Cauchy-Schwarz 不等式之本質與意義

林琦焜

學而不思則罔, 思而不學則殆。
— 論語 —

1. 前言:

一次偶然的機會, 看到一位學生很努力地讀數學, 花了九牛二虎之力嘗試證明 Cauchy-Schwarz 不等式, 看他苦思無助的樣子, 讓自己也回想在求學的經驗。隨後看了看其草稿紙, 我很坦白地告訴他「你不是在學數學, 而是在背書, 背定理證明。」隨後, 自己就很"雞婆"地從判別式與投影的角度解釋這不等式, 事後, 自己覺得意猶未竟,「師者, 所謂傳道、授業、解惑者也。」如果自己的一些心得能夠解某些人對於數學之困惑與誤解, 何樂而不爲呢? 這就是寫本文之動機與緣由。

其實我們不能完全責怪學生素質不好、用功程度不夠, 或怪罪社會風氣差, 影響著校園。如果數學的教學仍停留在定義、定理、證明, 就難怪學生是如此學習。如果這就是數學, 老師就沒有什麼好教的, 學生也沒有什麼好學(教學等於把課本念一遍!) 有誰會喜歡這種 "無血無目屎" 的怪胎呢?

在舊約聖經列王紀下記載著一段感人的師生故事, 當先知以利亞要被耶和華接去之前問他的學生以利沙說: 「我未曾被接去離開你, 你要我爲你作甚麼只管求我。」以利沙說:「願感動你的靈加倍的感動我。」教的人有如此之胸襟與眼光不忌賢不藏私, 學的人有如此之抱負與積極學習之渴望。否則一代不如一代, 教育就失去其價值與意義。

『一個理想的釋經學者應該是一位去過第一世紀那個奇異世界的人, 感覺到其中的一片陌生, 但卻在那裡逗留, 直到自己生活在其中, 他的思想和感受與初聽福音的人一樣爲止; 然後再回到今日的世界, 將所獲悉的眞理用我們今日的思想訴說出來。』

— Charles Harold Dodd —

Cauchy-Schwarz 不等式顧名思義與 Cauchy 有關, 我們從最常見的形式開始

1.1 定理 (Cauchy 不等式): 已知

$a_1,\ldots,a_n,b_1,\ldots,b_n$ 爲實數, 則

$$\left(\sum_{i=1}^n a_ib_i\right)^2\le\left(\sum_{i=1}^n a_i^2\right)\left(\sum_{i=1}^n b_i^2\right)\qquad(1.1)$$

等式成立之充分必要條件是 $a_i=\lambda b_i\,, i=1,\ldots,n$。

這是最常見的 Cauchy 不等式, 其實當 $n=3$ 可追朔至法國數學家 J. L. Lagrange (1736-1813)。Cauchy 不等式可推廣至複數。如何推廣呢? 不等式只有在實數時才有意義, 對於複數或向量要談大小關係, 自然的選擇就是其長度。對任意複數 $z=x+iy$, 其長度 $|z|=\sqrt{x^2+y^2}$, 因此對 (1.1) 而言我們只須將平方的意義, 更改爲複數之模數 (modulus) 的平方即可。

1.2 定理 (Cauchy 不等式): 已知 $a_1,\ldots,a_n,b_1\ldots,b_n$ 爲複數, 則

$$\left|\sum_{i=1}^n a_ib_i\right|^2\le\left(\sum_{i=1}^n |a_i|^2\right)\left(\sum_{i=1}^n |b_i|^2\right)\qquad(1.2)$$

等式成立之充分必要條件是$a_i=\lambda b_i$, $i=1,\ldots,n\,,\lambda$ 爲一複數。

令 $\vec{a}=(a_1,a_2,\ldots,a_n)$、$\vec{b}=(b_1,b_2,\ldots,b_n)$ 並且定義向量之長度,

$$||\vec{a}||=\sqrt{|a_1|^2+\cdots+|a_n|^2}$$

則 Cauchy 不等式可表爲

$$|\vec{a}\cdot\vec{b}|\le\|\vec{a}\|\|\vec{b}\|\qquad(1.3)$$

由線性代數之理論任意正定對稱矩陣都可以定義內積, 因此若 $A=(a_{ij})$ 爲一正定對稱矩陣, 則有底下之 Cauchy 不等式

1.3 定理 (Cauchy 不等式): 已知 $(a_{ij})_{ij}$ 爲一正定對稱矩陣 $(a_{ij}=a_{ji})$, $x_1,\ldots,x_n$、$y_1,\ldots,y_n$ 爲任意實數 (或複數) 則

$$\left|\sum_{i,j=1}^n a_{ij}x_iy_j\right|\le\sqrt{\sum_{i,j=1}^n a_{ij}x_ix_j}\sqrt{\sum_{i,j=1}^n a_{ij}y_iy_j}\qquad(1.4)$$

如果你覺得(1.4)不容易理解, 可利用向量來看

$$\vec{\xi}\cdot\vec{\eta}=\vec{x}A\vec{y}^{\,t}=\sum_{i,j=1}^n a_{ij}x_iy_j$$

$$\|\xi\|^2=\vec{\xi}\cdot\vec{\xi}=\vec{x}A\vec{x}^t=\sum_{i,j=1}^n a_{ij}x_ix_j$$

$$\|\eta\|^2=\vec{\eta}\cdot\vec{\eta}=\vec{y}A\vec{y}^{\,t}=\sum_{i,j=1}^n a_{ij}y_iy_j$$

我們在此略微休息、沈思, 對於學數學的人而言, 無限(infinite) 是一個很自然的概念, 並不需要藉助於宗教或哲學上深奧的術語。Cauchy不等式對任意自然數n都成立, 那麼請問$n=\infty$是否也成立呢? 要回答這個問題實際上等於問無窮級數$\sum_{i=1}^\infty|a_i|^2$、$\sum_{i=1}^\infty|b_i|^2$ 是否收斂的問題。

1.4 定理 (Cauchy 不等式): 已知 $a_i,b_i\in\boldsymbol{C}$ 則

$$\left|\sum_{i,j=1}^\infty a_ib_j\right|\le\left(\sum_{i=1}^\infty|a_i|^2\right)^{\frac12}\left(\sum_{i=1}^\infty|b_i|^2\right)^{\frac12}\qquad(1.5)$$

等式成立之充分必要條件是 $a_i=\lambda b_i\,, i=1,2,3\ldots,\lambda\in\boldsymbol{C}$。如果 $\sum_{i=1}^\infty|a_i|^2<\infty$、$\sum_{i=1}^\infty|b_i|^2<\infty$, 則 $|\sum_{i=1}^\infty a_ib_i|<\infty$。

從 Cauchy 不等式的角度而言, 無窮數列 $\{a_i\}_{i=1}^\infty$ 之平方和收斂, $\sum_{i=1}^\infty|a_i|^2<\infty$,

是很自然而然出現的空間, 在實變函數論或泛函分析我們稱之爲 l^2 空間。這是 n 維實數空間 $\boldsymbol{R}^n$ 最自然的推廣, 它是一個 Hilbert 空間, 最重要的應用就是量子力學。

在數學中尤其是分析學的思考過程通常是

$$\text{有限和} \iff \text{無窮級數} \iff \text{積分} \tag{1.6}$$

因此想當然耳 Cauchy 不等式是可以推廣至積分。

1.5 定理 (Cauchy-Schwarz 不等式): 已知 f,g 是區間 $[a,b]$ 上的連續函數, $f,g\in C[a,b]$, 則

$$\left|\int_a^b f(x)g(x)dx\right|^2 \le \int_a^b |f(x)|^2dx \int_a^b |g(x)|^2dx \tag{1.7}$$

從實變函數論的角度而言, 我們僅需要求 f、g 是平方可積分函數 $(L^2[a,b])$ 則 Cauchy-Schwarz 不等式仍然成立。其空間關係可對照前一式 (1.6):

$$\boldsymbol{R}^n \iff l^2 \iff L^2 \tag{1.8}$$

(1.7) 這個不等式通常稱爲 Schwarz 不等式、Cauchy-Schwarz 不等式或 Cauchy-Schwarz-Bunyakovsky 不等式, 是烏克蘭數學家 Viktor Yakovlerich Bunyakovsky (1804-1889) 與德國數學家 (原籍波蘭)Karl Herman Amandus Schwarz (1843-1921), 各自於 1861 年與 1885 年發現的。實際上 Bunyakovsky 比 Schwarz 還要早25年發現此不等式! 但他的名字卻常常被忽略甚至被遺忘! 有時候所謂的名譽、聲望其實是勢力分佈的結果。在19世紀數學的勢力基本上是在德國與法國, 德國則以格庭根大學與柏林大學爲中心。

Schwarz 起初是在柏林大學就讀化學, 而後受 Kummer 與 Weierstrass 之影響而轉讀數學。受業於 Weierstrass 門下, 據聞他於1861在 Weierstrass 課堂上 (Integral calculus) 所做之講義, 至今仍留著。Schwarz 是 Weierstrass 最出色的學生之一, 他於1892年接替 Weierstrass 在柏林大學的位置一直到1917年止。19世紀的數學基本上是複變函數論, 因此在Weierstrass 的指導下, Schwarz主要是研究保角變換 (conformal mapping) 並涉及變分學 (calculus of variations) 特別是最小曲面的問題。他發現如何將上半平面映至多邊形的公式, 今天我們稱之爲 Schwarz-Christoffel 公式。他更用保角變換的技巧來解偏微分方程 (Laplace 方程) 的 Dirichlet 問題, 這方法有效地避免了 Riemann 利用 Dirichlet 原則 (Dirichlet principle) 之爭議。Schwarz是第一個給出圓盤上調和函數之 Dirichlet 問題存在性的嚴格證明。這問題原先是由 Poisson 所解。

Schwarz最重要的工作則是於 Weierstrass 70歲生日之賀禮, 他解決了給定一最小曲面是否可得最小面積的問題。而在這文章中他發現了一個關於積分的不等式, 就是現在熟悉的 Schwarz 不等式。

Cauchy不等式最出名的推廣就是 Hölder 不等式

1.6 定理 (Hölder 不等式): 已知 $a_1,\ldots,a_n,b_1,\ldots,b_n$ 爲任意複數, 且$p,q\geq 1$, $\frac{1}{p}+\frac{1}{q}=1$, 則

$$\left|\sum_{i=1}^n a_ib_i\right|\leq\left(\sum_{i=1}^n|a_i|^p\right)^{\frac{1}{p}}\left(\sum_{i=1}^n|b_i|^q\right)^{\frac{1}{q}} \tag{1.9}$$

Hölder 不等式對$n=\infty$也成立。另外最著名的就是積分之形式

1.7 定理 (Hölder 不等式): 已知$f,g\in C[a,b]$, $\frac{1}{p}+\frac{1}{q}=1$且$p,q\geq 1$則

$$\left|\int_a^b f(x)g(x)dx\right| \leq\left(\int_a^b|f(x)|^pdx\right)^{\frac{1}{p}}\left(\int_a^b|g(x)|^qdx\right)^{\frac{1}{q}} \tag{1.10}$$

或者更一般的形式

1.8 定理 (Hölder 不等式): 已知 $f_1,\ldots,f_n\in C[a,b]$, 且 $\frac{1}{p_1}+\frac{1}{p_2}+\cdots+\frac{1}{p_n}=1$, $p_i\geq 1$ 則

$$\left|\int_a^b f_1(x)f_2(x)\cdots f_n(x)dx\right| \leq\left(\int_a^b|f_1(x)|^{p_1}dx\right)^{\frac{1}{p_1}}\cdots\left(\int_a^b|f_n(x)|^{p_n}dx\right)^{\frac{1}{p_n}} \tag{1.11}$$

這不等式是德國數學家 Otto Ludwig Hölder (1859-1937) 在 1884 年間研究 Fourier 級數之收歛性問題發現的。Hölder就讀於德國柏林大學, 受業於一代宗師 Weierstrass、Kronecker與 Kummer 等人之下。原先研究興趣是解析函數 (analytic function) 與利用算術平均求和的問題。而後受 Kronecker 與 Klein 的影響, 也投入群論 (group theory) 之研究, 其中最著名的是 Jordan-Hölder 定理。

Hölder 不等式也稱爲 Hölder-Riesz 不等式, 這個不等式經由匈牙利數學家 F. Riesz (1880-1956) 有系統地整理應用, 而成爲研究近代分析 (特別是泛函分析) 的主要工具。F. Riesz 可以說是泛函分析之創始者, 他利用 Frechet 的想法將 Lebesgue 之實變函數論與 Hilbert-Schmidt 的積分方程理論建立起連接之環節。在 1907-1909 年間他利用 Stieltjes 積分得平方可積分函數之線性泛函的表現定理。我們習慣稱爲 Riesz 表現理論 (Riesz representation theorem), 他進而更研究p次可積分函數 (L^p), 這可說是研究 normed space(賦範空間) 之開始, 其中他引進弱收歛 (weak convergence) 的概念。他也證明L^p空間的完備性, 今日稱爲 Riesz-Fisher 定理, 這定理成爲證明量子力學中矩陣力學 (matrix mechanics: Heisenberg) 與波動力學 (wave mechanics: Schrödinger) 兩個理論是全等的數學基礎。

2. 餘弦定律

談 Cauchy-Schwarz 不等式, 最好是從餘弦定律開始, 我們考慮三角形 $\triangle ABC$ 三邊的分量分別是

$$\vec{a}=\overrightarrow{AB},\ \vec{b}=\overrightarrow{AC},\ \vec{c}=\overrightarrow{CB}=\vec{a}-\vec{b}$$

並且假設 $\angle CAB = \theta$, 則由餘弦定律 $a^2 + b^2 - c^2 = 2ab\cos\theta$ 或者換成向量表示爲

$$|\vec{a}|^2 + |\vec{b}|^2 - |\vec{a} - \vec{b}|^2 = 2|\vec{a}||\vec{b}|\cos\theta \quad (2.1)$$

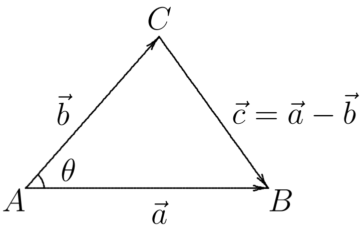
可得

$$\vec{a}\cdot\vec{b} = |\vec{a}||\vec{b}|\cos\theta \quad (2.2)$$

由餘弦的性質 $|\cos\theta| \le 1$, 可結論

$$|\vec{a}\cdot\vec{b}| \le |\vec{a}||\vec{b}| \quad (2.3)$$

這就是 Cauchy-Schwarz 不等式, 它告訴我們兩個向量的內積永遠小於這兩個向量長度的乘積。



圖一

這是對於 Cauchy-Schwarz 不等式最直接明瞭的闡述, 但是這裡面卻有件事值得商榷的 — 角度。如果我們的興趣只停留在二維或三維空間, 則Cauchy-Schwarz 不等式討論至此即足矣! 數學的探索有時是出於奇想。對於四維或任意N 維空間是否也有 Cauchy-Schwarz 不等式呢? 此時"角度"的意義馬上變得異常的曖昧, 對於四維以上之空間的角度是什麼呢? 這個問題我們回到(2.2) 式將之改寫爲

$$\cos\theta = \frac{\vec{a}\cdot\vec{b}}{|\vec{a}||\vec{b}|} \quad (2.4)$$

左邊是角度有維數 (dimension) 之限制然而右邊是內積則不受此限制。所以要將角度的觀念推廣至高維甚至無窮維空間勢必從右式著手, 我們暫且將 $\cos\theta$ 與 $\frac{\vec{a}\cdot\vec{b}}{|\vec{a}||\vec{b}|}$ 視爲兩個不同的事物。前面證明的邏輯

$$\text{餘弦定律} \Longrightarrow \text{Cauchy-Schwarz 不等式}$$

如果要推廣, 我們必須先放棄餘弦定律而直接證明 Cauchy-Schwarz 不等式:

$$|\vec{a}\cdot\vec{b}| \le |\vec{a}||\vec{b}|$$

然後再回到 (2.4) 來定義角度, 當然我們會問爲何是 $\cos\theta$ 而非$\sin\theta$ 呢? 由於絕對值小於等於1, 因此兩者都有可能, 但任何的推廣必須建立在既有的事實上面, 所以唯一的選擇便是 $\cos\theta$。

註解:

(1) "角度" 是數學中一很有趣的主題, 例如在球面上如何定義呢? 對於更一般的曲面讀者有興趣可參閱幾何學方面的專著。對複數內積空間, 角度有兩種定義方式

$$\cos\theta = \frac{|\vec{a}\cdot\vec{b}|}{\|\vec{a}\|\|\vec{b}\|} \quad (2.5)$$

$$\cos\theta = \frac{R(\vec{a}\cdot\vec{b})}{\|\vec{a}\|\|\vec{b}\|} \quad (2.6)$$

$R(\vec{a}\cdot\vec{b})$ 表內積 $\vec{a}\cdot\vec{b}$ 之實部。對於(2.5) 而言顯然

$$\theta = \frac{\pi}{2} \Longleftrightarrow \vec{a}\cdot\vec{b} = 0$$

但餘弦定律卻不成立, 但 (2.6) 則成立, 如果我們允許複數角的話則可定義

$$\cos\theta_c = \frac{\vec{a}\cdot\vec{b}}{|\vec{a}||\vec{b}|} \quad (2.7)$$

($\vec{a}\cdot\vec{b}$可以是複數),$\cos\theta_c$ 是一複數可以表爲

$$\cos\theta_c = \rho e^{i\varphi},\ \rho \le 1,\ -\pi \le \varphi \le \pi \tag{2.8}$$

其中

$$\rho = \cos\theta_H = |\cos\theta_c|, \quad 0 \le \theta_H \le \frac{\pi}{2}$$

θ_H 稱爲 Hermitan 角 (Hermitian angle), 而 φ 則稱爲擬角 (pseudo-angle)。

3. Cauchy-Schwarz 不等式之證明

3.1 判別式

關於這個不等式最常見的證明方法是利用判別式, 因爲向量$\vec{a}$、$\vec{b}$ 之夾角 $\theta \ne 0$, 即 $\vec{a}$、$\vec{b}$ 非平行的兩向量, 因此由向量$\vec{a}$、$\vec{b}$ 所形成之平面上任意向量可表爲

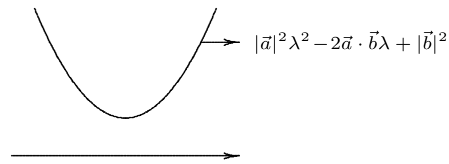
$$\vec{c} = \vec{b} - \lambda\vec{a}, \quad \lambda \in \boldsymbol{R} \tag{3.1}$$

我們考慮向量 $\vec{c}$ 的長度

$$\begin{aligned} |\vec{c}|^2 &= \vec{c}\cdot\vec{c} = (\vec{b}-\lambda\vec{a})\cdot(\vec{b}-\lambda\vec{a}) \\ &= \vec{b}\cdot\vec{b} - 2\vec{a}\cdot\vec{b}\lambda + \vec{a}\cdot\vec{a}\lambda^2 \\ &= |\vec{a}|^2\lambda^2 - 2\vec{a}\cdot\vec{b}\lambda + |\vec{b}|^2 \end{aligned} \tag{3.2}$$

(3.2) 可視爲 λ 之二次方程式, 由於 $|\vec{c}|^2 \ge 0$, 而且$|\vec{a}|^2 \ge 0$, 所以 (3.2) 所代表的是開口向上而且在 λ 軸上方的拋物線, 由於與 λ 軸不相交, 所以沒有實根, 因此判別式小於或等於0

$$\triangle = (2\vec{a}\cdot\vec{b})^2 - 4|\vec{a}|^2|\vec{b}|^2 \le 0 \Longrightarrow |\vec{a}\cdot\vec{b}| \le |\vec{a}||\vec{b}|$$



圖二

3.2 投影 — 最短距離

整個證明過程完全避開角度, 而只用到內積, 因此對於任意N維或無窮維空間都可適用, 這是該證明方法最大的優點。然而我們對此仍不滿足, 例如 λ 的意義是什麼? 難道它只是一過渡性的工具而已嗎? 我們再看另一種證明方法, 仍然仿前面的方法, 只是此時我們取

$$\lambda = \frac{\vec{a}\cdot\vec{b}}{|\vec{a}|^2}, \quad \vec{c} = \vec{b} - \lambda\vec{a} = \vec{b} - \frac{\vec{a}\cdot\vec{b}}{|\vec{a}|^2}\vec{a} \tag{3.3}$$

代回 (3.2)

$$0 \le |\vec{c}|^2 = |\vec{a}|^2\left(\frac{\vec{a}\cdot\vec{b}}{|\vec{a}|^2}\right)^2 - 2\vec{a}\cdot\vec{b}\left(\frac{\vec{a}\cdot\vec{b}}{|\vec{a}|^2}\right) + |\vec{b}|^2 \tag{3.4}$$

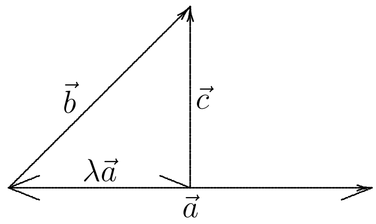
整理得

$$(\vec{a}\cdot\vec{b})^2 \le |\vec{a}|^2|\vec{b}|^2$$

正好是 Cauchy-Schwarz 不等式 (開根號)

$$|\vec{a}\cdot\vec{b}| \le |\vec{a}||\vec{b}|$$

但是這個證明方法反而更令人沮喪! 突然天外飛來一筆, 什麼都沒有交代。寫書或教書的人不負責任, 就如此含糊帶過去, 當學生的讀書不求甚解, 不想也不問最後就硬背起來, 學數學變成是背書, 這實在不是教育的目的, 其實稍用點心思, 從證明過程是可以領悟一些奧秘的。



圖三

爲何 λ 要如此取呢? 我們看向量 $\vec{a}$、$\vec{c}$ 之關係

$$\vec{a}\cdot\vec{c}=\vec{a}\cdot(\vec{b}-\lambda\vec{a})=\vec{a}\cdot\vec{b}-\lambda|\vec{a}|^2$$

因此

$$\vec{a}\cdot\vec{c}=0\Longleftrightarrow\vec{a}\perp\vec{c}\Longleftrightarrow\lambda=\frac{\vec{a}\cdot\vec{b}}{|\vec{a}|^2}\quad(3.5)$$

換句話說,(3.3) 中 λ 的取法是有幾何意義並非無中生有靠靈感而來的。此時 $\lambda\vec{a}$ 正是向量 $\vec{b}$ 在 $\vec{a}$ 之投影, 而 $|\vec{c}|$ 則是 $\vec{b}$ 至 $\vec{a}$ 的最短距離。λ還有一個很出名的解釋就是 Lagrange乘子 (Lagrange multiplier): 爲求$\vec{b}$ 至 $\vec{a}$ 的最短距離我們必須調整λ使得 $\vec{a}\perp(\vec{b}-\lambda\vec{a})$, 如果是函數則該條件換爲微分或梯度(gradient) 等於零。

如果我們並不是急著要完成證明, 而略爲享受證明的過程, 或許我們會有一些收穫。數學與人生一樣, 過程比答案要來的重要。回到 (3.4)

$$|\vec{a}|^2|\vec{b}|^2-(\vec{a}\cdot\vec{b})^2=|\vec{a}|^2|\vec{c}|^2\qquad(3.6)$$

右邊 $|\vec{a}|^2|\vec{c}|^2$ 就是 Cauchy-Schwarz 不等式的誤差項, 顯然 $|\vec{a}|\neq 0$, 所以 Cauchy-Schwarz 不等式要成爲等式若且唯若

$$|\vec{c}\,|=0\,,\qquad \vec{b}=\lambda\vec{a}$$

也就是說$\vec{a}$、$\vec{b}$是兩平行的向量。

研究學問就像釣魚必須耐心等待一段漫長枯燥的時間, 才會有有趣的事發生。我們並不急著要告訴有關Cauchy–Schwarz 不等式的應用 $\cdots$ 等等, 學數學要懂得如何去欣賞其結構美。由 (3.6) 不難想像其左式本質就是“行列式”, 因此我們引進Gram 行列式 (Gram determinant)

$$\Gamma(\vec{a},\vec{b})\equiv\begin{vmatrix}\vec{a}\cdot\vec{a} & \vec{a}\cdot\vec{b}\\ \vec{a}\cdot\vec{b} & \vec{b}\cdot\vec{b}\end{vmatrix},\ \Gamma(\vec{a})\equiv|\vec{a}|^2=\vec{a}\cdot\vec{a}\qquad(3.7)$$

因此由點 $\vec{b}$ 到直線 $\vec{a}$的最短距離 $\delta\equiv|\vec{c}|$可表爲

$$\delta^2=\frac{\Gamma(\vec{a},\vec{b})}{\Gamma(\vec{a})}\qquad(3.8)$$

又由於 $\Gamma(\vec{a})=|\vec{a}|^2>0$ (因爲 $\vec{a}\neq 0$) 所以馬上可以結論 Gram 行列式 $\Gamma(\vec{a},\vec{b})\geq 0$。當然利用 Gram 行列式也可以得到 Cauchy-Schwarz 不等式的另一種形式。

3.1 定理 (Cauchy-Schwarz 不等式):

$$|\vec{a}\cdot\vec{b}|\leq|\vec{a}||\vec{b}|\Longleftrightarrow\Gamma(\vec{a},\vec{b})\geq 0$$

由 Gram 行列式也可判定向量是否線性獨立。

3.2 定理: 向量 $\{\vec{a},\vec{b}\}$ 線性獨立若且唯若 $\Gamma(\vec{a},\vec{b})>0$。

註解:

(1) 3.1 定理告訴我們由矩陣的正定性可以導出 Cauchy-Schwarz 不等式。由矩陣理論來看是明顯的, 由於

$$\det(PP^t)=(\det P)^2\geq 0\qquad(3.9)$$

所以可推出 Cauchy-Schwarz 不等式

$$\Gamma(\vec{a},\vec{b})=\left|\begin{pmatrix}\vec{a}\\ \vec{b}\end{pmatrix}(\vec{a},\vec{b})\right|=\begin{vmatrix}\vec{a}\cdot\vec{a} & \vec{a}\cdot\vec{b}\\ \vec{b}\cdot\vec{a} & \vec{b}\cdot\vec{b}\end{vmatrix}=|\vec{a}|^2|\vec{b}|^2-(\vec{a}\cdot\vec{b})^2\geq 0 \quad (3.10)$$

甚至還有更廣的不等式

$$\Gamma(\vec{a},\vec{b},\vec{c})=\left|\begin{pmatrix}\vec{a}\\ \vec{b}\\ \vec{c}\end{pmatrix}(\vec{a},\vec{b},\vec{c})\right|=\begin{vmatrix}\vec{a}\cdot\vec{a} & \vec{a}\cdot\vec{b} & \vec{a}\cdot\vec{c}\\ \vec{b}\cdot\vec{a} & \vec{b}\cdot\vec{b} & \vec{b}\cdot\vec{c}\\ \vec{c}\cdot\vec{a} & \vec{c}\cdot\vec{b} & \vec{c}\cdot\vec{c}\end{vmatrix}\geq 0 \quad (3.11)$$

$$\Gamma(\vec{a}_1,\vec{a}_2,\ldots,\vec{a}_n)=\begin{vmatrix}\vec{a}_1\cdot\vec{a}_1 & \vec{a}_1\cdot\vec{a}_2 & \ldots & \vec{a}_1\cdot\vec{a}_n\\ \vec{a}_2\cdot\vec{a}_1 & \vec{a}_2\cdot\vec{a}_2 & \ldots & \vec{a}_2\cdot\vec{a}_n\\ \vdots & \vdots & \ddots & \vdots\\ \vec{a}_n\cdot\vec{a}_1 & \vec{a}_n\cdot\vec{a}_2 & \ldots & \vec{a}_n\cdot\vec{a}_n\end{vmatrix}\geq 0 \quad (3.12)$$

(2) 在函數空間例如$f_1,f_2,\ldots,f_k\in C[a,b]$ 則 Gram 行列式取底下之形式:

$$\Gamma(f_1,f_2,\ldots,f_k)=\begin{vmatrix}\int_a^b f_1^2(x)dx & \int_a^b f_1(x)f_2(x)dx & \ldots & \int_a^b f_1(x)f_k(x)dx\\ \int_a^b f_2(x)f_1(x)dx & \int_a^b f_2^2(x)dx & \ldots & \int_a^b f_2(x)f_k(x)dx\\ \vdots & \vdots & \ddots & \vdots\\ \int_a^b f_k(x)f_1(x)dx & \int_a^b f_k(x)f_2(x)dx & \ldots & \int_a^b f_k^2(x)dx\end{vmatrix}$$

則由 3.1, 3.2 定理可知 $\Gamma(f_1,f_2,\ldots,f_k)\geq 0$, $f_1,f_2,\ldots,f_k$爲線性相依, 若且唯若 $\Gamma(f_1,f_2,\ldots,f_k)=0$。

(3) 平面上任意點(x^*,y^*)至直線$ax+by=c$的距離可以藉由內積來認識: 首先連接(x^*,y^*)、(x,y)兩點且假設此線段與直線$ax+by=c$ 垂直, 因爲(a,b)爲直線$ax+by=c$ 的法向量故可假設

$$\begin{aligned}&(x^*-x,y^*-y)\\&=t\Big(\frac{a}{\sqrt{a^2+b^2}},\frac{b}{\sqrt{a^2+b^2}}\Big)\end{aligned}$$

上式右方取單位向量的好處是此時$\delta=|t|$是向量 (x^*-x,y^*-y)之長度因爲(x,y)在直線$ax+by=c$上故

$$ax^*+by^*-c=t\sqrt{a^2+b^2}$$
$$\Longrightarrow\delta=|t|=\frac{|ax^*+by^*-c|}{\sqrt{a^2+b^2}} \quad (3.13)$$

我們將直線 $ax+by=c$ 表爲內積:

$$\begin{aligned}ax+by&=(a,b)\cdot(x,y)\\&=\sqrt{a^2+b^2}\frac{c}{\sqrt{a^2+b^2}}\end{aligned}$$

則 $\frac{|c|}{\sqrt{a^2+b^2}}$ 正是原點 $(0,0)$ 至直線 $ax+by=c$ 之距離。同理三維空間上任意點 (x^*,y^*,z^*) 至平面 $ax+by+cz=c$ 的距離公式

$$\delta=\frac{|ax^*+by^*+cz^*-d|}{\sqrt{a^2+b^2+c^2}} \quad (3.14)$$

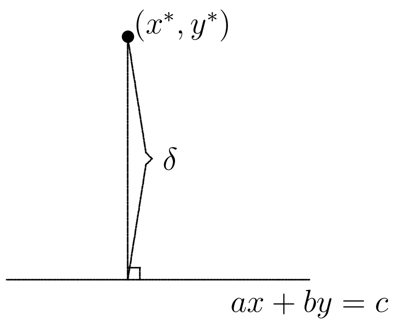
我們將平面 $ax+by+cz=d$ 表爲內積之形式

$$\begin{aligned}
&ax+by+cz\\
&=(a,b,c)\cdot(x,y,z)\\
&=\sqrt{a^2+b^2+c^2}\frac{d}{\sqrt{a^2+b^2+c^2}}
\end{aligned}$$

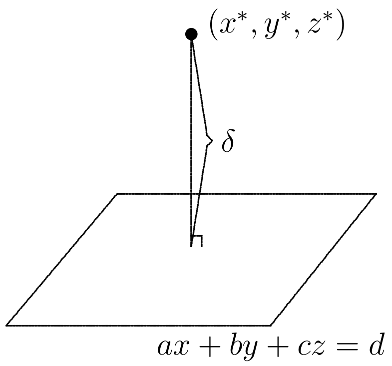
則 $\frac{|d|}{\sqrt{a^2+b^2+c^2}}$ 是原點 $(0,0,0)$ 至平面$ax+by+cz=d$之距離。N 維空間仍有相同公式

$$\vec{a}\cdot\vec{x}=|\vec{a}|p \tag{3.15}$$

$|p|$是原點$(0,\ldots,0)$至平面$\vec{a}\cdot\vec{x}=|\vec{a}|p$之距離。這種分解方法稱爲平面波 (plane wave) 是研究偏微分方程的重要方法。



圖四



圖五

3.3 面積

(3.6) 告訴我們的還不僅於此, 它同時也告訴我們由向量 $\vec{a}$、$\vec{b}$ 所形成之平行四邊形的面積 $A=|\vec{a}||\vec{c}|=\sqrt{\Gamma(\vec{a},\vec{b})}$, 當然由向量 $\vec{a}$、$\vec{b}$ 所夾成的三角形面積等於

$$\triangle=\frac{1}{2}A=\frac{1}{2}\sqrt{\Gamma(\vec{a},\vec{b})} \tag{3.16}$$

從餘弦定律來看也是很有意思的

$$\cos\theta=\frac{|\vec{a}-\vec{b}|^2-|\vec{a}|^2-|\vec{b}|^2}{2|\vec{a}||\vec{b}|}=\frac{\vec{a}\cdot\vec{b}}{|\vec{a}||\vec{b}|}$$

因此

$$\begin{aligned}
\triangle&=\frac{1}{2}|\vec{a}||\vec{b}||\sin\theta|\\
&=\frac{1}{2}\sqrt{|\vec{a}|^2|\vec{b}|^2-(\vec{a}\cdot\vec{b})^2}\\
&=\frac{1}{2}\sqrt{\Gamma(\vec{a},\vec{b})}
\end{aligned} \tag{3.17}$$

3.3 定理: 向量 $\vec{a}$、$\vec{b}$ 所夾的角爲 θ, 則

$$\begin{aligned}
\Gamma(\vec{a},\vec{b})&=|\vec{a}|^2|\vec{b}|^2\sin^2\theta\\
&=|\vec{a}|^2|\vec{b}|^2\sin^2(\vec{a},\vec{b})
\end{aligned} \tag{3.18}$$

這公式告訴我們 Gram 行列式的幾何意義: $\Gamma(\vec{a},\vec{b})$ 等於向量 $\vec{a}$、$\vec{b}$ 所夾之平行四邊形面積的平方。同理三維空間由向量 $\vec{a}=(a_1,a_2,a_3)$、$\vec{b}=(b_1,b_2,b_3)$、$\vec{c}=(c_1,c_2,c_3)$ 所夾之平行六面體 (parallepiped) 體積

$$V\equiv\begin{vmatrix}a_1&a_2&a_3\\b_1&b_2&b_3\\c_1&c_2&c_3\end{vmatrix} \tag{3.19}$$

由行列式之性質可得

$$V^2=\det\left(\begin{bmatrix}a_1&a_2&a_3\\b_1&b_2&b_3\\c_1&c_2&c_3\end{bmatrix}\begin{bmatrix}a_1&b_1&c_1\\a_2&b_2&c_2\\a_3&c_3&c_3\end{bmatrix}\right)$$

$$=\begin{vmatrix} \vec{a}\cdot\vec{a} & \vec{a}\cdot\vec{b} & \vec{a}\cdot\vec{c} \\ \vec{b}\cdot\vec{a} & \vec{b}\cdot\vec{b} & \vec{b}\cdot\vec{c} \\ \vec{c}\cdot\vec{a} & \vec{c}\cdot\vec{b} & \vec{c}\cdot\vec{c} \end{vmatrix} = \Gamma(\vec{a},\vec{b},\vec{c}) \tag{3.20}$$

換句話說$\Gamma(\vec{a},\vec{b},\vec{c})$等於向量$\vec{a}=(a_1,a_2,a_3)$、$\vec{b}=(b_1,b_2,b_3)$、$\vec{c}=(c_1,c_2,c_3)$ 所夾之平行六面體體積之平方。

由最短距離可知 $\delta=|\vec{c}\,|\le|\vec{b}|$, 所以由 (3.8) 可推得著名的 Hadamard 不等式。

3.4 定理 (Hadamard 不等式): Gram 行列式 $\Gamma(\vec{a},\vec{b})$滿足不等式

$$\Gamma(\vec{a},\vec{b})\le\Gamma(\vec{a})\Gamma(\vec{b}) \tag{3.21}$$

而且等號成立的充分必要條件是向量 $\vec{a}$、$\vec{b}$互相垂直

$$\Gamma(\vec{a},\vec{b})=\Gamma(\vec{a})\Gamma(\vec{b}) \Longleftrightarrow \vec{a}\perp\vec{b} \tag{3.22}$$

Hadamard不等式的幾何意義如下: 由向量 $\vec{a}$、$\vec{b}$ 所形成之平行四邊形的面積小於以向量 $\vec{a}$、$\vec{b}$爲兩邊所圍之長方形的面積, 而且唯有平行四邊形是長方形即 $\vec{a}\perp\vec{b}$ 時等式才成立。

註解:

(1) n 維空間的 Hadamard 不等式是

$$\Gamma(\vec{a}_1,\vec{a}_2,\ldots,\vec{a}_n)\le\Gamma(\vec{a}_1)\Gamma(\vec{a}_2)\cdots\Gamma(\vec{a}_n)$$
$$\vec{a}_i\ne 0 \quad \forall i=1,\ldots,n \tag{3.23}$$

而且等號要成立的充分必要條件是 $\vec{a}_1,\ldots,\vec{a}_n$ 兩兩要互相垂直, 其幾何意義是平行多面體的體積不超過其邊長的乘積。

3.4 Lagrange 恆等式

其實 (3.6) 就是 Lagrange 恆等式 (Lagrange identity):

3.5 定理 (Lagrange): 已知向量 $\vec{a}=(a_1,a_2,\ldots,a_n)$、$\vec{b}=(b_1,b_2,\ldots,b_n)$ 則

$$\left(\sum_{i=1}^n a_i^2\right)\left(\sum_{i=1}^n b_i^2\right)-\left(\sum_{i=1}^n a_ib_i\right)^2 = \sum_{1\le i<j\le n}(a_ib_j-a_jb_i)^2 \tag{3.24}$$

或表爲行列式

$$\Gamma(\vec{a},\vec{b})=\begin{vmatrix} \sum_{i=1}^n a_i^2 & \sum_{i=1}^n a_ib_i \\ \sum_{i=1}^n a_ib_i & \sum_{i=1}^n b_i^2 \end{vmatrix} = \sum_{1\le i<j\le n}\begin{vmatrix} a_i & a_j \\ b_i & b_j \end{vmatrix}^2 \tag{3.25}$$

證明該恆等式可以利用歸納法加上行列式的性質, 這是一個很好的題目 (如果你有足夠的耐性!) 這恆等式的幾何意義是很意思。先看 $n=2$

$$\Gamma(\vec{a},\vec{b})=(a_1^2+a_2^2)(b_1^2+b_2^2)-(a_1b_1+a_2b_2)^2 = (a_1b_2-a_2b_1)^2=\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}^2 \tag{3.26}$$

因爲行列式 $\begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix}$ 是向量 (a_1,a_2)、(b_1,b_2) 所形成之平行四邊形的面積, 因此 $\Gamma(\vec{a},\vec{b})$ 等於 (a_1,a_2),(b_1,b_2) 所圍成之平行四邊形面積的平方。

同理對任意 n 維我們可結論: $\Gamma(\vec{a},\vec{b})$ 等於由向量 $\vec{a}=(a_1,a_2,\ldots,a_n)$、$\vec{b}=(b_1,b_2,\ldots,b_n)$ 所形成之平行四邊形面積的平方且等於在所有二維座標之投影的平行四

邊形面積的平方和。對於更一般的 Gram 行列式 $\Gamma(\vec{a}_1, \vec{a}_2, \ldots, \vec{a}_n)$ 亦有相類似的公式 (請參考高等數學分析: 華羅庚著)。

註解:

(1) 向量除內積之外還有另一個重要的量 — 外積 (exterior product; vector product)。給定向量 $\vec{a} = (a_1, a_2, a_3)$、$\vec{b} = (b_1, b_2, b_3)$定義其外積

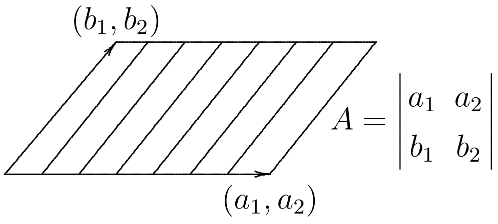
$$\begin{aligned}\vec{a} \times \vec{b} &\equiv \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix} \\ &= \left(\begin{vmatrix} a_2 & a_3 \\ b_2 & b_3 \end{vmatrix}, \begin{vmatrix} a_3 & a_1 \\ b_3 & b_1 \end{vmatrix}, \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix} \right)\end{aligned} \tag{3.27}$$

假設向量 $\vec{a}$、$\vec{b}$所夾的角爲 θ, 則

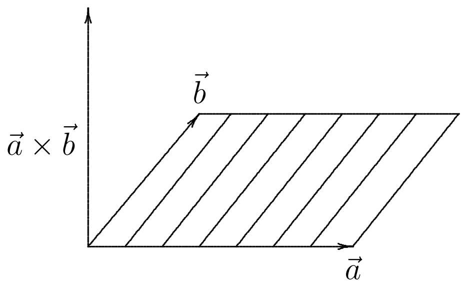
$$|\vec{a} \times \vec{b}| = |\vec{a}||\vec{b}| \sin\theta \tag{3.28}$$

所以 (3.6) 或 Lagrange 恆等式 (3.24) 告訴我們內積與外積之關係, 而其本質就是畢氏定理

$$\sin^2\theta + \cos^2\theta = 1$$



圖六



圖七

4. Cauchy-Schwarz 不等式之推廣 — Hölder 不等式:

談 Cauchy-Schwarz 不等式不得不提算數-幾何平均不等式, 因爲這是當初 Cauchy 的目的之一

$$\sqrt{ab} \le (a+b) \Longleftrightarrow ab \le \frac{1}{2}a^2 + \frac{1}{2}b^2 \tag{4.1}$$

利用這個不等式也可以證明 Cauchy-Schwarz 不等式: 令

$$\tilde{a}_i = \frac{a_i}{\sqrt{\sum_{i=1}^n a_i^2}}, \qquad \tilde{b}_i = \frac{b_i}{\sqrt{\sum_{i=1}^n b_i^2}}$$

代回 (4.1)

$$\tilde{a}_i \tilde{b}_i \le \frac{1}{2}\tilde{a}_i^2 + \frac{1}{2}\tilde{b}_i^2$$

或

$$\frac{a_i b_i}{\sqrt{\sum\limits_{i=1}^n a_i^2}\sqrt{\sum\limits_{i=1}^n b_i^2}} \le \frac{1}{2}\frac{a_i^2}{\sqrt{\sum\limits_{i=1}^n a_i^2}} + \frac{1}{2}\frac{b_i^2}{\sqrt{\sum\limits_{i=1}^n b_i^2}}$$

取有限和

$$\frac{\sum\limits_{i}^{n} a_i b_i}{\left(\sum\limits_{i=1}^n a_i^2\right)^{\frac{1}{2}} \left(\sum\limits_{i=1}^n b_i^2\right)^{\frac{1}{2}}} \le \frac{1}{2}\frac{\sum\limits_{i=1}^n a_i^2}{\sum\limits_{i=1}^n a_i^2} + \frac{1}{2}\frac{\sum\limits_{i=1}^n b_i^2}{\sum\limits_{i=1}^n b_i^2} = 1$$

因此可得

$$\sum_{i=1}^{n} a_i b_i \leq \left(\sum_{i=1}^{n} a_i^2\right)^{\frac{1}{2}} \left(\sum_{i=1}^{n} b_i^2\right)^{\frac{1}{2}}$$

因爲不等式 (4.1) 對任意a、b皆成立, 如此允許我們有極大的彈性做一些變換

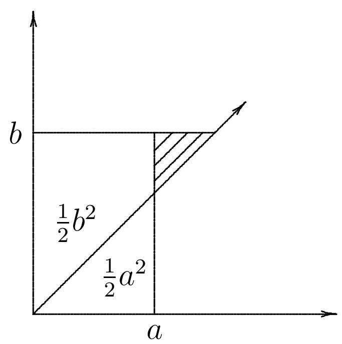
$$ab \leq \frac{\epsilon}{2}a^2 + \frac{1}{2\epsilon}b^2, \qquad \forall \epsilon > 0$$

4.1 Young's 不等式

(4.1) 不等式有時也稱爲 Cauchy 不等式, 其幾何意義如下: $\frac{1}{2}a^2$, $\frac{1}{2}b^2$ 分別表示直線 $y = x$ 與 x 軸與 y 軸所夾三角形面積。而 ab 則是長方形之面積, 從圖形上來看顯然長方形面積小於兩個三角形面積的和, 誤差項則是陰影的部分。由於是面積, 因此 (4.1) 可以表示爲積分

$$ab \leq \frac{1}{2}a^2 + \frac{1}{2}b^2 = \int_0^a x dx + \int_0^b y dy \tag{4.2}$$

而且由圖形容易"看"出來等式成立的充分必要條件是 $a = b$ (因爲$f = g$都是單位函數)。



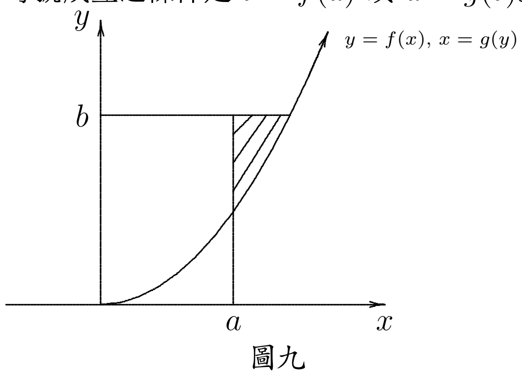
圖八

由圖形上分析, 我們總是有如此的誘惑, 不等式 (4.2) 是否能推廣至一般情形呢? 在數學上做任何的推廣之前提是對原先之結果有本質上的體會。對 (4.2) 而言 $g(y) = y$ 正好是 $f(x) = x$ 之反函數。因此 $\frac{1}{2}a^2$, $\frac{1}{2}b^2$ 分別是兩個三角形之面積。整個問題的核心就在於 f 之反函數 g 是否存在? 由函數之理論我們僅需要求 f 是遞增函數 (increasing function) 即可。

4.1 定理 (Young's 不等式): 若 $y = f(x)$ 是一遞增連續函數, 滿足 $f(0) = 0$, 而且 $x = g(y)$ 爲 f 之反函數 (因此 g 也是遞增連續函數) 則

$$ab \leq \int_0^a f(x)dx + \int_0^b g(y)dy \equiv \Phi(a) + \Phi^*(b) \tag{4.3}$$

Young's 不等式 (4.3) 的幾何意義與前面同。只是此時直線 $y = x$ 由曲線 $y = f(x)$ 所取代, 而積分 $\int_0^b f(x)dx$ $\int_0^b g(y)dy$ 分別是曲線 $y = f(x)$(或$x = g(y)$) 與 x 軸 y 軸所夾之面積陰影部分仍然是誤差項, 而且顯然等號成立之條件是 $b = f(a)$ 或 $a = g(b)$。



圖九

最常見的 Young's 不等式是當 $f(x) = x^{p-1}$, 因爲 f 必須是遞增函數, 因此要求 $p > 1$, 所以 $g(y) = y^{\frac{1}{p-1}}$,(4.3) 成爲

$$ab \leq \frac{a^p}{p} + \frac{b^{1+\frac{1}{p-1}}}{1+\frac{1}{p-1}} = \frac{a^p}{p} + \frac{b^q}{q} \tag{4.4}$$

其中 p, q 滿足關係式

$$q = 1 + \frac{1}{p-1} = \frac{p}{p-1} = \frac{1}{1-\frac{1}{p}}$$

$$\Longrightarrow \frac{1}{p} + \frac{1}{q} = 1 \tag{4.5}$$

而且等號成立的條件是$b = a^{p-1}$或$a^p = b^q$。利用不等式 (4.4) 我們馬上可得 Cauchy-Schwarz不等式的推廣。

4.2 定理 (Hölder 不等式): 給定任意實數 $a_1, a_2, \ldots, a_n, b_1, b_2, \ldots, b_n$ 且任意 $p, q \geq 1, \frac{1}{p} + \frac{1}{q} = 1$, 則

$$\sum_{i=1}^{n} a_i b_i \leq \left(\sum_{i=1}^{n} |a_i|^p\right)^{\frac{1}{p}} \left(\sum_{i=1}^{n} |b_i|^q\right)^{\frac{1}{q}} \tag{4.6}$$

證明: 仿 (4.1) 底下之證明, 取

$$\tilde{a}_i = \frac{a_i}{(\sum\limits_{i=1}^{n} |a_i|^p)^{\frac{1}{p}}}, \quad \tilde{b}_i = \frac{b_i}{(\sum\limits_{i=1}^{n} |b_i|^q)^{\frac{1}{q}}},$$

因爲 Young 不等式對任意 a, b 皆成立, 因此 (4.4) 可以改寫成底下之形式 ($\frac{1}{p} + \frac{1}{q} = 1$)

$$\begin{aligned} ab &\leq \varepsilon \frac{a^p}{p} + \epsilon^{-q/p} \frac{b^q}{q} \\ ab &\leq \frac{1}{p}(\varepsilon a)^p + \frac{1}{q}\left(\frac{b}{\varepsilon}\right)^q \\ ab &\leq \varepsilon a^{\frac{1}{\alpha}} + \left(\frac{\alpha}{\varepsilon}\right)^{\frac{\alpha}{1-\alpha}} (1-\alpha) b^{\frac{1}{1-\alpha}} \end{aligned}$$

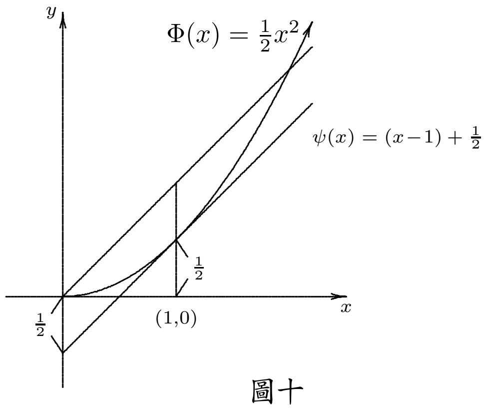
4.2 Legendre 變換

我們再從另一個角度來看 Cauchy 不等式 (4.1)

$$ab \leq \frac{1}{2}a^2 + \frac{1}{2}b^2 \Longleftrightarrow \frac{a}{b} \leq \frac{1}{2}(\frac{a}{b})^2 + \frac{1}{2} \tag{4.7}$$

因此不等式對任意實數 a、b 皆成立, 因此 (4.7) 式相當於

$$x \leq \frac{1}{2}x^2 + \frac{1}{2}, \qquad \forall x \in \boldsymbol{R} \tag{4.8}$$



圖十

我們考慮函數

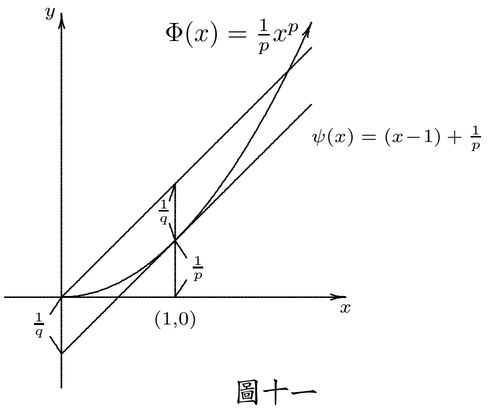
$$\Psi(x) = \frac{1}{2}x^2 - x + \frac{1}{2} \quad \forall x \in \boldsymbol{R} \tag{4.9}$$

$$\Psi'(x) = x - 1 \Longrightarrow \begin{cases} \Psi'(x) > 0, & x > 1 \\ \Psi'(x) = 0, & x = 1 \\ \Psi'(x) < 0, & x < 1 \end{cases}$$

因此 $\Psi(x)$ 之極小值 (最小值) 產生在 $x = 1$, 且 $\Psi(1) = 0$, 故

$$\Psi(x) = \frac{1}{2}x^2 - x + \frac{1}{2} \geq 0 \Longrightarrow x \leq \frac{1}{2}x^2 + \frac{1}{2}$$

再令 $x = \frac{a}{b}$ 可得 Cauchy 不等式 (4.1) 而且等號發生在 $\Psi(x)$ 之極小值 $x = \frac{a}{b} = 1$, 即 $a = b$。



圖十一

同理如果 $\Phi(x) = \frac{1}{p}x^p$, $\phi(x) = x - \frac{1}{q}$, $p > 1$; 我們考慮函數

$$\Psi(x) = \frac{1}{p}x^p - x + \frac{1}{q} \quad \forall x \in \boldsymbol{R} \quad (4.10)$$

$$\Psi'(x) = x^{p-1} - 1 \Longrightarrow \begin{cases} \Psi'(x) > 0, & x > 1 \\ \Psi'(x) = 0, & x = 1 \\ \Psi'(x) < 0, & x < 1 \end{cases}$$

故

$$\Psi(x) = \frac{1}{p}x^p - x + \frac{1}{q} \geq 0$$
$$\Longrightarrow x \leq \frac{1}{p}x^p + \frac{1}{q}$$

再令 $x = |a||b|^{-\frac{q}{p}}$ 則

$$|a||b|^{-\frac{q}{p}} \leq \frac{1}{p}|a|^p|b|^{-q} + \frac{1}{q}$$
$$\Longrightarrow |a||b| \leq \frac{1}{p}|a|^p + \frac{1}{q}|b|^q$$

而且等號發生在

$$x = |a||b|^{-\frac{q}{p}} = 1 \Longrightarrow |a|^p = |b|^q$$

由圖形之分析可知直線$x = 1$ 與$y = x$, $y = \frac{1}{p}x^p$ 相交的點是 $(1,1)$,$(1,\frac{1}{p})$。直線 $y = x$ 與曲線 $y = \frac{1}{p}x^p$ 相割之後最寬的距離是 $\frac{1}{q}$ 正好是$(1,1)$,$(1,\frac{1}{p})$兩點之距離, 因此 $\frac{1}{p} + \frac{1}{q} = 1$。

從幾何的角度而言上面兩個圖形告訴我們曲線 $y = \frac{1}{2}x^2$ (或$y = \frac{1}{p}x^p$) 永遠在直線 $y = x - \frac{1}{2}$ (或 $y = x - \frac{1}{q}$) 之上方, 且 $y = \phi(x)$ 與 $y = \Phi(x)$ 相切於 $(1,\Phi(1))$, 換言之直線 $y = \phi(x)$ 是 $y = \Phi(x)$ 之支撐線 (supporting line)。如果將切線 $y = x - \frac{1}{2}$ (或$y = x - \frac{1}{q}$) 平移為通過原點 (0,0) 之直線 $y = x$, 則 $\frac{1}{2}$(或$\frac{1}{q}$) 就是直線 $y = x$ 與 曲線 $y = \frac{1}{2}x^2$ (或$y = \frac{1}{p}x^p$) 相割之後垂直距離是寬者。如果不限定在$x = 1$, 而在任意點 $x = r$ 變動, 則支撐線為

$$\begin{aligned} y &= \Phi'(r)(x - r) + \Phi(r) \\ &= x\Phi'(r) - [r\Phi'(r) - \Phi(r)] \quad (4.11) \end{aligned}$$

此時y截距為

$$r\Phi'(r) - \Phi(r) = \max_x (x\Phi'(r) - \Phi(x)) \quad (4.12)$$

(4.11) 告訴我們 由點$(\Phi'(r), \Phi(r) - r\Phi'(r))$ 可唯一決定曲線 $y = \Phi(x)$上的點$(r, \Phi(r))$, 換句話說, 這兩個點之間可定義某種變換關係, 這就是著名的 Legendre 變換 (Legendre transformation)

$$\Phi^*(y) = xy - \Phi(x), \qquad y = \Phi'(x) \quad (4.13)$$

由 (4.12) 馬上可得 Young's 不等式

$$xy \leq \Phi(x) + \Phi^*(y) \quad (4.14)$$

我們回到前面兩個例子。若 $\Phi(x) = \frac{1}{2}x^2$ 則 $y = \Phi'(x) = x$

$$\Phi^*(y) = xy - \Phi(x) = y^2 - \frac{1}{2}y^2 = \frac{1}{2}y^2$$

而 Young's 不等式爲

$$xy \le \Phi(x)+\Phi^*(y)=\frac{1}{2}x^2+\frac{1}{2}y^2$$

同理若 $\Phi(x)=\frac{1}{p}x^p$, $y=\Phi'(x)=x^{p-1}$

$$\Phi^*(y)=xy-\Phi(x)=\frac{1}{q}y^q$$

則 Young's 不等式 (4.14) 取底下之形式

$$xy\le\Phi(x)+\Phi^*(y)=\frac{1}{p}x^p+\frac{1}{q}y^q,\ \frac{1}{p}+\frac{1}{q}=1$$

整個上面之論證所依賴的就是支撐線 (supporting line) 的概念, 因此並沒有侷限在$\Phi(x)=\frac{1}{2}x^2$。何時支撐線會存在呢 ? 我們回到 (4.11)

$$(x,\Phi(x))\longleftrightarrow(\Phi'(x),\Phi(x)-x\Phi'(x))$$

之間能建立起$1-1$ 的對應關係, 所以自然的要求就是$\Phi'(x)$ 是一增函數, 如此$\Phi'(x)$ 方可以與實數軸$\boldsymbol{R}$建立起$1-1$ 關係, 換句話說, $\Phi''>0$, 也就是Φ 是一凸函數 (convex function) 則 Young's 不等式成立。這提醒我們 Hölder 不等式可以從凸函數的角度來思考, 而其本質就是 Jensen 不等式。

我們可以刻意表示如下 (利用凸函數之性質)

$$\begin{aligned}ab&=e^{\log ab}=e^{\log a+\log b}=e^{\frac{1}{2}\log a^2+\frac{1}{2}\log b^2}\\&\le\frac{1}{2}e^{\frac{1}{2}\log a^2}+e^{\frac{1}{2}\log b^2}=\frac{1}{2}a^2+\frac{1}{2}b^2\end{aligned}\tag{4.15}$$

同理

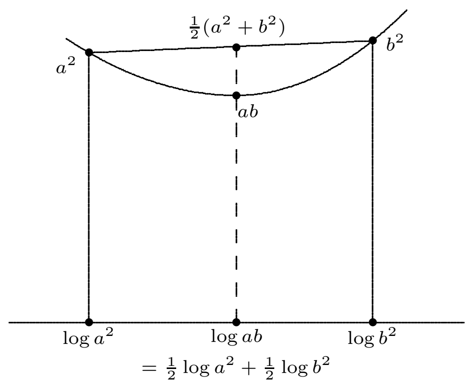
$$\begin{aligned}ab&=e^{\log ab}=e^{\log a+\log b}=e^{\frac{1}{p}\log a^p+\frac{1}{q}\log b^q}\\&\le\frac{1}{p}e^{\log a^p}+e^{\log b^q}=\frac{1}{p}a^p+\frac{1}{q}b^q\end{aligned}\tag{4.16}$$

這兩個不等式的幾何意義可以由指數函數e^x之圖形來理解, 爲了方便起見我們假設$a<b$, $\log a<\log b$ 由於e^x 是一凸函數, 所以連接 $(\log a^2,\ a^2)$、$(\log b^2,\ b^2)$ 或 $(\log a^p,\ a^p)$、$(\log b^q,\ b^q)$ 兩點的弦一定落在連接這兩點的弧 (函數e^x 圖形之一部份) 之上方。因爲我們希望不等式

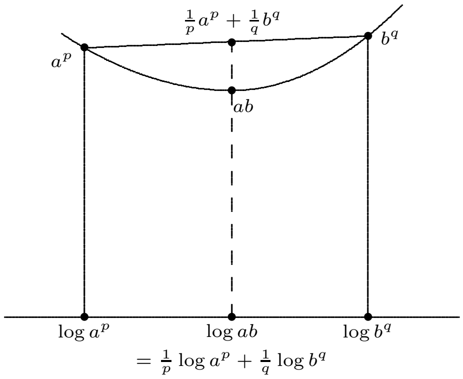
$$\log a^p\le\log ab\le\log b^q$$

成立, 因此p、q 必須滿足$p>1$、$q>1$。由分點公式 $\log ab=\frac{1}{p}\log a^p+\frac{1}{q}\log b^q$可以表爲 $\log a^p$、$\log b^q$ 的凸組合 (convex combination), 因此p、q還要滿足 $\frac{1}{p}+\frac{1}{q}=1$, 又由梯形 (包含弦之部份), 利用對應邊成比例, 所以再由分點公式可知在 ab正上方弦那點的值爲$\frac{1}{p}a^p+\frac{1}{q}b^q$, 因此可以結論

$$ab\le\frac{1}{p}a^p+\frac{1}{q}b^q$$



圖十二



圖十三

最後我們從量綱分析 (dimensional analysis) 的角度來說明為何p、q必須滿足 $\frac{1}{p}+\frac{1}{q}=1$。首先將 x 變換為 λx, 並定義函數 $f_\lambda(x)\equiv f(\lambda x)$ 則

$$\begin{aligned}\|f_\lambda\|&\equiv\Big(\int|f_\lambda(x)|^p dx\Big)^{\frac{1}{p}}\\&=\Big(\frac{1}{\lambda}\Big)^{\frac{1}{p}}\Big(\int f(\lambda x)d\lambda x\Big)^{\frac{1}{p}}=\frac{1}{\lambda^{\frac{1}{p}}}\|f\|_p\end{aligned}$$

我們刻意假設$\|f_\lambda\|_p=1$, 則

$$\|f\|_p\longleftrightarrow\lambda^{\frac{1}{p}}\longleftrightarrow\frac{1}{p}$$

同理

$$\|g\|_q\longleftrightarrow\lambda^{\frac{1}{q}}\longleftrightarrow\frac{1}{q}$$

因為

$$\begin{aligned}&\int|f_\lambda(x)g_\lambda(x)|dx\\&=\frac{1}{\lambda}\int f(\lambda x)g(\lambda x)d\lambda x\\&=\frac{1}{\lambda}\int f(y)g(y)dy\end{aligned}$$

因此 Hölder 不等式 (1.10) 要成立的話其左右兩式之量綱 (dimension) 必須相同

$$\begin{aligned}&\frac{1}{\lambda}=\Big(\frac{1}{\lambda}\Big)^{\frac{1}{p}}\Big(\frac{1}{\lambda}\Big)^{\frac{1}{q}}\\\Longleftrightarrow\ &\frac{1}{p}+\frac{1}{q}=1\end{aligned}$$

換句話說f之量綱可視為$\frac{1}{p}$, 而g 之量綱則為$\frac{1}{q}$, 積分$\int|fg|dx$ 要有意義, 也就是說積分是一實數, 而實數是一維故$\frac{1}{p}+\frac{1}{q}=1$。同理 (1.11) 左右兩式之量綱必須相同故

$$\frac{1}{p_1}+\frac{1}{p_2}+\cdots+\frac{1}{p_n}=1$$

Hölder 不等式另一個推廣之型式為

4.3 定理: 已知 $f,g\in C[a,b]$ 且 $1\le p,q,r<\infty$, $\frac{1}{p}+\frac{1}{q}=\frac{1}{r}$, 則

$$\begin{aligned}&\Big(\int_a^b|f(x)g(x)|^r dx\Big)^{\frac{1}{r}}\\\le&\Big(\int_a^b|f(x)|^p dx\Big)^{\frac{1}{p}}\Big(\int_a^b|g(x)|^q dx\Big)^{\frac{1}{q}}\end{aligned}$$

或

$$\|fg\|_r\le\|f\|_p\|g\|_q$$

這個證明是很好的習題, 我們留給讀者做練習。由量綱分析可得

$$\|f_\lambda g_\lambda\|_r=\Big(\frac{1}{\lambda}\Big)^{\frac{1}{r}}\|fg\|_r$$

所以不等式左邊的量綱 (dimension) 等於$\frac{1}{r}$, 而右邊是 $\frac{1}{p}+\frac{1}{q}$, 所以 p,q,r必須滿足關係式$\frac{1}{p}+\frac{1}{q}=\frac{1}{r}$。

最後我們要強調的是 Hölder 不等式之本質就是凸函數

4.4 定理: 定義函數

$$F(f)\equiv\ln\Big(\int_a^b e^{f(x)}dx\Big)$$

則F是一凸函數。

證明: 由 Hölder 不等式取 $p=\frac{1}{\theta}$ 則

$$
\begin{aligned}
&F(\theta f+(1-\theta)g)\\
=&\ln\Big[\int_a^b e^{\theta f(x)}e^{(1-\theta)g(x)}dx\Big]\\
\le&\ln\Big[\Big(\int_a^b e^{f(x)}dx\Big)^{\theta}\Big(\int_a^b e^{g(x)}dx\Big)^{1-\theta}\Big]\\
=&\theta F(f)+(1-\theta)F(g)
\end{aligned}
$$

註解:

(1) 從 Hölder 不等式出發可得到許多重要且有趣的不等式, 其中最出名的就是空間與算子的插值問題 (interpolation problem)。例如 Riesz convex 定理, Riesz-Torin 定理或 Riesz-Torin-Stein 定理 $\cdots$ 等。

參考資料:

1. A. Cauchy, Cours d'Analyse de Ecole Royal Polytechnique, 1821.
2. G. H. Hardy, J. E. Littlewood and G. Polya, Inequalities, Cambridge University Press, Cambridge, 1952.
3. O. Hölder, Über einen Mittelwertsatz, Gottingen Nachr., 1889.
4. F. Riesz, Untersuchungen über Systeme integrierbarer Funktionen, Math. Ann., 69 (1910).
5. H. A. Schwarz, Zur Integration der partiellen Differentialgleichung $\partial^2u/\partial x^2+\partial^2u/\partial y^2=0$, Coll. Works, Berlin, 1890.
6. T. Needham, A visual explanation of Jensen's inequality, American Math. Monthly 100 (1993), 768-771.
7. 華羅庚著, 高等數學分析, 凡異出版社, 1987。
8. 林琦焜著, 凸函數, Jensen 不等式與 Legendre 變換, 數學傳播 (中央研究院數學所), Vol. 76, p.51-57, 1995。

—本文作者現任教於國立成功大學數學系暨應用數學研究所—