mu,$$

而由牛頓實驗定律（分離速率等於接近速率 eu 的 e 倍），得

$$w_B - w_A = e \cdot eu = e^2 u.$$

兩式相減和相加，得

$$w_A = -\tfrac{1}{2}u(1+e^2), \qquad w_B = -\tfrac{1}{2}u(1-e^2).$$

所以緊接第二次碰撞之後，A 的速率為 $\frac{1}{2}u(1+e^2)$，正如題目所求，而 B 的速率為 $\frac{1}{2}u(1-e^2)$，兩者都朝着 A 最初被釋放的那一方運動。

驗算

- 完全彈性，$e = 1$：最後 A 的速率為 u，而 B 靜止。這是對的：第一次碰撞後 A 停下，B 取得速率 u；其後 B 以速率 u 返回，並把它全數交還給 A。
- $e = 0$：兩個速率都是 $\frac{1}{2}u$。兩球實際上合為一體，第一次碰撞後以速率 $\frac{1}{2}u$ 一起運動，所以這也是對的。
- 第二次碰撞後的動量為 $-\frac{1}{2}mu(1+e^2) - \frac{1}{2}mu(1-e^2) = -mu$，與原來的 mu 方向相反。
- 我們模擬了兩個真實（非線性）的單擺，取 $l = 1\,\text{m}$，初始角度為 0.01 rad。第二次碰撞在第一次碰撞之後 1.0001 個半周期發生，位置與最低點相差不到 3×10^{-6} rad；取 $e = 0.7$ 時，碰撞後的速率為 $0.745u$ 和 $0.255u$，與 $\frac{1}{2}(1+e^2)$ 和 $\frac{1}{2}(1-e^2)$ 吻合。

常見錯誤

- 忘記在第二次碰撞之前兩個速度都已反向，把情況當作 B 撞向靜止的 A。
- 在第二次碰撞時以 eu 代替 $e \cdot$(接近速率)。這裏接近速率剛好是 eu，所以分離速率是 e^2u。
- 說碰撞在最低點發生，是「因為球本來就懸在那裏」。真正的原因是周期與振幅無關，這也是為何球只可拉開很小的距離。

STEP 之外

等時性（isochronism，即周期與振幅無關）對單擺而言只是近似成立：擺幅較大時周期會增長，其準確值由橢圓積分（elliptic integral）給出，而碰撞位置亦會逐漸偏離最低點。Kleppner 與 Kolenkow 的 *An Introduction to Mechanics*（力學入門教科書）對單擺和碰撞都有嚴謹細緻的論述。

詳細解答

STEP I 1998，第 10 題

力學 • 拋體運動

STEP I 1998，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/98-s1-q10

關鍵思路

我們不知道拋射角，也不需要知道：答案只以 v、r 和 t 表示。因此要找一條能消去這個角的關係式。最簡潔的做法是用向量（vector）來思考：質點（particle）經過時間 t 後的落點，等於初速度乘以 t，再加上重力造成的下墜 $\frac{1}{2}gt^2$。這三個向量構成一個直角三角形，餘下的工夫交給畢氏定理（Pythagoras' theorem）就可以了。

解答

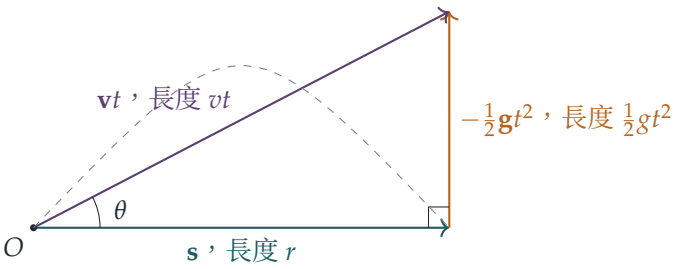
我們把每一粒泥土看成一個質點，只受重力作用而自由運動（不計空氣阻力），由位於地面的爆炸中心 O 出發。垂直向上射出的質點會落在 O，所以可取 $r > 0$，並設拋射角（angle of projection）θ 為與水平方向所成的仰角，其中 $0 < \theta < \frac{\pi}{2}$。

著地時間

在恆定加速度 $\mathbf{g}$（垂直向下）之下，經過時間 t 後的位移（displacement）為

$$\mathbf{s} = \mathbf{v}t + \tfrac{1}{2}\mathbf{g}t^2,$$

其中 $\mathbf{v}$ 是初速度，其大小為 v。在著地的一刻，$\mathbf{s}$ 是水平的，長度為 r。移項後得 $\mathbf{v}t = \mathbf{s} - \frac{1}{2}\mathbf{g}t^2$，即一個長度為 r 的水平向量與一個長度為 $\frac{1}{2}gt^2$ 的垂直（向上）向量之和。



這兩個向量互相垂直，所以根據畢氏定理，

$$v^2t^2 = r^2 + \left(\tfrac{1}{2}gt^2\right)^2 = r^2 + \tfrac{1}{4}g^2t^4.$$

兩邊乘以 g^2，並設 $u = \frac{1}{2}g^2t^2$（於是 $u^2 = \frac{1}{4}g^4t^4$ 及 $g^2v^2t^2 = 2v^2u$），上式變成

$$u^2 - 2v^2u + g^2r^2 = 0,$$

這是 u 的二次方程（quadratic）。它的根為

$$\tfrac{1}{2}g^2t^2 = u = v^2 \pm \sqrt{v^4 - g^2r^2},$$

這正是所要求的結果。

? **想一想** 如果不用向量，我們怎樣消去 θ？

用分量（components）。水平方向有 $r = vt\cos\theta$；垂直方向，落在地面即表示 $0 = vt\sin\theta - \frac{1}{2}gt^2$，所以 $vt\sin\theta = \frac{1}{2}gt^2$（因為 $t \neq 0$）。把兩式平方後相加，再利用 $\cos^2\theta + \sin^2\theta = 1$，便再次得到 $v^2t^2 = r^2 + \frac{1}{4}g^2t^4$。向量圖不過是把這個計算畫成一個三角形而已。

只有當 $v^4 \geq g^2r^2$，即 $r \leq v^2/g$ 時，兩個根才都是實數：這就是大家熟悉的最大射程（maximum range）。當 $r < v^2/g$ 時，兩個根都是正數（兩根之積為 $g^2r^2 > 0$，兩根之和為 $2v^2 > 0$），所以有兩個著地時間，一個屬於高拋，一個屬於平射。由於 $t = 2v\sin\theta/g$ 隨 θ 遞增，$+$ 號對應較陡的軌跡（trajectory），即 $\theta > \frac{\pi}{4}$，而 $-$ 號則對應較平的軌跡。

最大高度

整段飛行是對稱的：垂直速度 $v\sin\theta - gt'$ 在 $t' = \frac{1}{2}t$ 時為零，即飛行的中途。由最高點起，質點（在垂直方向上）由靜止下墜，歷時 $\frac{1}{2}t$，所以最大高度為

$$H = \tfrac{1}{2}g\left(\tfrac{1}{2}t\right)^2 = \tfrac{1}{8}gt^2 = \frac{1}{4g}\cdot\tfrac{1}{2}g^2t^2.$$

利用第一部分的結果，

$$H = \frac{v^2 \pm \sqrt{v^4 - g^2r^2}}{4g},$$

其中 $+$ 號對應較陡的軌跡，$-$ 號對應較平的軌跡（當 $r < v^2/g$ 時兩者皆有可能；當 $r = v^2/g$ 時兩者重合，此時 $H = v^2/(4g)$）。

驗算

- 垂直發射：$r = 0$ 給出 $\frac{1}{2}g^2t^2 = 2v^2$，所以 $t = 2v/g$，而 $H = v^2/(2g)$，與預期相符。另一個根 $t = 0$、$H = 0$，則是貼着地面掠過的一擊。
- 兩個高度之和為 $v^2/(2g)$。這是理所當然的：兩條軌跡的拋射角互為餘角，即 θ 和 $\frac{\pi}{2} - \theta$，而 $\frac{v^2\sin^2\theta}{2g} + \frac{v^2\cos^2\theta}{2g} = \frac{v^2}{2g}$。
- 數值上，取 $g = 9.8$、$v = 20$ 和 $\theta = 1$ 弧度，逐步模擬運動可得 $t \approx 3.4346$ 和 $H \approx 14.450$；取 $+$ 號的公式給出 3.4346 和 14.450。餘角的情況則與 $-$ 號吻合（$t \approx 2.2053$，$H \approx 5.958$）。

常見錯誤

- 只給出一個高度。速率相同的質點可以沿兩條不同的路徑到達同一點，而 $\pm$ 號會一直帶到 H 的式子裏。
- 把 H 寫成 θ 的式子便停下來。題目只要求以 v、r 和 g 表示，所以藉 $H = \frac{1}{8}gt^2$ 這條關係式連繫到第一部分，才是最省力的途徑。

STEP 之外

這個關於 t^2 的二次方程，是「軌跡的包絡線（envelope）」在時域上的版本：滿足 $v^4 = g^2r^2$ 的點（以及離開地面時的拋物面（paraboloid）$z = \frac{v^2}{2g} - \frac{gr^2}{2v^2}$）標示出泥土所能到達的區域的邊界，而在這個區域內，每一點都恰好有兩條軌跡經過。曲線族的包絡線，是大學一年級課程中偏微分（partial differentiation）的一個漂亮應用。Kleppner 與 Kolenkow 所著的 *An Introduction to Mechanics*（力學入門教科書）以恰到好處的程度講解拋體運動（projectile motion）和向量運動學。

詳細解答

STEP I 1998，第 11 題

力學 • 跨過滑輪的連接質點

STEP I 1998，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/98-s1-q11

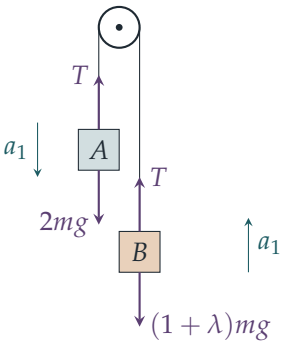
關鍵思路

每次往返分為兩個階段，而在每個階段中各質量保持不變，所以加速度是常數，可以使用 $s = \frac{1}{2}at^2$。下行時，盛滿水的桶比礦石連桶更重，兩者合起來就是一部阿特伍德機（Atwood machine）。回程上行時，桶 A 是空的而且沒有質量，所以沒有任何東西拉住 B：它只是自由下墜（falls freely）。至於第二部分，「在很長的時間內運上最多礦石」的意思是「每個循環運上的礦石量除以每個循環所需的時間要最大」。

解答

一個循環所需的時間

模型：兩個桶連同所盛之物視為質點（particle），鏈是輕而不可伸長的（light and inextensible），滑輪細小、光滑而且輕。因此鏈兩端的張力（tension）T 相同，而兩個桶以相同的速率運動。當 A 在頂部時，B 在底部（鏈的長度為 l），所以每個階段中每個桶都移動距離 l。



第 1 階段：A 下降。桶 A 盛有質量 $2m$ 的水；桶 B 的總質量為 $m+\lambda m=(1+\lambda)m$。設 a_1 為兩者共同的加速度（A 向下，B 向上）。對每個桶應用牛頓第二定律（Newton's second law），得

$$\begin{aligned} A\text{ (向下)}: &\quad 2mg-T=2ma_1,\\ B\text{ (向上)}: &\quad T-(1+\lambda)mg=(1+\lambda)ma_1. \end{aligned}$$

兩式相加消去 T：

$$(1-\lambda)mg=(3+\lambda)ma_1 \quad\Longrightarrow\quad a_1=\frac{(1-\lambda)g}{3+\lambda},$$

由於 $\lambda<1$，這個值是正數，所以 A 確實下降。由靜止出發，以恆加速度移動距離 l，有 $l=\frac{1}{2}a_1t_1^2$，所以

$$t_1=\sqrt{\frac{2l}{a_1}}=\sqrt{\frac{2l}{g}}\sqrt{\frac{3+\lambda}{1-\lambda}}.$$

想一想 在第 2 階段，桶 A 完全沒有質量。鏈的張力是多少？

第 2 階段：B 下降。現在 A 是空的，所以它的質量為零，而 B 的質量為 m。對沒有質量的桶 A，牛頓定律給出 $T-0\cdot g=0\cdot a$，所以 $T=0$：一條輕鏈若另一端繫着的東西沒有質量，就無法拉動任何物體。因此 B 由靜止以加速度 g 自由下墜，經過距離 l，由 $l=\frac{1}{2}gt_2^2$ 得

$$t_2=\sqrt{\frac{2l}{g}}.$$

（用阿特伍德機的公式，取質量為 0 和 m，也得到同樣的答案：$a=\frac{m-0}{m+0}g=g$。）這個階段結束時，A 第一次回到頂部。

總時間。由在頂部釋放 A 起，到它第一次回到頂部為止，所需的時間為

$$\boxed{t_1+t_2=\sqrt{\frac{2l}{g}}\left(1+\sqrt{\frac{3+\lambda}{1-\lambda}}\right)=\sqrt{\frac{2l}{g}}\,\frac{(1-\lambda)^{1/2}+(3+\lambda)^{1/2}}{(1-\lambda)^{1/2}}}$$

使運上的礦石量最大

以 $\tau(\lambda)=t_1+t_2$ 表示循環時間。每個循環運上質量為 λm 的礦石，而整個過程不斷以完全相同的方式重複。在一段很長的時間 S 內，共有 $\lfloor S/\tau\rfloor$ 個完整循環，所以運上的礦石量介乎 $\lambda m\,(S/\tau-1)$ 與 $\lambda m\,S/\tau$ 之間。當 S 遠大於 τ 時，那個未完成的循環所損失的部分可以忽略，運上的礦石量基本上就是

$$\frac{\lambda mS}{\tau(\lambda)}=mS\sqrt{\frac{g}{2l}}\,\frac{\lambda(1-\lambda)^{1/2}}{(1-\lambda)^{1/2}+(3+\lambda)^{1/2}}=mS\sqrt{\frac{g}{2l}}\,\mathrm{f}(\lambda).$$

因子 $mS\sqrt{g/2l}$ 與 λ 無關，所以我們必須選取 $\lambda\in(0,1)$，使 $\mathrm{f}(\lambda)$ 盡可能大。

想一想 為甚麼最佳的 λ 必定是駐點（stationary point），而不是端點？

函數 f 在 $(0,1)$ 上可導（differentiable），因為在該區間內 $1-\lambda>0$ 和 $3+\lambda>0$，而且分母是正數。它在 $(0,1)$ 上取正值，並且在兩端都趨向 0：當 $\lambda\to 0$ 時，因子 λ 趨向零（沒有運載礦石）；當 $\lambda\to 1$ 時，因子 $(1-\lambda)^{1/2}$ 趨向零（兩個桶幾乎平衡，所以第 1 階段需時極長）。因此 f 的最大值在區間內部的某一點取得，而可導函數若在內點取得極大值，該處的導數必為零。所以最佳的 λ 滿足 $f'(\lambda)=0$，這正是所要證明的。

驗算

- 極限情況：當 $\lambda\to 1$ 時，第 1 階段中兩個桶的質量分別為 $2m$ 和 $2m$，所以 $a_1\to 0$ 而 $t_1\to\infty$，與公式吻合。
- 當 $\lambda\to 0$ 時，得 $a_1=g/3$，正是質量為 $2m$ 和 m 的阿特伍德機的慣常數值。
- 數值上，取 $g=9.8$，對隨機選取的 λ 和 l 逐步模擬運動，結果與公式一致：例如 $\lambda=0.504$、$l=26.9$ 給出的循環時間為 8.568 s。以數值方法求 f 的最大值，得 $\lambda\approx 0.704$；而在一段很長的時間內點算完整循環的數目，也得到同一個最佳的 λ。

常見錯誤

- 忘記了桶 B 本身的質量 m，在第 1 階段用了 λm 而不是 $(1+\lambda)m$。
- 把第 2 階段當作帶有某個張力的阿特伍德機來處理：由於 A 沒有質量，B 是自由下墜的。
- 求每個循環運上的礦石量 λm 的最大值，而不是每單位時間運上的礦石量的最大值。載得愈重，循環愈慢。

STEP 之外

把一個速率（每個循環的產出除以每個循環的時間）最優化，正是概率論中*更新報酬定理*（renewal–reward theorem）背後的想法：在那裏循環的長度是隨機的，而長遠的報酬率仍然等於每個循環的平均報酬除以平均循環長度。這裏的力學是恆加速度運動，Kleppner 與 Kolenkow 所著的 *An Introduction to Mechanics*（力學入門）由此出發，發展成對質點系統的完整論述。

詳細解答

STEP I 1998，第 12 題

概率與統計 • 計數與獨立性

STEP I 1998，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/98-s1-q12

關鍵思路

這一題從頭到尾都是計數（counting）。只要知道解的總數，以及其中有多少個解的 X 等於某個指定的值，每個概率（probability）就是兩個數目之比。至於獨立性（independence），最快捷的檢驗方法是考慮一個極端情況：找出一對數值，使聯合概率（joint probability）根本不可能等於邊際概率（marginal probability）之積。最後一部分中，「XYZ 可被 5 整除」的意思是「X, Y, Z 之中最少有一個可被 5 整除」，這就要用上容斥原理（inclusion–exclusion）。

解答

數算解的數目

若固定 $X = x$（其中 $0 \le x \le 20$），則 Y 可取 $0, 1, \ldots, 20 - x$ 中的任何一個值，而 $Z = 20 - x - Y$ 隨之確定，並且是非負數。因此恰好有 $21 - x$ 個解滿足 $X = x$，而解的總數為

$$N = \sum_{x=0}^{20} (21 - x) = 21 + 20 + \cdots + 1 = \frac{21 \cdot 22}{2} = 231.$$

（這正是「隔板法」（stars and bars）的數目 $\binom{22}{2}$：把 20 顆星和 2 塊隔板排成一行。）每個解的概率均為 $1/231$，而

$$\mathrm{P}(X = x) = \frac{21 - x}{231}, \qquad x = 0, 1, \ldots, 20.$$

根據對稱性（symmetry），Y（以及 Z）有相同的分佈（distribution）。

X 與 Y 是否獨立？

不是。取 $X = 20$ 和 $Y = 20$。這兩個事件各自都有可能發生：

$$\mathrm{P}(X = 20) = \mathrm{P}(Y = 20) = \frac{1}{231} > 0,$$

但兩者不可能同時發生，因為 $X + Y \leq 20$。所以

$$\mathrm{P}(X = 20,\ Y = 20) = 0 \neq \frac{1}{231^2} = \mathrm{P}(X = 20)\,\mathrm{P}(Y = 20),$$

故 X 與 Y 並不獨立。直觀地說，X 的值越大，留給 Y 的空間就越少。

> ? **想一想** 只說「$X + Y \leq 20$，所以它們顯然是相關的」不就夠了嗎？

想到這樣的限制條件，直覺是對的；但獨立性是一個關於概率的精確陳述，所以論證時應該舉出一對數值，使得 $\mathrm{P}(X = x, Y = y) \neq \mathrm{P}(X = x)\mathrm{P}(Y = y)$。只需一個反例（counterexample）便已足夠，而 $(20, 20)$ 這對極端數值令計算變得輕而易舉。

X 可被 5 整除

範圍內 5 的倍數為 $X = 0, 5, 10, 15, 20$，對應的解分別有 $21, 16, 11, 6, 1$ 個。所以

$$\mathrm{P}(5 \mid X) = \frac{21 + 16 + 11 + 6 + 1}{231} = \frac{55}{231} = \frac{5}{21},$$

正如題目所求。

XYZ 可被 5 整除

由於 5 是質數（prime），$5 \mid XYZ$ 當且僅當 5 整除 X、Y、Z 之中最少一個。（謹記 0 也算是可被 5 整除。）設 A、B、C 分別為 X、Y、Z 可被 5 整除的解所組成的集合。由上一部分可知 $|A| = |B| = |C| = 55$。

考慮 $A \cap B$：設 $X = 5a$、$Y = 5b$，其中 $a, b \geq 0$ 且 $a + b \leq 4$。這樣的數對共有 $5 + 4 + 3 + 2 + 1 = 15$ 個。在每一個情況下，$Z = 20 - 5a - 5b$ 同樣是 5 的倍數，所以

$$|A \cap B| = |B \cap C| = |C \cap A| = |A \cap B \cap C| = 15.$$

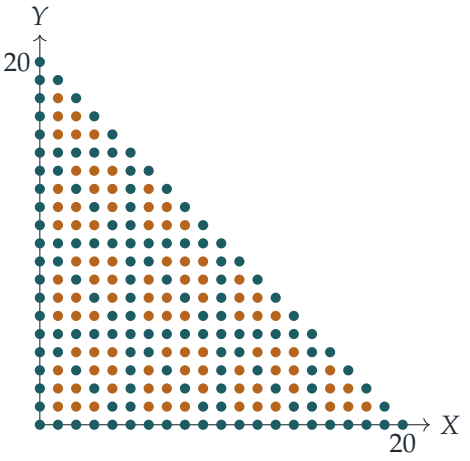
根據容斥原理，

$$\begin{aligned} |A \cup B \cup C| &= 3 \cdot 55 - 3 \cdot 15 + 15 \\ &= 135, \end{aligned}$$

因此

$$\mathrm{P}(5 \mid XYZ) = \frac{135}{231} = \frac{45}{77}.$$

下圖把各個解表示為滿足 $X+Y \le 20$ 的格點（lattice point）(X,Y)。藍綠色的點位於直線 $X=5k$、$Y=5k$ 或 $X+Y=20-5k$（即 $Z=5k$）其中之一上，而琥珀色的點則是 X,Y,Z 全都不是 5 的倍數的 96 個解。



驗算

- 補集（complement）應有 $231-135=96$ 個解，其中 X,Y,Z 全都不是 5 的倍數：這就是圖中琥珀色的點。用電腦列出全部 231 個解，並以精確分數計算，可確認 96、$|A\cap B|=15$、5/21 以及 $135/231=45/77\approx 0.584$ 均正確。
- 對 5/21 作一個合理性檢查：X 可被 5 整除的概率比 1/5 稍大，因為 X 取較小的值的可能性最高，而 $X=0$ 更是最有可能出現的值。

常見錯誤

- 忘記 0 可被 5 整除。漏掉 $X=0$ 會得出 34/231，而不是 5/21。
- 直接計算 $3\times\frac{5}{21}$，這樣會把重疊的部分重複計算多次。本題的重疊部分相當大：只要其中兩個變量是 5 的倍數，第三個也必然是。
- 論證兩者相關時，沒有舉出一對令乘法法則不成立的具體數值。

STEP 之外

從所有解中均勻隨機地選取一個，等同於隨機選取 20 分成三部分的一個有序分拆（composition）。若把 20 換成總和 n，並讓 n 增大，$(X/n,Y/n)$ 便會趨於在一個三角形上均勻分佈，這是狄利克雷分佈（Dirichlet distribution）最簡單的情況。容斥原理是處理「最

少一個」這類事件的標準工具，它還會在錯排（derangement）和歐拉函數（Euler's totient function）中再次出現。Grimmett 與 Welsh 合著的 *Probability: An Introduction*（概率論入門教材）對計數和獨立性這兩方面的概念均有論述。

詳細解答

STEP I 1998，第 13 題

概率與統計 • 放回抽樣

STEP I 1998，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/98-s1-q13

關鍵思路

這是一道條件概率（conditional probability）的題目：作為條件的事件是「在 n 次抽取中剛好吃掉一粒軟糖」，而我們按這粒軟糖是在何時吃掉的，把這個事件拆開。每一部分都是幾個簡單概率的乘積，因為袋中的內容只會在吃掉紅色軟糖的那一刻改變。把各部分加起來，就是一個等比級數（geometric series），兩個 n 次冪之差正是由此而來。

解答

設 $N = r + b$ 為開始時軟糖的總數。我們假設 $r \geq 1$ 及 $b \geq 1$（若 $b = 0$ 而 $n \geq 2$，則不可能剛好只吃掉一粒軟糖，也就沒有可作條件的事件）。

每次抽取時的袋子。 非紅色的軟糖會放回袋中，所以袋子只會在吃掉紅色軟糖時改變。吃掉紅色軟糖之前，袋中有 N 粒軟糖，其中 b 粒非紅色；之後袋中有 $N-1$ 粒軟糖，其中仍有 b 粒非紅色。由於每次抽取前都把袋子徹底搖勻，每次抽取都是從當時袋中的軟糖裏重新作一次均勻的隨機選取，所以我們可以把各次抽取的概率逐一相乘。

按吃掉的時間拆分。 對 $k = 1, 2, \ldots, n$，設 E_k 為以下事件：在 n 次抽取中唯一吃掉的那粒軟糖，是在第 k 次抽取時吃掉的。那麼第 $1, \ldots, k-1$ 次是從原來的袋中抽到非紅色，第 k 次是從原來的袋中抽到紅色，而第 $k+1, \ldots, n$ 次是從少了一粒的袋中抽到非紅色：

$$\mathrm{P}(E_k) = \left(\frac{b}{N}\right)^{k-1} \frac{r}{N} \left(\frac{b}{N-1}\right)^{n-k} = \frac{r\, b^{n-1}}{N^k\, (N-1)^{n-k}}.$$

事件 $E_1, \ldots, E_n$ 互斥（disjoint），而它們的併集（union）就是「剛好吃掉一粒軟糖」這個事件 E。所以

$$\mathrm{P}(E) = r\,b^{n-1} \sum_{k=1}^{n} \frac{1}{N^k (N-1)^{n-k}} = \frac{r\,b^{n-1}}{N^n (N-1)^n} \sum_{k=1}^{n} N^{n-k} (N-1)^k,$$

這裏我們把每一項的分子和分母同乘以 $N^n (N-1)^n$。

想一想 怎樣求 $\sum_{k=1}^{n} N^{n-k}(N-1)^k$ 的和？

這是一個等比級數，首項為 $N^{n-1}(N-1)$，公比（common ratio）為 $\frac{N-1}{N}$。更快捷的做法，是利用大家熟悉的因式分解（factorisation）

$$x^n - y^n = (x-y)\left(x^{n-1} + x^{n-2}y + \cdots + y^{n-1}\right)$$

取 $x = N$、$y = N-1$，於是 $x - y = 1$：

$$\sum_{k=1}^{n} N^{n-k}(N-1)^k = (N-1) \sum_{j=0}^{n-1} N^{n-1-j}(N-1)^j = (N-1)\left(N^n - (N-1)^n\right).$$

因此

$$\mathrm{P}(E) = \frac{r\,b^{n-1}\left(N^n - (N-1)^n\right)}{N^n (N-1)^{n-1}}.$$

條件概率。 「我在最後一次抽取時吃掉它，而且只吃了一粒」這個事件就是 E_n，其概率為 $\mathrm{P}(E_n) = \dfrac{r\,b^{n-1}}{N^n}$。由於 $E_n \subseteq E$，

$$\begin{aligned}\mathrm{P}(E_n \mid E) = \frac{\mathrm{P}(E_n)}{\mathrm{P}(E)} &= \frac{r\,b^{n-1}}{N^n} \cdot \frac{N^n (N-1)^{n-1}}{r\,b^{n-1}\left(N^n - (N-1)^n\right)} \\ &= \frac{(N-1)^{n-1}}{N^n - (N-1)^n} = \frac{(r+b-1)^{n-1}}{(r+b)^n - (r+b-1)^n},\end{aligned}$$

這正是所求的結果。

留意 r 和 b 完全消去，只剩下它們的總和。其實上面的計算還說明了更多：已知 E 發生，軟糖在第 k 次抽取時被吃掉的概率與 $\left(\frac{N-1}{N}\right)^k$ 成正比，所以較早的抽取比較後的抽取更有可能。這是合理的：早些吃掉紅色軟糖，袋中非紅色軟糖所佔的比例便較大，餘下各次「再沒有抽到紅色」的機會也就較高。

驗算

► $n = 1$：公式給出 $\dfrac{1}{N - (N-1)} = 1$，正確，因為唯一的一次抽取就是最後一次。

- $n = 2$：公式給出 $\dfrac{N-1}{2N-1}$，略小於 $\frac{1}{2}$，與「較早的抽取可能性稍高」一致。
- 取 $r = 3$、$b = 5$、$n = 4$，公式給出 $343/1695 \approx 0.2024$，而若四次抽取的可能性均等，則應為 $\frac{1}{4}$。模擬 4×10^5 次的結果為 0.2028；另外，逐一列舉所有抽取序列並追蹤袋中內容，所得的精確結果在 $1 \le r, b \le 4$ 及 $1 \le n \le 8$ 的所有情況下都與公式完全吻合。

常見錯誤

- 在紅色軟糖被吃掉之後，仍然用 b/N 作為抽到非紅色的概率。這時袋中只有 $N-1$ 粒軟糖，而這正是本題的關鍵所在：如果由始至終都用 b/N，每個 E_k 的概率便會相同，答案就變成 $1/n$。
- 忘記取條件：單是 $\mathrm{P}(E_n) = rb^{n-1}/N^n$ 並不是答案，因為題目已告訴我們剛好吃掉了一粒軟糖。

STEP 之外

這裏的袋子是一個簡單的馬可夫鏈（Markov chain）：它的狀態是餘下紅色軟糖的數目，而狀態向下移動的機會只取決於當前的狀態。以這類鏈的最終狀態為條件，反過來問它所經過的路徑，正是「隱馬可夫模型」（hidden Markov models）以及對歷史過程作貝葉斯推斷（Bayesian inference）背後的想法。Grimmett 與 Welsh 的 *Probability: An Introduction*（概率入門）正好以這個程度講解條件概率和馬可夫鏈，而 Blitzstein 與 Hwang 的 *Introduction to Probability*（概率導論）則有許多同類風格的題目。

詳細解答

STEP I 1998，第 14 題

概率與統計 • 幾何概率

STEP I 1998，第 14 題

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

step.maths.org/questions/98-s1-q14

關鍵思路

硬幣落在哪裏，完全取決於它的圓心落在哪裏。由於棋盤的圖案規則地重複，我們可以假設圓心在單獨一個 2×2 正方形內均勻分佈（uniformly distributed），於是每一種結果都對應該正方形內的一個區域。把「壓過一條線」和「蓋住一個交點」轉化為圓心與各條線及各個角的距離所滿足的條件，概率（probability）便變成面積，而這個面積是 a 的二次式，我們可以求其最大值。

解答

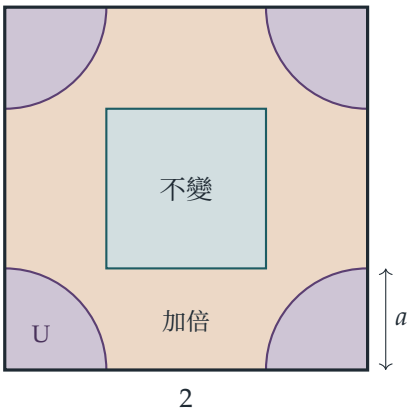
稅款加倍的概率

設硬幣的圓心均勻隨機地落在方格網的其中一個正方形內；棋盤很大，所以邊緣效應可以忽略，而且每個正方形看起來都一樣。在這個正方形上設立坐標，使它成為 $[0,2] \times [0,2]$，方格線沿着它的四邊。

設 U 為圓心到最近的垂直線的距離，V 為圓心到最近的水平線的距離。由於 x 坐標在 $[0,2]$ 上均勻分佈，$U = \min(x, 2-x)$ 在 $[0,1]$ 上均勻分佈；同理 V 在 $[0,1]$ 上均勻分佈，而且因為 x 與 y 獨立（independent），U 與 V 亦獨立。所以 (U,V) 在單位正方形 $[0,1]^2$ 上均勻分佈，而概率就是該正方形內的面積。

- 硬幣不壓過任何線，當且僅當 $U > a$ 且 $V > a$。其概率為 $(1-a)^2$。
- 硬幣蓋住一個交點，當且僅當正方形最近的一個角與圓心的距離小於 a，即 $U^2 + V^2 < a^2$。由於 $a < 1$，這個四分之一圓盤完全位於 $[0,1]^2$ 之內，所以概率為 $\frac{1}{4}\pi a^2$。

這兩個事件互斥（disjoint）：如果硬幣蓋住一個交點，它必定壓過通過該交點的線。（硬幣剛好碰到一條線之類的邊界情況，概率為零，所以用 $<$ 還是 $\leq$ 都沒有分別。）



圖中顯示圓心在一個正方形內的可能位置：藍綠色（沒有壓過任何線）和紫紅色（蓋住一個交點，在其中一個角標示為 U）的區域令稅款不變；琥珀色的區域令稅款加倍。

所以稅款加倍的概率為

$$\begin{aligned} p(a) &= 1-(1-a)^2-\tfrac{1}{4}\pi a^2 \\ &= 2a-\left(1+\frac{\pi}{4}\right)a^2. \end{aligned}$$

使稅收最大

要徵得最多的稅，就要令加倍的機會盡可能大，所以我們在 $0<a<1$ 的範圍內求 $p(a)$ 的最大值。它是一個首項係數為負的二次函數，而

$$p'(a)=2-2\left(1+\frac{\pi}{4}\right)a=0 \iff a=\left(1+\frac{\pi}{4}\right)^{-1}.$$

這個值是 $\frac{4}{4+\pi}\approx 0.560$，確實位於 $(0,1)$ 之內，而且由於 $p''(a)<0$，它給出最大值，正如題目所求。以 a^* 表示這個值，則有 $(1+\frac{\pi}{4})a^{*2}=a^*$，所以最大值為

$$p_{\max}=2a^*-a^*=a^*=\frac{4}{4+\pi}.$$

n 年後稅款的期望值

想一想 為甚麼每年的期望增長倍數是 $1+p$ 而不是 $2p$？

在稅款沒有加倍的年份，稅款維持不變，所以它是乘以 1，而不是乘以 0。設 M_k 為第 k

年的倍數：$M_k = 2$ 的概率為 p，$M_k = 1$ 的概率為 $1 - p$。於是

$$\mathrm{E}(M_k) = 2p + (1 - p) = 1 + p = 1 + \frac{4}{4 + \pi} = \frac{8 + \pi}{4 + \pi}.$$

假設每年的投擲互相獨立。n 年後的稅款為每頭 yak $T_n = M_1 M_2 \cdots M_n$ Slihings（由 1 開始），而獨立隨機變量（random variables）之積的期望值（expectation）等於它們各自期望值之積，所以

$$\mathrm{E}(T_n) = \mathrm{E}(M_1)\,\mathrm{E}(M_2) \cdots \mathrm{E}(M_n) = \left(\frac{8 + \pi}{4 + \pi}\right)^n,$$

正如題目所求。

驗算

- 當 $a \to 0$ 時，硬幣幾乎從不碰到任何線；當 $a = 1$ 時，它必定壓過一條線，所以 $p(1) = 1 - \frac{\pi}{4} \approx 0.215$，即避開角上圓盤的機會。兩者都與公式吻合，而最大值位於兩者之間。
- 以 400 000 個隨機圓心進行模擬，在 $a = 0.56$ 時得出 $p \approx 0.5620$，而公式給出 0.5601。對 a 作精細搜尋，得出最大值位於 $a \approx 0.5601 = 4/(4 + \pi)$。
- 以 $p = 4/(4 + \pi)$ 模擬每年的加倍，得出 3 年後稅款的平均值約為 3.805，而 $\left(\frac{8+\pi}{4+\pi}\right)^3 \approx 3.797$。

常見錯誤

- 從「不壓線」的面積中減去「蓋住交點」的面積，或者把它計算了兩次。這兩個事件互斥，兩者都要從 1 中減去。
- 在每個角使用整個圓盤的面積 πa^2，忘記了只有四分之一位於正方形之內（等價地說，每個角上的圓盤由四個正方形共用）。
- 把稅款的期望值寫成 $(2p)^n$，忘記了沒有加倍的年份稅款維持不變。

STEP 之外

這是蒲豐投針問題（Buffon's needle problem）的近親：把隨機放置的物件化為基本區域（fundamental region）內的一個隨機點，概率便成為面積。這個想法是幾何概率（geometric probability）和積分幾何（integral geometry）的起點，由此可得出一些出人意表的結果，例如以投針來估算 π。最後一部分用到獨立隨機變量的期望值可以相乘這個

事實，這正是金融學和種羣生物學中乘法增長模型的基礎。Grimmett 與 Welsh 合著的 *Probability: An Introduction*（概率入門）涵蓋了這兩個想法。

詳細解答

STEP II 1998，第 1 題

純數學 • 丟番圖方程與奇偶性

STEP II 1998，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/98-s2-q1

關鍵思路

把立方展開後，大部分的項會互相抵消，剩下一個關於 n 的三次式，其常數項很小。比較等式兩邊的奇偶性（parity），可知 n 必為偶數。再把這個三次方程寫成 n^2(某個式子) $=54$ 的形式，問題便化為一個關於因數（divisor）的問題。至於 $(**)$，利用偶數代換 $n=2m$，便直接回到 $(*)$。

解答

方程 $(*)$

利用 $(a+b)^3-(a-b)^3=6a^2b+2b^3$，並取 $a=n$，$b=3$，可得

$$(n+3)^3-(n-3)^3=18n^2+54,$$

所以 $(*)$ 等價於

$$n^3=18n^2+54, \qquad \text{即} \qquad n^2(n-18)=54. \tag{1}$$

n 是偶數。對於任何整數 n，右邊 $18n^2+54=2(9n^2+27)$ 都是偶數。假如 n 是奇數，則 n^3 也是奇數，與 $n^3=18n^2+54$ 矛盾。因此 n 是偶數。

n^2 整除 54。在 (1) 的第二種寫法中，$n-18$ 是整數，所以 54 等於 n^2 乘以一個整數。因此 n^2 是 54 的因數。

? **想一想** 這兩項事實合起來，怎樣把所有整數都排除掉？

由於 n 是偶數，可設 $n=2k$。於是 $n^2=4k^2$，它是 4 的倍數（而且 $n\neq 0$，因為把 $n=0$ 代入 (1) 會得出 $0=54$）。但 $54=2\times 27$ 不能被 4 整除，所以除了 0 之外，沒有任何 4

的倍數能整除它。這個矛盾說明沒有整數 n 滿足 $(*)$。

（另一個看法：能整除 $54 = 2 \cdot 3^3$ 的平方數只有 1 和 9，所以 $n \in \{\pm 1, \pm 3\}$，而這些全是奇數。）

方程 $(**)$

現在取 $b = 6$：

$$(n+6)^3 - (n-6)^3 = 36n^2 + 432,$$

所以 $(**)$ 等價於

$$n^3 = 36n^2 + 432.$$

對於任何整數 n，右邊都是偶數，所以與之前完全一樣，n^3 是偶數。由於奇數的立方是奇數，n 必為偶數。

> **想一想** 為甚麼不能直接照搬第一部分的因數論證？

我們固然可以寫成 $n^2(n-36) = 432$，但 $432 = 2^4 \cdot 3^3$ 有不少偶數的平方因數（4、16、36、144），所以單靠整除性的論證並不能完成證明。我們改為利用偶數這個性質。設 $n = 2m$，其中 m 為整數。於是

$$8m^3 = 36 \cdot 4m^2 + 432 \quad \Longleftrightarrow \quad m^3 = 18m^2 + 54,$$

這正正是整數 m 所滿足的方程 (1)。換句話說，m 滿足 $(*)$。事實上，把 $n = 2m$ 代入 $(**)$，所得的不過是把 $n = m$ 代入 $(*)$ 後再全式乘以 8：

$$(2m-6)^3 + (2m)^3 - (2m+6)^3 = 8\big[(m-3)^3 + m^3 - (m+3)^3\big].$$

我們已證明沒有整數滿足 $(*)$，所以這樣的 m 並不存在。因此沒有整數 n 滿足 $(**)$。

驗算

- 把 $n = 1$ 代入展開式試試：$(4)^3 - (-2)^3 = 64 + 8 = 72 = 18 + 54$。✓
- 三次式 $n^3 - 18n^2 - 54$ 恰好有一個實根，約為 $n \approx 18.164$，並不是整數。（用電腦直接搜尋 $|n| < 10^5$ 的範圍，兩個方程都找不到解。）
- 對於 $(**)$，實數解為 $n \approx 2 \times 18.164 = 36.33$，與 $n = 2m$ 這個縮放關係吻合。

常見錯誤

- 忘記 $n = 0$ 需要另外交代一句：0 是偶數，但 0 不是 54 的因數。在這裏，(1) 已即時把它排除。

- 把立方完全展開，結果在運算中出錯。恆等式 $(a+b)^3-(a-b)^3=6a^2b+2b^3$ 可以省掉一半工夫。
- 在 $(**)$ 中，嘗試證明 $n^2 \mid 432$ 是行不通的。「n 是偶數」的真正用處在於代換 $n=2m$。

STEP 之外

最後一步由 $(**)$ 的一個解，得出一條形式相同但數字較小的方程的解，這是費馬（Fermat）遞降法（method of descent）的一個簡單形式。反覆運用這個方法，可以證明 $x^3+y^3=z^3$ 沒有正整數解，這就是費馬最後定理（Fermat's Last Theorem）的三次情形，最先由歐拉（Euler）證明。本題的三項構成等差數列（arithmetic progression），令問題容易得多。事實上，$(n-3k)^3+n^3=(n+3k)^3$ 表示 $t=n/k$ 是 t^3-18t^2-54 的根，而根據有理根定理（rational root theorem），這個三次式沒有有理根，所以對任何 $k\neq 0$ 都沒有整數解。Davenport 的 *The Higher Arithmetic* 是一本介紹整除性、奇偶性論證和丟番圖方程（Diophantine equations）的出色入門書。

詳細解答

STEP II 1998，第 2 題

純數學 • 二項式級數近似

STEP II 1998，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/98-s2-q2

關鍵思路

二項式級數（binomial series）在變量很小時收斂（converge）得很快，所以我們希望把 $\sqrt{2}$（或 $\sqrt[3]{2}$）寫成一個簡單的有理數乘以 $(1+x)^a$，而其中的 x 極小。就平方根而言，$49/50$ 與 1 只差一點，而且含有因數 2；就立方根而言，我們要找一個幾乎等於 $2\times$ 另一個立方數的立方數，而 $128 = 2\times 4^3$ 正好緊貼着 $125 = 5^3$。

解答

2 的平方根

首先，我們把題目給出的式子與 $\sqrt{2}$ 聯繫起來。由於 $1-\frac{1}{50} = \frac{49}{50}$，

$$\left(1-\frac{1}{50}\right)^{1/2} = \frac{7}{\sqrt{50}} = \frac{7}{5\sqrt{2}} = \frac{7\sqrt{2}}{10}, \qquad \text{所以} \qquad \sqrt{2} = \frac{10}{7}\left(1-\frac{1}{50}\right)^{1/2}.$$

$(1+x)^{1/2}$ 的二項式級數的首幾項為

$$(1+x)^{1/2} = 1 + \tfrac{1}{2}x - \tfrac{1}{8}x^2 + \tfrac{1}{16}x^3 - \cdots,$$

對 $|x|<1$ 成立。設 $x = -\frac{2}{100} = -2y$，其中 $y = \frac{1}{100}$。2 的冪剛好可以約去：

$$\begin{aligned}\left(1-\frac{2}{100}\right)^{1/2} &\approx 1 - \tfrac{1}{2}(2y) - \tfrac{1}{8}(2y)^2 - \tfrac{1}{16}(2y)^3 \\ &= 1 - y - \tfrac{1}{2}y^2 - \tfrac{1}{2}y^3 \\ &= 1 - 0.01 - 0.00005 - 0.0000005 \\ &= 0.9899495.\end{aligned}$$

這就是把 $1/50$ 寫成 $2/100$ 的好處：每一項都變成 10 的冪乘以 1 或 $\frac{1}{2}$，所以運算全是精確的小數。

因此

$$\sqrt{2} \approx \frac{10 \times 0.9899495}{7} = \frac{9.899495}{7} = 1.41421357\ldots,$$

準確至小數點後六位，$\sqrt{2} \approx 1.414214$，正如題目所求。

2 的立方根

想一想 應取哪個 N 和 a？怎樣才能令 $(1+N/125)^a$ 與 $\sqrt[3]{2}$ 扯上關係？

由於 $125 = 5^3$，$(125+N)/125$ 的立方根就是 $\sqrt[3]{125+N}/5$。因此我們希望 $125+N$ 是某個完全立方數（perfect cube）的 2 倍，而且 N 要小。從 2×1^3, 2×2^3, 2×3^3, $2\times 4^3 = 2, 16, 54, 128$ 可見，128 與 125 只相差 3。所以取 $N=3$ 和 $a=\frac{1}{3}$：

$$\left(1+\frac{3}{125}\right)^{1/3} = \frac{\sqrt[3]{128}}{5} = \frac{4\sqrt[3]{2}}{5}, \qquad \text{所以} \qquad \sqrt[3]{2} = \frac{5}{4}\left(1+\frac{3}{125}\right)^{1/3}.$$

$(1+x)^{1/3}$ 的二項式級數的首幾項為

$$(1+x)^{1/3} = 1+\tfrac{1}{3}x+\frac{\frac{1}{3}\cdot\left(-\frac{2}{3}\right)}{2!}x^2+\frac{\frac{1}{3}\cdot\left(-\frac{2}{3}\right)\cdot\left(-\frac{5}{3}\right)}{3!}x^3+\cdots = 1+\tfrac{1}{3}x-\tfrac{1}{9}x^2+\tfrac{5}{81}x^3-\cdots.$$

和之前一樣，作一個代換（substitution）便可令運算變得簡潔。設 $x=3y$，其中 $y=\frac{1}{125}=0.008$；於是 $\frac{1}{3}x=y$、$\frac{1}{9}x^2=y^2$ 及 $\frac{5}{81}x^3=\frac{5}{3}y^3$。所以

$$\begin{aligned}\left(1+\frac{3}{125}\right)^{1/3} &\approx 1+y-y^2+\tfrac{5}{3}y^3\\ &= 1+0.008-0.000064+\tfrac{5}{3}(0.000000512)\\ &= 1.007936+0.00000085333\ldots\\ &= 1.00793685333\ldots.\end{aligned}$$

乘以 $\frac{5}{4}$，得

$$\sqrt[3]{2} \approx 1.25\times 1.00793685333\ldots = 1.25992106\ldots,$$

所以準確至小數點後六位，$\sqrt[3]{2}\approx 1.259921$。

驗算

計算機給出 $\sqrt{2}=1.41421356\ldots$ 和 $\sqrt[3]{2}=1.25992105\ldots$，所以兩個近似值都準確至小數點後六位；我們求得的值是 1.41421357 和 1.25992107。我們已用電腦以精確分數核對過這兩個級數。從第一個被略去的項可以大致看出誤差的大小：$-\frac{5}{128}x^4$ 對 $\sqrt{2}$ 的估值貢獻約 -9×10^{-9}，而 $-\frac{10}{243}x^4$ 對 $\sqrt[3]{2}$ 的估值貢獻約 -1.7×10^{-8}。兩者都遠小於足以改變小數點後第六位的 5×10^{-7}。

常見錯誤

- 忘記還原縮放的倍數：級數給出的是 $\frac{7\sqrt{2}}{10}$ 和 $\frac{4\sqrt[3]{2}}{5}$，而不是 $\sqrt{2}$ 和 $\sqrt[3]{2}$ 本身。
- 級數中的正負號出錯。在第一部分中 x 是負數，所以 1 之後的四項全是負的。在第二部分中，正負號則交替出現。
- 太早捨入。在最後一步除法或乘法之前，應至少保留八位小數。

STEP 之外

這道題的背後是帶餘項的泰勒定理（Taylor's theorem），它把「下一項極小」這個說法變成對誤差的嚴格界限。它解釋了為何一個公比約為 0.02 的級數，每多取一項便大約多得一位半至兩位小數的準確度。同一個技巧，即找一個有理數 p/q 使 p^n/q^n 接近目標數的某個簡單倍數，與數論中的連分數（continued fractions）和有理逼近（rational approximation）息息相關。Davenport 的 *The Higher Arithmetic*（數論入門名著）是閱讀這些課題的好去處，而 Abbott 的 *Understanding Analysis*（數學分析入門教材）則仔細講解了泰勒餘項。

詳細解答

STEP II 1998，第 3 題

純數學 • 裂項相消級數

STEP II 1998，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/98-s2-q3

關鍵思路

通項的分母若是幾個連續因子的乘積，自然應該把它拆成部分分式（partial fractions）。拆開後的各項會在求和時互相抵消，即所謂裂項相消（telescoping），最後只剩下首尾兩端的幾項。在第二部分，題目給出比值 a_n/a_{n-1}，這表示 a_n/a_1 是一連串比值的乘積，而這個乘積同樣可以逐項相消。運氣好的話，a_n 會變回第一部分的級數（series），而事實正是如此。

解答

和 S_N

設

$$\frac{2n-1}{n(n+1)(n+2)} = \frac{A}{n} + \frac{B}{n+1} + \frac{C}{n+2}.$$

利用掩蓋法（cover-up rule）：取 $n=0$，得 $A = \frac{-1}{1\cdot 2} = -\frac{1}{2}$；取 $n=-1$，得 $B = \frac{-3}{(-1)(1)} = 3$；取 $n=-2$，得 $C = \frac{-5}{(-2)(-1)} = -\frac{5}{2}$。因此

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

留意 $A+B+C=0$。正因如此，這個和才可以裂項相消。

設 $H_N = 1 + \frac{1}{2} + \cdots + \frac{1}{N}$。則

$$\sum_{n=1}^{N} \frac{1}{n+1} = H_N - 1 + \frac{1}{N+1}, \qquad \sum_{n=1}^{N} \frac{1}{n+2} = H_N - 1 - \frac{1}{2} + \frac{1}{N+1} + \frac{1}{N+2}.$$

於是

$$\begin{aligned} S_N &= -\tfrac{1}{2}H_N + 3\left(H_N - 1 + \frac{1}{N+1}\right) - \tfrac{5}{2}\left(H_N - \tfrac{3}{2} + \frac{1}{N+1} + \frac{1}{N+2}\right) \\ &= \left(-\tfrac{1}{2} + 3 - \tfrac{5}{2}\right)H_N + \left(-3 + \tfrac{15}{4}\right) + \left(3 - \tfrac{5}{2}\right)\frac{1}{N+1} - \frac{5}{2(N+2)} \\ &= \frac{3}{4} + \frac{1}{2(N+1)} - \frac{5}{2(N+2)} = \frac{1}{2}\left(\frac{3}{2} + \frac{1}{N+1} - \frac{5}{N+2}\right), \end{aligned}$$

這正是所需的結果。由於 $A + B + C = 0$，H_N 的係數為零。

當 $N \to \infty$ 時，$\frac{1}{N+1}$ 和 $\frac{5}{N+2}$ 都趨向 0，所以

$$S_N \to \frac{3}{4}.$$

數列 a_n

對於 $n \geq 2$，把題目給出的比值逐個相乘：

$$\frac{a_n}{a_1} = \prod_{k=2}^{n} \frac{a_k}{a_{k-1}} = \prod_{k=2}^{n} \frac{k-1}{k+2} \times \prod_{k=2}^{n} \frac{2k-1}{2k-3}.$$

想一想 這兩個乘積可以怎樣化簡？

把它們逐項寫出來。第二個乘積可以直接相消，因為每一項的分子都與下一項的分母約去：

$$\prod_{k=2}^{n} \frac{2k-1}{2k-3} = \frac{3}{1} \cdot \frac{5}{3} \cdot \frac{7}{5} \cdots \frac{2n-1}{2n-3} = 2n - 1.$$

在第一個乘積中，分子依次為 $1, 2, \ldots, n-1$，分母依次為 $4, 5, \ldots, n+2$。由 4 至 $n-1$ 的所有因子全部約去，所以

$$\prod_{k=2}^{n} \frac{k-1}{k+2} = \frac{(n-1)!}{(n+2)!/3!} = \frac{6}{n(n+1)(n+2)}.$$

這個階乘（factorial）形式對所有 $n \geq 2$ 都成立，即使 $n < 5$、根本沒有因子可約時亦然。因此

$$\frac{a_n}{a_1} = \frac{6(2n-1)}{n(n+1)(n+2)}.$$

當 $n = 1$ 時，右邊等於 $\frac{6}{6} = 1$，所以這條公式對所有 $n \geq 1$ 都成立。作為驗算，取 $n = 2$ 得 $\frac{18}{24} = \frac{3}{4}$，與題目給出的比值 $\frac{1 \cdot 3}{4 \cdot 1}$ 吻合。

當 $a_1 = \frac{2}{9}$ 時，

$$a_n = \frac{2}{9} \cdot \frac{6(2n-1)}{n(n+1)(n+2)} = \frac{4}{3} \cdot \frac{2n-1}{n(n+1)(n+2)}.$$

所以 a_n 是第一個級數第 n 項的 $\frac{4}{3}$ 倍。它的部分和（partial sums）為 $\frac{4}{3}S_N \to \frac{4}{3} \cdot \frac{3}{4}$，因此

$$\sum_{n=1}^{\infty} a_n = 1.$$

驗算

- $N=1$：公式給出 $\frac{1}{2}\left(\frac{3}{2}+\frac{1}{2}-\frac{5}{3}\right)=\frac{1}{6}$，正是首項 $\frac{1}{1\cdot2\cdot3}$。$N=2$ 給出 $\frac{1}{2}\left(\frac{3}{2}+\frac{1}{3}-\frac{5}{4}\right)=\frac{7}{24}=\frac{1}{6}+\frac{1}{8}$。兩者皆正確。
- 數值上，首 1999 項之和為 0.74900…，從下方趨近 $\frac{3}{4}$，與公式的預測一致（$S_N\approx\frac{3}{4}-\frac{2}{N}$）。
- 取 $a_1=\frac{2}{9}$，由遞推關係（recurrence）逐項算出 a_n，首 299 項之和為 0.9911…。餘下的尾項之和約為 $\frac{4}{3}\cdot\frac{2}{300}\approx0.009$，與總和為 1 相符。

常見錯誤

- 處理裂項相消的首尾兩端時出錯。像上面那樣用 H_N 來表示，就能把各項算得清清楚楚。
- 階乘乘積錯了一位。約分之前，先把開頭和結尾的幾個因子寫出來。
- 忘記檢查 a_n/a_1 的公式在 $n=1$ 時是否也成立，此時乘積是空的。

STEP 之外

裂項相消是微積分基本定理（fundamental theorem of calculus）的離散版本：$\sum(f(n+1)-f(n))=f(N+1)-f(1)$，正如 $\int f'=f(b)-f(a)$。第二部分的數列是一個超幾何（hypergeometric）數列，即相鄰兩項之比是 n 的有理函數（rational function）。Gosper 演算法（Gosper's algorithm）能夠準確判定這類和何時具有閉合形式（closed form），電腦代數系統正是用這個方法來求級數之和。至於部分和極限的分析，可參閱 Abbott 的 *Understanding Analysis*。

詳細解答

STEP II 1998，第 4 題

純數學 • 分部積分法

> **STEP II 1998，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/98-s2-q4

關鍵思路

那個棘手的因子 $\operatorname{cosec}(x/2)$ 無法直接積分，但相鄰兩個被積函數（integrand）之差卻好處理得多：正弦的差化積公式會產生一個因子 $\sin(x/2)$，剛好把餘割完全約去。剩下的是多項式乘以餘弦，用分部積分法（integration by parts）便可按標準做法處理。然後把這些差逐項相消（telescope），最終歸結到一個可以直接手算的初始值。

解答

差 $I_n - I_{n-1}$

由於 $nx + x/2 = (n + \frac{1}{2})x$，兩個被積函數之差為

$$\left(\tfrac{\pi}{2} - x\right)\left[\sin(n+\tfrac{1}{2})x - \sin(n-\tfrac{1}{2})x\right]\operatorname{cosec}\tfrac{x}{2}.$$

根據和差化積公式（factor formula）$\sin A - \sin B = 2\cos\frac{A+B}{2}\sin\frac{A-B}{2}$，取 $\frac{A+B}{2} = nx$ 及 $\frac{A-B}{2} = \frac{x}{2}$，得

$$\sin(n+\tfrac{1}{2})x - \sin(n-\tfrac{1}{2})x = 2\cos nx\,\sin\tfrac{x}{2}.$$

$\sin\frac{x}{2}$ 與餘割互相約去，所以

$$I_n - I_{n-1} = 2\int_0^{\pi}\left(\tfrac{\pi}{2} - x\right)\cos nx\,\mathrm{d}x.$$

> **想一想** 在 $x = 0$ 處 $\operatorname{cosec}(x/2)$ 為無窮大，這樣約去是否合理？

合理。每個原來的被積函數在 $x \to 0$ 時都有有限的極限，因為 $\sin(n+\frac{1}{2})x \big/ \sin\frac{x}{2} \to$

$2n+1$，所以這些積分都是再普通不過的積分。當 $0 < x \le \pi$ 時，約去只是嚴格的代數運算，而只在 $x = 0$ 這一點改變被積函數的值，並不會改變積分的值。

現在用分部積分法，取 $u = \frac{\pi}{2} - x$ 及 $\frac{\mathrm{d}v}{\mathrm{d}x} = \cos nx$，於是 $v = \frac{1}{n}\sin nx$（因 $n \ge 1$，故成立）：

$$\begin{aligned}\int_0^\pi \left(\tfrac{\pi}{2} - x\right)\cos nx\,\mathrm{d}x &= \left[\left(\tfrac{\pi}{2} - x\right)\frac{\sin nx}{n}\right]_0^\pi + \frac{1}{n}\int_0^\pi \sin nx\,\mathrm{d}x \\ &= 0 + \frac{1}{n}\left[-\frac{\cos nx}{n}\right]_0^\pi = \frac{1-\cos n\pi}{n^2} = \frac{1-(-1)^n}{n^2},\end{aligned}$$

其中邊界項為零，因為 $\sin 0 = \sin n\pi = 0$。因此

$$I_n - I_{n-1} = \frac{2\big(1-(-1)^n\big)}{n^2} = \begin{cases} \dfrac{4}{n^2}, & n \text{ 為奇數}, \\ 0, & n \text{ 為偶數}. \end{cases}$$

求 I_n 的值

要逐項相消，我們需要一個初始值。這個積分同樣可以用來定義 I_0：當 $n = 0$ 時，被積函數是 $(\frac{\pi}{2} - x)\sin\frac{x}{2}\operatorname{cosec}\frac{x}{2} = \frac{\pi}{2} - x$，所以

$$I_0 = \int_0^\pi \left(\tfrac{\pi}{2} - x\right)\mathrm{d}x = \left[\tfrac{\pi}{2}x - \tfrac{1}{2}x^2\right]_0^\pi = 0,$$

而上面的差公式對 $n = 1$ 也成立（推導過程只用到 $n \ge 1$）。把 $1, 2, \ldots, n$ 各項的差相加：

$$I_n = I_0 + \sum_{k=1}^{n}\big(I_k - I_{k-1}\big) = \sum_{k=1}^{n}\frac{2\big(1-(-1)^k\big)}{k^2}.$$

只有奇數的 k 才有貢獻，每項為 $4/k^2$，所以

$$\boxed{I_n = 4\sum_{\substack{k=1 \\ k \text{ 為奇數}}}^{n}\frac{1}{k^2} = 4\Big(1 + \frac{1}{3^2} + \frac{1}{5^2} + \cdots\Big),}$$

若 n 為奇數，最後一項是 $1/n^2$；若 n 為偶數，最後一項是 $1/(n-1)^2$。等價地，$I_n = 4\sum_{r=1}^{m}\frac{1}{(2r-1)^2}$，其中 $m = \lceil n/2 \rceil$。

如果你對使用 $n = 0$ 感到不安（題目說明 n 是正數），可以改由 I_1 開始。由於 $\sin\frac{3x}{2} = \sin\frac{x}{2}(1 + 2\cos x)$，利用上面 $n = 1$ 時的分部積分結果，可得 $I_1 = \int_0^\pi(\frac{\pi}{2} - x)(1 + 2\cos x)\,\mathrm{d}x = 0 + 2\cdot 2 = 4$。這與公式吻合。

驗算

- $I_1 = I_2 = 4$ 及 $I_3 = I_4 = 4 + \frac{4}{9} = \frac{40}{9} \approx 4.4444$。對原來的積分使用辛普森法則（Simpson's rule）（在 $x = 0$ 處取被積函數的極限 $\frac{\pi}{2}(2n+1)$），算得 I_3 為 $4.444444\ldots$，而且對每個 $n \le 12$，結果與上述和式的誤差都在 10^{-8} 之內。

- 對應 n 的被積函數是 $\frac{\pi}{2} - x$ 乘以 $1 + 2\cos x + \cdots + 2\cos nx$，而 $\frac{\pi}{2} - x$ 關於 $x = \frac{\pi}{2}$ 是奇函數。當 k 為偶數時（常數 1 亦然），$\cos kx$ 關於 $x = \frac{\pi}{2}$ 是偶函數，所以乘積是奇函數，在 $[0, \pi]$ 上的積分為零。這就是 $I_0 = 0$ 以及偶數 k 沒有任何貢獻的原因。

常見錯誤

- 把正弦差化積公式中的正負號或因子 2 弄錯；可代入一組數值檢驗，例如 $x = \pi$, $n = 1$。
- 把 $\cos n\pi = 1$ 寫了出來，而沒有寫成 $(-1)^n$，結果令每一個差都變成零。
- 忘記初始值，或把答案寫成對所有 k 求和，而沒有留意到偶數項全部為零。

STEP 之外

函數 $\sin(n + \frac{1}{2})x \big/ \sin\frac{x}{2} = 1 + 2\sum_{k=1}^{n}\cos kx$ 稱為狄利克雷核（*Dirichlet kernel*），是證明傅立葉級數（Fourier series）收斂的核心工具。當 $n \to \infty$ 時，我們的和趨向 $4 \cdot \frac{\pi^2}{8} = \frac{\pi^2}{2}$，這裏用到 $\sum_{\text{奇數}k} 1/k^2 = \pi^2/8$，它是歐拉（Euler）的 $\sum 1/k^2 = \pi^2/6$ 的近親（數值上，$I_{200} \approx 4.925$，而 $\pi^2/2 \approx 4.935$）。關於傅立葉級數和這類和式，Boyce 與 DiPrima 的 *Elementary Differential Equations*（常微分方程入門教材）有很好的講解；至於收斂性背後的分析學，可參考 Abbott 的 *Understanding Analysis*（數學分析入門教材）。

詳細解答

STEP II 1998，第 5 題

純數學 • 複數的三角不等式

> **STEP II 1998，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/98-s2-q5

關鍵思路

$|z_1 - z_2|$ 是阿根圖（Argand diagram）上兩點之間的距離，所以關於 $|z_1 + z_2|$ 的不等式，其實就是說三角形的任何一邊都不會比另外兩邊之和更長。把這個結果推廣到 n 項之後，最後一部分只需用每一項的模（modulus）來估計該項的大小，再認出一個等比級數（geometric series）：$|z|$ 太小的話，左邊就小得永遠達不到 1。

解答

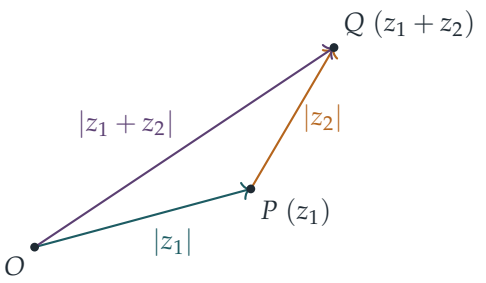
模、距離與三角不等式

若 $z = x + \mathrm{i}y$，其中 x, y 為實數，則 z 的模為

$$|z| = \sqrt{x^2 + y^2},$$

即原點到代表 z 的點 (x, y) 的距離。若 $z_1 = x_1 + \mathrm{i}y_1$ 及 $z_2 = x_2 + \mathrm{i}y_2$，則 $|z_1 - z_2| = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$，這正是代表 z_1 和 z_2 的兩點之間的距離。

現在設 O、P 和 Q 分別為代表 0、z_1 和 $z_1 + z_2$ 的點。



根據上述的幾何解釋，

$$OP = |z_1 - 0| = |z_1|, \qquad PQ = |(z_1 + z_2) - z_1| = |z_2|, \qquad OQ = |z_1 + z_2|.$$

直線段 OQ 是由 O 到 Q 的最短路徑，所以它不會比路徑 $O \to P \to Q$ 更長；即 $OQ \le OP + PQ$，而等號成立當且僅當 P 位於線段 OQ 上（包括各點重合的退化情況）。因此

$$|z_1 + z_2| \le |z_1| + |z_2|.$$

這就是三角不等式（triangle inequality）。

n 項的不等式

我們用歸納法（induction）證明：對所有 $n \ge 1$ 及所有複數 $z_1, \ldots, z_n$，

$$|z_1 + z_2 + \cdots + z_n| \le |z_1| + |z_2| + \cdots + |z_n|.$$

當 $n = 1$ 時，兩邊都等於 $|z_1|$。假設結果對某個 $n = k \ge 1$ 成立。給定 $z_1, \ldots, z_{k+1}$，把兩項的不等式應用於 $w = z_1 + \cdots + z_k$ 和 z_{k+1} 這兩個複數：

$$\begin{aligned} |z_1 + \cdots + z_k + z_{k+1}| = |w + z_{k+1}| &\le |w| + |z_{k+1}| \\ &\le |z_1| + \cdots + |z_k| + |z_{k+1}|, \end{aligned}$$

其中最後一步用了歸納假設（inductive hypothesis）。所以結果對 $n = k + 1$ 成立，由歸納法可知它對所有 $n \ge 1$ 都成立。

最後一部分

假設 $|z| \le \frac{1}{4}$，並記 $r = |z|$，於是 $0 \le r \le \frac{1}{4}$。每一項都滿足 $|a_k z^k| = |a_k|\,|z|^k \le 3r^k$，所以由 n 項的不等式可得

$$\begin{aligned} |a_1 z + a_2 z^2 + \cdots + a_n z^n| &\le |a_1|\, r + |a_2|\, r^2 + \cdots + |a_n|\, r^n \\ &\le 3\big(r + r^2 + \cdots + r^n\big). \end{aligned}$$

> **想一想** 怎樣把 $3(r + r^2 + \cdots + r^n)$ 化成可以與 1 比較的形式？又怎樣確保不等式是嚴格（strict）的？

這是一個有限等比級數，而由於 $r \le \frac{1}{4} < 1$，我們可以使用求和公式：

$$3\big(r + r^2 + \cdots + r^n\big) = \frac{3r(1 - r^n)}{1 - r}.$$

若 $r = 0$，上式等於 0。若 $0 < r \le \frac{1}{4}$，則 $0 < 1 - r^n < 1$，所以

$$\frac{3r(1 - r^n)}{1 - r} < \frac{3r}{1 - r} \le \frac{3 \cdot \frac{1}{4}}{1 - \frac{1}{4}} = 1,$$

其中最後一步成立，是因為 $\dfrac{3r}{1-r}$ 在 $[0,1)$ 上隨 r 遞增（分子遞增，而取正值的分母遞減）。無論是哪一種情況，都有

$$|a_1 z + a_2 z^2 + \cdots + a_n z^n| < 1,$$

所以左邊不可能等於 1。因此方程沒有滿足 $|z| \le \frac{1}{4}$ 的解。

驗算

- 這個界是最佳的：取每個 $a_k = 3$ 及 $z = \frac{1}{4}$，左邊就是 $3\left(\frac{1}{4} + \cdots + \frac{1}{4^n}\right) = 1 - 4^{-n}$，例如當 $n = 1, 2, 3$ 時分別為 0.75、0.9375、0.984375。它可以任意接近 1，卻永遠達不到，這正是嚴格不等式如此重要的原因。
- 在一次隨機測試中，我們選取了 200 000 組數據，其中 $n \le 15$、a_k 滿足 $|a_k| \le 3$、z 滿足 $|z| \le \frac{1}{4}$，所得 $|a_1 z + \cdots + a_n z^n|$ 的最大值約為 0.996，始終小於 1。

常見錯誤

- 單靠無窮級數之和 $3r/(1-r) \le 1$，只能得出 $|\text{左邊}| \le 1$，這並不足以排除方程有解的可能。嚴格不等式要靠有限項之和（或者 $r > 0$ 時 $r^n > 0$ 這個事實）才能得到。
- 在歸納法中，兩項的不等式必須應用於 $z_1 + \cdots + z_k$，並把它視為單一個複數；這正是該步驟的關鍵所在。

STEP 之外

三角不等式是度量（metric）的定義性質，而 $|z_1 - z_2|$ 令 $\mathbb{C}$ 成為一個度量空間（metric space）；分析學中幾乎每一個關於收斂的論證都建基於它。最後一部分是一類定理的入門版本，這類定理用來確定多項式和冪級數（power series）的零點（zeros）位置：這裏 $f(z) = a_1 z + \cdots + a_n z^n - 1$ 在圓盤 $|z| \le \frac{1}{4}$ 內沒有零點，與儒歇定理（Rouché's theorem）的精神一脈相承。想了解複數的幾何，可讀 Needham 的 *Visual Complex Analysis*（以圖像闡釋複分析）；想了解分析學中的度量概念，可讀 Abbott 的 *Understanding Analysis*（實分析入門）。

詳細解答

STEP II 1998，第 6 題

純數學 • 擺線與法線

> **STEP II 1998，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/98-s2-q6

關鍵思路

兩條曲線都是擺線（cycloid），即滾動的圓的圓周上一點所描出的路徑。只要用上半角公式（half-angle formulae）$1+\cos\theta = 2\cos^2\frac{\theta}{2}$、$1-\cos\theta = 2\sin^2\frac{\theta}{2}$ 和 $\sin\theta = 2\sin\frac{\theta}{2}\cos\frac{\theta}{2}$，有關它們的一切都會變得簡單。斜率（gradient）會化簡成 $\frac{\theta}{2}$ 的單一個正切或餘切，而法線（normal）的方程亦會縮成簡潔的一行。要證明一條直線是曲線 (2) 的切線（tangent），只需在 (2) 上找出一點，使該點在直線上，而且曲線在該點的斜率等於直線的斜率。

（題目中第二條曲線本應寫 y_2 的地方寫成了 y_1；這是印刷錯誤。）

解答

斜率

對於曲線 (1)，$\frac{\mathrm{d}x_1}{\mathrm{d}\theta} = 1+\cos\theta$ 及 $\frac{\mathrm{d}y_1}{\mathrm{d}\theta} = -\sin\theta$，所以

$$\frac{\mathrm{d}y_1}{\mathrm{d}x_1} = \frac{-\sin\theta}{1+\cos\theta} = \frac{-2\sin\frac{\theta}{2}\cos\frac{\theta}{2}}{2\cos^2\frac{\theta}{2}} = -\tan\frac{\theta}{2},$$

此式在 $\cos\frac{\theta}{2} \neq 0$ 時成立，即 $\theta \neq \pi$。

對於曲線 (2)，$\frac{\mathrm{d}x_2}{\mathrm{d}\theta} = 1-\cos\theta$ 及 $\frac{\mathrm{d}y_2}{\mathrm{d}\theta} = \sin\theta$，所以

$$\frac{\mathrm{d}y_2}{\mathrm{d}x_2} = \frac{\sin\theta}{1-\cos\theta} = \frac{2\sin\frac{\theta}{2}\cos\frac{\theta}{2}}{2\sin^2\frac{\theta}{2}} = \cot\frac{\theta}{2},$$

此式在 $\theta \neq 0, 2\pi$ 時成立。

當 $\theta = \pi/2$ 時，$\frac{\theta}{2} = \frac{\pi}{4}$：曲線 (1) 的斜率為 -1，曲線 (2) 的斜率為 1。

當 $\theta = 3\pi/2$ 時，$\frac{\theta}{2} = \frac{3\pi}{4}$，此時 $\tan = -1$ 且 $\cot = -1$：曲線 (1) 的斜率為 1，曲線 (2) 的斜率為 -1。

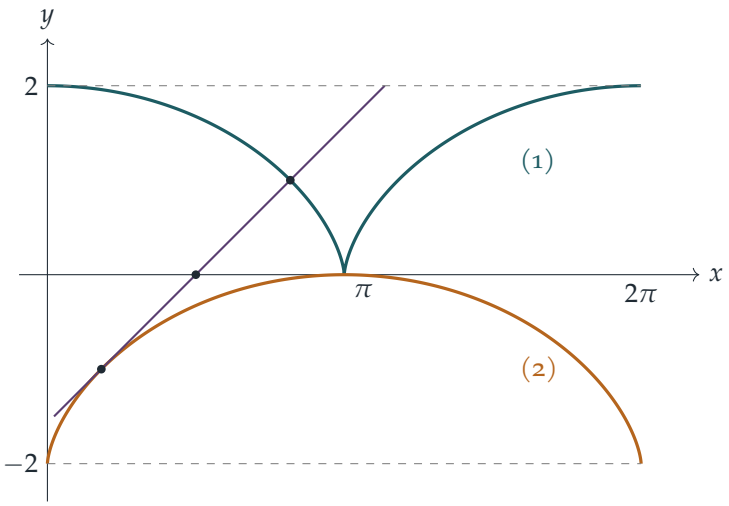
略圖

x_1 和 x_2 都是 θ 的非遞減函數（它們的導數 $1 \pm \cos\theta$ 永不為負），所以每條曲線都由左至右，從 $x = 0$ 延伸到 $x = 2\pi$。

曲線 (1) 由 $(0, 2)$ 出發，起點處斜率為 0，下降至 $(\pi, 0)$，然後對稱地上升至 $(2\pi, 2)$。當 $\theta = \pi$ 時，兩個導數同時為零，而 $|\mathrm{d}y_1/\mathrm{d}x_1| = |\tan\frac{\theta}{2}| \to \infty$，所以在 $(\pi, 0)$ 有一個尖點（cusp），該處的切線是鉛垂的。由於 $0 \le y_1 \le 2$，曲線位於 $y = 0$ 與 $y = 2$ 之間。

曲線 (2) 在 $(0, -2)$ 有一個尖點（切線鉛垂，因為 $\cot\frac{\theta}{2} \to \infty$），上升至 $(\pi, 0)$，該處的斜率為 $\cot\frac{\pi}{2} = 0$，然後下降至尖點 $(2\pi, -2)$。它是一個立於直線 $y = -2$ 上的普通擺線拱。

兩條曲線在 $(\pi, 0)$ 相接：曲線 (1) 鉛垂地到達該點，曲線 (2) 則水平地越過該點。整幅圖關於 $x = \pi$ 對稱，因為 $\theta \mapsto 2\pi - \theta$ 會令 $x \mapsto 2\pi - x$，而兩條曲線上的 y 都保持不變。圖中紫紅色的直線是 (1) 在 $\theta = \pi/2$ 處的法線，下文會再提到它。



曲線 (1) 的法線

(1) 的切線斜率為 $-\tan\frac{\theta}{2}$，所以法線的斜率為 $\cot\frac{\theta}{2}$（其中 $\theta \neq 0, \pi, 2\pi$；這些情況留待下文處理）。在點 $(\theta + \sin\theta,\ 1 + \cos\theta)$ 處的法線為

$$y - (1 + \cos\theta) = \frac{\cos\frac{\theta}{2}}{\sin\frac{\theta}{2}}\left(x - \theta - \sin\theta\right).$$

全式乘以 $\sin\frac{\theta}{2}$，並使用半角形式：

$$y\sin\tfrac{\theta}{2} - 2\cos^2\tfrac{\theta}{2}\sin\tfrac{\theta}{2} = (x-\theta)\cos\tfrac{\theta}{2} - 2\sin\tfrac{\theta}{2}\cos^2\tfrac{\theta}{2}.$$

兩邊最後一項相等，可以互相抵消，剩下

$$\boxed{y\sin\tfrac{\theta}{2} = (x-\theta)\cos\tfrac{\theta}{2}}$$

這就是法線的方程。

這個形式同時涵蓋了那些例外的值。當 $\theta = 0$ 時，它變成 $x = 0$，即通過 $(0,2)$ 的鉛垂線；這是正確的，因為該處的切線是水平的。當 $\theta = 2\pi$ 時，同理可得 $x = 2\pi$。當 $\theta = \pi$ 時，它變成 $y = 0$，即通過尖點的水平線，與尖點處的鉛垂切線互相垂直。

留意法線必定通過 x 軸上的點 $(\theta, 0)$。這正是滾動的圓與它所滾過的直線相切的那一點，是擺線的一個經典性質。

法線是曲線 (2) 的切線

想一想 法線應該與曲線 (2) 相切於哪一點？

(2) 在參數 (parameter) 為 ϕ 處的切線斜率為 $\cot\frac{\phi}{2}$，而法線的斜率為 $\cot\frac{\theta}{2}$。對於 ϕ, θ 在 $(0, 2\pi)$ 內的情況，兩者只在 $\phi = \theta$ 時相等，因為 cot 在 $(0,\pi)$ 上是一對一 (one-to-one) 的。所以唯一的候選點就是 (2) 上參數同為 θ 的那一點。

該點是 $(x_2, y_2) = (\theta - \sin\theta,\ -(1+\cos\theta))$。代入法線的方程：

$$\begin{aligned}
\text{左方} &= y_2\sin\tfrac{\theta}{2} = -2\cos^2\tfrac{\theta}{2}\sin\tfrac{\theta}{2},\\
\text{右方} &= (x_2-\theta)\cos\tfrac{\theta}{2} = -\sin\theta\cos\tfrac{\theta}{2} = -2\sin\tfrac{\theta}{2}\cos^2\tfrac{\theta}{2}.
\end{aligned}$$

兩者相等，所以 (1) 在 θ 處的法線通過 (2) 上參數為 θ 的點。在該點，(2) 的斜率為 $\cot\frac{\theta}{2}$，與法線的斜率相同。一條直線若通過曲線上某一點，而且斜率等於曲線在該點的斜率，它就是曲線在該點的切線，所以 (1) 的法線是 (2) 的切線，證明完畢。

特殊值的情況同樣成立。當 $\theta = \pi$ 時，法線 $y = 0$ 與 (2) 相切於其頂點 $(\pi, 0)$，該處 (2) 的斜率為 0。當 $\theta = 0$（或 2π）時，法線 $x = 0$（或 $x = 2\pi$）就是 (2) 在尖點處的鉛垂切線。

驗算

- 當 $\theta = \pi/2$ 時，法線為 $y = x - \frac{\pi}{2}$。它通過 (1) 上的 $(\frac{\pi}{2}+1, 1)$，通過 $(\frac{\pi}{2}, 0)$，亦通過 (2) 上的 $(\frac{\pi}{2}-1, -1)$，而 (2) 在該點的斜率為 1，與第一部分的結果一致。這就是略圖中紫紅色的直線。

- 以有限差分（finite differences）對斜率作數值驗算，結果與 $-1, 1$ 及 $1, -1$ 吻合至小數點後六位。在 1000 個隨機選取的 θ 值上，法線都通過兩個點，並且與 (2) 的切線平行，誤差在 10^{-12} 之內。

常見錯誤

- 把斜率保留為 $-\sin\theta/(1+\cos\theta)$ 的形式，然後在求法線時硬闖繁複的代數運算。用上半角形式，求法線只需兩行計算。
- 只證明了法線與曲線 (2) 相交。相交並不足夠：還須證明在交點處兩者的斜率相同。

STEP 之外

一條曲線的各個曲率中心（centres of curvature）所描出的曲線，稱為它的*漸屈線*（evolute），它同時也是該曲線所有法線的包絡線（envelope）。這道題說明了擺線的漸屈線是一條完全相同的擺線，只是平移了半個拱。沿法線由 (1) 到它與 (2) 的切點的距離為 $4|\cos\frac{\theta}{2}|$，剛好等於 (2) 由該點到其頂點的弧長（arc length）。所以 (1) 可以看成是一條由 (2) 上逐漸解開的繩子的末端所描出的曲線。惠更斯（Huygens）利用這一點造出了一個週期與擺幅無關的擺。想了解曲率、漸屈線和漸伸線（involutes），可參閱 Pressley 的 *Elementary Differential Geometry*（初等微分幾何）。

詳細解答

STEP II 1998，第 7 題

純數學 • 以導數證明不等式

STEP II 1998，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/98-s2-q7

關鍵思路

每個不等式 (inequality) 的證法都一樣：函數在 $x = 0$ 處為零，而它的導數 (derivative) 為正，所以函數由零開始上升。巧妙之處在於這些函數是環環相扣的。g 的導數原來與 f 有關，而 h 的導數原來就是 g 在 $2x$ 處的值。因此每一部分都為下一部分提供所需的正導數。在第 (iv) 部分，取對數（logarithm）可把商（quotient）F 化為和，而 F'/F 的正負號就是第 (iii) 部分那個式子的正負號。

解答

全題所用的工具是：若 ϕ 可微(differentiable)，$\phi(0) = 0$，且當 $0 < x < a$ 時 $\phi'(x) > 0$，則 ϕ 在 $[0, a)$ 上嚴格遞增（strictly increasing），所以當 $0 < x < a$ 時 $\phi(x) > \phi(0) = 0$。（根據中值定理（mean value theorem），存在某個 $c \in (0, x)$ 使得 $\phi(x) = \phi(x) - \phi(0) = x\,\phi'(c)$，而這是正數。）

第 (i) 部分

我們有 $\mathrm{f}(0) = \tan 0 - 0 = 0$，而且

$$\mathrm{f}'(x) = \sec^2 x - 1 = \tan^2 x.$$

當 $0 < x < \frac{1}{2}\pi$ 時，$\tan x$ 有定義且為正，所以 $\mathrm{f}'(x) > 0$。因此當 $0 < x < \frac{1}{2}\pi$ 時 $\mathrm{f}(x) > 0$；也就是說，在該範圍內 $\tan x > x$。

第 (ii) 部分

同樣地，$g(0) = 2 - 2 - 0 = 0$。求導，並對 $x \sin x$ 使用積法則（product rule），

$$\begin{aligned} g'(x) &= 2\sin x - (\sin x + x\cos x) \\ &= \sin x - x\cos x \\ &= \cos x\,(\tan x - x) = \cos x\, f(x). \end{aligned}$$

當 $0 < x < \frac{1}{2}\pi$ 時，我們有 $\cos x > 0$，而由第 (i) 部分知 $f(x) > 0$。所以 $g'(x) > 0$，又因為 $g(0) = 0$，可知當 $0 < x < \frac{1}{2}\pi$ 時 $g(x) > 0$。

第 (iii) 部分

這次同樣有 $h(0) = 0$。求導，

$$\begin{aligned} h'(x) &= 2 + \cos 2x - 2x\sin 2x - 3\cos 2x \\ &= 2 - 2\cos 2x - 2x\sin 2x. \end{aligned}$$

想一想 你以前見過這個式子嗎？

它就是把 g 中的 x 換成 $2x$：

$$h'(x) = 2 - 2\cos(2x) - (2x)\sin(2x) = g(2x).$$

若 $0 < x < \frac{1}{4}\pi$，則 $0 < 2x < \frac{1}{2}\pi$，所以由第 (ii) 部分知 $g(2x) > 0$。這就是為甚麼範圍現在是 $\frac{1}{4}\pi$ 而不是 $\frac{1}{2}\pi$。所以在 $0 < x < \frac{1}{4}\pi$ 上 $h'(x) > 0$，再加上 $h(0) = 0$，便得出當 $0 < x < \frac{1}{4}\pi$ 時 $h(x) > 0$。

至於「由此」的部分，用二倍角（double angle）公式改寫題目所給的式子。由於 $\sin^2 x + 3\cos^2 x = 1 + 2\cos^2 x = 2 + \cos 2x$ 及 $3\sin x\cos x = \frac{3}{2}\sin 2x$，

$$\begin{aligned} x(\sin^2 x + 3\cos^2 x) - 3\sin x\cos x &= x(2 + \cos 2x) - \tfrac{3}{2}\sin 2x \\ &= 2x + x\cos 2x - \tfrac{3}{2}\sin 2x = h(x). \end{aligned}$$

所以該式子正正就是 $h(x)$，而它在 $0 < x < \frac{1}{4}\pi$ 時為正，這正是所需的結果。

第 (iv) 部分

當 $0 < x < \frac{1}{4}\pi$ 時，x、$\sin x$ 和 $\cos x$ 這幾個量全為正，所以 $F(x) > 0$，我們可以取對數：

$$\ln F(x) = \ln x + \tfrac{1}{3}\ln\cos x - \ln\sin x.$$

兩邊對 x 求導，

$$\frac{\mathrm{F}'(x)}{\mathrm{F}(x)} = \frac{1}{x} - \frac{\sin x}{3\cos x} - \frac{\cos x}{\sin x}.$$

想一想 怎樣看出它的正負號？

把它通分，公分母（common denominator）為 $3x\sin x\cos x$，在這個區間上是正數：

$$\begin{aligned}\frac{\mathrm{F}'(x)}{\mathrm{F}(x)} &= \frac{3\sin x\cos x - x\sin^2 x - 3x\cos^2 x}{3x\sin x\cos x} \\ &= -\frac{x(\sin^2 x + 3\cos^2 x) - 3\sin x\cos x}{3x\sin x\cos x} = -\frac{\mathrm{h}(x)}{3x\sin x\cos x}.\end{aligned}$$

由第 (iii) 部分，分子 $\mathrm{h}(x)$ 為正，分母亦為正，所以 $\mathrm{F}'(x)/\mathrm{F}(x) < 0$。由於 $\mathrm{F}(x) > 0$，兩邊乘以它便得出當 $0 < x < \frac{1}{4}\pi$ 時 $\mathrm{F}'(x) < 0$，這正是所需的結果。

驗算

- 抽查數值：$\mathrm{F}(0.1) \approx 0.999998$、$\mathrm{F}(0.5) \approx 0.99849$ 及 $\mathrm{F}(\frac{1}{4}\pi) \approx 0.98954$，所以 F 是遞減的，只是減得很慢。
- 在 0 附近，這些不等式非常緊。展開成麥克勞林級數（Maclaurin series）可得 $\mathrm{h}(x) = \frac{4}{15}x^5 + \cdots$，而 x, x^3 兩項互相抵消。這就是為甚麼你不能單憑觀察就「看出」$\mathrm{h} > 0$，而需要這一連串導數。數值上，$\mathrm{h}(0.01)/0.01^5 \approx 0.26666$。
- 我們在 20 000 個隨機點上測試了全部四個不等式，以及恆等式 $\mathrm{g}' = \mathrm{f}\cos x$、$\mathrm{h}'(x) = \mathrm{g}(2x)$ 和 $\mathrm{F}'/\mathrm{F} = -\mathrm{h}/(3x\sin x\cos x)$。

常見錯誤

- 只論證 $\mathrm{f}'(x) \geq 0$，或忘記指出 $\mathrm{f}(0) = 0$。兩項事實都不可缺少：起點為零，以及導數嚴格為正。
- 在第 (iii) 部分使用第 (ii) 部分的結果，卻沒有檢查 $2x$ 是否落在 $(0, \frac{1}{2}\pi)$ 之內。$x < \frac{1}{4}\pi$ 這個限制正是由此而來。
- 由 $\mathrm{F}'/\mathrm{F} < 0$ 推出 $\mathrm{F}' < 0$，卻沒有指出 $\mathrm{F} > 0$。

STEP 之外

「起點為零而導數為正，則函數為正」這個原理是中值定理的推論，而中值定理是大學分析（analysis）課程中最早遇到的重要定理之一。像本題這樣反覆運用這個原理，是證明

$\sin x > x - \frac{1}{6}x^3$ 這類不等式的標準方法，而且它與帶餘項的泰勒定理（Taylor's theorem）關係密切。Abbott 的 *Understanding Analysis*（分析學入門教科書）對中值定理及其應用有仔細的講解。

詳細解答

STEP II 1998，第 8 題

純數學 • 向量幾何

> **STEP II 1998，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/98-s2-q8

關鍵思路

這裏的一切，都在於把每條直線寫成「點 + 參數（parameter）× 方向」的形式，然後巧妙地選取參數。各式子對 **a**、**b**、**c** 是對稱（symmetric）的，所以如果某個參數值給出的答案對這三個向量對稱，那麼同一點就位於全部三條直線上。這正是題目中「憑觀察」（by inspection）一語要我們去找的東西。

解答

三條中線（median）交於 R

BC 的中點（mid-point）的位置向量（position vector）為 $\frac{1}{2}(\mathbf{b}+\mathbf{c})$，所以由 A 到該中點的直線為

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

我們希望三個係數相等，使結果對稱：由 $1-s = \frac{s}{2}$ 得 $s = \frac{2}{3}$，而該點為

$$\mathbf{r} = \tfrac{1}{3}(\mathbf{a}+\mathbf{b}+\mathbf{c}).$$

由 B 和由 C 出發的直線，可把 $\mathbf{a}, \mathbf{b}, \mathbf{c}$ 循環置換而得，而同樣取 $s = \frac{2}{3}$，便在每條直線上得出同一個對稱的點。因此三條直線都通過點 R，其位置向量為

$$\mathbf{r}_R = \tfrac{1}{3}(\mathbf{a}+\mathbf{b}+\mathbf{c}),$$

即三角形的形心（centroid）。（由於 A、B、C 構成一個真正的三角形，這些直線中沒有兩條重合，所以 R 是它們唯一的公共點。）

三條平行線交於 Q

設 P 的位置向量為 $\mathbf{p}$。三條直線分別是

$$\begin{aligned}
&\text{通過}BC\text{ 的中點，且平行於}PA: \quad \mathbf{r} = \tfrac{1}{2}(\mathbf{b}+\mathbf{c}) + \lambda(\mathbf{a}-\mathbf{p}),\\
&\text{通過}CA\text{ 的中點，且平行於}PB: \quad \mathbf{r} = \tfrac{1}{2}(\mathbf{c}+\mathbf{a}) + \mu(\mathbf{b}-\mathbf{p}),\\
&\text{通過}AB\text{ 的中點，且平行於}PC: \quad \mathbf{r} = \tfrac{1}{2}(\mathbf{a}+\mathbf{b}) + \nu(\mathbf{c}-\mathbf{p}).
\end{aligned}$$

想一想 λ 取甚麼值，才會使第一個式子對 $\mathbf{a}$、$\mathbf{b}$、$\mathbf{c}$ 對稱？

在第一條直線中，$\mathbf{b}$ 和 $\mathbf{c}$ 的係數各為 $\frac{1}{2}$，而 $\mathbf{a}$ 的係數是 λ；要令它們一致，便需要 $\lambda = \frac{1}{2}$。於是

$$\mathbf{r} = \tfrac{1}{2}(\mathbf{a}+\mathbf{b}+\mathbf{c}-\mathbf{p}).$$

在另外兩條直線中取 $\mu = \frac{1}{2}$ 和 $\nu = \frac{1}{2}$，得到的是完全相同的向量。因此三條直線都通過點 Q，其中

$$\mathbf{q} = \tfrac{1}{2}(\mathbf{a}+\mathbf{b}+\mathbf{c}-\mathbf{p}).$$

PQ 與平面 ABC 交於 R

首先，$Q \neq P$：假如 $\mathbf{q} = \mathbf{p}$，則 $\mathbf{a}+\mathbf{b}+\mathbf{c} = 3\mathbf{p}$，所以 $\mathbf{p} = \mathbf{r}_R$，$P$ 便會是點 R，而這一點位於平面 ABC 上。所以 PQ 確實是一條直線，即

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

要消去含 $\mathbf{p}$ 的項，我們需要 $1 - \frac{3}{2}t = 0$，即 $t = \frac{2}{3}$，這時

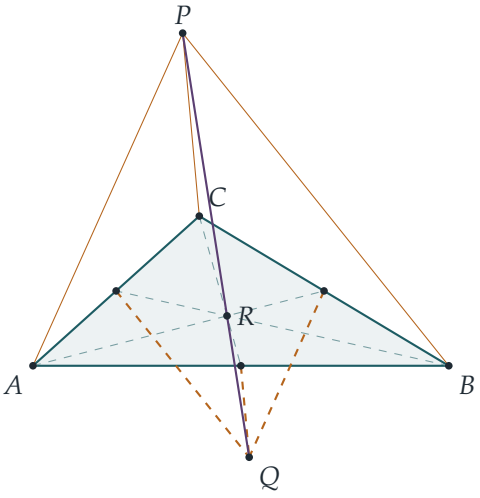
$$\mathbf{r} = \mathbf{p} + \tfrac{1}{3}(\mathbf{a}+\mathbf{b}+\mathbf{c}) - \mathbf{p} = \tfrac{1}{3}(\mathbf{a}+\mathbf{b}+\mathbf{c}) = \mathbf{r}_R.$$

所以 R 位於直線 PQ 上。它同時位於平面 ABC 上（它在由 A 出發的中線上，而這條中線位於該平面內）。

想一想 PQ 會不會還在別處與這個平面相交？

不會。一條直線與一個平面或者沒有交點，或者恰好有一個交點，或者整條位於平面之內。直線 PQ 並不在平面 ABC 內，因為 P 在這條直線上，而 P 不在該平面上。所以它與平面最多只有一個交點，而這個交點就是 R。這就證明了 PQ 與平面 ABC 交於 R。

額外的收穫是，$t = \frac{2}{3}$ 告訴我們 R 把 PQ 按比 $PR : RQ = 2 : 1$ 分割；等價地，$\mathbf{q} - \mathbf{r}_R = -\frac{1}{2}(\mathbf{p} - \mathbf{r}_R)$。



驗算

- 每個聲稱的交點都可以代回去檢驗：例如 $\mathbf{q} - \frac{1}{2}(\mathbf{b}+\mathbf{c}) = \frac{1}{2}(\mathbf{a}-\mathbf{p})$，顯然平行於 PA。
- 取 $A = (0,0,0)$、$B = (3,0,0)$、$C = (0,3,0)$ 和 $P = (0,0,3)$，我們得到 $R = (1,1,0)$ 和 $Q = (\frac{3}{2}, \frac{3}{2}, -\frac{3}{2})$；由 P 到 Q 的三分之二處的點是 $(1,1,0) = R$，位於平面 $z = 0$ 上。
- 我們還用精確的有理數運算，對 2000 組隨機選取的 A、B、C、P 檢驗了所有結果：直線 PQ 每次都與平面 ABC 交於 R，參數為 $t = \frac{2}{3}$。

常見錯誤

- 解方程求出其中兩條直線上的公共點後，忘記檢查它是否也在第三條直線上。
- 證明了 R 位於 PQ 上，卻沒有解釋為何它就是 PQ 與平面唯一的交點；「P 不在平面上」這個條件正是用在這裏。

STEP 之外

這個結果背後有一個簡潔的幾何原因。以 R 為中心、比例因子（scale factor）為 $-\frac{1}{2}$ 的放大（enlargement）（即映射 $\mathbf{x} \mapsto \mathbf{r}_R - \frac{1}{2}(\mathbf{x} - \mathbf{r}_R)$）把 A、B、C 分別送到 BC、CA、AB 的中點，並把任何直線送到與之平行的直線。所以它把交於 P 的直線 PA、PB、PC 變成那三條新直線，因而這三條新直線必定交於 P 的像；這個像就是 Q，而 P、R、Q 共線（collinear），因為放大的中心與每一點及其像都共線。透過變換（仿射映射（affine

map）及其不動點（fixed point））來思考幾何，是大學一年級線性代數（linear algebra）的一個核心思想；Strang 的 *Introduction to Linear Algebra*（線性代數入門教材）是了解這個思想如何展開的好去處。

詳細解答

STEP II 1998，第 9 題

力學 • 桁架的靜力學

> **STEP II 1998，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/98-s2-q9

關鍵思路

框架是輕質的（light），所以只有兩個外力作用在它上面：A 處的支承力和 C 處的重量。兩個力要達致平衡（equilibrium），必須大小相等、方向相反，並且沿同一直線作用；這就同時決定了支承力和六邊形的懸掛方式。之後便是節點法（method of joints）：從只有兩枝桿相交的節點入手，分解（resolve）各力，然後沿六邊形逐個節點推進，每次都移到最多只有兩個未知力的節點。

解答

支承力

把六邊形的各邊和三枝對角桿視為輕桿（light rod），在六個頂點處以光滑鉸接（smoothly pinned）相連，並以 W 表示 1 kg 質量的重量，所以 $W = g \approx 9.8\,\mathrm{N}$。各桿都是輕質的，因此作用於整個框架（framework）的外力只有 A 處支承所施的力 $\mathbf{R}$，以及在 C 處鉛垂向下作用的重量 W。

> ? **想一想** 框架可以隨我們喜歡以任何角度懸掛嗎？

不可以。就整個框架而言，各力之和必須為零，而且總力矩（moment）也必須為零。對 A 取力矩，支承力沒有力矩，所以 C 處的重量對 A 的力矩也必須為零。這只有在 C 位於 A 正下方時才會發生。因此框架懸掛時，短對角線 AC 是鉛垂的，而分解各力可得

$$\mathbf{R} = W \text{ 鉛垂向上}, \qquad |\mathbf{R}| = g \approx 9.8\,\mathrm{N},$$

即 1 kg 的重量，沿 CA 作用。

各桿中的力

取邊長為 1（各力與邊長無關）。當 AC 鉛垂時，B 位於 AC 的一半高度處，在它的一側，而六邊形的其餘部分則在另一側：F 與 A 同高，E 與 B 同高，D 與 C 同高。所以 AF、BE 和 CD 是水平的，邊 AB、BC、DE、EF 與水平線成 $60°$，而短對角線 BD、BF 與水平線成 $30°$。

設 T_{XY} 為桿 XY 中的張力（tension）；在桿的每一端，它都把節點拉向另一端。張力為負，表示該桿受推力（thrust）。

想一想 共有九個未知的桿力。我們應從哪裏入手？

從只有兩枝桿相交的節點入手，因為這樣兩條方程（水平和鉛垂方向）便可以立即解出。節點 C 只連着 CB 和 CD，而且負載就掛在這裏。

節點 C。桿 CB 指向水平線以上 $60°$，而 CD 是水平的，背向 B 的一側。分解各力，

$$\begin{aligned} &\text{鉛垂方向：} \quad T_{BC}\sin 60° = W \;\Rightarrow\; T_{BC} = \tfrac{2W}{\sqrt{3}}, \\ &\text{水平方向：} \quad T_{CD} = T_{BC}\cos 60° = \tfrac{W}{\sqrt{3}}. \end{aligned}$$

節點 D。現在只有 DE 和 DB 是未知的。從 D 看，桿 DC 沿水平方向指向 B 的一側，DE 指向水平線以上 $60°$ 並背向該側，而 DB 指向水平線以上 $30°$ 並朝向該側。分解各力，

$$\begin{aligned} &\text{鉛垂方向：} \quad T_{DE}\sin 60° + T_{BD}\sin 30° = 0 \;\Rightarrow\; T_{BD} = -\sqrt{3}\,T_{DE}, \\ &\text{水平方向：} \quad T_{CD} - T_{DE}\cos 60° + T_{BD}\cos 30° = 0 \;\Rightarrow\; T_{CD} - \tfrac{1}{2}T_{DE} - \tfrac{3}{2}T_{DE} = 0, \end{aligned}$$

所以 $T_{DE} = \frac{1}{2}T_{CD} = \frac{W}{2\sqrt{3}}$，而 $T_{BD} = -\frac{W}{2}$。

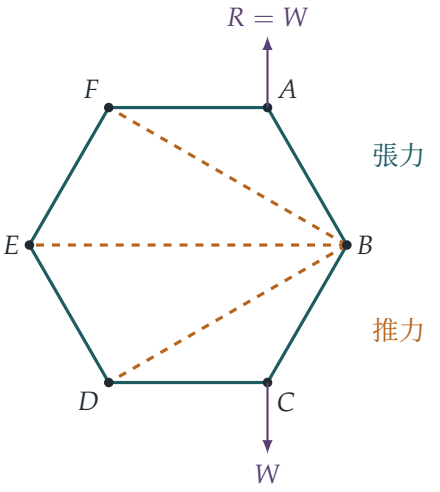
節點 E。桿 ED 指向水平線以下 $60°$，EF 指向水平線以上 $60°$，而 EB 是水平的，三者都朝向 B 的一側。沿鉛垂方向分解，得 $T_{EF} = T_{DE} = \frac{W}{2\sqrt{3}}$。沿水平方向分解，

$$\tfrac{1}{2}T_{DE} + \tfrac{1}{2}T_{EF} + T_{BE} = 0 \;\Rightarrow\; T_{BE} = -\tfrac{W}{2\sqrt{3}}.$$

節點 F。這是節點 D 對直線 BE 的鏡像。沿鉛垂方向分解，$T_{BF} = -\sqrt{3}\,T_{EF} = -\frac{W}{2}$；沿水平方向，$T_{FA} = 2T_{EF} = \frac{W}{\sqrt{3}}$。

節點 A。水平方向，$T_{AB}\cos 60° = T_{FA}$，所以 $T_{AB} = \frac{2W}{\sqrt{3}}$。鉛垂方向，$R = T_{AB}\sin 60° = W$，與第一部分的結果一致。

節點 B 還會給出兩條方程，但它們自動成立：整個框架本身已經處於平衡。驗證這兩條方程，是檢查運算有否出錯的好方法。



總結如下，其中 $W = g$：

- 六條邊全部受張力：AB 和 BC 承受 $\frac{2W}{\sqrt{3}}$，CD 和 FA 承受 $\frac{W}{\sqrt{3}}$，DE 和 EF 承受 $\frac{W}{2\sqrt{3}}$；
- 三枝內桿全部受推力：BD 和 BF 承受 $\frac{W}{2}$，而 BE 承受 $\frac{W}{2\sqrt{3}}$。

所以 BE 中的力是推力，大小為

$$\frac{g}{2\sqrt{3}} = \frac{\sqrt{3}g}{6} \approx 2.83\,\text{N}.$$

驗算

- 對稱性：對直線 BE 作反射，會把 $A \leftrightarrow C$ 和 $F \leftrightarrow D$ 互換，並把 A 處向上的力 W 變成 C 處向下的力 W。所以負載是對稱的，答案也應該對稱：$T_{AB} = T_{BC}$，$T_{FA} = T_{CD}$，$T_{EF} = T_{DE}$，$T_{BF} = T_{BD}$。結果確實如此。
- 節點 B：水平方向，兩條邊向一方給出 $2 \cdot \frac{2W}{\sqrt{3}} \cdot \frac{1}{2}$，而 BD、BF、BE 向另一方給出 $2 \cdot \frac{W}{2} \cdot \frac{\sqrt{3}}{2} + \frac{W}{2\sqrt{3}} = \frac{2W}{\sqrt{3}}$；鉛垂方向，$AB$ 和 BC 互相抵消，BD 和 BF 亦然。
- 我們取 $g = 9.8$，以數值方法解出全部十二條節點方程：它們互相一致，給出 $R = 9.8\,\text{N}$ 向上，以及 $|T_{BE}| \approx 2.829\,\text{N}$ 的推力；而若 AC 傾斜，方程便無解。

常見錯誤

- 把六邊形畫成有一條邊或一條長對角線是鉛垂的。懸掛位置是被迫確定的：C 在 A 的下方。

- 在不同節點之間混淆正負號的約定。一開始就定下「張力把節點拉向桿的另一端」，然後讓負的答案告訴你哪些桿受推力。

STEP 之外

一個有 j 個節點的框架，當它恰好有 $2j-3$ 枝桿時，便是剛性（rigid）而且靜定（statically determinate）的：這裏 $2\cdot 6-3=9$，這就是為何單靠靜力學便可求出各力。若桿的數目更多，結構便有「冗餘」，這時就需要考慮桿的彈性；這是結構力學的起點，也是圖的剛性理論這門數學的起點。想了解背後的向量力學，可讀 Kleppner and Kolenkow, *An Introduction to Mechanics*（力學入門教材）。

詳細解答

STEP II 1998，第 10 題

力學 • 移動楔上的質點

STEP II 1998，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/98-s2-q10

關鍵思路

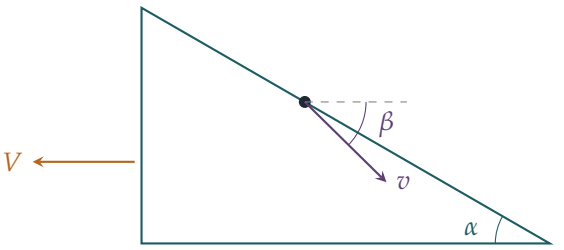
整道題由兩個事實推動。由於質點（particle）一直留在斜面上，它相對於楔塊（*wedge*）的速度必定沿着斜面的方向。另外，由於系統以外沒有任何力在水平方向推動它，而且除重力外沒有其他力做功，所以水平動量（momentum）和能量都守恆。(i)–(iii) 每一部分，都是把其中一個事實寫成方程。之後第 (ii) 部分說明 β 始終不變，於是最後一部分就變成勻加速運動。

解答

模型。質點視為質點；楔塊是只滑動而不轉動的剛體（rigid body）；斜面和地面都是光滑的；質點一直與斜面保持接觸。取斜面向右下方傾斜，於是質點向右下方運動，而楔塊被推向左方，如圖所示。

取水平分量向右為正、垂直分量向上為正，

$$\text{質點的速度} = (v\cos\beta,\ -v\sin\beta), \qquad \text{楔塊的速度} = (-V,\ 0).$$



第 (i) 部分

質點相對於楔塊的速度是

$$(v\cos\beta,\ -v\sin\beta)-(-V,\ 0)=(v\cos\beta+V,\ -v\sin\beta).$$

質點在斜面上時，這個相對速度（relative velocity）沿着斜面，方向在水平線以下 α 角（指向右下方）。所以

$$\frac{v\sin\beta}{v\cos\beta+V}=\tan\alpha=\frac{\sin\alpha}{\cos\alpha}.$$

交叉相乘，

$$\begin{aligned} & v\sin\beta\cos\alpha=v\cos\beta\sin\alpha+V\sin\alpha \\ \Longrightarrow\quad & V\sin\alpha=v(\sin\beta\cos\alpha-\cos\beta\sin\alpha)=v\sin(\beta-\alpha), \end{aligned}$$

這正是所要證明的結果。這純粹是運動學（kinematics）：不需要任何力學原理，只需要「質點不離開斜面」這個事實。

第 (ii) 部分

原理：水平方向的線動量守恆（conservation of linear momentum）。把質點和楔塊視為一個系統。作用在系統上的外力是兩者的重量和地面的法向反作用力（normal reaction），由於地面光滑，這些力全都是垂直的。（質點與斜面之間的反作用力是內力。）所以水平總動量是常數，而由於一切都從靜止開始，它起初是零：

$$mv\cos\beta-MV=0,\qquad \text{所以}\qquad V=\frac{m}{M}v\cos\beta.$$

代入 (i)，並展開 $\sin(\beta-\alpha)$，

$$\frac{m}{M}v\cos\beta\sin\alpha=v(\sin\beta\cos\alpha-\cos\beta\sin\alpha).$$

當 $t>0$ 時質點正在運動，所以 $v\neq 0$。另外 $\cos\alpha\neq 0$，而且 $\cos\beta\neq 0$：假如 β 是 90°，由動量可得 $V=0$，於是 (i) 給出 $\sin(\beta-\alpha)=0$，所以 $\beta=\alpha<90^\circ$，矛盾。兩邊除以 $v\cos\alpha\cos\beta$，

$$\frac{m}{M}\tan\alpha=\tan\beta-\tan\alpha\quad\Longrightarrow\quad\tan\beta=\left(1+\frac{m}{M}\right)\tan\alpha.$$

> ? **想一想** 為甚麼 β 不是直接等於 α？

因為斜面本身正在向左退開。相對於楔塊，質點以 α 角沿斜面下滑，但楔塊在後退，所以相對於地面，質點下降得更陡。留意 (ii) 的右邊是一個常數：β 在每一刻都相同，所以質點沿直線運動。如果 $M\to\infty$，楔塊就留在原位，而 $\beta\to\alpha$，正如所料。

第 (iii) 部分

原理：能量守恆（conservation of energy）。所有接觸面都是光滑的，所以對整個系統做功的力就只有重量。（質點與斜面之間的法向反作用力會分別對兩個物體做功，但這兩份功互相抵消，因為相對速度沿着斜面，所以與反作用力垂直。地面的反作用力則與楔塊的運動方向垂直。）楔塊的質心（centre of mass）在垂直方向沒有移動，所以勢能（potential energy）的損失是 mgy。由靜止開始，

$$mgy = \tfrac{1}{2}mv^2 + \tfrac{1}{2}MV^2.$$

利用 (ii) 中的 $V = \dfrac{m}{M}v\cos\beta$，

$$mgy = \tfrac{1}{2}mv^2 + \frac{m^2v^2\cos^2\beta}{2M}.$$

兩邊乘以 $2/m$，

$$2gy = v^2\left(1 + \frac{m\cos^2\beta}{M}\right) = \frac{v^2\left(M + m\cos^2\beta\right)}{M},$$

這正是所要證明的結果。

關於 y 的微分方程

質點下降的速率就是它速度的向下分量：

$$\frac{\mathrm{d}y}{\mathrm{d}t} = v\sin\beta.$$

把 $v = \dot{y}/\sin\beta$ 代入 (iii)（注意由於 $\beta > \alpha > 0$，所以 $\sin\beta \neq 0$）。這就得到微分方程（differential equation）

$$\left(\frac{\mathrm{d}y}{\mathrm{d}t}\right)^2 = k\,y, \qquad \text{其中} \qquad k = \frac{2gM\sin^2\beta}{M + m\cos^2\beta}.$$

關鍵在於 k 是一個常數，因為由 (ii) 可知 β 是常數。

要解這個方程，先對 t 求導：

$$2\dot{y}\,\ddot{y} = k\,\dot{y}.$$

當 $t > 0$ 時質點正在下降，所以 $\dot{y} > 0$，我們可以除以它：$\ddot{y} = \tfrac{1}{2}k$。質點的垂直運動有恆定的加速度

$$\ddot{y} = \frac{gM\sin^2\beta}{M + m\cos^2\beta}.$$

積分兩次，並利用 $t = 0$ 時 $y = 0$ 及 $\dot{y} = 0$（它由靜止開始），

$$y = \frac{gMt^2\sin^2\beta}{2\left(M + m\cos^2\beta\right)},$$

這就是所求的結果。

（另一個做法是取正平方根 $\dot{y} = \sqrt{ky}$，然後分離變數（separation of variables）：$\int y^{-1/2}\,\mathrm{d}y = \sqrt{k}\,t$ 給出 $2\sqrt{y} = \sqrt{k}\,t$，得到相同的答案。先求導的好處，是不必顧慮 $y \equiv 0$ 這個解；它同樣滿足 $\dot{y}^2 = ky$ 和 $y(0) = 0$，卻不是這裏的運動。）

驗算

- 固定的楔塊。令 $M \to \infty$。這時 $\beta \to \alpha$，而 $y \to \frac{1}{2}gt^2\sin^2\alpha$，這是正確的：在固定的光滑斜面上，沿斜面的加速度是 $g\sin\alpha$，而它的垂直分量是 $g\sin^2\alpha$。
- 一個數值例子。取 $M = 2$、$m = 1$、$\alpha = 30°$，由 (ii) 得 $\tan\beta = \frac{3}{2}\cdot\frac{1}{\sqrt{3}} = \frac{\sqrt{3}}{2}$，所以 $\sin^2\beta = \frac{3}{7}$、$\cos^2\beta = \frac{4}{7}$，而公式給出 $y = \frac{1}{6}gt^2$。取 $g = 9.8$ 和 $t = 1$，即 $y \approx 1.6333$。直接對楔塊和質點作逐步模擬（用牛頓定律（Newton's laws），並由接觸條件求出法向反作用力），得到相同的數值；而且在 2000 個隨機例子中，(i)–(iii) 都在捨入誤差範圍內成立。

常見錯誤

- 說質點的運動方向與水平線成 α 角。這只是相對於楔塊而言才成立。
- 沿斜面方向或垂直方向使用動量守恆。地面的反作用力是垂直方向的外力，所以只有水平分量守恆。
- 在能量方程中漏掉楔塊的動能（kinetic energy）$\frac{1}{2}MV^2$。

STEP 之外

被約束在運動曲面上的質點，是拉格朗日力學（Lagrangian mechanics）的經典試金石，你會在大學一年級或二年級學到這門學問。在那裏，你選取兩個坐標（例如楔塊的位置和質點沿斜面下滑的距離），約束力（constraint forces）就會從方程中完全消失，所以只需幾行就能得出那些恆定的加速度。關於帶約束的牛頓力學，以及許多正是這一類的題目，可參閱 Kleppner and Kolenkow, *An Introduction to Mechanics*（一本力學入門教科書）。

詳細解答

STEP II 1998，第 11 題

力學 • 拋體運動

STEP II 1998，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/98-s2-q11

本題在 *Beyond STEP* 第 11.11 節外野手如何接球中有完整的引導式講解：https://beyondstep.keithmcnulty.org/zh-hant/ch11/fielders.html

關鍵思路

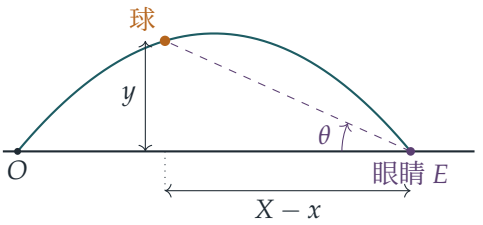
把原點設在擊球點，並取其高度為眼睛的高度，然後把 $\tan\theta$ 寫成「球的高度除以球與外野手眼睛之間的水平距離」。一個由眼睛高度出發、最後又回到眼睛高度的拋體（projectile），其高度可以因式分解為 $\frac{1}{2}g\,t(T-t)$，其中 T 是飛行時間（time of flight）；而如果外野手最後身處正確的位置，水平距離同樣是 $T-t$ 的倍數。因子 $T-t$ 互相約去，餘下的 $\tan\theta$ 便與 t 成正比。

解答

我們把球視為只受重力作用（沒有空氣阻力（air resistance））的質點（particle），並把外野手的眼睛視為一個與擊球點等高的點。取 x 軸為水平方向，由擊球點 O 指向外野手，y 軸鉛垂向上。設球的初速度的水平分量為 $U>0$，鉛垂分量為 $W>0$。那麼在時刻 t

$$x = Ut, \qquad y = Wt - \tfrac{1}{2}gt^2 = \tfrac{1}{2}g\,t\,(T-t), \qquad \text{其中} T = \frac{2W}{g}$$

是球回到眼睛高度的時刻。



如果外野手的眼睛位於水平位置 X（球在飛行期間有 $X > x$），視線（line of sight）便與水平線成仰角 θ，其中

$$\tan\theta = \frac{y}{X-x}.$$

站立不動的外野手

「位置恰到好處」的意思是球剛好飛到她的眼睛：她的眼睛位於球回到眼睛高度的那一點，即 $X = UT$。於是當 $0 < t < T$ 時

$$\tan\theta = \frac{\frac{1}{2}gt(T-t)}{UT-Ut} = \frac{\frac{1}{2}gt(T-t)}{U(T-t)} = \frac{g}{2U}t,$$

這裏約去了 $T-t$，因為球在飛行期間它不等於零。因此

$$\frac{\mathrm{d}}{\mathrm{d}t}(\tan\theta) = \frac{g}{2U},$$

這是一個常數，正是所求。（在接球前的一刻，$\tan\theta \to gT/(2U) = W/U$：她看見球飛來的角度，恰好就是球被擊出時的角度，這與拋物線的對稱性吻合。）

奔跑中的外野手

現在設下一球的速度分量為 U、W（所以同樣有 $y = \frac{1}{2}gt(T-t)$，其中 $T = 2W/g$），並設外野手在球被擊出的一刻與 O 相距 D，然後以恆速 v 跑向擊球手。她的眼睛位於 $X = D - vt$，所以

$$\tan\theta = \frac{\frac{1}{2}gt(T-t)}{D-(U+v)t}. \qquad (*)$$

她能接到球，當且僅當在時刻 T 她的眼睛正好位於球回到眼睛高度的那一點，即

$$D - vT = UT \quad\Longleftrightarrow\quad D = (U+v)T.$$

我們要證明，這種情況發生當且僅當 $\frac{\mathrm{d}}{\mathrm{d}t}(\tan\theta)$ 為常數，因此「選擇 v 使 $\frac{\mathrm{d}}{\mathrm{d}t}(\tan\theta)$ 保持不變」確是正確的法則。

如果她選對了速率，則 $D = (U+v)T$，而 $(*)$ 變成

$$\tan\theta = \frac{\frac{1}{2}gt(T-t)}{(U+v)(T-t)} = \frac{g}{2(U+v)}t, \qquad \text{所以} \qquad \frac{\mathrm{d}}{\mathrm{d}t}(\tan\theta) = \frac{g}{2(U+v)},$$

這是一個常數，與第一部分完全一樣，只是 U 換成了球相對於她的速率 $U+v$。

想一想 這不是只完成了一半嗎？會不會有另一個 v 同樣令變化率保持不變？

不會。假設在整個飛行過程中 $\frac{\mathrm{d}}{\mathrm{d}t}(\tan\theta) = k$ 為常數。在 $t = 0$ 時球位於眼睛高度，所以 $\tan\theta = 0$，因而 $\tan\theta = kt$。把這個結果代入 $(*)$ 並除以 t（其中 $0 < t < T$），得

$$\tfrac{1}{2}g\,(T - t) = k\big(D - (U + v)t\big) \quad \text{對所有} t \in (0, T).$$

兩個關於 t 的線性函數如果在一個區間上相等，它們的係數必定相同，所以

$$\tfrac{1}{2}g = k(U + v) \qquad \text{及} \qquad \tfrac{1}{2}gT = kD.$$

由第一式可知 $k \neq 0$；把第二式除以第一式便得 $D = (U + v)T$，這正是接到球的條件。

所以，能把她帶到球那裏的速率 $v = D/T - U$，恰好就是令 $\frac{\mathrm{d}}{\mathrm{d}t}(\tan\theta)$ 保持不變的那個速率。實際上，球還在空中時她無法計算 $D/T - U$，但她可以觀察：如果球在她眼中上升得愈來愈快，她便站得太後，應該跑快一點；如果球看來上升得愈來愈慢，她便應該放慢腳步。（由於她是*朝着*擊球手跑的，球的落點比她的起始位置近，即 $D > UT$，所以 $v > 0$。）

驗算

- 在第二部分取 $v = 0$，便回到第一部分的結果：$D = UT$，變化率為 $g/(2U)$。
- 單位：g/U 的單位是 $\mathrm{m\,s^{-2}}/\mathrm{m\,s^{-1}} = \mathrm{s^{-1}}$，正好是一個純數的變化率應有的單位。
- 一個數值例子：球以 $20\ \mathrm{m\,s^{-1}}$ 的速率、45° 的仰角被擊出，取 $g = 9.8\ \mathrm{m\,s^{-2}}$，則 $T \approx 2.886\ \mathrm{s}$，射程為 $40.82\ \mathrm{m}$；站在該處的外野手會看見 $\tan\theta$ 以 $g/(2U) \approx 0.346\ \mathrm{s^{-1}}$ 的速率增長。如果她改為從 $49.47\ \mathrm{m}$ 外起步，並以 $3\ \mathrm{m\,s^{-1}}$ 的速率奔跑，變化率則為 $0.286\ \mathrm{s^{-1}}$。在多個時刻對 $(*)$ 作數值求導，可確認這兩個變化率都是常數，而選錯了 v 則會得到一個不斷改變的變化率。

常見錯誤

- 從擊球手而不是從外野手的眼睛量度 θ：$\tan\theta$ 中的水平距離是 $X - x$，而不是 x。
- 只證明了正確的 v 會給出恆定的變化率，卻沒有證明恆定的變化率必然導致正確的 v；正是這個逆命題（converse）才令這條法則真正可用。

STEP 之外

這就是接球心理學中的「光學加速度抵消」（optical acceleration cancellation）概念：外野手似乎是這樣跑的——讓球的影像以穩定的速率上升，而從不需要預測球會落在哪裏。

這是反饋控制律（feedback control law）的一個例子：把一個觀察到的量保持不變，而不是去計算軌跡。關於拋體運動、相對速度（relative velocity）以及這種思考方式的入門，可參閱 Kleppner and Kolenkow, *An Introduction to Mechanics*（力學入門教材）。

詳細解答

STEP II 1998，第 12 題

概率與統計 • 貝葉斯定理

> **STEP II 1998，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/98-s2-q12

本題在 *Beyond STEP* 第 12.2 節基本比率謬誤中有完整的引導式講解：https://beyondstep.keithmcnulty.org/zh-hant/ch12/base-rate.html

關鍵思路

這一題是貝葉斯定理（Bayes' theorem）的兩次應用。第一部分是經典的「罕見疾病」意外：當疾病很罕見時，即使檢測相當準確，大部分陽性結果仍可能是假陽性（false positive）。至於第二部分，竅門在於看出接受 STEP 檢測的人並不是整體人口的隨機樣本。他們是 AL 陽性的那一羣人，所以第一部分的答案就成為第二次檢測的先驗概率（prior）。

解答

設 D 為某人患有 mathematitis 的事件，A 為 AL 檢測呈陽性的事件，B 為 STEP 檢測呈陽性的事件。已知

$$\mathrm{P}(D) = \tfrac{1}{100}, \qquad \mathrm{P}(A \mid D) = \tfrac{9}{10}, \qquad \mathrm{P}(A \mid D') = \tfrac{1}{11}.$$

Frankie 的概率

根據全概率公式（law of total probability），

$$\begin{aligned}\mathrm{P}(A) &= \mathrm{P}(A \mid D)\,\mathrm{P}(D) + \mathrm{P}(A \mid D')\,\mathrm{P}(D') \\ &= \frac{9}{10}\cdot\frac{1}{100} + \frac{1}{11}\cdot\frac{99}{100} = \frac{9}{1000} + \frac{9}{100} = \frac{99}{1000}.\end{aligned}$$

再根據貝葉斯定理，

$$\mathrm{P}(D \mid A) = \frac{\mathrm{P}(A \mid D)\,\mathrm{P}(D)}{\mathrm{P}(A)} = \frac{9/1000}{99/1000} = \frac{1}{11}.$$

所以即使 AL 檢測結果呈陽性，Frankie 是患者的概率也只有 $\frac{1}{11}$。在 1000 人當中，約有 9 名患者檢測呈陽性，但 990 名非患者中也約有 90 人呈陽性，假陽性的數目遠遠蓋過了真陽性。

求 p

現在只考慮 AL 檢測呈陽性的人。在這一羣人之中，

$$\mathrm{P}(D \mid A) = \tfrac{1}{11}, \qquad \mathrm{P}(D' \mid A) = \tfrac{10}{11}.$$

我們把題目所給有關 STEP 的資料理解為 $\mathrm{P}(B \mid D \cap A) = 0.9$ 和 $\mathrm{P}(B \mid D' \cap A) = p$：STEP 檢測的表現只取決於該人是否患者，而與 AL 的結果無關。

> ? **想一想** 為甚麼我們不從 $\mathrm{P}(D) = \frac{1}{100}$ 重新開始？

因為只有 AL 呈陽性的人才會接受 STEP 檢測，而在這一羣人之中，患者所佔的比例已不再是 1%，而是 $\frac{1}{11}$，這正是第一部分求得的結果。第一次檢測的後驗概率（posterior）成為第二次檢測的先驗概率。

在 AL 陽性的一羣人之中，

$$\mathrm{P}(D \cap B \mid A) = \frac{1}{11} \cdot \frac{9}{10} = \frac{9}{110},$$
$$\mathrm{P}(D' \cap B \mid A) = \frac{10}{11} \cdot p = \frac{10p}{11}.$$

題目告訴我們，兩項檢測皆呈陽性的人當中有一半是患者，即 $\mathrm{P}(D \mid A \cap B) = \frac{1}{2}$。這表示上面兩個概率相等：

$$\frac{9}{110} = \frac{10p}{11} \quad \Longrightarrow \quad p = \frac{9}{100} = 0.09.$$

驗算

- 賠率形式（odds form）：患病的先驗賠率為 1 : 99。AL 把它乘以 $\frac{0.9}{1/11} = 9.9$，得 $9.9 : 99 = 1 : 10$，與 $\frac{1}{11}$ 吻合。要使賠率達到 1 : 1，STEP 必須把它乘以 10，所以 $\frac{0.9}{p} = 10$，即 $p = 0.09$。
- 模擬二百萬人先後接受兩項檢測（取 $p = 0.09$），結果 AL 陽性者當中患者所佔的比例為 0.0909，而兩項檢測皆呈陽性者當中患者所佔的比例為 0.5003。

常見錯誤

- 以 $\mathrm{P}(D) = 0.01$ 作為 STEP 檢測的先驗概率，這樣會得出 $p = 0.0009$，比正確答案小了一百倍。
- 把 $\mathrm{P}(A \mid D) = 0.9$ 與 $\mathrm{P}(D \mid A)$ 混淆。第一部分的重點正正在於這兩者相差甚遠。

STEP 之外

把貝葉斯定理寫成賠率形式，即後驗賠率等於先驗賠率乘以似然比（likelihood ratio），連續檢測的計算就變得近乎機械化：每一項獨立的證據只不過是把賠率乘以它自己的因子。這是貝葉斯推論（Bayesian inference）的基礎，也是現實中設計醫學篩查計劃的依據。我們所作的假設，即在給定患病狀況下各項檢測條件獨立（conditionally independent），與「樸素貝葉斯」（naive Bayes）分類器背後的假設相同。Blitzstein 與 Hwang 的 *Introduction to Probability*（概率論入門教科書）對條件概率和貝葉斯法則有極佳的講解，並附有許多與本題類似的例子。

詳細解答

STEP II 1998，第 13 題

概率與統計 • 指數分佈

STEP II 1998，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/98-s2-q13

關鍵思路

第一個結果是指數分佈（exponential distribution）的無記憶性（memoryless property）：如果一次服務已經持續了時間 t，餘下所需的時間與一次全新服務的時間有相同的分佈。有了這個結果，店舖這個情境幾乎不用再作任何計算。就概率而言，服務到一半的顧客與剛剛開始接受服務的顧客並無分別。因此，在每一個關鍵時刻，仍在接受服務的人之間的競賽，都是相同而且獨立（independent）的指數時間之間的競賽。

解答

無記憶性

對於 $x \geq 0$，

$$\mathrm{P}(X > x) = \int_x^\infty \lambda \mathrm{e}^{-\lambda u}\,\mathrm{d}u = \Big[-\mathrm{e}^{-\lambda u}\Big]_x^\infty = \mathrm{e}^{-\lambda x}.$$

取 $s, t \geq 0$。事件 $\{X > s+t\}$ 包含於 $\{X > t\}$ 之內，所以兩者的交集就是 $\{X > s+t\}$。根據條件概率（conditional probability）的定義，

$$\begin{aligned}\mathrm{P}(X > s+t \mid X > t) &= \frac{\mathrm{P}(X > s+t \text{ 及} X > t)}{\mathrm{P}(X > t)} = \frac{\mathrm{P}(X > s+t)}{\mathrm{P}(X > t)} \\ &= \frac{\mathrm{e}^{-\lambda(s+t)}}{\mathrm{e}^{-\lambda t}} = \mathrm{e}^{-\lambda s} = \mathrm{P}(X > s),\end{aligned}$$

這就是所要求的結果。（我們可以作除法，是因為 $\mathrm{P}(X > t) = \mathrm{e}^{-\lambda t} > 0$。）

兩位顧客都仍在接受服務

模型：兩個服務時間互相獨立，而且各自服從題目所給的指數分佈。設 A 和 B 為我到達時正在接受服務的兩位顧客。

> **想一想** 顧客 A 已經接受了一段長度未知的服務。這難道不會改變他還需要多少時間的分佈嗎？

不會，而這正是第一部分告訴我們的。假設 A 的服務至今已持續了 u。那麼它最少再持續 t 的概率是

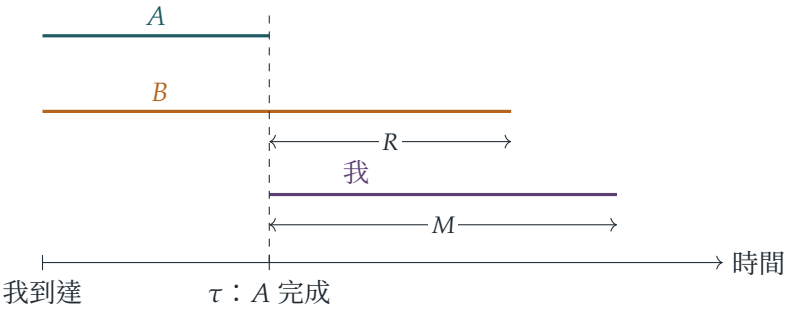
$$\mathrm{P}(X > u + t \mid X > u) = \mathrm{P}(X > t) = \mathrm{e}^{-\lambda t},$$

不論 u 是多少。所以，由我走進店舖的一刻起計，A 的剩餘服務時間服從同一個參數（parameter）為 λ 的指數分佈，B 的情況亦然。由於完整的服務時間互相獨立，這兩個剩餘時間也互相獨立。因此

$$\mathrm{P}(\text{兩人均仍在接受服務，於時刻}t) = \mathrm{e}^{-\lambda t} \cdot \mathrm{e}^{-\lambda t} = \mathrm{e}^{-2\lambda t}.$$

另一位顧客完成時，我是否仍在接受服務？

假設 A 先完成，時刻為我到達後的 τ，而我隨即到 A 的店員那裏接受服務。



在時刻 τ，設 R 為 B 仍需要的時間，並設 M 為我的服務時間。

- M 服從參數為 λ 的指數分佈，因為我的服務在 τ 才重新開始。它與之前發生的一切獨立，因為不同顧客的服務時間互相獨立。
- R 同樣服從參數為 λ 的指數分佈。在時刻 τ，B 已經接受了一段時間的服務（我到達之前的時間再加上 τ），而再次根據無記憶性，剩餘時間服從原來的分佈。它與 M 獨立。

B 完成時我仍在接受服務，當且僅當 $M > R$。現在 M 和 R 互相獨立，並且有相同的連續分佈（continuous distribution），所以根據對稱性（symmetry），$\mathrm{P}(M > R) = \mathrm{P}(R >$

M)。此外，對於連續隨機變量（continuous random variables），$\mathrm{P}(M = R) = 0$，所以這兩個概率之和為 1。因此

$$\mathrm{P}(\text{另一位顧客完成時我仍在接受服務}) = \tfrac{1}{2}.$$

如果你寧願用計算而不用對稱性的論證，可以以 R 為條件：

$$\mathrm{P}(M > R) = \int_0^\infty \lambda \mathrm{e}^{-\lambda r}\,\mathrm{P}(M > r)\,\mathrm{d}r = \int_0^\infty \lambda \mathrm{e}^{-2\lambda r}\,\mathrm{d}r = \tfrac{1}{2}.$$

驗算

- 當 $t = 0$ 時，$\mathrm{e}^{-2\lambda t} = 1$：我到達時，兩位顧客必定都正在接受服務。當 $t \to \infty$ 時，概率趨向 0，這也是應有的結果。
- 答案 $\frac{1}{2}$ 與 λ 無關，這正如我們所料，因為改變 λ 只是改變了時間的單位。
- 以 $\lambda = 0.7$ 和 $t = 0.9$ 進行模擬，得出兩位顧客都仍在接受服務的比例為 0.2831，而 $\mathrm{e}^{-1.26} \approx 0.2837$。在另一個模擬中，兩位顧客已經各自接受了隨機長度的服務，結果我在 0.5000 比例的試驗中仍在接受服務。

常見錯誤

- 以為正在接受服務的顧客「快要完成」，所以他們的剩餘時間會較短。對指數分佈而言，已經發生的事情不會造成任何分別。
- 把 $\mathrm{P}(X > s + t \text{ 及 } X > t)$ 寫成 $\mathrm{P}(X > s + t)\,\mathrm{P}(X > t)$。這兩個事件並不獨立；其中一個包含另一個。

STEP 之外

指數分佈是唯一具有無記憶性的連續分佈。這個事實是*泊松過程*（Poisson process）和*排隊論*（queueing theory）的基礎；在排隊論中，一間有 c 位店員、服務時間服從指數分佈的店舖，就是經典的 M/M/c 隊列。兩位顧客的計算可以推廣至 n 位：n 個參數為 λ 的獨立指數變量之中的最小值，服從參數為 $n\lambda$ 的指數分佈（把尾概率相乘即可），而根據對稱性，每個變量成為最小者的可能性均等。Grimmett 與 Welsh 所著的 *Probability: An Introduction*（概率入門）以這個程度講解指數變量和泊松過程。

詳細解答

STEP II 1998，第 14 題

概率與統計 • 二項分佈與泊松分佈的期望值

STEP II 1998，第 14 題

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

step.maths.org/questions/98-s2-q14

關鍵思路

雖然資料庫把這題另作歸類，但它其實是一道概率（probability）題：只要 Doom 博士在新方案下的期望薪金高於固定薪金 A，她就應該接受新方案。不合格人數服從二項分佈（binomial distribution），所以我們需要求 $\mathrm{E}\big[1/(1+X)\big]$。棘手的因子 $\frac{1}{k+1}$ 可以用恆等式 $\frac{1}{k+1}\binom{N}{k} = \frac{1}{N+1}\binom{N+1}{k+1}$ 吸收掉，於是該和式就變成 $(p+q)^{N+1}$ 幾乎完整的二項展開式。同樣的技巧配合 $\frac{\lambda^k}{(k+1)!}$，便可處理泊松（Poisson）模型。

解答

何時接受新方案

設 X 為 Doom 博士的學生中不合格的人數，並記 $q = 1 - p$。各學生是否不合格互相獨立（independent），每人不合格的概率均為 p，所以 $X \sim \mathrm{B}(N, p)$：

$$\mathrm{P}(X = k) = \binom{N}{k} p^k q^{N-k}, \qquad k = 0, 1, \ldots, N.$$

我們假設 Doom 博士願意以期望薪金來衡量兩個方案。在新方案下，她的薪金是隨機變量（random variable）$S = B/(1+X)$，所以

$$\mathrm{E}(S) = B \sum_{k=0}^{N} \frac{1}{k+1} \binom{N}{k} p^k q^{N-k}.$$

想一想 怎樣消去和式中的 $\frac{1}{k+1}$？

利用階乘（factorial）：

$$\frac{1}{k+1}\binom{N}{k} = \frac{N!}{(k+1)!\,(N-k)!} = \frac{1}{N+1}\cdot\frac{(N+1)!}{(k+1)!\,\big((N+1)-(k+1)\big)!}$$
$$= \frac{1}{N+1}\binom{N+1}{k+1}.$$

把這個結果代入，並同時乘以和除以 p，使 p 的冪次與新的二項式係數（binomial coefficient）相配，

$$\begin{aligned}\mathrm{E}(S) &= \frac{B}{(N+1)p}\sum_{k=0}^{N}\binom{N+1}{k+1}p^{k+1}q^{N-k}\\ &= \frac{B}{(N+1)p}\sum_{j=1}^{N+1}\binom{N+1}{j}p^{j}q^{N+1-j}\\ &= \frac{B}{(N+1)p}\Big[(p+q)^{N+1}-q^{N+1}\Big]\end{aligned}$$

其中先設 $j = k+1$，再運用二項式定理（binomial theorem），因為該和式正是 $(p+q)^{N+1}$ 的完整展開式，只欠 $j=0$ 的一項 q^{N+1}。由於 $p+q=1$，

$$\mathrm{E}(S) = \frac{B\big(1-(1-p)^{N+1}\big)}{(N+1)p}.$$

若 $\mathrm{E}(S) > A$，她就應該接受。兩邊同乘 $(N+1)p$，而此數為正（公式要有意義，須有 $p>0$；若 $p=0$，她只需直接比較 B 與 A），便得

$$A(N+1)p < B\big(1-(1-p)^{N+1}\big),$$

這就是所求的條件。

泊松模型

當 N 很大而 p 很小，且 Np 大小適中時，二項分佈 $\mathrm{B}(N,p)$ 可以用平均值為 $\lambda = Np$ 的泊松分佈（Poisson distribution）作良好的近似。因此，如果 Doom 博士任教的班人數眾多，每名學生不合格的機會都很低，而且各人是否不合格仍然互相獨立，她的學生的不合格人數就可以用泊松分佈作模型。（更一般地說，若不合格的情況逐一出現、互相獨立，且平均發生率恆定，便適合採用泊松模型。）

設 $X \sim \mathrm{Po}(\lambda)$，$\lambda = Np$，同樣的想法依然可行，這次用的是 $\frac{1}{k+1}\cdot\frac{1}{k!} = \frac{1}{(k+1)!}$：

$$\mathrm{E}\Big(\frac{B}{1+X}\Big) = B\sum_{k=0}^{\infty}\frac{1}{k+1}\cdot\frac{\mathrm{e}^{-\lambda}\lambda^{k}}{k!} = \frac{B\mathrm{e}^{-\lambda}}{\lambda}\sum_{k=0}^{\infty}\frac{\lambda^{k+1}}{(k+1)!}$$
$$= \frac{B\mathrm{e}^{-\lambda}}{\lambda}\big(\mathrm{e}^{\lambda}-1\big) = \frac{B\big(1-\mathrm{e}^{-\lambda}\big)}{\lambda},$$

因為最後的和式正是指數級數（exponential series），只欠首項 1。所以在這個模型下，Doom 博士的期望薪金為

$$\frac{B\left(1-\mathrm{e}^{-Np}\right)}{Np} \text{ 英鎊.}$$

驗算

- $N=1$：直接計算得 $\mathrm{E}(S)=B\left(q+\frac{1}{2}p\right)=B\left(1-\frac{1}{2}p\right)$，而公式給出 $B\,\frac{1-q^2}{2p}=B\,\frac{1+q}{2}=B\left(1-\frac{1}{2}p\right)$。
- $p\to 0$：二項分佈的表達式趨向 B（沒有人不合格），而當 $\lambda\to 0$ 時，泊松分佈的表達式亦然。
- 在近似理應準確的情況下，兩個答案相當吻合：取 $N=1000$，$p=0.003$，由二項分佈公式得 $\mathrm{E}(S)/B$ 為 0.31655，由泊松公式則得 0.31674。取 $N=20$，$p=0.3$ 時，逐項精確計算與模擬的結果均為 $\mathrm{E}(S)/B\approx 0.1586$。

常見錯誤

- 寫成 $\mathrm{E}\left(\frac{1}{1+X}\right)=\frac{1}{1+\mathrm{E}(X)}=\frac{1}{1+Np}$。期望值（expectation）不能直接穿過非線性函數；事實上 $\frac{1}{1+x}$ 是凸函數（convex function），所以根據延森不等式（Jensen's inequality），真正的值比這個大。
- 忘記了重新編排指標後的和式欠了 $j=0$ 的一項，結果得出 $\mathrm{E}(S)=B/((N+1)p)$。

STEP 之外

這裏的技巧其實是一個特例：計算 $\mathrm{E}\big(1/(1+X)\big)$ 時可以利用 $\frac{1}{1+x}=\int_0^1 t^x\,\mathrm{d}t$，從而把它與概率生成函數（probability generating function）$G(t)=\mathrm{E}(t^X)$ 聯繫起來：答案就是 $\int_0^1 G(t)\,\mathrm{d}t$。試以 $G(t)=(q+pt)^N$ 和 $G(t)=\mathrm{e}^{\lambda(t-1)}$ 驗證一下。關於生成函數，以及二項分佈收斂（convergence）至泊松分佈，Grimmett 與 Welsh 所著的 *Probability: An Introduction*（概率論入門教材）有很好的講解。

詳細解答

STEP III 1998，第 1 題

純數學 • 曲線描繪

> **STEP III 1998，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/98-s3-q1

關鍵思路

把所有式子都用 $\cos x$ 表示：利用 $\sin^2 x = 1 - \cos^2 x$，函數 f 便成為 $\cos x$ 的二次式。配方（completing the square）之後，f 的值域（range）一目了然，而最後一部分正是由這個值域決定。至於 g，用商法則（quotient rule）求得的導數（derivative）是 $\mathrm{f}'(x)$ 的常數倍再除以一個平方。

解答

描繪 $y = \mathrm{f}(x)$ 的圖像

利用 $\sin^2 x = 1 - \cos^2 x$，

$$\mathrm{f}(x) = 2 + 2\cos x - \cos^2 x = 3 - (1 - \cos x)^2.$$

由於 $0 \le 1 - \cos x \le 2$，可得 $-1 \le \mathrm{f}(x) \le 3$。當 $\cos x = 1$ 時取得值 3，當 $\cos x = -1$ 時取得值 -1。

求導得

$$\mathrm{f}'(x) = -2(1 - \cos x)\sin x = 2\sin x\,(\cos x - 1).$$

當 $\sin x = 0$ 或 $\cos x = 1$ 時，上式為零；在 $[0, 2\pi]$ 上，即 $x = 0,\ \pi,\ 2\pi$。駐點（stationary point）為

$$(0,\ 3), \qquad (\pi,\ -1), \qquad (2\pi,\ 3).$$

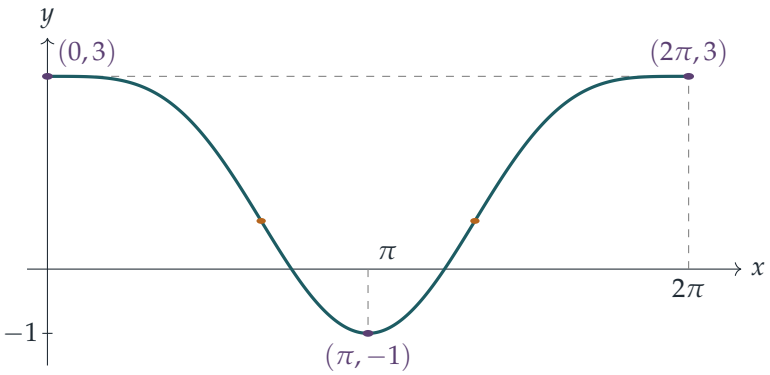
在 $(0, \pi)$ 上，$\sin x > 0$ 而 $\cos x - 1 < 0$，所以 $\mathrm{f}' < 0$，f 遞減。在 $(\pi, 2\pi)$ 上，$\sin x < 0$ 而 $\cos x - 1$ 仍然是負數，所以 $\mathrm{f}' > 0$，f 遞增。因此 $(\pi, -1)$ 是極小點（minimum），而端點 $(0, 3)$ 和 $(2\pi, 3)$ 是極大點（maximum）。

想一想 曲線在 $x = 0$ 附近是甚麼樣子？像拋物線嗎？

不像，它比拋物線更平坦。在 $x = 0$ 附近，$1 - \cos x \approx \frac{1}{2}x^2$，所以 $\mathrm{f}(x) \approx 3 - \frac{1}{4}x^4$。曲線在頂部非常平坦，這與 $\mathrm{f}''(0) = 0$ 吻合。在 $x = \pi$ 附近，設 $x = \pi + t$，則 $1 - \cos x = 1 + \cos t \approx 2 - \frac{1}{2}t^2$，所以 $\mathrm{f} \approx -1 + 2t^2$：這是一個普通的拋物線形極小點。

描繪圖像時還有以下幾個有用的特徵：

- 曲線關於 $x = \pi$ 對稱，因為 $\cos(2\pi - x) = \cos x$。
- 曲線與 x 軸相交於 $(1 - \cos x)^2 = 3$ 之處，即 $\cos x = 1 - \sqrt{3}$。由此得 $x = \arccos(1 - \sqrt{3}) \approx 2.39$ 和 $x = 2\pi - \arccos(1 - \sqrt{3}) \approx 3.89$。
- $\mathrm{f}''(x) = 2\cos 2x - 2\cos x = 2(2\cos x + 1)(\cos x - 1)$。它在 $\cos x = -\frac{1}{2}$ 處變號，所以拐點（point of inflection）位於 $\left(\frac{2\pi}{3}, \frac{3}{4}\right)$ 和 $\left(\frac{4\pi}{3}, \frac{3}{4}\right)$。



琥珀色的點是拐點。

g 的駐點

根據商法則，凡是 $c\,\mathrm{f}(x) + d \neq 0$ 之處，

$$\mathrm{g}'(x) = \frac{a\mathrm{f}'(x)\big(c\mathrm{f}(x) + d\big) - c\mathrm{f}'(x)\big(a\mathrm{f}(x) + b\big)}{\big(c\mathrm{f}(x) + d\big)^2} = \frac{(ad - bc)\,\mathrm{f}'(x)}{\big(c\mathrm{f}(x) + d\big)^2}.$$

由於 $ad \neq bc$，在 g 有定義的每一點上，$\mathrm{g}'(x) = 0$ 當且僅當 $\mathrm{f}'(x) = 0$。所以除了 f 的駐點之外，g 沒有其他駐點。

我們還需要 g 在 $x = 0, \pi, 2\pi$ 處有定義，另外兩個條件正是為此而設。在 $x = 0$ 和 $x = 2\pi$ 處，分母是 $3c + d$，因為 $d \neq -3c$，所以它不等於零。在 $x = \pi$ 處，分母是 $d - c$，因為 $d \neq c$，所以它不等於零。因此 $y = \mathrm{g}(x)$ 的駐點恰好位於 $x = 0,\ \pi,\ 2\pi$，正

如題目所求，並且

$$\mathrm{g}(0) = \mathrm{g}(2\pi) = \frac{3a+b}{3c+d}, \qquad \mathrm{g}(\pi) = \frac{b-a}{d-c}.$$

$|\mathrm{g}|$ 何時有界

假設 $d/c < -3$ 或 $d/c > 1$。特別地，$c \neq 0$，所以可以寫成

$$c\,\mathrm{f}(x) + d = c\left(\mathrm{f}(x) + \frac{d}{c}\right).$$

由於 $\mathrm{f}(x)$ 在 $[-1,3]$ 內取值，括號內的式子落在 $\left[\frac{d}{c} - 1,\ \frac{d}{c} + 3\right]$ 之內。

- 若 $d/c > 1$，括號內的式子至少是 $\frac{d}{c} - 1 > 0$。
- 若 $d/c < -3$，括號內的式子至多是 $\frac{d}{c} + 3 < 0$。

無論哪種情況，都存在常數 $k > 0$，使得對所有 x 都有 $\left|c\,\mathrm{f}(x) + d\right| \geq |c|\,k$。換句話說，分母始終與零保持一段距離。分子是有界（bounded）的，因為 $|a\mathrm{f}(x) + b| \leq 3|a| + |b|$。因此對 $[0, 2\pi]$ 內的每個 x，都有

$$|\mathrm{g}(x)| \leq \frac{3|a| + |b|}{|c|\,k}$$

所以 $|\mathrm{g}(x)|$ 不可能任意大。

> **想一想** 如果 $-3 < d/c < 1$，會出甚麼問題？

這時 $-d/c$ 嚴格介乎 -1 與 3 之間。由於 f 連續（continuous），並且取遍 $[-1,3]$ 內的每一個值，所以存在某個 x 使得 $\mathrm{f}(x) = -d/c$，分母在該處等於零。分子在同一點卻不等於零：假如兩者同時為零，便有 $a(-d/c) + b = 0$，即 $ad = bc$。所以 g 有一條鉛垂漸近線（vertical asymptote），$|\mathrm{g}|$ 無界。由此可見，最後一部分的條件恰到好處，不多也不少。

驗算

- 以數值計算，在 $[0, 2\pi]$ 上取一個細密的網格，f 的最小值是 -1，最大值是 3。此外 $3 - \mathrm{f}(0.1) \approx 2.496 \times 10^{-5}$，而 $\frac{1}{4}(0.1)^4 = 2.5 \times 10^{-5}$，印證了曲線在 $x = 0$ 處的四次平坦性。
- 一個特殊情況：取 $a = 1$、$b = 0$、$c = 0$、$d = 1$，則 $\mathrm{g} = \mathrm{f}$，而公式給出 $\mathrm{g}(0) = 3$ 和 $\mathrm{g}(\pi) = -1$，與預期相符。
- 隨機選取 a, b, c, d，以有限差分（finite difference）計算 g 的導數，結果與 $(ad - bc)\mathrm{f}'/(c\mathrm{f} + d)^2$ 一致。

常見錯誤

- 遺漏端點 $x = 0$ 和 $x = 2\pi$ 這兩個駐點。f′ 在這兩處確實等於零。
- 忘記檢查 g 在 f 的駐點處是否有定義。條件 $d \neq -3c$ 和 $d \neq c$ 正是為此而設。
- 在 $x = 0$ 處把頂部畫成尖削的拋物線形狀。曲線在那裏其實平坦得多。

STEP 之外

當 $ad \neq bc$ 時，映射 $t \mapsto \dfrac{at + b}{ct + d}$ 稱為莫比烏斯變換（*Möbius transformation*）。它是可逆（invertible）的，而且它的導數 $(ad - bc)/(ct + d)^2$ 永不為零，所以與它複合既不會產生新的駐點，也不會令原有的駐點消失。莫比烏斯映射在複分析（complex analysis）和幾何中佔有核心地位，並且在複合運算下構成一個羣（group）。想看一本以圖像為主、賞心悅目的介紹，可參閱 Needham 的 *Visual Complex Analysis*（《複分析：可視化方法》）。

詳細解答

STEP III 1998，第 2 題

純數學 • Beta 積分的遞推關係

> **STEP III 1998，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/98-s3-q2

關鍵思路

這個積分其實是喬裝了的貝塔函數（beta function）。每一部分都只需一個標準手法。第 (i) 部分用反射 $t \mapsto 1-t$，把 t 與 $1-t$ 的角色對調。第 (ii) 部分把被積函數（integrand）乘以 $1 = t + (1-t)$。第 (iii) 部分用分部積分法（integration by parts），把 $(1-t)$ 的一個冪次轉移到 t 上。重複使用第 (iii) 部分，直至 $(1-t)$ 的冪次變成零，剩下的積分便可立即算出。

解答

第 (i) 部分

作代換（substitution）$u = 1-t$，則 $\mathrm{d}t = -\mathrm{d}u$，而積分限 $t = 0, 1$ 變成 $u = 1, 0$：

$$\mathrm{I}(a,b) = \int_1^0 (1-u)^a u^b \,(-\mathrm{d}u) = \int_0^1 u^b (1-u)^a \,\mathrm{d}u = \mathrm{I}(b,a).$$

$\mathrm{d}t = -\mathrm{d}u$ 帶來的負號，剛好在把積分限換回 0 和 1 時抵銷。

第 (ii) 部分

由於 $t + (1-t) = 1$，被積函數可以寫成

$$t^a(1-t)^b = t^a(1-t)^b\big(t + (1-t)\big) = t^{a+1}(1-t)^b + t^a(1-t)^{b+1}.$$

把每一項在 $[0,1]$ 上積分，便得 $\mathrm{I}(a,b) = \mathrm{I}(a+1,b) + \mathrm{I}(a,b+1)$，正是所求。

第 (iii) 部分

我們對 $\mathrm{I}(a,b)$ 作分部積分，對 $(1-t)^b$ 求導，對 t^a 積分：

$$\begin{aligned}\mathrm{I}(a,b) &= \left[\frac{t^{a+1}}{a+1}(1-t)^b\right]_0^1 - \int_0^1 \frac{t^{a+1}}{a+1}\cdot\left(-b(1-t)^{b-1}\right)\mathrm{d}t \\ &= \left[\frac{t^{a+1}}{a+1}(1-t)^b\right]_0^1 + \frac{b}{a+1}\int_0^1 t^{a+1}(1-t)^{b-1}\,\mathrm{d}t.\end{aligned}$$

? **想一想** 為甚麼邊界項（boundary term）會消失？哪裏需要用到 a 和 b 是正數？

在 $t=1$ 處，因為 $b>0$，因子 $(1-t)^b$ 等於 0。在 $t=0$ 處，因為 $a+1>0$，因子 t^{a+1} 等於 0。所以方括號內的項為零，兩邊同乘以 $a+1$ 便得

$$(a+1)\,\mathrm{I}(a,b) = b\,\mathrm{I}(a+1,b-1),$$

這就是所求的結果。（若 $0<b<1$，最後那個積分中 $(1-t)$ 的冪次是負數。但由於 $b-1>-1$，它仍然收斂（converge）。至於下面的整數情形，反正總有 $b-1\geq 0$。）

對整數 a,b 計算 $\mathrm{I}(a,b)$。把結果寫成 $\mathrm{I}(a,b)=\dfrac{b}{a+1}\mathrm{I}(a+1,b-1)$，然後反覆使用。每一步都令第一個參數增加 1，第二個參數減少 1。當第二個參數降至 0 時便停止：

$$\begin{aligned}\mathrm{I}(a,b) &= \frac{b}{a+1}\mathrm{I}(a+1,b-1) = \frac{b}{a+1}\cdot\frac{b-1}{a+2}\mathrm{I}(a+2,b-2) = \cdots \\ &= \frac{b(b-1)\cdots 1}{(a+1)(a+2)\cdots(a+b)}\mathrm{I}(a+b,0).\end{aligned}$$

每一步都是合理的，因為每次使用公式時，第二個參數都是正整數。現在 $\mathrm{I}(a+b,0)=\int_0^1 t^{a+b}\,\mathrm{d}t=\dfrac{1}{a+b+1}$。另外 $(a+1)(a+2)\cdots(a+b)=\dfrac{(a+b)!}{a!}$，所以

$$\boxed{\mathrm{I}(a,b)=\frac{a!\,b!}{(a+b+1)!}}$$

這條公式對 a 和 b 是對稱的，與第 (i) 部分的要求一致。當 a 或 b 等於 0 時，公式同樣成立。

驗算

- $\mathrm{I}(1,1)=\int_0^1(t-t^2)\,\mathrm{d}t=\frac{1}{2}-\frac{1}{3}=\frac{1}{6}$，而公式給出 $\frac{1!\,1!}{3!}=\frac{1}{6}$。
- 公式滿足第 (ii) 部分：$\frac{(a+1)!\,b!}{(a+b+2)!}+\frac{a!\,(b+1)!}{(a+b+2)!}=\frac{a!\,b!\,(a+b+2)}{(a+b+2)!}=\frac{a!\,b!}{(a+b+1)!}$。
- 數值方面，用辛普森法則（Simpson's rule）得 $\mathrm{I}(2,3)\approx 0.0166667$，與 $\frac{2!\,3!}{6!}=\frac{1}{60}$ 吻合。我們亦隨機取非整數的 a,b，以數值積分（quadrature）驗算了第 (i) 至 (iii) 部分，結果吻合至約 10^{-11}。

常見錯誤

- 分部積分時把方向弄反。對 t^a 求導會令 $(1-t)$ 的冪次升高，方向錯了。
- 忘記說明邊界項為何消失。它之所以為零，全因 $a+1>0$ 及 $b>0$。
- 遞推（recursion）早了一步停下，或者漏掉來自 $\mathrm{I}(a+b,0)$ 的最後一個因子 $\frac{1}{a+b+1}$。

STEP 之外

$\mathrm{I}(a,b)$ 就是貝塔函數 $B(a+1,b+1)$。對所有正實數參數，它與伽瑪函數（gamma function）有以下關係：$B(x,y)=\Gamma(x)\Gamma(y)/\Gamma(x+y)$。我們的答案正是 $\Gamma(n+1)=n!$ 這個特殊情形。在概率論中，$t^a(1-t)^b/\mathrm{I}(a,b)$ 是貝塔分佈（beta distribution）的密度函數（density），而貝塔分佈在貝葉斯統計學（Bayesian statistics）中佔有核心地位。Blitzstein 與 Hwang 合著的 *Introduction to Probability*（概率論入門）對此有很精彩的論述，當中包括一個計數論證（counting argument），解釋了為甚麼答案是 $a!\,b!/(a+b+1)!$。

詳細解答

STEP III 1998，第 3 題

純數學 • 線性遞推關係

STEP III 1998，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/98-s3-q3

關鍵思路

這是一個帶常數項的一階線性遞推關係（first-order linear recurrence），可說是微分方程（differential equation）$y' = cy - d$ 的離散版本。處理手法與常微分方程（ODE）相同：先找出常數解（即「平衡」(equilibrium) 解），然後留意到與平衡解之間的偏差每天只是乘以同一個因子。至於極限，真正要花工夫的只有一件事：弄清楚 $\big((1+c)^N - 1\big)/c$ 在 $c \to 0$ 時的變化，而二項式定理（binomial theorem）正好解決這個問題。

解答

$c > 0$ 時求解

我們嘗試對每個 T 設 $V_T = Ak^T + B$。代入遞推關係，得

$$Ak^{N+1} + B = (1+c)\big(Ak^N + B\big) - d \quad \Longleftrightarrow \quad Ak^N\big(k - (1+c)\big) = cB - d.$$

右邊與 N 無關，而左邊除非恆為零，否則必隨之變化（我們要求 $A \neq 0$ 及 $k \neq 0$，否則這個形式只是一個常數）。因此對每個 N，兩邊都必須為零：

$$k = 1 + c, \qquad B = \frac{d}{c},$$

這裏可以除以 c，因為 $c > 0$。取這些值後，無論常數 A 是甚麼，$V_T = A(1+c)^T + d/c$ 都滿足遞推關係。代入 $T = 0$ 得 $V_0 = A + d/c$，所以 $A = V_0 - d/c$，於是

$$\boxed{V_N = \Big(V_0 - \frac{d}{c}\Big)(1+c)^N + \frac{d}{c}}$$

想一想 我們猜了一個形式，結果行得通。但怎樣知道沒有其他解呢？

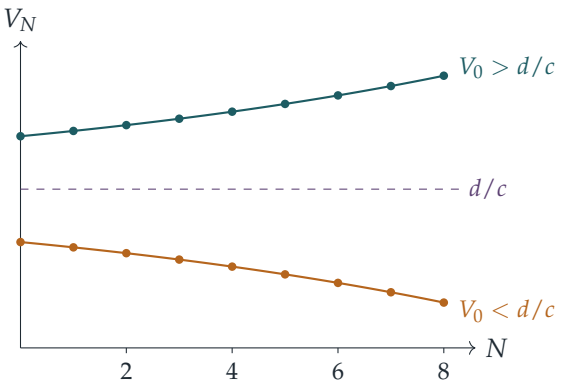
遞推關係由 V_0 決定 V_1，再由 V_1 決定 V_2，如此類推，所以給定 V_0 後，恰好只有一個數列符合條件。我們的公式滿足遞推關係，而且 V_0 的值也正確，所以它就是那個數列。（如果你想要一個直接的論證，可設 $U_N = V_N - d/c$；遞推關係變成 $U_{N+1} = (1+c)U_N$，所以 $U_N = (1+c)^N U_0$，答案相同。）

把答案改寫成以下形式，稍後會很有用：

$$V_N = V_0(1+c)^N - d\,\frac{(1+c)^N - 1}{c}. \qquad (*)$$

這個式子有一個很好的解讀：第一項是 V_0 以每天 c 的比率增長，第二項是 $d\big(1 + (1+c) + \cdots + (1+c)^{N-1}\big)$，即每天的扣減額各自按餘下的日數增長後，以等比級數（geometric series）求和的結果。

常數 d/c 是平衡值。若 $V_0 > d/c$，債券的價值會無止境地增長；若 $V_0 = d/c$，價值維持不變；若 $V_0 < d/c$，價值會愈跌愈快：



（圖中取 $c = 0.1$ 及 $d/c = 5$，並分別取 $V_0 = 6$ 和 $V_0 = 4$。）

$c = 0$ 的情形

這時遞推關係是 $V_{N+1} = V_N - d$：價值每天減少 d，是一個等差數列（arithmetic progression）。所以

$$\boxed{V_N = V_0 - Nd}$$

留意 $c > 0$ 時的公式不能直接代入 $c = 0$ 求值，因為式中有 d/c；這正是題目要問極限的原因。

$c \to 0$ 時的極限

固定 N，並使用 $(*)$。第一項趨向 V_0，因為對固定的 N 有 $(1+c)^N \to 1$。至於第二項，用二項式定理展開：

$$(1+c)^N - 1 = Nc + \binom{N}{2}c^2 + \cdots + c^N,$$

$$\text{所以} \quad \frac{(1+c)^N - 1}{c} = N + \binom{N}{2}c + \cdots + c^{N-1}.$$

由於 N 是固定的，這是一個關於 c 的多項式，只有有限多項，而且第一項之後的每一項都含有 c 的正數次冪。因此當 $c \to 0$ 時它趨向 N，於是

$$\lim_{c\to 0} V_N = V_0 - Nd,$$

這正是 $c = 0$ 時的解，符合題目要求。（等價地說，$\big((1+c)^N - 1\big)/c$ 是 x^N 在 $x = 1$ 處的差商（difference quotient），所以它的極限就是導數 N。）

「對固定的 N」這幾個字很重要：若 c 很小但 N 很大，像 $\binom{N}{2}c$ 這樣的項不一定小，而指數增長最終會勝過線性變化。

驗算

- $N = 1$：公式給出 $(V_0 - d/c)(1+c) + d/c = (1+c)V_0 - d$，這就是遞推關係本身。
- $V_0 = d/c$ 時，對所有 N 都有 $V_N = d/c$，即平衡狀態。
- 數值上，取 $V_0 = 7$、$d = 2$、$N = 12$：公式在 $c = 0.01$ 時給出 $V_{12} \approx -17.48$，在 $c = 10^{-4}$ 時給出 -17.005，逐漸接近 $V_0 - 12d = -17$。在許多隨機選取的有理數例子中，這個閉合公式也與直接迭代的結果完全吻合。

常見錯誤

- 取 $B = -d/c$：正負號來自 $cB - d = 0$，所以 $B = +d/c$。
- 用 $A = V_0$ 而不是 $A = V_0 - d/c$：常數 B 在 $T = 0$ 時同樣有貢獻。
- 試圖在 $c > 0$ 的公式中直接代入 $c = 0$，但該式此時沒有定義；必須取極限。

STEP 之外

這類遞推關係稱為*線性差分方程*（linear difference equations），其理論與線性常微分方程並行：通解（general solution）等於一個特解（particular solution，這裏是常數解）加

上齊次方程（homogeneous equation）$U_{N+1} = (1+c)U_N$ 的通解。在金融學中，這就是定額還款的貸款或年金（annuity）的公式，而 $c \to 0$ 的極限則是由複利變化過渡到單利變化。Boyce 與 DiPrima 的 *Elementary Differential Equations*（初等微分方程教科書）在常微分方程理論之外，另有一節講述一階差分方程。

詳細解答

STEP III 1998，第 4 題

純數學 • 極坐標玫瑰線

> **STEP III 1998，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/98-s3-q4

關鍵思路

要理解極坐標曲線（polar curve），方程簡單時最容易的做法是把它化為直角坐標（Cartesian）形式；方程不簡單時，則追蹤 r 的正負號。整道題的關鍵在於以下約定：$r<0$ 的點要畫在相反方向 $\theta+\pi$ 上、與原點距離 $|r|$ 之處。對於 $r=\cos k\theta$，這個約定決定了負的迴圈是落在新的位置（形成新的花瓣），還是疊在已畫好的花瓣之上。

解答

圓 $r=\cos\theta$

兩邊同乘 r，並利用 $x=r\cos\theta$、$y=r\sin\theta$、$r^2=x^2+y^2$：

$$\begin{aligned} r^2=r\cos\theta \quad &\Longrightarrow \quad x^2+y^2=x \\ &\Longleftrightarrow \quad \left(x-\tfrac{1}{2}\right)^2+y^2=\tfrac{1}{4}. \end{aligned}$$

因此曲線上每一點都在圓心為 $(\frac{1}{2},0)$、半徑為 $\frac{1}{2}$ 的圓上。

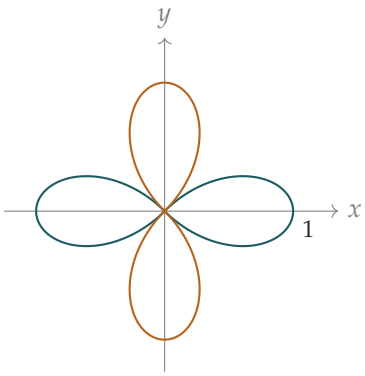
反過來，取這個圓上原點以外的任意一點 (x,y)。由於 $x=x^2+y^2>0$，它的極角滿足 $-\frac{1}{2}\pi<\theta<\frac{1}{2}\pi$，而把 $r^2=x=r\cos\theta$ 除以 $r>0$ 便得 $r=\cos\theta$。至於原點，則在 $\theta=\pm\frac{1}{2}\pi$ 時到達。所以這條曲線正正就是這個圓，當 θ 由 $-\frac{1}{2}\pi$ 走到 $\frac{1}{2}\pi$ 時恰好描畫一次；在這個範圍內始終有 $r=\cos\theta\geq 0$，故此毋須用到正負號的約定。

曲線 $r=\cos 2\theta$

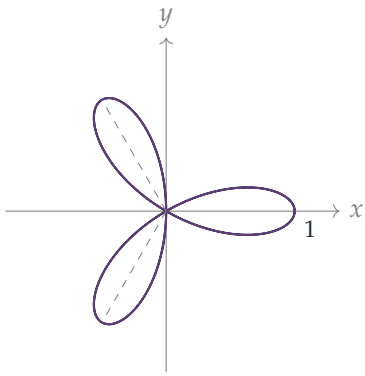
現在 r 會變號，而當 $r<0$ 時，我們把 (r,θ) 畫在方向 $\theta+\pi$ 上、距離為 $|r|$ 之處。

- 當 $-\frac{1}{4}\pi < \theta < \frac{1}{4}\pi$ 時，$r > 0$，由 0 上升至 $\theta = 0$ 時的 1，再回落到 0：得出一片沿正 x 軸的花瓣。
- 當 $\frac{1}{4}\pi < \theta < \frac{3}{4}\pi$ 時，$r < 0$，所以這些點畫在 $\frac{5}{4}\pi$ 至 $\frac{7}{4}\pi$ 的方向上：得出一片沿負 y 軸的花瓣。
- 同理，$\frac{3}{4}\pi < \theta < \frac{5}{4}\pi$（$r > 0$）給出一片沿負 x 軸的花瓣，而 $\frac{5}{4}\pi < \theta < \frac{7}{4}\pi$（$r < 0$）給出一片沿正 y 軸的花瓣。

結果是一條四葉玫瑰線（four-leaved rose），每片花瓣長度為 1，關於兩條坐標軸都對稱，而且花瓣之間只在原點相接。在草圖中，藍綠色的花瓣來自 $r > 0$，琥珀色的花瓣來自 $r < 0$。



$r = \cos 2\theta$



$r = \cos 3\theta$

曲線 $r = \cos 2n\theta$

由於 $\cos(-2n\theta) = \cos 2n\theta$，$n$ 與 $-n$ 給出同一條曲線，所以取 $n \geq 1$（$n = 0$ 的情況稍後再談）。

想一想 當 $r < 0$ 時，那些迴圈是疊在已畫好的迴圈之上，還是落在新的位置？

在以 $\theta = k\pi/n = 2k \cdot \frac{\pi}{2n}$（$k = 0, 1, \ldots, 2n-1$）為中心、長度為 $\pi/(2n)$ 的各個區間上，我們有 $r > 0$。每個區間給出一片指向 $2k \cdot \frac{\pi}{2n}$ 方向的花瓣，這個方向是 $\frac{\pi}{2n}$ 的偶數倍。

在以 $\theta = (2k+1)\frac{\pi}{2n}$ 為中心的各個區間上，我們有 $r < 0$，而這些點畫在方向 $\theta + \pi$ 上，所以花瓣指向

$$(2k+1)\frac{\pi}{2n} + \pi = (2k+1+2n)\frac{\pi}{2n},$$

即 $\frac{\pi}{2n}$ 的奇數倍。因此負的迴圈落在新的位置，也就是正花瓣之間的空隙。

所以 $r = \cos 2n\theta$ 是一條有 $4n$ 片全等（congruent）花瓣的玫瑰線，每片花瓣長度為 1，分別指向 $m\pi/(2n)$（$m = 0, 1, \ldots, 4n-1$）的方向，每片花瓣佔的角為 $\pi/(2n)$，而相鄰

的花瓣只在原點相接。（當 $n = 1$ 時，這就是上面的四葉玫瑰線。）各花瓣所佔的角合共為 $4n \cdot \frac{\pi}{2n} = 2\pi$，所以它們剛好填滿一整個周角，互不重疊。

至於退化情況 $n = 0$，曲線是 $r = 1$，即單位圓。

面積與 n 無關

由於花瓣互不重疊，當 θ 由 0 走到 2π 時，所圍區域的每一部分都恰好被掃過一次，於是面積公式 $\frac{1}{2}\int r^2\,\mathrm{d}\theta$（它只涉及 r^2，所以 r 的正負號並無影響）給出：當 $n \geq 1$ 時，

$$\begin{aligned} A &= \frac{1}{2}\int_0^{2\pi} \cos^2 2n\theta\,\mathrm{d}\theta = \frac{1}{4}\int_0^{2\pi} (1+\cos 4n\theta)\,\mathrm{d}\theta \\ &= \frac{1}{4}\left[\theta + \frac{\sin 4n\theta}{4n}\right]_0^{2\pi} = \frac{\pi}{2}, \end{aligned}$$

因為 $\sin 8n\pi = 0$。這個結果與 n 無關，正如題目所求：花瓣愈多，每片花瓣就愈窄，兩者恰好互相抵銷。（被排除的情況 $n = 0$，即單位圓，面積為 π；公式中 $\sin 4n\theta/(4n)$ 一項正是用到 $n \neq 0$ 的地方。）

曲線 $r = \cos 3\theta$

> **想一想** 為甚麼這條曲線有三片花瓣，而不是六片？

現在 $r > 0$ 的區間以 $\theta = 0, \frac{2}{3}\pi, \frac{4}{3}\pi$ 為中心，給出指向這些方向的花瓣。負的迴圈以 $\theta = \frac{1}{3}\pi, \pi, \frac{5}{3}\pi$ 為中心，並畫在方向 $\frac{4}{3}\pi$、2π 和 $\frac{8}{3}\pi \equiv \frac{2}{3}\pi$ 上：正正就是正花瓣的方向。事實上，把 θ 換成 $\theta + \pi$，便得到相反方向上的

$$r = \cos(3\theta + 3\pi) = -\cos 3\theta$$

，那是同一點。所以 $r = \cos 3\theta$ 是一條三葉玫瑰線，花瓣長度為 1，沿 $\theta = 0, \frac{2}{3}\pi, \frac{4}{3}\pi$ 的方向，每片佔的角為 $\frac{1}{3}\pi$，而當 θ 由 0 走到 2π 時，整條曲線被描畫了兩次（$0 \leq \theta \leq \pi$ 時描畫一次）。這就是上面右邊的草圖。

同樣的論證說明：對任何奇數 k，$r = \cos k\theta$ 是一條描畫兩次的 k 葉玫瑰線；而對偶數 k，它是一條描畫一次的 $2k$ 葉玫瑰線。

驗算

- 在 $r = \cos\theta$ 中代入 $\theta = 0$：點 $(1, 0)$ 在 $(x - \frac{1}{2})^2 + y^2 = \frac{1}{4}$ 上，原點（在 $\theta = \pm\frac{1}{2}\pi$ 時）也在其上。
- 圓 $r = \cos\theta$ 的面積為 $\pi(\frac{1}{2})^2 = \frac{1}{4}\pi$，而 $\frac{1}{2}\int_{-\pi/2}^{\pi/2} \cos^2\theta\,\mathrm{d}\theta = \frac{1}{4}\pi$ 與此吻合。同樣，

對於三葉玫瑰線，$\frac{1}{2}\int_0^{\pi}\cos^2 3\theta\,\mathrm{d}\theta = \frac{1}{4}\pi$。若在這裏用 0 至 2π，便會重複計算，得出 $\frac{1}{2}\pi$。

- 數值驗證：對 $n = 1,\ldots,5$ 繪畫 $r = \cos 2n\theta$，可找到 $4n$ 個花瓣尖端，而以蒙地卡羅方法（Monte Carlo）估算所圍面積，每次都得出 $1.57 \approx \pi/2 = 1.5708$。

常見錯誤

- 忽略負的 r，結果 $r = \cos 2\theta$ 只畫出兩片花瓣。
- 以為 $r = \cos 3\theta$ 有六片花瓣，理由是 r 有六個迴圈：其實負的迴圈疊在正的迴圈之上。
- 對描畫兩次的曲線（例如 $r = \cos 3\theta$）使用 $\frac{1}{2}\int_0^{2\pi} r^2\,\mathrm{d}\theta$，令面積變成兩倍。

STEP 之外

這些曲線稱為 *rhodonea*，即玫瑰線（rose curve）。把一點寫成複數（complex number）$z = r\,\mathrm{e}^{\mathrm{i}\theta}$，玫瑰線 $r = \cos k\theta$ 就是 $z = \frac{1}{2}\left(\mathrm{e}^{\mathrm{i}(k+1)\theta} + \mathrm{e}^{-\mathrm{i}(k-1)\theta}\right)$，即兩個方向相反的圓周運動之和。這個觀點把玫瑰線與本輪（epicycle）及傅立葉級數（Fourier series）聯繫起來，而面積公式 $\frac{1}{2}\int r^2\,\mathrm{d}\theta$ 則是向量微積分（vector calculus）中格林定理（Green's theorem）的特殊情況。Needham 的 *Visual Complex Analysis*（《複分析：可視化方法》）把這種以幾何眼光看待複數的方法發揮得淋漓盡致。

詳細解答

STEP III 1998，第 5 題

純數學 • 矩陣指數

STEP III 1998，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/98-s3-q5

關鍵思路

矩陣（matrix）的無窮級數（infinite series）看來嚇人，但這裏兩個矩陣的冪都按非常簡單的規律重複：$\mathbf{M}$ 的冪像 i 的冪一樣循環，而 $\mathbf{N}$ 的冪由平方起全部為零。一旦知道每一個冪，級數便會分拆成熟悉的純量級數（或者乾脆終止），於是可以直接讀出相應的變換（transformation）。最後一部分則是直接的矩陣計算，而它的幾何意義，來自追問每個變換會令哪些直線保持不變。

解答

矩陣 $\mathbf{M}$

乘開，得

$$\mathbf{M}^2 = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} = -\mathbf{I},$$

正如所求。因此，對每個整數 $k \geq 0$，

$$\mathbf{M}^{2k} = (\mathbf{M}^2)^k = (-1)^k\mathbf{I}, \qquad \mathbf{M}^{2k+1} = (-1)^k\mathbf{M}.$$

把級數按 $\theta\mathbf{M}$ 的偶次冪和奇次冪分開，

$$\begin{aligned} \exp(\theta\mathbf{M}) &= \sum_{k=0}^{\infty} \frac{\theta^{2k}}{(2k)!}\mathbf{M}^{2k} + \sum_{k=0}^{\infty} \frac{\theta^{2k+1}}{(2k+1)!}\mathbf{M}^{2k+1} \\ &= \Big(\sum_{k=0}^{\infty} \frac{(-1)^k\theta^{2k}}{(2k)!}\Big)\mathbf{I} + \Big(\sum_{k=0}^{\infty} \frac{(-1)^k\theta^{2k+1}}{(2k+1)!}\Big)\mathbf{M} \\ &= \cos\theta\,\mathbf{I} + \sin\theta\,\mathbf{M} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}. \end{aligned}$$

（這樣分拆是合理的，因為矩陣級數的每個元素都是一個實數級數，其各項的大小以 $|\theta|^r/r!$ 為界，所以它絕對收斂（converges absolutely），可以重新組合。）

想一想 我在哪裏見過 $\cos\theta\,\mathbf{I}+\sin\theta\,\mathbf{M}$？

由於 $\mathbf{M}^2 = -\mathbf{I}$，矩陣 $\mathbf{M}$ 的表現與 i 完全一樣，而這個計算正是 $e^{i\theta} = \cos\theta + i\sin\theta$ 的矩陣版本。從幾何上看，$\exp(\theta\mathbf{M})$ 把 $(1,0)$ 送到 $(\cos\theta,\sin\theta)$，把 $(0,1)$ 送到 $(-\sin\theta,\cos\theta)$：它是**平面繞原點逆時針旋轉（rotation）角** θ。（事實上 $\mathbf{M}$ 本身就是轉角為 $\frac{\pi}{2}$ 的旋轉，正如乘以 i 一樣。）

矩陣 N

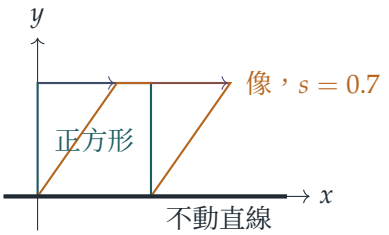
這裏

$$\mathbf{N}^2 = \begin{pmatrix}0&1\\0&0\end{pmatrix}\begin{pmatrix}0&1\\0&0\end{pmatrix} = \begin{pmatrix}0&0\\0&0\end{pmatrix},$$

所以對所有 $r \geq 2$ 都有 $\mathbf{N}^r = \mathbf{0}$，級數在兩項之後便終止：

$$\exp(s\mathbf{N}) = \mathbf{I} + s\mathbf{N} = \begin{pmatrix}1&s\\0&1\end{pmatrix}.$$

它把 (x,y) 送到 $(x+sy,\,y)$。每一點都平行於 x 軸移動，移動的距離與它的高度 y 成正比；x 軸上的點則完全不動。所以 $\exp(s\mathbf{N})$ 是一個**平行於 x 軸的錯切（shear）**，以 x 軸為其不動點（fixed point）所成的直線，並把單位正方形送到一個面積相同的平行四邊形（行列式（determinant）為 1）。



它們何時可交換？

記 $c = \cos\theta$ 及 $\sigma = \sin\theta$。則

$$\exp(s\mathbf{N})\exp(\theta\mathbf{M}) = \begin{pmatrix}1&s\\0&1\end{pmatrix}\begin{pmatrix}c&-\sigma\\\sigma&c\end{pmatrix} = \begin{pmatrix}c+s\sigma&-\sigma+sc\\\sigma&c\end{pmatrix},$$

$$\exp(\theta\mathbf{M})\exp(s\mathbf{N}) = \begin{pmatrix}c&-\sigma\\\sigma&c\end{pmatrix}\begin{pmatrix}1&s\\0&1\end{pmatrix} = \begin{pmatrix}c&sc-\sigma\\\sigma&s\sigma+c\end{pmatrix}.$$

兩式相減，

$$\exp(s\mathbf{N})\exp(\theta\mathbf{M}) - \exp(\theta\mathbf{M})\exp(s\mathbf{N}) = s\sin\theta\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}.$$

這個矩陣對每個 s 都是零矩陣，當且僅當 $\sin\theta = 0$（若 $\sin\theta \neq 0$，取 $s = 1$ 即可）。所以這兩個矩陣對所有 s 都可交換（commute），恰好是在

$$\theta = n\pi, \qquad n \in \mathbb{Z}.$$

的時候。對這些值，$\exp(\theta\mathbf{M}) = (-1)^n\mathbf{I}$，即不是單位矩陣（identity matrix）就是 $-\mathbf{I}$。

想一想 為甚麼與這個錯切可交換的旋轉，必定是半周旋轉或者根本不轉？

設 R 為該旋轉，S 為該錯切（其中 $s \neq 0$），並假設 $SR = RS$。在 x 軸上任取一點 P，則 $S(P) = P$。於是

$$S\big(R(P)\big) = R\big(S(P)\big) = R(P),$$

所以 $R(P)$ 也是錯切的不動點。但錯切所固定的點只有 x 軸上的點。所以這個旋轉必須把 x 軸映射到它自身，而繞原點的旋轉之中能做到這一點的，只有轉角為 0（恆等變換）或 π（半周旋轉 $(x, y) \mapsto (-x, -y)$）的兩個。反過來，這兩個變換與每一個線性映射（linear map）都可交換，因為它們就是 $\pm\mathbf{I}$：把平面轉半周和對它作錯切，不論先後次序，結果都一樣。任何其他旋轉都會令錯切的不動直線傾斜，所以先錯切後旋轉與先旋轉後錯切會得出不同的結果。

驗算

- 當 $\theta = 0$ 時旋轉矩陣是 $\mathbf{I}$，當 $s = 0$ 時錯切是 $\mathbf{I}$，與級數的要求一致。
- 兩個行列式都是 1（分別為 $\cos^2\theta + \sin^2\theta$ 和 $1 \cdot 1$），旋轉和錯切都保持面積不變，理應如此。
- 數值上，取 $\theta = 1$ 並把級數的首 80 項相加，所得的元素與 $\cos 1 \approx 0.5403$ 和 $\sin 1 \approx 0.8415$ 吻合。取 $\theta = 1$、$s = 1$ 時，交換子（commutator）的對角線元素為 $\pm\sin 1 \approx \pm 0.8415$，與 $s\sin\theta\,\mathrm{diag}(1, -1)$ 吻合。

常見錯誤

- 以為對矩陣也有 $\exp(\mathbf{A})\exp(\mathbf{B}) = \exp(\mathbf{A} + \mathbf{B})$。這條法則需要 $\mathbf{AB} = \mathbf{BA}$，而這正是此處不成立的條件，除非 $\sin\theta = 0$。
- 只給出 $\theta = 0$（或 $\theta = 0, \pi$），而沒有給出 π 的所有整數倍。

- 把 $\exp(s\mathbf{N})$ 稱為拉伸（stretch）或放大（enlargement）：它保持面積不變，並且逐點固定 x 軸，這正是它成為錯切的原因。

STEP 之外

矩陣指數（matrix exponential）是*李代數*（*Lie algebra*）與*李羣*（*Lie group*）之間的橋樑：$\mathbf{M}$ 生成旋轉羣，$\mathbf{N}$ 生成錯切羣，而 $\exp(t\mathbf{A})$ 是線性微分方程（linear differential equation）$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{x} = \mathbf{A}\mathbf{x}$ 的解。對於不可交換的矩陣，$\exp(\mathbf{A})\exp(\mathbf{B}) = \exp(\mathbf{A}+\mathbf{B})$ 不成立，其偏差的程度由交換子 $\mathbf{AB}-\mathbf{BA}$ 來量度，這個想法你在最後一部分已經遇過。Strang 的 *Introduction to Linear Algebra*（線性代數入門）涵蓋矩陣指數及其在微分方程中的應用。

詳細解答

STEP III 1998，第 6 題

純數學 • 多面體的體積

> **STEP III 1998，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/98-s3-q6

關鍵思路

把正方體放在坐標軸上，使它的頂點（vertex）就是坐標全為 0 或 a 的那些點。這樣，距離可由三維的畢氏定理（Pythagoras' theorem）求得，而那些不易處理的立體，則可以從正方體切去簡單的小塊來量度。正四面體（regular tetrahedron）就是正方體切去四個角後剩下的部分，而正八面體（regular octahedron）則是兩個底對底拼合的正方形底角錐（pyramid）。

解答

第 (i) 部分

取正方體 $[0,a]^3$。它的頂點是八個點 (x,y,z)，其中每個坐標為 0 或 a；兩個頂點相鄰（adjacent，即由一條稜（edge）相連）當且僅當它們恰好有一個坐標不同。

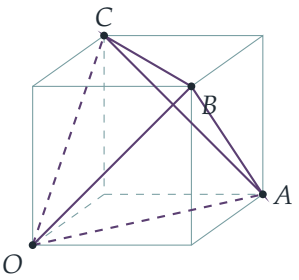
可以選哪四個頂點？按頂點的坐標中有多少個等於 a，把頂點稱為偶頂點或奇頂點。沿一條稜移動會改變一個坐標，所以必定令奇偶互換。假設我們的四個頂點中包括一個偶頂點 P 和一個奇頂點 Q。它們並不相鄰，所以它們有三個坐標不同：Q 就是與 P 相對的頂點。其餘每個頂點都與 P 或 Q 相鄰（三個奇頂點與 P 相鄰，三個偶頂點與 Q 相鄰），所以無法再加入第三個頂點。因此這四個頂點全是偶頂點，

$$O=(0,0,0),\quad A=(a,a,0),\quad B=(a,0,a),\quad C=(0,a,a),$$

或者全是奇頂點（即這組頂點對平面 $x=\frac{1}{2}a$ 的反射，所得的立體與原來的全等（congruent））。

正則性。O,A,B,C 中任意兩點都恰好有兩個坐標不同，每個相差 a，所以六個距離全都

是 $\sqrt{a^2+a^2}=\sqrt{2}\,a$。因此四個面都是邊長為 $\sqrt{2}\,a$ 的等邊三角形（equilateral triangle），而 $OABC$ 是一個正四面體。它的每條稜都是正方體某一個面的對角線。



體積。從正方體中移走這個四面體，剩下四個角塊，每個奇頂點處各有一個。例如，位於 $(a,0,0)$ 的角塊，就是以該頂點及其三個相鄰頂點 O、A、B 為頂點的四面體。它在 $(a,0,0)$ 處的三條稜互相垂直，長度均為 a，所以它是一個角錐，底是面積為 $\frac{1}{2}a^2$ 的直角三角形，高為 a，體積為 $\frac{1}{3}\cdot\frac{1}{2}a^2\cdot a=\frac{1}{6}a^3$。因此

$$V_{\text{tet}}=a^3-4\cdot\tfrac{1}{6}a^3=\tfrac{1}{3}a^3.$$

要令稜長為單位長度，我們需要 $\sqrt{2}\,a=1$，即 $a=1/\sqrt{2}$ 及 $a^3=1/(2\sqrt{2})$。所以稜長為單位長度的正四面體的體積是

$$\frac{1}{3}\cdot\frac{1}{2\sqrt{2}}=\frac{1}{6\sqrt{2}}=\frac{\sqrt{2}}{12}.$$

想一想 為甚麼四個角塊和這個四面體恰好填滿正方體，而且互不重疊？

四面體的每個面，例如 OAB，都恰好切去正方體的一個頂點，這裏是 $(a,0,0)$：通過 O、A、B 的平面 $x-y-z=0$ 把 $(a,0,0)$ 嚴格地分在一側，而其餘七個頂點則在另一側或在平面上。所以正方體分成四個角塊（每個面之外各有一個）和位於中間的四面體。

第 (ii) 部分

稜長為單位長度的正八面體，可以把六個頂點放在

$$(\pm h,0,0),\quad (0,\pm h,0),\quad (0,0,\pm h),\qquad h=\tfrac{1}{\sqrt{2}},$$

其中每個頂點都與不和它相對的四個頂點相連。兩個這樣的頂點，例如 $(h,0,0)$ 和 $(0,h,0)$，相距 $\sqrt{2}\,h=1$。所以十二條稜的長度全是 1，而八個面（每個卦限（octant）各有一個）都是等邊三角形。位於平面 $z=0$ 上的四個頂點組成一個邊長為 1 的正方形。這個八面體就是以這個正方形為底的兩個正方形底角錐，頂點為 $(0,0,\pm h)$，高各為 h。因此

$$V_{\text{oct}}=2\cdot\tfrac{1}{3}\cdot 1^2\cdot\frac{1}{\sqrt{2}}=\frac{\sqrt{2}}{3}.$$

更一般地，把長度放大 e 倍，體積就放大 e^3 倍，所以稜長為 e 的正八面體的體積是 $\frac{\sqrt{2}}{3}e^3$。

第 (iii) 部分

現在把邊長為 a 的正方體的中心放在原點會較為簡潔，即取 $[-\frac{1}{2}a, \frac{1}{2}a]^3$。它各個面的中心是

$$(\pm\tfrac{1}{2}a, 0, 0), \quad (0, \pm\tfrac{1}{2}a, 0), \quad (0, 0, \pm\tfrac{1}{2}a).$$

這正是第 (ii) 部分的佈局，其中 $h = \frac{1}{2}a$。相對兩面的中心相距 a。相鄰兩面的中心相距 $\sqrt{(\frac{1}{2}a)^2 + (\frac{1}{2}a)^2} = a/\sqrt{2}$。以這六點為頂點的凸（convex）立體有八個面，每個卦限各有一個：例如，在卦限 $x, y, z \geq 0$ 內的面是以 $(\frac{1}{2}a, 0, 0)$、$(0, \frac{1}{2}a, 0)$、$(0, 0, \frac{1}{2}a)$ 為頂點的三角形。每個面都是邊長為 $a/\sqrt{2}$ 的等邊三角形，所以各面的中心組成一個稜長為 $e = a/\sqrt{2}$ 的正八面體。

根據第 (ii) 部分，它的體積是

$$\frac{\sqrt{2}}{3}\left(\frac{a}{\sqrt{2}}\right)^3 = \frac{\sqrt{2}}{3} \cdot \frac{a^3}{2\sqrt{2}} = \frac{a^3}{6}.$$

第 (i) 部分的四面體，其頂點是同一個正方體的其中四個頂點，體積為 $\frac{1}{3}a^3$。由於 $\frac{1}{6}a^3 = \frac{1}{2} \cdot \frac{1}{3}a^3$，八面體的體積是四面體體積的一半，正如題目所求。

驗算

- 利用「底面積乘高」的公式，以底面積 $\frac{\sqrt{3}}{4}$ 和高 $\sqrt{2/3}$ 計算，得出單位正四面體的體積為 $\frac{1}{3} \cdot \frac{\sqrt{3}}{4} \cdot \sqrt{\frac{2}{3}} = \frac{\sqrt{2}}{12} \approx 0.11785$，與第 (i) 部分吻合。
- 稜長相等時，$V_{\text{oct}} = 4V_{\text{tet}}$，因為 $\frac{\sqrt{2}}{3} = 4 \cdot \frac{\sqrt{2}}{12}$。這與一個經典的分割相符：把稜長為 2 的四面體分割成四個稜長為 1 的四面體（在它的四個角）和一個稜長為 1 的八面體（在中間）；以單位四面體為量度單位，就是 $8 = 4 + 4$。
- 數值上，對所有四個互不相鄰的正方體頂點的選法作行列式（determinant）計算，得出單位正方體的情況下體積為 $\frac{1}{3}$；而對區域 $|x| + |y| + |z| \leq 1/\sqrt{2}$ 作蒙地卡羅（Monte Carlo）估算，得出 0.4715，對照 $\sqrt{2}/3 \approx 0.47140$。

常見錯誤

- 在第 (i) 部分把正方體的邊長取為 1：這時四面體的稜長是 $\sqrt{2}$ 而不是 1，它的體積須按 $(1/\sqrt{2})^3$ 的比例重新縮放。

- 減去錯誤的角塊：角塊共有四個（每個沒有用到的頂點處各有一個），每個體積為 $\frac{1}{6}a^3$，而不是八個。
- 在第 (iii) 部分，把正方體的邊長當作八面體的稜長。相鄰兩面的中心相距 $a/\sqrt{2}$。

STEP 之外

正方體內的兩個四面體，以及由各面中心組成的八面體，是柏拉圖立體（Platonic solids）及其對稱羣（symmetry group）這個故事的開端：正方體的旋轉之中，保持其中一個內接四面體不變的那些旋轉組成一個階為 12 的羣，是正方體全部 24 個旋轉的一半。正方體與八面體互為對偶（dual）（其中一個的各面中心就是另一個的頂點），而這類立體的體積可以由行列式簡潔地求出，因為行列式量度的正是線性映射（linear map）把體積縮放多少。Strang 的 *Introduction to Linear Algebra*（線性代數入門）把行列式視為體積來處理，而 Armstrong 的 *Groups and Symmetry*（羣與對稱）則研究這些立體的對稱羣。

詳細解答

STEP III 1998，第 7 題

純數學 • 曲線描繪與積分

> **STEP III 1998，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/98-s3-q7

關鍵思路

第一個函數並非隨意安排的熱身題：把 $\mathrm{b}(x)$ 求導並令導數為零時，所得的方程正好是 $\mathrm{f}(x/T)=0$。因此，我們對 f 的一切認識（只有一個正根，且位於 2 與 3 之間）都告訴我們 b 在何處轉向。至於 B 如何依賴 T，則來自自然的代換（substitution）$x = Tu$，它把 T 從被積函數（integrand）中抽走。

解答

繪畫 f 的草圖並求 m

我們有 $\mathrm{f}(0) = 1\cdot(-3)+3 = 0$，而且

$$\mathrm{f}'(s) = \mathrm{e}^{s}(s-3)+\mathrm{e}^{s} = \mathrm{e}^{s}(s-2).$$

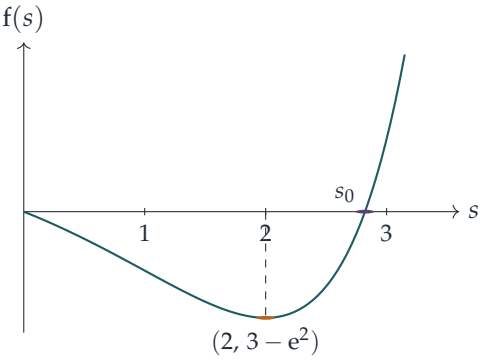
由於 $\mathrm{e}^{s}>0$，f 在 $[0,2]$ 上遞減，在 $[2,\infty)$ 上遞增，並有極小值 $\mathrm{f}(2) = 3-\mathrm{e}^{2} \approx 3-7.29 = -4.29$。當 $s\to\infty$ 時，$\mathrm{e}^{s}(s-3)\to\infty$，所以 $\mathrm{f}(s)\to\infty$。

因為 $\mathrm{f}(0)=0$ 且 f 在 $[0,2]$ 上嚴格遞減（strictly decreasing），所以當 $0<s\le 2$ 時 $\mathrm{f}(s)<0$。在 $[2,\infty)$ 上，它由一個負值嚴格遞增至 $+\infty$，故此在該處恰好穿過零一次，設在某個 $s_0>2$。因此 s_0 是 f 唯一的正根。

取 $\mathrm{e}\approx 2.7$：

$$\mathrm{f}(1) = 3-2\mathrm{e} \approx -2.4, \qquad \mathrm{f}(2)\approx -4.29, \qquad \mathrm{f}(3) = 0+3 = 3.$$

所以滿足 $\mathrm{f}(m)>0$ 的最小正整數是 $m=3$，而且 $2<s_0<3$（事實上 $s_0\approx 2.82$）。



b 的唯一轉向點

設 $u = x/T$。根據商法則（quotient rule），

$$\mathrm{b}'(x) = \frac{3x^2(\mathrm{e}^{x/T}-1) - x^3\cdot\frac{1}{T}\mathrm{e}^{x/T}}{(\mathrm{e}^{x/T}-1)^2} = \frac{x^2\left(3\mathrm{e}^{u}-3-u\,\mathrm{e}^{u}\right)}{(\mathrm{e}^{u}-1)^2} = -\frac{x^2\,\mathrm{f}(u)}{(\mathrm{e}^{u}-1)^2}.$$

當 $x > 0$ 時，我們有 $x^2 > 0$ 及 $(\mathrm{e}^u - 1)^2 > 0$，所以 $\mathrm{b}'(x) = 0$ 當且僅當 $\mathrm{f}(x/T) = 0$。由第一部分可知，這只會在 $x/T = s_0$ 時發生，所以 b 在 $0 < x < \infty$ 內只有一個轉向點（turning point），位於

$$x = s_0 T \approx 2.82\,T.$$

此外，當 $0 < u < s_0$ 時 $\mathrm{f}(u) < 0$，當 $u > s_0$ 時 $\mathrm{f}(u) > 0$，所以在這一點之前 $\mathrm{b}' > 0$，之後 $\mathrm{b}' < 0$：它是一個極大點。

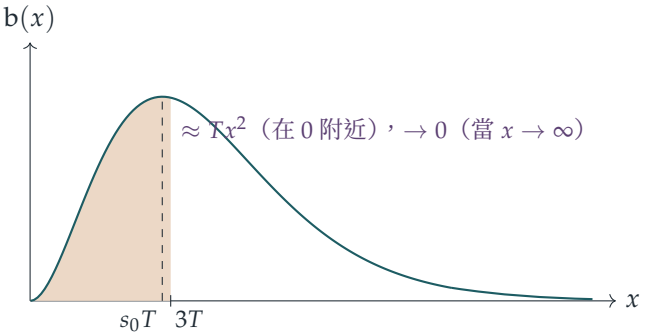
? **想一想** 在 $x = 0$ 附近，分子和分母同時趨於零，b 的圖像是甚麼樣子的？

當 x 很小時，$\mathrm{e}^{x/T} - 1 \approx x/T$，所以

$$\mathrm{b}(x) \approx \frac{x^3}{x/T} = Tx^2,$$

圖像離開原點時形如拋物線，斜率為零。（所以 $\mathrm{b}(0) = 0$ 是自然的取值。）當 x 很大時，指數函數壓倒任何冪函數：$\mathrm{b}(x) \approx x^3\mathrm{e}^{-x/T} \to 0$，所以 x 軸是漸近線（asymptote）。由於對所有 $x > 0$ 都有 $\mathrm{b}(x) > 0$，圖像由原點上升至位於 $x = s_0 T$ 的單一個峯，然後向橫軸衰減。

利用第一部分，峯的高度是 $\mathrm{b}(s_0T) = T^3 s_0^3/(\mathrm{e}^{s_0}-1) \approx 1.42\,T^3$，而 $\mathrm{b}(3T) = 27T^3/(\mathrm{e}^3 - 1)$ 只是略小一點。



（草圖以 $T = 1$ 繪畫；陰影區域是最後一部分所用的區域。）

B 的形式

代入 $x = Tu$，則 $\mathrm{d}x = T\,\mathrm{d}u$，而由於 $T > 0$，積分上下限 $0, \infty$ 不變：

$$B = \int_0^\infty \frac{T^3u^3}{\mathrm{e}^u - 1}\,T\,\mathrm{d}u = T^4\int_0^\infty \frac{u^3}{\mathrm{e}^u - 1}\,\mathrm{d}u.$$

最後這個積分不含 T，所以它是一個常數 K（我們可假設它是有限的）。因此 $B = KT^4$，其中 $n = 4$。

估算 K

用同一個代換，並取 $m = 3$，

$$KT^4 = B \approx 2\int_0^{3T} \mathrm{b}(x)\,\mathrm{d}x = 2T^4\int_0^3 \frac{u^3}{\mathrm{e}^u - 1}\,\mathrm{d}u, \qquad \text{所以} \qquad K \approx 2\int_0^3 \frac{u^3}{\mathrm{e}^u - 1}\,\mathrm{d}u.$$

這是草圖中陰影面積的兩倍（取 $T = 1$）。從圖像讀出高度，並取 $\mathrm{e} \approx 2.7$：

$$\mathrm{b}(0) = 0, \quad \mathrm{b}(1) \approx \frac{1}{1.7} \approx 0.59, \quad \mathrm{b}(2) \approx \frac{8}{6.29} \approx 1.27, \quad \mathrm{b}(3) \approx \frac{27}{18.7} \approx 1.45.$$

以闊度為 1 的長條使用梯形法則（trapezium rule），得

$$\int_0^3 \approx \tfrac{1}{2}(0) + 0.59 + 1.27 + \tfrac{1}{2}(1.45) \approx 2.58,$$

所以

$$K \approx 5.$$

更粗略的估算是把陰影區域當作底為 3、高約 1.45 的三角形，得面積約 2.2，即 $K \approx 4.3$。無論用哪種方法，K 大約是 4 至 5。（三角形的估算必定偏低，因為曲線位於由原點至 $(3, \mathrm{b}(3))$ 的直線之上。）

驗算

- 以數值二分法（bisection）求得 f 的根為 $s_0 \approx 2.8214$；取 $T = 1.7$ 對 b′ 作數值掃描，恰好找到一次變號，位於 $x \approx 4.797 \approx 1.7 s_0$。
- 用辛普森法則（Simpson's rule）計算，當 $T = 0.5$、1 及 2.3 時都得 $B/T^4 \approx 6.4939$：每次的值相同，印證了 $n = 4$。這與已知的精確值 $K = \pi^4/15 \approx 6.494$ 吻合。
- 積分 $\int_0^3 u^3/(\mathrm{e}^u - 1)\,\mathrm{d}u \approx 2.552$，所以題目容許的近似給出 $K \approx 5.10$。我們估算的 5 大致正確。與 6.49 之間的差距，是題目的近似 $B \approx 2\int_0^{3T}$ 本身的誤差，而不是我們計算上的誤差。

常見錯誤

- 忘記 $\mathrm{f}(0) = 0$：$s = 0$ 是 f 的一個根，但它對應 $x = 0$，不在 $0 < x < \infty$ 之內，所以轉向點仍然只有一個。
- 在 B 中作變數代換時，忘記由 $\mathrm{d}x = T\,\mathrm{d}u$ 而來的因子 T，結果得出 $n = 3$。

STEP 之外

函數 b 就是黑體輻射（black-body radiation）的普朗克定律（Planck's law），其中 x 擔當頻率的角色，T 則是溫度。峯的位置 $x = s_0 T$ 是維恩位移定律（Wien's displacement law），而 $B \propto T^4$ 是斯特藩–玻爾茲曼定律（Stefan–Boltzmann law）。精確值 $K = \pi^4/15$ 來自展開式 $1/(\mathrm{e}^u - 1) = \sum_{k\ge 1} \mathrm{e}^{-ku}$，由此得 $K = 6\sum_{k\ge 1} 1/k^4 = 6\zeta(4)$。論證求和與積分可以這樣互換，是典型的大學一年級分析（analysis）題目；可參閱 Abbott 的 *Understanding Analysis*。

詳細解答

STEP III 1998，第 8 題

純數學 • 球面的向量幾何

STEP III 1998，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/98-s3-q8

關鍵思路

這裏的一切都歸結為同一個做法：把「點 $\mathbf{r}$ 位於球面上」寫成純量方程（scalar equation）$|\mathbf{r}-\mathbf{c}|^2 = a^2$，用點積（dot product）展開，然後讀出所需的結果。對於 (i) 中的直線，這樣會得到一個關於 λ 的二次方程，交點的數目由其判別式（discriminant）決定。對於 (ii)，我們需要球心到平面的距離；而對於 (iii)，把兩個球面方程相減，得到的正是「垂直」所問及的那個點積。

解答

第 (i) 部分

點 $\mathbf{b}+\lambda\mathbf{m}$ 位於球面上，當且僅當 $(\mathbf{b}+\lambda\mathbf{m})\cdot(\mathbf{b}+\lambda\mathbf{m}) = a^2$。展開，並利用 $\mathbf{m}\cdot\mathbf{m} = 1$（因為 $\mathbf{m}$ 是單位向量（unit vector））得

$$\lambda^2 + 2(\mathbf{b}\cdot\mathbf{m})\lambda + \mathbf{b}\cdot\mathbf{b} - a^2 = 0.$$

每一個實根 λ 給出一個交點，而不同的 λ 值給出不同的點（因為 $\mathbf{m} \neq \mathbf{0}$）。所以，有兩個交點當且僅當這個二次方程有兩個相異實根，即其判別式為正：

$$4(\mathbf{b}\cdot\mathbf{m})^2 - 4(\mathbf{b}\cdot\mathbf{b} - a^2) > 0 \iff a^2 > \mathbf{b}\cdot\mathbf{b} - (\mathbf{b}\cdot\mathbf{m})^2,$$

這就是所要求的條件。

當判別式為零時，恰好有一個交點：

$$a^2 = \mathbf{b}\cdot\mathbf{b} - (\mathbf{b}\cdot\mathbf{m})^2.$$

這時重根為 $\lambda = -\mathbf{b}\cdot\mathbf{m}$（即 λ 的係數的一半再取負號），所以該點為

$$\mathbf{p} = \mathbf{b} - (\mathbf{b}\cdot\mathbf{m})\,\mathbf{m}, \qquad \text{而} \qquad \mathbf{m}\cdot\mathbf{p} = \mathbf{b}\cdot\mathbf{m} - (\mathbf{b}\cdot\mathbf{m})(\mathbf{m}\cdot\mathbf{m}) = 0,$$

這裏再次用到 $\mathbf{m}\cdot\mathbf{m} = 1$。所以切線（tangent line）垂直於通往切點的半徑，這正是應有的結果。

第 (ii) 部分

垂直於 $\mathbf{n}$ 的平面，其方程為 $\mathbf{r}\cdot\mathbf{n} = k$，其中 k 為常數；由於 $\mathbf{n}$ 是單位向量，該平面與原點的距離為 $|k|$。選取 $\mathbf{n}$ 的正負號使 $k \geq 0$（即 $\mathbf{n}$ 由原點指向平面），則平面為

$$\mathbf{r}\cdot\mathbf{n} = l.$$

想一想 球心 $\mathbf{d}$ 與這個平面相距多遠？

由 $\mathbf{d}$ 出發沿法線（normal）方向走：點 $\mathbf{d} + \mu\mathbf{n}$ 位於平面上，當 $\mathbf{d}\cdot\mathbf{n} + \mu = l$，所以 $\mu = l - \mathbf{d}\cdot\mathbf{n}$。這一點就是由 $\mathbf{d}$ 所作垂線的垂足（foot of the perpendicular），所以球心到平面的距離為 $|\mu| = |l - \mathbf{d}\cdot\mathbf{n}|$。

平面與球面相切，當且僅當這個距離等於半徑（距離較小時，平面與球面相交於一個圓；距離較大時，兩者不相交）。所以條件是

$$|\mathbf{d}\cdot\mathbf{n} - l| = a, \qquad \text{等價於} \qquad (\mathbf{d}\cdot\mathbf{n} - l)^2 = a^2,$$

即 $\mathbf{d}\cdot\mathbf{n} = l \pm a$。（若 $\mathbf{n}$ 改為背向平面，則平面的方程為 $\mathbf{r}\cdot\mathbf{n} = -l$，而條件變成 $|\mathbf{d}\cdot\mathbf{n} + l| = a$。）

第 (iii) 部分

第一個球面的球心為 O，第二個的球心為 D（位置向量（position vector）為 $\mathbf{d}$），兩者半徑均為 a。點 $\mathbf{r}$ 同時位於兩個球面上，當

$$\mathbf{r}\cdot\mathbf{r} = a^2 \qquad \text{及} \qquad (\mathbf{r}-\mathbf{d})\cdot(\mathbf{r}-\mathbf{d}) = a^2.$$

把第二式展開再減去第一式，得 $-2\,\mathbf{r}\cdot\mathbf{d} + \mathbf{d}\cdot\mathbf{d} = 0$，所以每一個交點都滿足

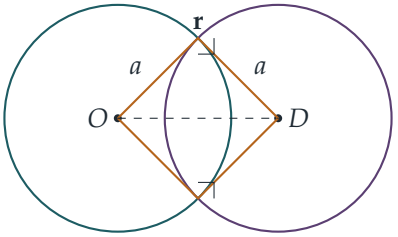
$$\mathbf{r}\cdot\mathbf{d} = \tfrac{1}{2}\,\mathbf{d}\cdot\mathbf{d}.$$

通往點 $\mathbf{r}$ 的兩條半徑是向量 $\mathbf{r}$（由 O 出發）和 $\mathbf{r}-\mathbf{d}$（由 D 出發）。兩者垂直，當

$$\mathbf{r}\cdot(\mathbf{r}-\mathbf{d}) = \mathbf{r}\cdot\mathbf{r} - \mathbf{r}\cdot\mathbf{d} = a^2 - \tfrac{1}{2}\,\mathbf{d}\cdot\mathbf{d}$$

等於零。右邊的式子與我們所選的交點無關，所以若 $\mathbf{d}\cdot\mathbf{d} = 2a^2$，則在每一個交點處兩條半徑都互相垂直，這正是我們要證明的。（這個論證也可以倒過來：只要在其中一個公共點處兩條半徑互相垂直，便有 $\mathbf{d}\cdot\mathbf{d} = 2a^2$。）

在這個情況下，兩個球面確實相交：$|\mathbf{d}| = \sqrt{2}\,a$ 嚴格介乎 0 與 $2a$ 之間，所以兩者相交於一個圓。通過 O 和 D 的截面（cross-section）顯示了其中的直角。



在平面上，兩個半徑為 a 的圓若在公共點處的半徑互相垂直，則 OD 是一個等腰直角三角形的斜邊，所以 $OD^2 = a^2 + a^2 = 2a^2$：用畢氏定理（Pythagoras' theorem）得到同一答案。

驗算

- 在 (i) 中，取 $\mathbf{b} = \mathbf{0}$：直線通過球心，條件變成 $a^2 > 0$，而這樣的直線確實總與球面相交於兩點。
- 在 (ii) 中，若球心在原點（$\mathbf{d} = \mathbf{0}$），條件就是 $l = a$：平面與球心的距離為 a 時，便與球面相切。
- 數值上，取 $a = 1$ 及 $|\mathbf{d}| = 1.2$（而非 $\sqrt{2}$），其中一個交點為 $(0.6, 0.8, 0)$，而 $\mathbf{r}\cdot(\mathbf{r} - \mathbf{d}) = 1 - 0.72 = 0.28 \neq 0$，與 $a^2 - \frac{1}{2}\mathbf{d}\cdot\mathbf{d}$ 吻合。對全部三個部分所作的隨機測試，結果的誤差均在 10^{-13} 以內。

常見錯誤

- 忘記 $\mathbf{m}\cdot\mathbf{m} = 1$，以致 λ^2 帶着一個不必要的係數。
- 把 (ii) 中的相切條件只寫成 $\mathbf{d}\cdot\mathbf{n} - l = a$：球心可以在平面的任何一側。

STEP 之外

$\mathbf{b}\cdot\mathbf{b} - (\mathbf{b}\cdot\mathbf{m})^2$ 這個量就是 $|\mathbf{b}\times\mathbf{m}|^2$，即原點到直線的距離的平方；而 $\mathbf{b} - (\mathbf{b}\cdot\mathbf{m})\mathbf{m}$ 是 $\mathbf{b}$ 除去 $\mathbf{m}$ 方向分量後的正交投影（orthogonal projection），這個概念正是線性代數（linear algebra）的核心。以直角相交的球面稱為正交（orthogonal）球面，它們在反演幾何（inversive geometry）和雙曲幾何（hyperbolic geometry）的模型中都佔有中心地位。Strang 的 *Introduction to Linear Algebra*（線性代數入門）講述投影；Needham 的 *Visual Complex Analysis*（複分析的視覺化講解）則展示正交圓的實際應用。

詳細解答

STEP III 1998，第 9 題

力學 • 轉動慣量

STEP III 1998，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/98-s3-q9

關鍵思路

實心圓錐可以看成由一疊薄圓盤堆疊而成，圓盤的半徑隨着與頂點的距離線性增長。只要知道薄圓盤的轉動慣量（moment of inertia）（繞其軸線的，以及繞一條直徑的），便可以用一個積分把各薄片的貢獻加起來。至於第二條軸，我們需要垂直軸定理（perpendicular axis theorem）（把圓盤繞軸線的轉動慣量化為繞直徑的轉動慣量）和平行軸定理（parallel axis theorem）（把該軸平移到頂點）。之後的擺動就是一個複擺（compound pendulum），其周期所需要的正是繞頂點的轉動慣量。

解答

建立薄片

把頂點置於原點，圓錐的軸線沿 z 軸，於是底面就是平面 $z = h$。設 ρ 為（恆定的）密度。在高度 z 處的橫截面是一個圓盤，由相似三角形可知其半徑為

$$r = \frac{az}{h},$$

厚度為 δz 的薄片是一個薄圓盤，質量為 $\delta m = \rho\pi r^2\,\delta z = \rho\pi\dfrac{a^2z^2}{h^2}\,\delta z$。積分得

$$m = \int_0^h \rho\pi\frac{a^2z^2}{h^2}\,\mathrm{d}z = \tfrac{1}{3}\rho\pi a^2 h, \qquad \text{所以} \qquad \rho\pi = \frac{3m}{a^2h}.$$

繞軸線的轉動慣量

質量為 M、半徑為 r 的均勻圓盤，繞其軸線的轉動慣量為 $\frac{1}{2}Mr^2$。（把它分割成半徑為 s、質量為 $\frac{2Ms}{r^2}\,\delta s$ 的圓環；每個圓環的貢獻是其質量的 s^2 倍，而 $\int_0^r \frac{2M}{r^2}s^3\,\mathrm{d}s = \frac{1}{2}Mr^2$。）

圓錐的每一塊薄片的中心都在軸線上，所以各薄片的貢獻可以直接相加：

$$\begin{aligned} I_{\text{軸}} &= \int_0^h \tfrac{1}{2}r^2\,\rho\pi r^2\,\mathrm{d}z = \tfrac{1}{2}\rho\pi\frac{a^4}{h^4}\int_0^h z^4\,\mathrm{d}z = \tfrac{1}{10}\rho\pi a^4 h \\ &= \frac{1}{10}\cdot\frac{3m}{a^2h}\cdot a^4 h = \frac{3}{10}ma^2, \end{aligned}$$

正如所求。

繞過頂點的軸的轉動慣量

我們要用到兩個定理。

- **垂直軸定理**。對於位於某平面內的薄片（lamina），繞一條垂直於該平面的軸的轉動慣量，等於繞平面內任意兩條互相垂直、並與該軸交於同一點的軸的轉動慣量之和。
- **平行軸定理**。若質量為 M 的物體繞一條通過其質心（centre of mass）G 的軸的轉動慣量為 I_G，則它繞一條與之平行、相距 d 的軸的轉動慣量為 $I_G + Md^2$。

對於薄圓盤，由對稱性可知兩條直徑是等價的，所以垂直軸定理給出 $2I_{\text{直徑}} = \frac{1}{2}Mr^2$，即 $I_{\text{直徑}} = \frac{1}{4}Mr^2$。

取過頂點且平行於底面的軸 ℓ（例如沿 x 軸）。在高度 z 處的薄片，其質心在 z 軸上，與 ℓ 相距 z，而它本身平行於 ℓ 的那條直徑正通過該質心。由平行軸定理，這塊薄片的貢獻為 $\left(\frac{1}{4}r^2 + z^2\right)\delta m$。因此

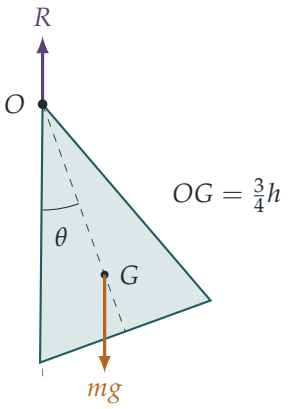
$$\begin{aligned} I_{\text{頂點}} &= \int_0^h \left(\frac{a^2z^2}{4h^2} + z^2\right)\rho\pi\frac{a^2z^2}{h^2}\,\mathrm{d}z = \rho\pi\frac{a^2}{h^2}\left(\frac{a^2}{4h^2} + 1\right)\frac{h^5}{5} \\ &= \frac{3m}{a^2h}\cdot\frac{a^2h^3}{5}\cdot\frac{a^2+4h^2}{4h^2} = \frac{3m}{20}\left(a^2+4h^2\right). \end{aligned}$$

想一想 為甚麼不能一次過對整個圓錐應用平行軸定理，由質心平移過去？

其實可以，但我們暫時還不知道繞過 G 的平行軸的轉動慣量；那正是需要計算的東西。分割成薄片就可以避開這一點，因為每個圓盤繞自身直徑的轉動慣量是已知的，而且每個圓盤的中心都在軸線上，與 ℓ 的距離是已知的 z。（順帶可得 $I_G = I_{\text{頂點}} - m\left(\frac{3}{4}h\right)^2 = \frac{3m}{80}(4a^2+h^2)$。）

繞頂點的小幅擺動

把圓錐視為剛體，可繞一條通過其頂點、垂直於運動平面的光滑水平軸自由轉動。這樣的軸平行於底面，所以繞它的轉動慣量就是 $I_{\text{頂點}}$。設 θ 為圓錐軸線與鉛垂向下方向之間的夾角。



軸上的反作用力對 O 沒有力矩，所以唯一的力矩來自重力；重力作用於 G，與 O 相距 $\frac{3}{4}h$。它對 O 的力矩為 $mg \cdot \frac{3}{4}h\sin\theta$，其作用是使 θ 減小。對該軸取力矩，

$$I_{\text{頂點}}\,\ddot{\theta} = -\tfrac{3}{4}mgh\sin\theta.$$

對於小幅擺動，$\sin\theta \approx \theta$，所以

$$\ddot{\theta} = -\omega^2\theta, \qquad \omega^2 = \frac{\frac{3}{4}mgh}{\frac{3m}{20}(a^2+4h^2)} = \frac{5gh}{4h^2+a^2}.$$

這是簡諧運動（simple harmonic motion），周期為

$$\frac{2\pi}{\omega} = 2\pi\sqrt{\frac{4h^2+a^2}{5gh}},$$

正如所求。

驗算

- 單位：$\frac{4h^2+a^2}{gh}$ 的單位是 $\mathrm{m}^2/(\mathrm{m\,s^{-2}}\cdot\mathrm{m}) = \mathrm{s}^2$，與周期的平方應有的單位一致。
- 細長圓錐（$a \to 0$）：全部質量都集中在軸線上，而 $\int z^2\,\mathrm{d}m = \frac{3}{5}mh^2$，與 $I_{\text{頂點}} \to \frac{3}{5}mh^2$ 吻合。周期變成 $2\pi\sqrt{4h/(5g)}$，即長度為 $\frac{4}{5}h$ 的單擺（simple pendulum）的周期；這個長度比 $\frac{3}{4}h$ 長，對於質量沿長度方向分佈的物體來說，理應如此。
- 數值驗算：取 $m = 1$、$a = 0.7$、$h = 2.3$，用辛普森法則（Simpson's rule）對薄片求和，以及在圓錐內以蒙地卡羅（Monte Carlo）方法抽樣取點，兩者都得出 $I_{\text{軸}} = 0.147$ 和 $I_{\text{頂點}} \approx 3.2475$，與公式吻合；取 $g = 9.81$，由小振幅出發對 $I\ddot{\theta} = -\frac{3}{4}mgh\sin\theta$ 逐步作數值計算，得出周期為 $2.7525\,\mathrm{s}$，與公式的預測一致。

常見錯誤

- 在第二部分對每塊薄片使用 $\frac{1}{2}r^2\,\delta m$：繞一條平行於底面的軸時，圓盤是繞其直徑轉動的，貢獻為 $\frac{1}{4}r^2\,\delta m$，而不是 $\frac{1}{2}r^2\,\delta m$。
- 由頂點而不是由質心出發應用平行軸定理：只有當兩條軸的其中一條通過質心時，該定理才成立。
- 把圓錐當作長度為 $\frac{3}{4}h$ 的單擺；決定周期的是轉動慣量，而不僅僅是 G 的位置。

STEP 之外

繞過某一點的任意軸的轉動慣量，全部記錄在一個對稱的 3×3 矩陣之中，這個矩陣稱為慣性張量（inertia tensor）；垂直軸定理和平行軸定理，正是它在旋轉和平移下如何變化的特殊情況。它的特徵向量（eigenvectors）稱為主軸（principal axes），可以解釋為甚麼旋轉的物體繞某些軸轉動時穩定，繞另一些軸時卻會搖晃。Kleppner 與 Kolenkow 合著的 *An Introduction to Mechanics*（力學入門）以這個程度講述剛體轉動和複擺。

詳細解答

STEP III 1998，第 10 題

力學 • 斜碰撞

> **STEP III 1998，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/98-s3-q10

關鍵思路

由於兩球質量相等，質量在兩條守恆定律中都可以約去，剩下兩個簡潔的向量關係：速度之和不變，以及速率平方之和不變。比較 $|\mathbf{a}+\mathbf{b}|^2$ 與 $|\mathbf{a}|^2+|\mathbf{b}|^2$，便可分離出純量積（scalar product）$\mathbf{a}\cdot\mathbf{b}$，而整條題目的每一部分都由這一個量帶動。在最後一部分，只要知道第二個球的速度，動量便決定了第一個球的速度，於是能量方程變成只含一個未知速率的二次方程，而有實根的二次方程，其判別式（discriminant）必為非負。

解答

把兩球看作質量同為 m 的光滑球體，在光滑桌面上運動，因此碰撞期間水平方向的力只有兩球互相施加的力。

守恆定律與相對速率

由動量（momentum）和動能（kinetic energy）守恆，得

$$m\mathbf{v}_1+m\mathbf{v}_2=m\mathbf{v}_1'+m\mathbf{v}_2',$$
$$\tfrac{1}{2}m|\mathbf{v}_1|^2+\tfrac{1}{2}m|\mathbf{v}_2|^2=\tfrac{1}{2}m|\mathbf{v}_1'|^2+\tfrac{1}{2}m|\mathbf{v}_2'|^2,$$

即

$$\mathbf{v}_1+\mathbf{v}_2=\mathbf{v}_1'+\mathbf{v}_2' \quad (1), \qquad |\mathbf{v}_1|^2+|\mathbf{v}_2|^2=|\mathbf{v}_1'|^2+|\mathbf{v}_2'|^2 \quad (2).$$

把 (1) 的兩邊各自與自身取純量積：

$$|\mathbf{v}_1|^2+|\mathbf{v}_2|^2+2\,\mathbf{v}_1\cdot\mathbf{v}_2=|\mathbf{v}_1'|^2+|\mathbf{v}_2'|^2+2\,\mathbf{v}_1'\cdot\mathbf{v}_2'.$$

減去 (2)，餘下

$$\mathbf{v}_1 \cdot \mathbf{v}_2 = \mathbf{v}_1' \cdot \mathbf{v}_2'. \qquad (3)$$

相對速率（relative speed）的平方是

$$|\mathbf{v}_1 - \mathbf{v}_2|^2 = |\mathbf{v}_1|^2 + |\mathbf{v}_2|^2 - 2\,\mathbf{v}_1 \cdot \mathbf{v}_2,$$

而根據 (2) 和 (3)，右邊每一項在碰撞前後都相同。所以 $|\mathbf{v}_1 - \mathbf{v}_2|^2 = |\mathbf{v}_1' - \mathbf{v}_2'|^2$，又因為速率為非負，故 $|\mathbf{v}_1 - \mathbf{v}_2| = |\mathbf{v}_1' - \mathbf{v}_2'|$：相對速率守恆。

其中一球起初靜止

設 $\mathbf{v}_2 = \mathbf{0}$。則由 (3) 得

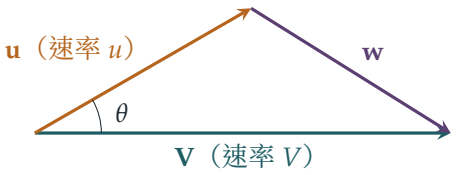
$$\mathbf{v}_1' \cdot \mathbf{v}_2' = \mathbf{v}_1 \cdot \mathbf{0} = 0.$$

由於碰撞後兩球都在運動，$\mathbf{v}_1' \neq \mathbf{0}$ 且 $\mathbf{v}_2' \neq \mathbf{0}$，所以純量積為零即表示兩個末速度互相垂直。（若沒有這個條件，純量積為零可能只是因為其中一個速度為零，例如正面碰撞（head-on collision）的情況，此時兩球只是互換速度。）

損失能量的碰撞

現在 (1) 仍然成立，但 (2) 不再成立。把靜止的球稱為 A，運動的球稱為 B。設 B 的初速度為 $\mathbf{V}$，其中 $|\mathbf{V}| = V$，末速度為 $\mathbf{u}$，速率為 u，與 $\mathbf{V}$ 成角 θ。設 A 的末速度為 $\mathbf{w}$。

由動量守恆得 $\mathbf{V} = \mathbf{u} + \mathbf{w}$，所以三個速度構成一個三角形：



因此 $\mathbf{w} = \mathbf{V} - \mathbf{u}$，而由於 $\mathbf{V} \cdot \mathbf{u} = Vu\cos\theta$，

$$|\mathbf{w}|^2 = V^2 - 2Vu\cos\theta + u^2$$

（即三角形中的餘弦定理（cosine rule））。最終動能是初始動能的 $1 - k$ 倍。約去 $\frac{1}{2}m$，

$$u^2 + |\mathbf{w}|^2 = (1-k)V^2 \quad \Longrightarrow \quad 2u^2 - 2Vu\cos\theta + V^2 = (1-k)V^2,$$

所以第二個球的末速率 u 滿足

$$\boxed{2u^2 - 2Vu\cos\theta + kV^2 = 0.}$$

想一想 題目只要求我們證明 $k \le \frac{1}{2}$。我們手上只有一條方程，不等式從何而來？

速率 u 是一個實際出現的實數，所以這條關於 u 的二次方程確實有實根。因此它的判別式必為非負：

$$4V^2\cos^2\theta - 8kV^2 \ge 0 \quad \Longrightarrow \quad k \le \tfrac{1}{2}\cos^2\theta \le \tfrac{1}{2},$$

當中我們除以 $8V^2 > 0$（第二個球在運動），並用了 $\cos^2\theta \le 1$。所以 $k \le \frac{1}{2}$，正是所求。

等號成立需要 $\cos^2\theta = 1$，並且有重根 $u = \frac{1}{2}V\cos\theta$。由於 $u \ge 0$，這表示 $\theta = 0$ 及 $u = \frac{1}{2}V$，於是 $\mathbf{w} = \frac{1}{2}\mathbf{V}$ 亦然：兩球以一半的速率一同離開，這正是完全非彈性（perfectly inelastic）的正面碰撞。動能最多只能損失一半。

驗算

- $k = 0$ 時得 $2u^2 = 2Vu\cos\theta$，所以 $u = V\cos\theta$（或 $u = 0$）。此時 $\mathbf{w} = \mathbf{V} - \mathbf{u}$ 與 $\mathbf{u}$ 垂直，因為 $\mathbf{u}\cdot(\mathbf{V}-\mathbf{u}) = Vu\cos\theta - u^2 = 0$。這與第二部分吻合。
- 恢復係數（coefficient of restitution）為 e 的正面碰撞給出 $k = \frac{1}{2}(1-e^2)$，其值介乎 0 與 $\frac{1}{2}$ 之間。取 $e = 0.3$，得 $k = 0.455$。
- 數值方面，我們模擬了 10^5 次質量相等的光滑球體的隨機彈性碰撞(elastic collision)。動量、動能、$\mathbf{v}_1\cdot\mathbf{v}_2$ 和相對速率全部吻合，誤差在 10^{-13} 以內。在 2×10^5 個隨機的能量損失結果中，該二次方程成立，誤差在 10^{-14} 以內，而所得最大的 k 為 0.49999。

常見錯誤

- 把「相對速率」當作 $\big||\mathbf{v}_1| - |\mathbf{v}_2|\big|$。它其實是相對速度的大小，即 $|\mathbf{v}_1 - \mathbf{v}_2|$。
- 在第二部分，忘記使用兩球都在運動這一事實。只有當兩個向量都不是零向量時，純量積為零才表示互相垂直。
- 在最後一部分，再次使用 (2) 或 (3)。它們來自動能守恆，而動能守恆已不再成立。只有動量守恆仍然有效。

STEP 之外

這個純量積論證，其實是一個標準技巧的不依賴參考系的版本：在零動量參考系（zero-momentum frame）中，彈性碰撞只是把兩個速度旋轉，而大小保持不變。最大損失為 $\frac{1}{2}$，反映的是一個普遍事實：只有相對於質心（centre of mass）的動能才可以損失，在這

裏就是 $\frac{1}{2} \cdot \frac{1}{2}mV^2$。這些想法通往散射理論（scattering theory），以及狹義相對論（special relativity）中的碰撞。Kleppner 與 Kolenkow 所著的 *An Introduction to Mechanics*（力學入門教科書）對質心參考系中的碰撞有仔細的論述。

詳細解答

STEP III 1998，第 11 題

力學 • 大振幅擺

> **STEP III 1998，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/98-s3-q11

本題在 *Beyond STEP* 第 11.3 節 *真實的擺* 中有完整的引導式講解：https://beyondstep.keithmcnulty.org/zh-hant/ch11/real-pendulum.html

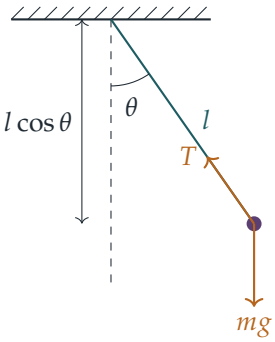
關鍵思路

能量守恆（energy conservation）給出擺在每個角度的角速率（angular speed），而週期（period）等於四分之一次擺動所需時間的四倍，$\int \mathrm{d}t = \int \mathrm{d}\theta/\dot{\theta}$。半角恆等式（half-angle identity）$\cos\theta = 1 - 2\sin^2(\theta/2)$ 把能量方程化成被積函數（integrand）的形式。題目建議的代換（substitution）消去了在 $\theta = \gamma$ 處那個棘手的無窮大，剩下的被積函數形如 $(1-x)^{-1/2}$，其中 x 很小，這正是二項展開式（binomial expansion）大派用場的地方。

解答

能量方程

把擺看成一個質量為 m 的質點，繫在一條長度為 l 的不可伸長輕繩上，在鉛垂平面內擺動，不計空氣阻力。角 θ 由向下的鉛垂線量起。



擺錘（bob）位於支點下方 $l\cos\theta$ 的深度。張力（tension）沿繩的方向作用，與擺錘的速度（沿圓弧方向）垂直，所以張力不作功；只有重力作功，因此機械能（mechanical energy）守恆。擺錘的速率為 $l\dot{\theta}$。

在最大擺角 $\theta = \gamma$ 處，擺錘瞬時靜止。由該處到一般的角度 θ，擺錘下降的高度為 $l\cos\theta - l\cos\gamma$，所以

$$\tfrac{1}{2}ml^2\dot{\theta}^2 = mgl(\cos\theta - \cos\gamma) \quad \Longrightarrow \quad \tfrac{1}{2}l\left(\frac{\mathrm{d}\theta}{\mathrm{d}t}\right)^2 = g(\cos\theta - \cos\gamma),$$

這就是第一個結果。

週期

對兩個餘弦都使用 $\cos\alpha = 1 - 2\sin^2(\alpha/2)$，得

$$\cos\theta - \cos\gamma = 2\big(\sin^2(\gamma/2) - \sin^2(\theta/2)\big),$$

於是能量方程變成

$$\left(\frac{\mathrm{d}\theta}{\mathrm{d}t}\right)^2 = \frac{4g}{l}\big(\sin^2(\gamma/2) - \sin^2(\theta/2)\big).$$

根據對稱性（運動是可逆的，而且能量方程在 $\theta \mapsto -\theta$ 之下不變），四段四分之一擺動 $0 \to \gamma \to 0 \to -\gamma \to 0$ 所需的時間相等。在由 $\theta = 0$ 到 $\theta = \gamma$ 的那四分之一擺動中，θ 遞增，所以取正平方根：

$$\frac{\mathrm{d}t}{\mathrm{d}\theta} = \frac{1}{2}\sqrt{\frac{l}{g}}\left(\sin^2(\gamma/2) - \sin^2(\theta/2)\right)^{-1/2}.$$

由 0 積分至 γ，再乘以 4，得

$$P = 2\sqrt{\frac{l}{g}}\int_0^{\gamma}\left(\sin^2(\gamma/2) - \sin^2(\theta/2)\right)^{-1/2}\mathrm{d}\theta,$$

這正是所求。（被積函數在 $\theta = \gamma$ 處趨向無窮大，因為擺錘在該處靜止，但積分是收斂（converge）的；下面的代換會清楚顯示這一點。）

小振幅近似

記 $k=\sin(\gamma/2)$，其中 $0<\gamma<\pi$，因而 $0<k<1$，並設 $\sin(\theta/2)=k\sin\phi$。當 θ 由 0 變到 γ 時，$\sin\phi$ 由 0 變到 1，所以 ϕ 由 0 變到 $\pi/2$。求導後得

$$\tfrac{1}{2}\cos(\theta/2)\,\mathrm{d}\theta = k\cos\phi\,\mathrm{d}\phi.$$

由於 $0\le\theta/2<\pi/2$ 及 $0\le\phi\le\pi/2$，兩個餘弦都非負，所以

$$\cos(\theta/2)=\sqrt{1-k^2\sin^2\phi},\qquad \left(k^2-k^2\sin^2\phi\right)^{1/2}=k\cos\phi.$$

> **想一想** 在這個代換之下，那個麻煩的因子 $(\sin^2(\gamma/2)-\sin^2(\theta/2))^{-1/2}$ 會變成甚麼？

它變成 $1/(k\cos\phi)$，而 $\mathrm{d}\theta$ 中的 $k\cos\phi$ 恰好把它完全約去：

$$\frac{\mathrm{d}\theta}{\left(k^2-\sin^2(\theta/2)\right)^{1/2}}=\frac{1}{k\cos\phi}\cdot\frac{2k\cos\phi\,\mathrm{d}\phi}{\sqrt{1-k^2\sin^2\phi}}=\frac{2\,\mathrm{d}\phi}{\sqrt{1-k^2\sin^2\phi}}.$$

因此

$$P=4\sqrt{\frac{l}{g}}\int_0^{\pi/2}\left(1-k^2\sin^2\phi\right)^{-1/2}\mathrm{d}\phi,$$

這個積分的被積函數性質十分良好。（到此為止，一切都是精確的。）

當 γ 很小時，$x=k^2\sin^2\phi$ 滿足 $0\le x\le k^2$，而這個上界很小，所以由二項展開式 $(1-x)^{-1/2}=1+\frac{1}{2}x+\frac{3}{8}x^2+\cdots$ 得

$$\left(1-k^2\sin^2\phi\right)^{-1/2}\approx 1+\tfrac{1}{2}k^2\sin^2\phi,$$

誤差的階為 k^4。由於 $\int_0^{\pi/2}\sin^2\phi\,\mathrm{d}\phi=\int_0^{\pi/2}\frac{1}{2}(1-\cos 2\phi)\,\mathrm{d}\phi=\frac{\pi}{4}$，

$$P\approx 4\sqrt{\frac{l}{g}}\left(\frac{\pi}{2}+\frac{k^2}{2}\cdot\frac{\pi}{4}\right)=2\pi\sqrt{\frac{l}{g}}\left(1+\frac{k^2}{4}\right).$$

最後，當 γ 很小時，$k=\sin(\gamma/2)\approx\gamma/2$（誤差的階為 γ^3），所以 $k^2\approx\gamma^2/4$，只相差階為 γ^4 的項，於是

$$P\approx 2\pi\sqrt{\frac{l}{g}}\left(1+\frac{\gamma^2}{16}\right),$$

這正是所求。首項是大家熟悉的小幅振盪週期；修正項則說明擺幅愈大，擺動所需的時間略長。

驗算

- 當 $\gamma \to 0$ 時，我們重新得到 $2\pi\sqrt{l/g}$，即由線性化方程 $\ddot{\theta} = -(g/l)\theta$ 所得的週期。
- 量綱（dimensions）：$\sqrt{l/g}$ 是時間，而 γ 是無量綱的。
- 數值上，取 $l/g = 1.7/9.81$ 及 $\gamma = 0.5$，以步長很小的龍格－庫塔（Runge–Kutta）方法逐步求解方程 $\ddot{\theta} = -(g/l)\sin\theta$，得 $P = 2.65705$，與精確的 ϕ 積分在所示的各位數字上完全吻合；近似公式則給出 2.65646，相對誤差約為 2×10^{-4}。該誤差與 γ^4 成比例（約為 $0.0036\gamma^4$），由此確認 γ^2 項是正確的。

常見錯誤

- 忘記因子 4：由 0 到 γ 的積分只是四分之一個週期。
- 在展開之前，就在代換中使用 $\cos(\theta/2) \approx 1$。這樣會丟掉一個 $k^2\sin^2\phi$ 項，而它與你要找的修正項同階；應把 $\sqrt{1 - k^2\sin^2\phi}$ 保持精確，直到使用二項展開的那一步。
- 停在 $1 + k^2/4$，沒有把 $k = \sin(\gamma/2)$ 換成 $\gamma/2$。

STEP 之外

精確積分 $\int_0^{\pi/2}(1 - k^2\sin^2\phi)^{-1/2}\,\mathrm{d}\phi$ 就是**第一類完全橢圓積分**（complete elliptic integral of the first kind）$K(k)$，所以精確的週期是 $4\sqrt{l/g}\,K(\sin(\gamma/2))$。把二項級數繼續展開下去，可得 $1 + \frac{1}{16}\gamma^2 + \frac{11}{3072}\gamma^4 + \cdots$，而高斯（Gauss）發現，利用算術幾何平均（arithmetic–geometric mean）可以極快地計算 K。週期取決於振幅這一事實，是非線性振盪（nonlinear oscillation）最簡單的例子。Kleppner 與 Kolenkow 合著的 *An Introduction to Mechanics*（力學入門教材）以大學一年級的程度討論了擺與振盪。

詳細解答

STEP III 1998，第 12 題

概率與統計 • 網絡可靠度

STEP III 1998，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/98-s3-q12

關鍵思路

道路只有六條，所以道路開通與封閉的組合只有 $2^6 = 64$ 種，而最簡潔的做法是數算壞的組合：即各村莊分裂成若干組、組與組之間無法互通的那些組合。把壞的組合按仍然開通的道路數目分類，數算就變得很短，因為四條村莊之間若有四條或以上的道路開通，網絡（network）絕不可能不連通。

動筆之前先提醒一點：試卷上印出的公式，括號前面是 p^2，但正確的因子是 p^3。印出的版本不可能正確，我們會在「驗算」一節說明。我們要證明的是修正後的結果

$$P = 1 - p^3(6p^3 - 12p^2 + 3p + 4).$$

解答

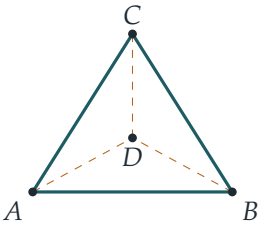
設 $q = 1 - p$ 為某一條道路開通的概率（probability）。由於各條道路的情況互相獨立（independent），任何一個有 k 條道路開通、$6 - k$ 條道路封閉的特定組合，其概率為 $q^k p^{6-k}$。設 D 為網絡不連通（disconnected）這一事件，即有某條村莊無法到達另一條村莊。於是 $P = 1 - \mathbb{P}(D)$，而 $\mathbb{P}(D)$ 就是對所有不連通的組合把 $q^k p^{6-k}$ 加起來的總和。我們按開通道路的數目 k 來數算這些組合。

四條或以上道路開通：絕不會不連通。 如果網絡不連通，各村莊便分裂成至少兩組，組與組之間沒有開通的道路。可能的分法只有 $1 + 3$（一條孤立的村莊加上其餘三條）和 $2 + 2$ 兩種。所有開通的道路都在組內。在 $1 + 3$ 的分法中，三條村莊的那一組只有 3 條道路；在 $2 + 2$ 的分法中，每一對只有 1 條道路，合共 2 條。無論哪一種，最多只能有 3 條道路開通。所以凡是有 $k \geq 4$ 條道路開通的組合都是連通（connected）的。

恰好三條道路開通。

? **想一想** 有三條道路開通時，哪些組合無法把四條村莊全部連接起來？

根據上面的論證，有三條開通道路而又不連通的組合，必定是 $1+3$ 的分法，而且三條村莊那一組之內的三條道路全部開通：即一個三角形，第四條村莊被隔絕。孤立的村莊有 4 種選法，所以這樣的組合有 4 個。（其餘 $\binom{6}{3}-4=16$ 個組合是四面體的生成樹（spanning tree），它們都是連通的。）



開通的道路 AB, BC, CA（實線）；
通往 D 的道路封閉（虛線）：
D 被隔絕

兩條或以下道路開通：必定不連通。 要把四條村莊連接起來，至少需要三條道路（每條道路最多只能把一條新的村莊，接到從 A 出發已能到達的那一組）。所以有兩條道路開通的全部 $\binom{6}{2}=15$ 個組合、有一條開通的全部 6 個，以及完全沒有道路開通的那 1 個，都是不連通的。

加起來。 因此

$$\mathbb{P}(D)=4p^3q^3+15p^4q^2+6p^5q+p^6=p^3\left(4q^3+15pq^2+6p^2q+p^3\right).$$

現在以 $q=1-p$ 展開括號：

$$\begin{aligned}4(1-p)^3&=4-12p+12p^2-4p^3,\\15p(1-p)^2&=15p-30p^2+15p^3,\\6p^2(1-p)&=6p^2-6p^3,\\p^3&=p^3.\end{aligned}$$

相加後，常數項是 4，p 的各項給出 $-12p+15p=3p$，p^2 的各項給出 $12-30+6=-12$，而 p^3 的各項給出 $-4+15-6+1=6$。所以

$$\mathbb{P}(D)=p^3\left(6p^3-12p^2+3p+4\right),\qquad\text{因此}\qquad P=1-p^3\left(6p^3-12p^2+3p+4\right).$$

驗算

► $p=0$（沒有道路會封閉）給出 $P=1$，而 $p=1$ 給出 $P=1-(6-12+3+4)=0$。兩者都正確，但 p^2 的版本同樣滿足這兩點，所以它們無法分辨兩個版本。

- p 很小時就能把兩者區分開來。網絡要不連通，至少要有三條道路封閉（例如連接同一條村莊的那三條），所以當 p 很小時，$\mathbb{P}(D)$ 應該大約是 $4p^3$。我們的公式給出 $\mathbb{P}(D) \approx 4p^3$；印出的公式則會給出 $4p^2$，這是不可能的。
- 當 $p = \frac{1}{3}$ 時，我們的公式給出 $P = 1 - \frac{1}{27} \cdot \frac{105}{27} = \frac{208}{243} \approx 0.856$。用電腦在 21 個 p 值上枚舉全部 64 個組合，結果與 p^3 的公式完全吻合；另外在 $p = 0.4$ 時模擬 200 000 場降雪，得到 0.767，而公式的值是 0.7655。

常見錯誤

- 只數算有孤立村莊的情況，而忘記檢查 $2+2$ 的分法。在這一題它們沒有帶來新的組合：$2+2$ 的分法最多只留下兩條開通的道路，所以那些組合已經包括在「兩條或以下」的數算之中。但你仍然需要把這一點說明。
- 把互相重疊的事件（例如「A 被隔絕」和「B 被隔絕」）的概率直接相加，而沒有修正重疊的部分。以整個組合為單位來數算，就可避免重複計算。
- 相信印出的 p^2，並硬把代數運算湊成那個樣子。如果老老實實算出來的結果與「show that」題目所給的不符，就用一個特殊情況來檢驗。

STEP 之外

這是**網絡可靠性**（network reliability）的一個小例子：圖（graph）的每條邊各自獨立地失效，而我們想求餘下的圖仍然連通的概率。答案必定是 p 的多項式，其係數數算的是連通生成子圖（connected spanning subgraph）的數目（這裏對應 $6, 5, 4, 3$ 條開通道路的是 $1, 6, 15, 16$）。對於大型的圖，這個多項式在計算上很難求出，並且與 Tutte 多項式（Tutte polynomial）和滲流理論（percolation theory）密切相關。Grimmett 與 Welsh 的 *Probability: An Introduction*（概率論入門）是打好概率基礎的好選擇；Bollobás 的 *Modern Graph Theory*（現代圖論）則涵蓋隨機圖（random graph）和 Tutte 多項式。

詳細解答

STEP III 1998，第 13 題

概率與統計 • 最大似然估計

> **STEP III 1998，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/98-s3-q13

關鍵思路

這裏的未知數是 n，而它是整數，所以微積分（calculus）派不上用場：我們無法對 n 求導。要求一個數列（sequence）的最大值，標準的工具是考察相鄰兩項的比值（ratio）。比值大於 1 時數列上升，一旦跌到 1 以下數列便下降；而出現並列的情況（即有兩個最大似然估計，maximum likelihood estimate，簡稱 MLE），正正就是比值在某處等於 1 的時候。

解答

擲一枚公平硬幣（fair coin）n 次，所得正面的次數服從二項分佈（binomial distribution），所以

$$\mathrm{P}(k \text{ 次正面，共擲} n \text{ 次}) = \binom{n}{k}\frac{1}{2^n}, \qquad n \geq k,$$

而若 $n < k$，這個概率是 0。以下一律取 $k \geq 1$（那位朋友至少見過一次正面），所以只有 $n \geq k$ 才是可能的答案。

公平硬幣的一般計算

對 $n \geq k$，記 $L(n) = \binom{n}{k}2^{-n}$，把它看成 k 固定時 n 的函數。對 $n \geq k$，

$$\frac{L(n+1)}{L(n)} = \frac{\binom{n+1}{k}}{\binom{n}{k}} \cdot \frac{1}{2} = \frac{n+1}{n+1-k} \cdot \frac{1}{2},$$

其中用到 $\binom{n+1}{k} = \frac{n+1}{n+1-k}\binom{n}{k}$。由於 $n+1-k > 0$，我們可以消去分母而無須改變不等號的方向：

$$\frac{L(n+1)}{L(n)} \geq 1 \iff n+1 \geq 2(n+1-k) \iff n \leq 2k-1,$$

等號恰好在 $n = 2k-1$ 時成立。因此

$$L(k) < L(k+1) < \cdots < L(2k-1) = L(2k) > L(2k+1) > L(2k+2) > \cdots .$$

想一想 比值怎樣告訴我們最大值在哪裏？

當 $n < 2k-1$ 時，比值大於 1，所以每一項都比前一項大：數列嚴格遞增，直至 $n = 2k-1$。在 $n = 2k-1$ 時，比值恰好是 1，所以 $L(2k) = L(2k-1)$。當 $n > 2k-1$ 時，比值小於 1，數列自此嚴格遞減。因此最大值恰好在兩個 n 的值上取得。

一次正面

取 $k = 1$，兩個 MLE 是 $n = 2k-1 = 1$ 和 $n = 2k = 2$。直接計算，

$$L(1) = \tfrac{1}{2}, \quad L(2) = 2 \cdot \tfrac{1}{4} = \tfrac{1}{2}, \quad L(3) = 3 \cdot \tfrac{1}{8} = \tfrac{3}{8},$$

而上面的論證表明 $L(n)$ 在 $n = 2$ 之後一直遞減。所以有兩個 MLE，即 1 和 2，正如題目所要求。

k 次正面

根據一般計算，擲硬幣次數的兩個 MLE 是

$$n = 2k-1 \quad 及 \quad n = 2k.$$

偏差硬幣

現在設 $0 < p < 1$ 及 $q = 1-p$，於是

$$L(n) = \binom{n}{k} p^k q^{n-k}, \qquad n \geq k.$$

對 $n \geq k$，因子 p^k 互相抵消，並多出一個因子 q：

$$\frac{L(n+1)}{L(n)} = \frac{(n+1)q}{n+1-k}.$$

同樣有 $n+1-k > 0$，所以

$$\frac{L(n+1)}{L(n)} \gtreqless 1 \iff (n+1)(1-p) \gtreqless n+1-k \iff k \gtreqless (n+1)p \iff n+1 \lesseqgtr \frac{k}{p}.$$

因此 L 在 $n+1 < k/p$ 時遞增，在 $n+1 = k/p$ 時持平，而一旦 $n+1 > k/p$ 便遞減。留意 $k/p > k$，因為 $p < 1$，所以轉折點落在 $n \geq k$ 這個範圍之內。

- 若 k/p 不是整數，設 $m = \lfloor k/p \rfloor$。那麼 $n+1 < k/p$ 當且僅當 $n+1 \le m$，所以 $L(k) < \cdots < L(m)$，而 $L(m+1) < L(m)$，其後繼續遞減。此時只有一個 MLE，
$$n = \left\lfloor \frac{k}{p} \right\rfloor.$$
- 若 $k/p = N$ 是整數，則 L 一直遞增至 $L(N-1)$，並有 $L(N) = L(N-1)$，然後遞減。此時有兩個 MLE，
$$n = \frac{k}{p} - 1 \quad \text{及} \quad n = \frac{k}{p}.$$
兩者都不小於 k：由於 N 是整數且 $N > k$，我們有 $N - 1 \ge k$。

所以 MLE 唯一**當且僅當** k/p **不是整數**。對公平硬幣而言，$k/p = 2k$ 總是整數，這就是前面各部分總是得到兩個 MLE，即 $2k-1$ 和 $2k$ 的原因。

驗算

- 答案合乎情理：若有 k 次正面，而每次擲出正面的概率是 p，我們預計大約擲了 k/p 次，而 MLE 正是 k/p 向下取整（若它是整數，則是其左右並列的兩個值）。
- 取 $p = \frac{1}{3}$，$k = 1$，於是 $k/p = 3$。那麼 $L(2) = 2 \cdot \frac{1}{3} \cdot \frac{2}{3} = \frac{4}{9}$ 而 $L(3) = 3 \cdot \frac{1}{3} \cdot \frac{4}{9} = \frac{4}{9}$，兩者並列，而 $L(4) = 4 \cdot \frac{1}{3} \cdot \frac{8}{27} = \frac{32}{81} < \frac{4}{9}$。
- 我們用電腦以分數精確計算 $L(n)$，並找出取得最大值的位置：公平硬幣取 $k = 1, \ldots, 39$（結果總是 $\{2k-1, 2k\}$），另外還有 924 對 (p,k)，其中 p 是分母不超過 12 的分數，且 $k \le 14$。每一個情況都與上述規則吻合。

常見錯誤

- 對 n 求導。這裏 n 是整數，而若沒有額外的工具，$\binom{n}{k}$ 對其他的 n 值並無定義；比值判別法（ratio test）才是合適的工具。
- 把不等號的方向弄反，或者忘記了正是比值等於 1 才產生第二個 MLE。
- 在所有情況下都答 $\lfloor k/p \rfloor$，而遺漏了 k/p 是整數時的並列情況。

STEP 之外

最大似然法（maximum likelihood）是統計學中最核心的估計方法：選取令觀察所得數據出現的概率最大的參數值。參數通常是連續的，我們會用微積分求對數似然（log-likelihood）的最大值。估計像 n 這樣的整數（即「試驗次數」問題），則是這個做法失效

的經典例子，而本題更顯示 MLE 甚至不一定唯一。Blitzstein 與 Hwang 的 *Introduction to Probability*（概率入門）對二項分佈講解得很好，而 Grimmett 與 Welsh 的 *Probability: An Introduction*（概率導論）則涵蓋了相關的概率背景知識。

詳細解答

STEP III 1998，第 14 題

概率與統計 • 無偏估計量

STEP III 1998，第 14 題

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

step.maths.org/questions/98-s3-q14

本題在 *Beyond STEP* 第 13.3 節德國坦克問題中有完整的引導式講解：https://beyondstep.keithmcnulty.org/zh-hant/ch13/german-tanks.html

關鍵思路

Z_2 是若干獨立（independent）部分之和，所以它的平均值和方差（variance）可直接由單一個均勻分佈（uniform distribution）變量的平均值和方差得出。Z_1 取決於最大值，而處理最大值最自然的方法，是從它的分佈函數（distribution function）入手：各個 X_i 中最大的一個不超過 m，當且僅當每一個 X_i 都不超過 m。這樣便化成若干概率的乘積，再求導就得到密度函數（density）。

解答

Z_2 的平均值

每個 X_i 在 $[0, N]$ 上的密度為 $1/N$，所以

$$\mathrm{E}(X_i) = \frac{N}{2}, \qquad \mathrm{E}(X_i^2) = \frac{1}{N}\int_0^N x^2\,\mathrm{d}x = \frac{N^2}{3}, \qquad \mathrm{Var}(X_i) = \frac{N^2}{3} - \frac{N^2}{4} = \frac{N^2}{12}.$$

由期望值的線性性質（linearity of expectation），

$$\mathrm{E}(Z_2) = \frac{2}{n}\sum_{i=1}^{n}\mathrm{E}(X_i) = \frac{2}{n}\cdot n\cdot\frac{N}{2} = N.$$

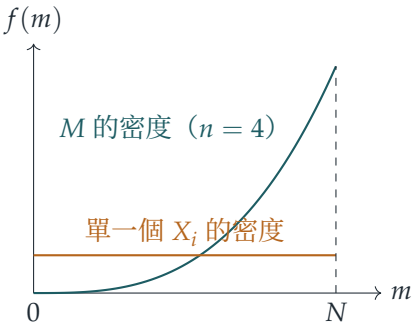
最大值的分佈

設 $M = \max\{X_1, \ldots, X_n\}$。對於 $0 \le m \le N$，事件 $M \le m$ 發生，當且僅當對每個 i 都有 $X_i \le m$；而各個 X_i 互相獨立，所以

$$\mathrm{P}(M \le m) = \prod_{i=1}^{n} \mathrm{P}(X_i \le m) = \left(\frac{m}{N}\right)^n.$$

求導後，得 M 的密度為

$$f(m) = \frac{n\,m^{n-1}}{N^n}, \qquad 0 \le m \le N.$$



最大值集中在 N 附近，所以只需乘上一個很小的修正因子，便可成為無偏（unbiased）估計量。它的矩（moments）為

$$\mathrm{E}(M) = \int_0^N m \cdot \frac{n\,m^{n-1}}{N^n}\,\mathrm{d}m = \frac{n}{N^n} \cdot \frac{N^{n+1}}{n+1} = \frac{n}{n+1}N,$$
$$\mathrm{E}(M^2) = \int_0^N m^2 \cdot \frac{n\,m^{n-1}}{N^n}\,\mathrm{d}m = \frac{n}{n+2}N^2.$$

Z_1 的平均值

$$\mathrm{E}(Z_1) = \frac{n+1}{n}\,\mathrm{E}(M) = \frac{n+1}{n} \cdot \frac{n}{n+1}N = N.$$

因此兩個估計量（estimator）的平均值都是 N，正如題目所求。

方差

對於 Z_1，首先

$$\mathrm{Var}(M) = \frac{n}{n+2}N^2 - \frac{n^2}{(n+1)^2}N^2 = \frac{n\left[(n+1)^2 - n(n+2)\right]}{(n+2)(n+1)^2}N^2 = \frac{n}{(n+2)(n+1)^2}N^2,$$

因為 $(n+1)^2 - n(n+2) = 1$。乘以 $\left(\frac{n+1}{n}\right)^2$ 後，

$$\boxed{\mathrm{Var}(Z_1) = \frac{(n+1)^2}{n^2} \cdot \frac{n}{(n+2)(n+1)^2}N^2 = \frac{N^2}{n(n+2)}.}$$

對於 Z_2，各個 X_i 互相獨立，所以它們的方差可以相加：

$$\boxed{\operatorname{Var}(Z_2) = \frac{4}{n^2}\sum_{i=1}^{n}\operatorname{Var}(X_i) = \frac{4}{n^2}\cdot n\cdot\frac{N^2}{12} = \frac{N^2}{3n}.}$$

想一想 哪一個較小？是否永遠都是同一個較小？

兩者之比為

$$\frac{\operatorname{Var}(Z_1)}{\operatorname{Var}(Z_2)} = \frac{3n}{n(n+2)} = \frac{3}{n+2}.$$

當 $n = 1$ 時，這個比等於 1（此時 $Z_1 = Z_2 = 2X_1$，兩個估計量其實是同一個）；而對每個 $n \geq 2$，它都小於 1。

應選哪一個

兩者都是無偏的，所以我們比較它們的離散程度。我們選擇 Z_1：

- 當 $n \geq 2$ 時，它的方差嚴格較小，而且 n 越大，優勢越明顯：$\operatorname{Var}(Z_1)$ 以 $1/n^2$ 的速度下降，而 $\operatorname{Var}(Z_2)$ 只以 $1/n$ 的速度下降。例如當 $n = 10$ 時，Z_1 的標準差（standard deviation）約為 $0.09N$，而 Z_2 的則約為 $0.18N$。
- Z_1 必定不小於所見到的最大識別號碼，而 N 亦必須如此。Z_2 卻可能比我們實際觀察到的某個號碼還要小（例如讀數為 $1, 2, 30$ 時，得 $Z_2 = 22$），給出一個我們明知不可能的估計值。

驗算

- 當 $n = 1$ 時，兩條公式都給出 $N^2/3$；這是必然的，因為 $Z_1 = Z_2 = 2X_1$，而 $\operatorname{Var}(2X_1) = 4\cdot N^2/12$。
- 以 $N = 10$ 進行 200 000 次模擬試驗，兩個估計量的樣本平均值與 10 相差都不超過 0.01；當 $n = 4$ 時，樣本方差分別為 4.16 和 8.33，而理論值為 $\frac{100}{24} \approx 4.17$ 和 $\frac{100}{12} \approx 8.33$。

常見錯誤

- 試圖直接處理最大值的密度。應先求 $\mathrm{P}(M \leq m)$；由於互相獨立，它只是一個簡單的乘積。
- 忘記把常數平方：$\operatorname{Var}(cY) = c^2\operatorname{Var}(Y)$，所以兩個因子分別是 $\left(\frac{n+1}{n}\right)^2$ 和 $\frac{4}{n^2}$。

STEP 之外

這就是經典的「德國坦克問題」（German tank problem）：第二次世界大戰期間，盟軍的統計學家利用坦克的序號估計德國的坦克產量，結果準確得驚人。在這裏，最大值是一個充分統計量（sufficient statistic）：一旦知道了它，樣本的其餘部分就再不能提供關於 N 的任何額外資訊。萊曼–謝費定理（Lehmann–Scheffé theorem）進一步指出，在所有無偏估計量之中，Z_1 的方差最小。充分性和有效估計量（efficient estimator）是大學一年級統計學的課題；若想認識順序統計量（order statistics），Blitzstein 與 Hwang 合著的 *Introduction to Probability*（概率論入門）是很好的起點。

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

1998 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。