凸函數、Jensen不等式與 Legendre變換

林琦焜

1. 前言

凸函數的出現絕非偶然, 在古典力學中的動能 $E=\frac{1}{2}mv^2$, 就是最自然直接的凸函數, 其他如熵 (entropy) $\cdots$ 等皆是, 當然從幾何的角度而言就是拋物線。近代分析由於受凸分析研究所得之進展的影響, 使得在非線性分析, 非線性微分方程皆有長足之進展與突破, 其中較重要的就是逐漸將非線性(nonlinearity) 視爲一個體, 而非只是線性化 (linearization) 而已。凸函數是如此地美麗且重要, 而一般教科書只是提個定義然後定理之後便是習題。對於這樣的數學我們實在不滿足也無法忍受, 畢竟數學要教導我們聰明並學習如何去思考。因此本文秉持此原則將著重於幾何與物理直觀並與一些相關聯的領域作一些對應以思索 在我們前面的那些數學巨人是如何思考問題。

2. 凸函數

我們從凸函數之定義開始

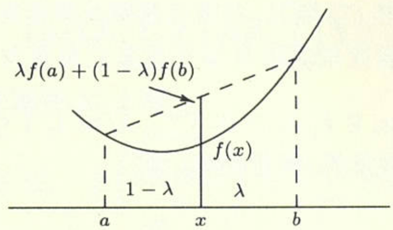
定義: f 爲一定義在區間 $I\subseteq\mathbb{R}$ 上之一實值函數 (real-valued function)

$$f:\ I\longrightarrow\mathbb{R}$$

若對任意的 $0<\lambda<1$, $a,b\in I$, f 滿足下式

$$f(\lambda a+(1-\lambda)b)\le\lambda f(a)+(1-\lambda)f(b)\quad(1)$$

則稱 f 爲一凸函數(convex function)。



圖一

其幾何意義爲連接 $(a,f(a))$, $(b,f(b))$ 兩點的弦永遠在弧 $y=f(x)$ 之上 (圖一)。

利用分點公式我們可將 (1) 式表爲下列之形式:

$$\begin{aligned}f(x)&\le\frac{b-x}{b-a}f(a)+\frac{x-a}{b-a}f(b)\\&=f(a)+\frac{f(b)-f(a)}{b-a}(x-a)\quad(2)\\&=f(b)+\frac{f(b)-f(a)}{b-a}(x-b)\end{aligned}$$

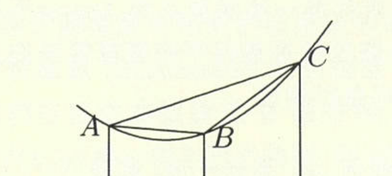
由 (2) 式可得

$$f(x)-f(a)\le\frac{f(b)-f(a)}{b-a}(x-a)$$

$$\frac{f(b)-f(a)}{b-a}(b-x)\le f(b)-f(x)$$

即

$$\frac{f(x)-f(a)}{x-a}\le\frac{f(b)-f(a)}{b-a}\le\frac{f(b)-f(x)}{b-x}\quad(3)$$



圖二

其幾何意義從圖形上之斜率可知。

我們的主要目的在於如何將 (1) 式推廣至一般情形。首先同時也是自然而然地 (在數學上2與 n 是沒有差別的) 將 (1) 式推廣至 n 個點 $x_1, \ldots, x_n$。(可用歸納法)

$$f\left(\sum_{i=1}^{n} \lambda_i x_i\right) \leq \sum_{i=1}^{n} \lambda_i f(x_i) \tag{4}$$

其中 $x_i \in I$, $\lambda_i \geq 0$, $\sum_{i=1}^{n} \lambda_i = 1$, $1 \leq i \leq n$ 有時候我們 (有目的地) 令

$$\lambda_i = \frac{p_i}{\sum_{j=1}^{n} p_j} \tag{5}$$

則 (4) 式可改寫爲

$$f\left(\frac{\sum_{i=1}^{n} p_i x_i}{\sum_{j=1}^{n} p_j}\right) \leq \frac{\sum_{i=1}^{n} p_i f(x_i)}{\sum_{j=1}^{n} p_j} \tag{6}$$

這就是 Jensen不等式之一形式。若取特殊的 p_i 例如:

$$p_i = 1, \qquad i = 1, 2, \ldots n$$

則 (6) 式可表爲

$$f\left(\frac{\sum_{i=1}^{n} x_i}{n}\right) \leq \frac{1}{n} \sum_{i=1}^{n} f(x_i) \tag{7}$$

典型的凸函數有底下之類型:

$$f(x) = x^n,\ e^x,\ x \log x,\ -\log x, \quad n > 1 \tag{8}$$

在尚未作進一步推廣之前,Jensen不等式最直接的應用就是幾何平均與算術平均之關係: 讀者可自行練習

例題1: (幾何 —— 算術平均) 試證

(a) $\left(\prod_{i=1}^{n} a_i\right)^{\frac{1}{n}} \leq \frac{1}{n} \sum_{i=1}^{n} a_i, \quad a_i > 0$

(b) $\prod_{i=1}^{n} y_i^{\alpha_i} \leq \sum_{i=1}^{n} \alpha_i y_i, \quad \alpha_i > 0, y_i > 0,$ $\sum_{i=1}^{n} \alpha_i = 1$

3. Jensen 不等式之意義

我們感興趣的問題是關於 Jensen不等式 (6) 式或 (7) 式之幾何意義與物理意義, 首先介紹 **質量中心**:

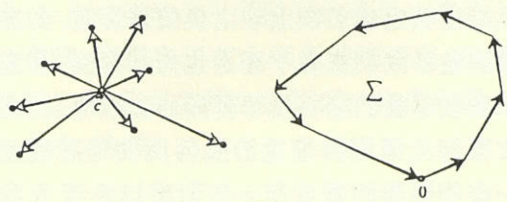
假設平面上有 n 個點且它們皆有相同之質量, 其位置向量爲 $\overrightarrow{\alpha}_i$, $1 \leq i \leq n$, 則質量中心之位置向量爲

$$\overrightarrow{c} = \frac{1}{n} \sum_{i=1}^{n} \overrightarrow{\alpha}_i \tag{9}$$

或

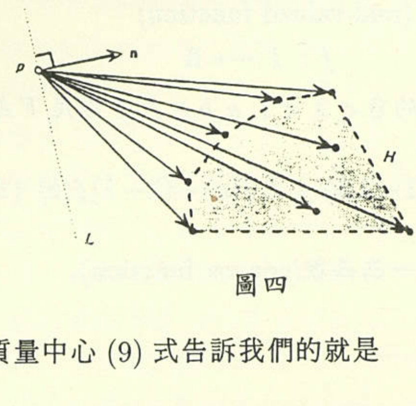
$$\sum_{i=1}^{n} \left(\overrightarrow{\alpha}_i - \overrightarrow{c}\right) = 0$$

這意思是從 $\overrightarrow{c}$ 點到各點之向量彼此互相抵消。



圖三

我們可以這麼想像: 在每一點 $\overrightarrow{\alpha}_i$ 爲一釘有木樁 而後用一條橡皮筋連接各點 $\overrightarrow{\alpha}_i$, 則如此可形成一多邊形 H (陰影區域) 而這就是 $\{\overrightarrow{\alpha}_1 \ldots \overrightarrow{\alpha}_n\}$ 的"凸包"(convex hull)。



圖四

質量中心 (9) 式告訴我們的就是

$$\overrightarrow{c} \text{ 必定落在} H \text{之內部 (陰影區域)} \tag{10}$$

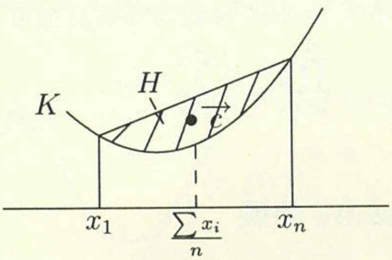
這點可由圖形直觀而得。通過任意一點 P, P 在該集合之外部, 我們可劃一直線 L 使得 H 及其所圍區域完全落在 L 之一邊。當然這些向量不可能互相抵消, 因為它們在法向量 $\overrightarrow{n}$ 上均有正的分量。

註: 上面所談的這個概念其實就是泛函分析中 Banach Separation 定理之一雛形。

有了這個預備工作之後, 我們回到原來的點:

$$(x_1, f(x_1)) \ldots (x_n, f(x_n)),$$

$$x_1 \le x_2 \le \ldots \le x_n$$



圖五

令 $K = \{(x, f(x))\}$ 為函數 f 之圖形 (graph) 同時我們也連接兩端點$(x_1, f(x_1))$,$(x_n, f(x_n))$ 則由質量中心為

$$\overrightarrow{c} = \left(\frac{\sum_{i=1}^n x_i}{n}, \frac{\sum_{i=1}^n f(x_i)}{n}\right)$$

必定落在陰影區域 H 之內部, 即

$$f\left(\frac{\sum_{i=1}^n x_i}{n}\right) \le \frac{1}{n}\sum_{i=1}^n f(x_i)$$

這就是 (7) 式, 其意義為: 質量中心 $\overrightarrow{c}$ 必定在圖形 K 之上方。而通過$(x_1, f(x_1))$,$(x_n, f(x_n))$ 兩點之弦方程式為

$$y = g(x) = \frac{f(x_n) - f(x_1)}{x_n - x_1}(x - x_1) + f(x_1) \tag{11}$$

由圖形亦知

$$\frac{1}{n}\sum_{i=1}^n f(x_i) \le g\left(\frac{\sum_{i=1}^n x_i}{n}\right) \tag{12}$$

而且對所有 $x \in I = [x_1, x_n]$ 下式成立

$$f(x) \le g(x) \tag{13}$$

這個不等式我們可視為比較定理 (Comparison 定理), 最簡單的形式, 而這在微分方程理論中扮演著舉足輕重的角色。比較 (7) 與 (12) 式各等式要成立其充分必要條件為質量中心 $\overrightarrow{c}$ 落在圖形 K 上即

$$\overrightarrow{c} \in K, \quad f\left(\frac{\sum_{i=1}^n x_i}{n}\right) = \frac{1}{n}\sum_{i=1}^n f(x_i) \tag{14}$$

這相當於

$$x_1 = x_2 = \cdots = x_n \tag{14$'$}$$

如果將 $\frac{1}{n}$ 視為 x_i 之機率分配 (一致分配), 則 Jensen 不等式 (7), 也可以用機率的角度來看

$$f(E(x)) \le E(f(x)) \tag{15}$$

E 為期望值。

對於較一般的 (6) 式其意義仍是一樣的, 即視 $x_1, \ldots x_n$ 為 n 個點但其質量分別為 p_i 而 $\sum_{i=1}^n p_i$ 為其總質量, 故有

$$f\left(\frac{\sum_{i=1}^n p_i x_i}{\sum_{j=1}^n p_j}\right) \le \frac{\sum_{i=1}^n p_i f(x_i)}{\sum_{j=1}^n p_j}$$

若視 $p_i/\sum_{j=1}^n p_j$ 為點 x_i 之機率分配, 則上式可以期望值之形式表達出來其形式與 (15) 式同。

若仔細推敲可知我們前面這些推導的過程中對維數 (dimension) 之依賴並不深, 因此我們可自然地推廣至 n 維空間。例如設 $z = f(x, y)$ 為一向上凹之曲面, 則 (7) 式可推廣為

$$f\left(\frac{\sum_{i=1}^n x_i}{n}, \frac{\sum_{i=1}^n y_i}{n}\right) \le \frac{1}{n}\sum_{i=1}^n f(x_i, y_i) \tag{16}$$

或用向量之形式 $\overrightarrow{x}_i=(x_i,y_i)$

$$f\left(\frac{1}{n}\sum_{i=1}^{n}\overrightarrow{x}_i\right)\le\frac{1}{n}\sum_{i=1}^{n}f(\overrightarrow{x}_i)\qquad(16)'$$

另一個方向的推廣則是想像粒子數目增加至無窮多個 ($n\to\infty$), 如此我們便可以從離散型過渡到連續型, 表記如下:

$$\text{有限和}\longleftrightarrow\text{無窮級數}\longleftrightarrow\text{積分}\qquad(17)$$

這就是我們在數學上尤其是分析學思想的過程而需要克服的問題是 — "收斂性", 即無窮級數或積分是否有意義 (即是否收斂)。

在區間 $[a,b]$ 我們可以取分割點

$$x_k=a+\frac{k}{n}(b-a),\quad k=0,\ldots,n\qquad(18)$$

則由 (6) 式知

$$f\left(\frac{p(x_0)\varphi(x_0)+\cdots+p(x_{n-1})\varphi(x_{n-1})}{p(x_0)+p(x_1)+\cdots+p(x_{n-1})}\right)$$
$$\le\frac{p(x_0)f(\varphi(x_0))+\cdots+p(x_{n-1})f(\varphi(x_{n-1}))}{p(x_0)+p(x_1)+\cdots+p(x_{n-1})}\qquad(19)$$

將上式表為 Riemann和之形式為

$$f\left[\frac{\sum p(x_k)\varphi(x_k)\Delta x_k}{\sum p(x_k)\Delta x_k}\right]$$
$$\le\frac{\sum p(x_k)f(\varphi(x_k))\Delta x_k}{\sum p(x_k)\Delta x_k}\qquad(20)$$

再取極限 $n\to\infty$, 我們就有積分形式的Jensen不等式。

定理 (Jensen不等式一):

若 p 滿足 $\int_a^b p(x)dx>0$, 且 f 為一凸函數, 則

$$f\left(\frac{\int_a^b p(x)\varphi(x)dx}{\int_a^b p(x)dx}\right)\le\frac{\int_a^b p(x)f(\varphi(x))dx}{\int_a^b p(x)dx}\qquad(21)$$

更一般情形則將區間 $[a,b]$ 代換為任意可測集合 A ($[a,b]\to A$)

定理 (Jensen不等式二):

$$f\left(\frac{\int_A p\varphi\,dx}{\int_A pdx}\right)\le\frac{\int_A pf(\varphi)dx}{\int_A pdx}\qquad(22)$$

讀者若有機率或測度 (measure) 之概念, 則可將 p 視為一密度函數, 故有

定理 (Jensen不等式三):

$$f\left(\frac{\int_A\varphi\,du}{\int_A du}\right)\le\frac{\int_A f(\varphi)du}{\int_A du}\qquad(23)$$

作個簡單的習題, 其實就是例題1之推廣

例題2: $\alpha_i>0$, $\xi_i>0$, $\sum_{i=1}^{\infty}\alpha_i=1$, 試證

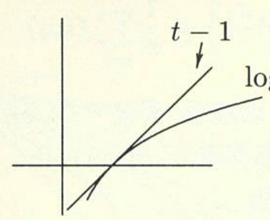
$$\prod_{i=1}^{\infty}\xi_i^{\alpha_i}\le\sum_{i=1}^{\infty}\alpha_i\xi_i。$$

4. Legendre 變換

關於 Jensen不等式之證明最簡單直接的方法就是用支撐線 (supporting line) 之概念, 而這方法在 F. Riesz 寫給 Hardy 的信中 (1930年) 就曾提過關於幾何 — 算術平均不等式的證明就是利用底下之不等式

$$\log\ t\le t-1,\qquad t>0\qquad(24)$$

這就是支撐線 (supporting line) 之概念。



圖六

若 f 為區間 $(0,1)$ 上的一個正的且可積函數, 則由 (24) 式知 ($t\to f(x)/A(f)$)

$$\log\frac{f(x)}{A(f)}\le\frac{f(x)}{A(f)}-1$$

其中 $A(f)=\int_0^1 f(x)dx$ 爲 f 之算術平均, 將上式積分一次得

$$\int_0^1 \log\ \frac{f(x)}{A(f)}dx \le \int_0^1 \frac{f(x)}{A(f)}dx-1=0$$

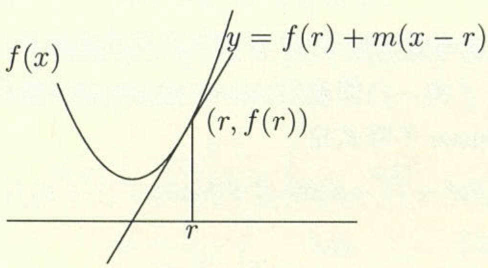
由對數函數之性質知

$$\int_0^1 \log\ f(x)dx \le \log\ A(f)=\log \int_0^1 f(x)dx$$

或者表爲

$$\underset{\text{(幾何平均)}}{\exp(\int_0^1 \log\ f(x)dx)} \le \underset{\text{(算術平均)}}{\int_0^1 f(x)dx} \quad (25)$$

仿此精神我們證明 Jensen 不等式



圖七

由圖形知 $y=f(r)+m(x-r)$, $m=f'(r)$ 爲凸函數 $f(x)$ 之支撐線 (supporting line), 即

$$f(r)+m(x-r)\le f(x) \qquad (26)$$

現在取 r 爲質量中心

$$r=\frac{\int_A p\,\varphi\,dx}{\int_A p\,dx} \qquad (27)$$

而 x 則取爲 $\varphi(x)$, 則 (26) 式成爲

$$f(r)+m(\varphi(x)-r)\le f(\varphi(x))$$

兩邊同時乘 p 並積分得

$$f(r)\int_A pdx+m(\int_A p\varphi dx-r\int_A pdx) \le \int_A f(\varphi)pdx$$

但由 r 之選法知

$$\int_A p\,\varphi\,dx-r\int_A p\,dx=0$$

故得

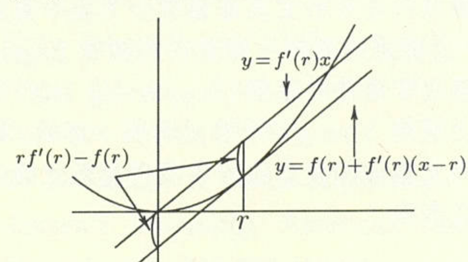
$$f(r)\int_A p\,dx\le\int_A f(\varphi)\,dx$$

這就是 Jensen 不等式。 ■

在尚未作進一步論述之前, 我們不禁要對 F. Riesz 的想法獻上我們的敬意。所謂的"好數學"便是以簡單的方法來解決困難的問題, 而不是學了很深的數學然後再說"Trivial" 簡單、容易這基本上是對數學的無知。另外一門好的數學就是其本身有"將來性"而非解完一個問題便壽終正寢。我們要特別強調的是 Riesz 所提支撐線的概念實際上就是 Legendre 變換之化身。不失一般性可設函數 f 通過原點, $f(0)=0$, 因此通過 $(r,f(r))$ 之切線方程式 (即支撐線) 爲

$$\begin{aligned} y&=f'(r)(x-r)+f(r)\\ &=x\cdot f'(r)-[rf'(r)-f(r)] \end{aligned} \qquad (28)$$

這式子告訴我們 $(f'(r),\ f(r)-rf'(r))$ 唯一決定點 $(r,f(r))$ 即這兩者之間可定義某種變換關係, 而這就是我們要談的 Legendre 變換。在還沒有正式談 Legendre 變換之前, 我們先看看 (28) 式之幾何意義。



圖八

首先將切線平移爲通過原點斜率爲 $f'(r)$ 之直線

$$y=f'(r)x \qquad (29)$$

因此 $[rf'(r)-f(r)]$ 爲直線 $y=f'(r)x$ 之 y 截距, 由圖形可知其實

$$rf'(r)-f(r)=\max_x(f'(r)x-f(x)) \quad (30)$$

即直線 $y=f'(r)x$ 與曲線 $y=f(x)$ 相割後垂直距離最寬者, 而這就是 Legendre 變換。記爲

$$f^*(p)=xp-f(x), \quad p=f'(x) \quad (31)$$

直接由 (31) 式, 即 Legendre 變換之定義可得的就是 Young's 不等式

$$xy\le f(x)+f^*(y) \quad (32)$$

一般我們所熟知的形式爲 (利用 Jensen 不等式)

$$\begin{aligned} xy &=\exp(\log\ xy)\\ &=\exp(\frac{1}{p}\log\ x^p+\frac{1}{q}\log\ y^q) \quad (33)\\ &\le\frac{1}{p}x^p+\frac{1}{q}x^q \quad (\frac{1}{p}+\frac{1}{q}=1,\ p>1)\end{aligned}$$

有時候我們可略作變化

$$x\to\epsilon^{\frac{1}{p}}x, \qquad y\to\epsilon^{-\frac{1}{p}}y$$

則 (33) 式可改寫爲

$$xy\le\epsilon\ \frac{x^p}{p}+\epsilon^{-\frac{q}{p}}\ \frac{y^q}{q} \quad (34)$$

這個技巧在分析尤其是偏微分方程中是常用的。上面這些探討主要是告訴讀者 Legendre 變換之本質是支撐線 (supporting line) 而實際上就是 Young's 不等式的另一形式。除此之外支撐線的概念也提供我們重新定義凸函數之方法:

定義: f 爲一定義在區間 $[a,b]$ 之一連續函數, 若對任意的點 $\xi\in[a,b]$, 皆存在一相應之值 $\lambda=\lambda(\xi)$, 滿足下式

$$f(\xi)+\lambda(x-\xi)\le f(x) \quad \forall x\in[a,b] \quad (35)$$

則稱 f 爲一凸函數。

這個定義可由 Taylor 展開式來看。f 在 ξ 點之 Taylor 展開式爲

$$f(x)=f(\xi)+f'(\xi)(x-\xi)+f''(\tau)(x-\xi)^2 \quad (36)$$

若 f 爲一凸函數, 則 $f''>0$ 故有

$$f(\xi)+f'(\xi)(x-\xi)\le f(x)$$

因此通常 (35) 式中之 λ 是取 $\lambda=f'(\xi)$。

5. Jensen 不等式之應用

應用一:

任給兩個正數 a,b, 其 p 階平均爲

$$N_p\equiv[\theta\,a^p+(1-\theta)b^p]^{\frac{1}{p}} \quad (37)$$

現在考慮函數 $f(x)=x^{\frac{q}{p}}$, $p<q$, 因爲 $\frac{q}{p}>1$, 故 f 爲一凸函數 (convex function)。因此由 Jensen 不等式知

$$f(\theta a^p+(1-\theta)b^p)\le\theta\,f(a^p)+(1-\theta)f(b^p)$$

即

$$\begin{aligned}[\theta a^p+(1-\theta)b^p]^{\frac{q}{p}} &\le\theta\,a^{p\cdot\frac{q}{p}}+(1-\theta)b^{p\cdot\frac{q}{p}}\\ &=\theta\,a^q+(1-\theta)b^q\end{aligned}$$

故

$$\begin{aligned}N_p &=[\theta\ a^p+(1-\theta)b^p]^{\frac{1}{p}}\\ &\le[\theta\ a^q+(1-\theta)b^q]^{\frac{1}{q}}=N_q \quad (38)\end{aligned}$$

即如果將 N_p 視爲 p 之函數, 則 N_p 爲 p 之增函數。同理可得積分形式的 p 階平均:

$$N_p[f]\equiv\left(\frac{1}{|\Omega|}\int_\Omega|f|^p\right)^{\frac{1}{p}},\ 0<|\Omega|<\infty$$

則

$$N_{p_1}[f]\le N_{p_2}[f], \quad p_1<p_2 \quad (39)$$

其中 $|\Omega|$ 表示 Ω 之面積或體積。讀者若有實變函數論的觀念, 則 (39) 式所表示的函數空間之關係爲

$$L^{p_2}(\Omega)\subseteq L^{p_1}(\Omega), \quad \frac{1}{p_2}\le\frac{1}{p_1}, \quad |\Omega|<\infty \quad (40)$$

其中函數空間 $L^p(\Omega)$ 表示 p 次方後可積分之函數所形成之集合

$$L^P(\Omega)=\left\{f\middle|\int_\Omega |f|^P dx<\infty\right\} \tag{41}$$

要特別叮嚀的是 (40) 式之關係, 只有在 $|\Omega|<\infty$ 之條件下才成立, 因爲此時質量中心才有定義。

應用二:

凸函數在二維或更高維數的空間, 例如複變函數, 所對應的便是*次調合函數* (subharmonic function)

$$\Delta u \ge 0 \tag{42}$$

對於此類函數具有非常重要地位的*平均值不等式* (mean-value inequality) 爲

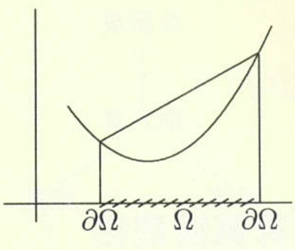
$$\begin{aligned} u(y) &\le \frac{1}{nw_n R^{n-1}}\int_{\partial B_R(y)} udS \\ &= \frac{1}{w_n R^n}\int_{B_R(y)} udx \end{aligned} \tag{43}$$

$B_R(y)$ 表示以 y 爲圓心, 半徑爲 R 之 n 維球, $\partial B_R(y)$ 則表示其球面, w_n 爲 n 維單位球之體積。(43) 式實際上就是 Jensen 不等式之一特例, 但要特別叮嚀的是 (41) 式之積分區域務必要取均勻的球, $B_R(y)$ 或球面 $\partial B_R(y)$, 因爲此時 y 是 $B_R(y)$ 或 $\partial B_R(y)$ 的質量中心。由 (43) 式可推得最大值原理 (maximum principle)。

定理 *最大值原理*(maximum principle):

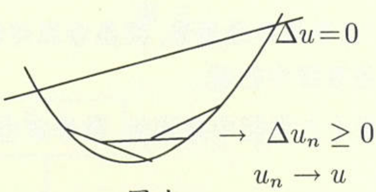
$$u\in C^2(\Omega)\cap C^\circ(\bar\Omega),\quad \Delta u\ge 0,\quad \text{則}$$
$$\max_{\bar\Omega} u=\max_{\partial\Omega} u \tag{44}$$

這定理告訴我們一個定義在有界區域 Ω 之次調合函數其最大值必定發生在邊界 $\partial\Omega$ 上。關於這件事實, 我們亦可以凸函數之性質來想像。讀者可參考底下之圖形



圖九

另外在偏微分方程中的 Laplace 方程 $\Delta u=0$, 解之存在性證明方法中的 Perron 方法, 也可由此角度來思考。



圖十

參考資料

[1] D. Gilbarg and N. S. Trudinger, *Elliptic Partial Differential Equations of Second Order,* 2nd ed., Springer-Verlag, (1983).

[2] G. H. Hardy, J. E. Littlewood and G. Pólya, *Inequalities,* Cambridge University Press, Cambridge (1952).

[3] Fritz John, *Partial Differential Equations,* 4th ed., Appl. Math. Sci., 1, Springer-Verlag, (1982).

[4] T. Needham, *A visual Explanation of Jensen's Inequality,* American Math. Monthly 100, 768-771 (1993).

—本文作者任教於成功大學數學系—