Weierstrass 逼近定理

“暫時的沒落, 不是永遠的埋沒。”

謹以此文紀念 Karl Weierstrass (1815-1897) 逝世一百週年

林琦焜

1、近代分析之父—Karl Weierstrass：

“英雄出少年”這似乎是數學史甚至是整個科學史的普遍定律, 從牛頓、Euler、 Lagrange、Laplace、高斯 ··· 等這些大數學家, 在很年輕的時候便已展露其才華。對於這些天才, 我們除了讚嘆之外, 坦白而言他們離我們這些凡夫俗子是太遙遠了。如果要像這些人的成就才算偉大的話, 那麼我們沒有幾個人的生命是有價值有意義, 更遑論我們所作一點點的數學工作。

在數學史上出現 Karl Weierstrass, 讓我們感到非常的欣慰, 因爲一個人的成功不見得是牛頓或高斯的方式, 也可以是 Weierstrass 的方式。

Karl Weierstrass 1815年生於德國的 Ostenfelde ， 大學時代學的是法律與財經, 後來改讀數學但沒有拿到博士學位。在四十歲以前都還在中學教書, 後來由於所發表的文章關於 Abel 函數理論, 才受到大學的注目, 他在普魯士高等學校任教多年之後, 才在 1856年成爲柏林大學數學系教授, 一直到去世。

正如古希臘哲學家入門的教訓 — “認識你自己”, 一個人只有在正確認識自己之後, 才有可能獲得成功。我相信 Weierstrass 是認識他自己, 不像高斯、 Cauchy、 Riemann 等數學家那麼天才橫溢有快速的直覺, Weierstrass 對數學的研究注重方法處事嚴謹。Weierstrass 的名聲主要是基於他極端細心的推論方法, 人們稱之爲 “Weierstrass 的嚴格”。

Weierstrass 早期的工作是關於超橢圓積分 (hyperelliptic integrals)、Abel 函數

與代數微分方程。但他最廣爲人知的則是利用冪級數 (power series) 來建構複變函數理論, 而其中他最有興趣的部份則是整函數 (entire function) 與函數的無窮乘積。均勻收斂 (uniform convergence) 是他發現的、平面上的解析函數稱爲整函數就是他取名的、線性代數的行列式 (determinant) 是他首次以矩陣之元素所形成的齊次多項式來定義。另外他對於雙線性型式 (bilinear form) 與二次型 (quadratic form) 也有著不可抹滅的貢獻。對二十世紀的數學而言, 他和他的學派所引介的"分析算術化"(arithmetization of analysis) 無疑是最重要的。

他從不諱言曾經在中學教過書。由於這樣的經驗, 就算是在大學教書也是一位好老師, 他的講義都是經過細心準備, 所以享有極高的聲譽, 而他的觀念主要是通過這些講義而成爲數學家共有的財富。分析中 ϵ-δ 的觀念是他所提出的。由於他作學問的才華和教書的本事吸引了很多學生到柏林來, 這些弟子中最出名的有 H. A. Schwarz(1843-1921)、Sofya Kovalevskaya (1850-1891)、G. Cantor (1845-1918)、Mittag–Leffler (1846-1927), 還有 D. Hilbert(1862-1943) —— 本世紀最偉大的數學家之一。其中最令人稱道的是, 在他那個時代破例收 Sofya Kovalevskaya 或 (Sophie Kowalevski) 爲學生 (由於柏林大學拒收女性, Sofya 無法入學且進入課堂聽講), 他自己每星期日下午撥出一段時間在家裡特別指導她學習數學, 最後並推薦她獲得德國哥庭根大學的博士學位 (Sofya Kovalevskaya 是哥庭根大學有史以來第二個女博士)。Weierstrass大概是數學史上高等數學最偉大的教師吧!

Weierstrass 對於連續函數的研究有絕對性的影響。其中之一就是他造了第一個*處處連續處處不可微分的函數*, 由於這個結果使得前面諸多對連續函數的誤解得以釐清。另外則是這篇文章所要探討的— Weierstrass 逼近定理: *任意定義在實數軸上之閉區間的連續函數都都可藉由絕對且均勻收斂的多項式表示之*。這個定理Weierstrass 在中學教書時就發現。可見環境並不是作研究的唯一決定因素, 個人的毅力與執著才是更重要的動力。

回顧一下; 在高等微積分關於一致收斂或均勻收斂 (uniformly convergent) 有一個很重要的定理: *若有一序列均勻收斂的連續函數, 則其極限必定是連續函數*。換句話說連續函數在均勻收斂下有完備性。通常我們最有把握也最有興趣的是多項式函數, 因此上面這段敘述可換爲: *若有一序列均勻收斂的多項式函數, 則其極限一定是連續函數*。由此自然而然可以問其逆敘述是否成立呢? 換句話說, 在區間 $[a,b]$ 的任意連續函數是否可以取多項式函數來逼近它呢? 這就是著名的Weierstrass *逼近定理* (*Weierstrass approximation theorem*)。

Weierstrass 逼近定理一: 任意定義在有界閉區間 $[a,b]$ 的連續函數 f, 總是可以用多項式 P_n 來逼近而且兩者之誤差

$$E_n(f)=\max_{a\le x\le b}|f(x)-P_n(x)|=||f-P_n||_\infty$$

當 $n\to\infty$ 時趨近於零。

這就是 Weierstrass 在上個世紀末證明的定理。其重要性在於保證任意的連續函數都可以用多項式來逼近, 而且可以達到任意要求的精度。這個結果對函數的近似理論有很大的幫助, 同時也說明了爲何多項式可以形成通用的近似函數之理由。

除了代數多項式之外, 另一近似函數的形式是三角函數, 或者更正確地說; 用 “Fourier 級數” 來逼近連續函數, 這也是 Weierstrass 的貢獻。

Weierstrass 逼近定理二: 任意定義在 $[-\pi, \pi]$ 且週期等於 2π 的週期連續函數 f 永遠可以找到三角多項式

$$S_n = a_0 + (a_1 \cos x + b_1 \sin x) + \cdots + (a_n \cos nx + b_n \sin nx)$$

使得兩者的誤差 $||f - S_n||$ 是任意小。

這個定理可視爲定理一的系 (corollary), 原因是我們可以利用三角變換將 S_n 轉換爲 $\cos x$ 與 $\sin x$ 的多項式。Weierstrass逼近定理的直接應用首先就是函數的近似理論 (approximation theory)。這個定理對函數論而言是個不可或缺的工具。在複數域也有類似的定理。1937年 M. H. Stone 爲了解決拓樸空間中之連續實値函數中某些代數性質之問題, 發現 Weierstrass 逼近定理之一般化, 今天我們稱爲 Stone–Weierstrass定理, 應用此定理使得一些重要的古典定理之證明變得簡單。對於 Stone–Weierstrass 定理有興趣的讀者可參閱實變函數論與泛函分析的書, 或者直接看 Stone 的文章 (請參考 [2]、[7]) 。

這文章是作者探索數學中之某一主題時經驗的記錄與讀書心得, 這是一次令人興奮之旅而我希望能把它寫得明簡, 同時也希望透過 Weierstrass 的簡介, 能再次激勵我們學習數學的心志。

『數學是打開科學大門的鑰匙, ⋯ , 輕視數學必將造成對一切知識的損害, 因爲輕視數學的人不可能掌握其它科學和理解宇宙萬物。』 — Roger Bacon —

2、Bernstein 的方法 :

Bernstein(1912) 的證明方法是建構性的 (constructive): 直接造一個多項式來逼近連續函數, 這個多項式我們現在稱爲 Bernstein 多項式。

2.1 Bernstein多項式

2.1 定義: 任意連續函數 $f \in C[0,1]$, 其對應的 Berstein 多項式如下:

$$B_n(x; f) \equiv \sum_{k=0}^{n} \binom{n}{k} f(\frac{k}{n}) x^k (1-x)^{n-k} \tag{2.1}$$

這個多項式當然不會憑空想像就出現 (靈感?) 數學最重要的是思維的方法, 如果你要等待“靈感”, 那麼你就永遠播不了種也無法收割, 無緣享受知識的樂趣, 明白這當中的奧妙。Bernstein 多項式的起源就是二項式定理:

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k} \tag{2.2}$$

上式對 x 微分後乘 $\frac{x}{n}$ 得

$$x(x+y)^{n-1}=\sum_{k=0}^{n}\binom{n}{k}\frac{k}{n}x^k y^{n-k} \quad (2.3)$$

再重複前面 (2.3) 的計算過程可得

$$\begin{aligned} & x^2(x+y)^{n-2}+\frac{xy}{n}(x+y)^{n-2} \\ =& \sum_{k=0}^{n}\binom{n}{k}(\frac{k}{n})^2 x^k y^{n-k} \end{aligned} \quad (2.4)$$

令 $y=1-x$, 則由 (2.2)-(2.4) 可分別得連續函數 1、x、x^2 的 Bernstein 多項式

$$\begin{aligned} B_n(x;1) &= \sum_{k=0}^{n}\binom{n}{k}x^k(1-x)^{n-k}=1 \\ B_n(x;x) &= \sum_{k=0}^{n}\binom{n}{k}(\frac{k}{n})x^k(1-x)^{n-k}=x \\ B_n(x;x^2) &= \sum_{k=0}^{n}\binom{n}{k}(\frac{k}{n})^2 x^k(1-x)^{n-k} \\ &= x^2+\frac{1}{n}x(1-x) \end{aligned} \quad (2.5)$$

對於第三項作個估計:

$$\left|x^2-B_n(x;x^2)\right|=\frac{x(1-x)}{n}\leq\frac{1}{4n} \quad \forall n \quad (2.6)$$

因此當 n 很大時 $B_n(x;x^2)$ 與 x^2 是非常接近, 換句話說可用 Bernstein 多項式來逼近而這正是 Bernstein 的想法。我們看特殊的函數 $f(x)=e^x$, 則

$$\begin{aligned} B_n(x;e^x) &= \sum_{k=0}^{n}\binom{n}{k}e^{\frac{k}{n}}x^k(1-x)^{n-k} \\ &= \sum_{k=0}^{n}\binom{n}{k}(e^{\frac{1}{n}}x)^k(1-x)^{n-k} \\ &= \left[e^{\frac{1}{n}}x+1-x\right]^n \\ &= \left[1+(e^{\frac{1}{n}}-1)x\right]^n\approx\left[1+\frac{x}{n}\right]^n \end{aligned}$$

我們有足夠理由相信

$$\lim_{n\to\infty}B_n(x;e^x)=\lim_{n\to\infty}(1+\frac{x}{n})^n=e^x$$

2.2 定理之證明

我們先直接相減看看:

$$\begin{aligned} & |f(x)-B_n(x;f)| \\ =& \left|f(x)-\sum_{k=0}^{n}\binom{n}{k}f(\frac{k}{n})x^k(1-x)^{n-k}\right| \\ =& \left|\sum_{k=0}^{n}\binom{n}{k}\left[f(x)-f(\frac{k}{n})\right]x^k(1-x)^{n-k}\right| \end{aligned} \quad (2.7)$$

對 (2.7) 式而言, 我們期待絕對值會趨近於 0, 乍看之下似乎 $|f(x)-f(\frac{k}{n})|$ 這項就夠了, 但這只是假象, 因爲還有另一個困難就是 (2.7) 的右式是有限項和, 這兩個困難可分別由一致有界 (uniformly bounded) 與一致連續 (uniformly continuous) 來克服。因爲$f\in C[0,1]$, 而 [0,1]是一緊緻集, 因此 f 在 [0,1]是有界且一致連續, 若$|f(x)|<M$, 則

$$\left|f(x)-f(\frac{k}{n})\right|<2M, \quad x\in[0,1] \quad (2.8)$$

另外由一致連續可知 $\forall\epsilon>0$, $\exists\delta(\epsilon)\ni$

$$\left|f(x)-f(\frac{k}{n})\right|<\frac{\epsilon}{2}, \quad \left|x-\frac{k}{n}\right|<\delta(\epsilon) \quad (2.9)$$

我們把 (2.7) 拆成兩部份

$$\begin{aligned} |f(x)-B_n(x;f)| &\leq \sum_{|x-\frac{k}{n}|<\delta(\epsilon)}+\sum_{|x-\frac{k}{n}|\geq\delta(\epsilon)} \\ &= S_1+S_2 \end{aligned} \quad (2.10)$$

由一致連續性 (2.9) 可得

$$S_1 < \frac{\epsilon}{2} \sum_{|x-\frac{k}{n}|<\delta(\epsilon)} \binom{n}{k} x^k (1-x)^{n-k}$$
$$\leq \frac{\epsilon}{2} \sum_{k=0}^{n} \binom{n}{k} x^k (1-x)^{n-k} = \frac{\epsilon}{2} \quad (2.11)$$

而利用一致有界 (2.8) 並 $|x-\frac{k}{n}| \geq \delta(\epsilon)$, 可得

$$\begin{aligned} S_2 &< 2M \sum_{|x-\frac{k}{n}|\geq\delta(\epsilon)} \binom{n}{k} x^k (1-x)^{n-k} \\ &\leq \frac{2M}{\delta^2(\epsilon)} \sum_{k=0}^{n} \binom{n}{k} (x-\frac{k}{n})^2 x^k (1-x)^{n-k} \\ &= \frac{2M}{\delta^2(\epsilon)} \Big[x^2 B_n(x;1) - 2xB_n(x;x) \\ &\quad +B_n(x;x^2)\Big] \\ &= \frac{2M}{\delta^2(\epsilon)} \left[x^2 - 2x^2 + x^2 + \frac{x(1-x)}{n}\right] \\ &\leq \frac{2M}{4\delta^2(\epsilon)n} \end{aligned} \quad (2.12)$$

因此只需要取 n 相當大,$n > \frac{M}{\delta^2(\epsilon)\epsilon}$, 則由 (2.11)-(2.12) 可得

$$|f(x) - B_n(x;f)| < \frac{\epsilon}{2} + \frac{2M}{4\delta^2(\epsilon)n} < \epsilon \quad (2.13)$$

證明完畢。

2.2 定理之意義

Weierstras 逼近定理如果從機率的角度來看是很有意思的。x爲已知, 假設由$[0,1]$這區間任意取 n 個數, 令k表示落在 $[0,x]$ 的數而 $n-k$ 是落在 $(x,1]$ 的數, 而賠率 (pay-off) 爲 $f(\frac{k}{n})$, 則 k 之機率爲

$$\binom{n}{k} x^k (1-x)^{n-k} \quad (2.14)$$

而期望值正是 Bernstein 多項式

$$B_n(x;f) = \sum_{k=0}^{n} \binom{n}{k} f(\frac{k}{n}) x^k (1-x)^{n-k} \quad (2.15)$$

或者以丟銅板爲例

$$\begin{aligned} x: &\quad \text{正面的機率} \\ 1-x: &\quad \text{負面的機率} \\ f(\frac{k}{n}): &\quad \text{付錢數} \\ B_n(x;f): &\quad \text{平均所付} \end{aligned}$$

當 n 很大時、$k \approx nx$、因此由大數法則 (law of large number) 我們期待

$$B_n(x;f) \approx f(\frac{nx}{n}) = f(x)$$

所以我們可結論: Weierstras 逼近定理正是二項分配的大數法則。

註: Bernstein多項式雖然形式比較簡單, 意義也很淸楚, 但在實際的應用上卻有極大的困難, 主要原因是收斂速度太慢, 甚至對於很好的函數, 例如可微分函數也不例外。

3、Landau 的方法:

好的定理必定伴隨不同漂亮的證明方法。Landau(1877–1938) 的證明方法不像 Bernstein 那麼直接、基本上就是 Weierstrass 的想法, 他所使用的函數以現代的語言來說就是 Dirac 序列 (Dirac sequence)。

3.1 Dirac序列

Landau 考慮底下之函數族

$$\phi_n(x)=\begin{cases} c_n(1-x^2)^n, & -1\le x\le 1\\ 0, & \text{其它}\end{cases} \tag{3.1}$$

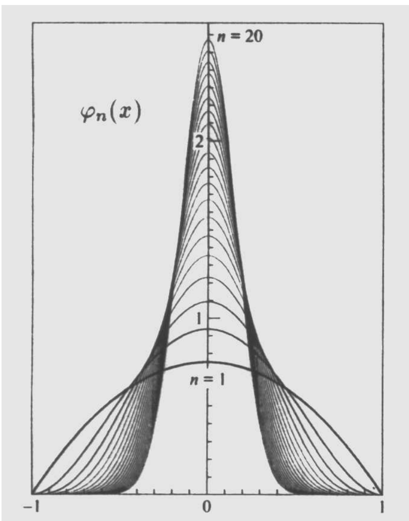
係數c_n 的選取必須滿足

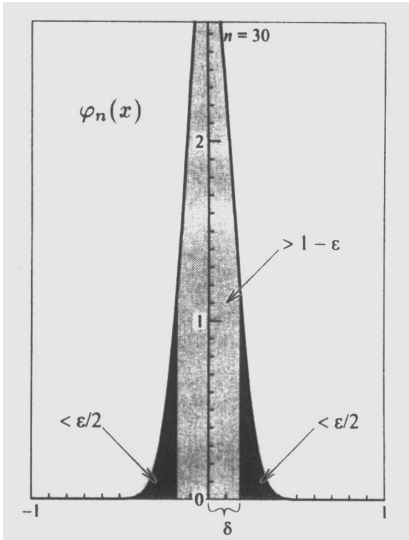
$$\int_{-\infty}^{\infty}\phi_n(x)dx=1 \tag{3.2}$$

因此由分部積分或者 Beta 函數的理論可知

$$c_n=\frac{1\cdot 3\cdot 5\cdots(2n+1)}{2\cdot 2\cdot 4\cdot 6\cdots 2n} \tag{3.3}$$

由圖形可知函數族 $\{\phi_n\}_{n=1}^{\infty}$ 有一些特性; 這些函數與 x 軸所夾的面積始終等於1, 當n越來越大的時候, 整個面積 (可視爲質量) 越往這些函數的山峰處集中好像一根釘子狀 (spike-like)。





3.1 引理: $\forall\epsilon>0,\ \ \forall\delta>0,\ \ 0<\delta<1$, 恆存在自然數 $N\in\mathbf{N}$ 使得 $\forall n\ge N$,ϕ_n 滿足

$$1-\epsilon<\int_{-\delta}^{\delta}\phi_n(x)dx\le 1 \tag{3.4}$$

$$\int_{-1}^{-\delta}\phi_n(x)dx+\int_{\delta}^{1}\phi_n(x)dx<\epsilon \tag{3.5}$$

證明方法要利用 Stirling 公式與這些函數的特性。我們並不打算給一個嚴謹的證明, 而寧願讀者有直觀的概念。

詳解: 3.1引理以另一形式表達爲

3.1′ 引理: $\forall\delta>0,\ \delta\in(0,1)$

$$\lim_{n\to\infty}\frac{\int_\delta^1(1-x^2)^n dx}{\int_0^1(1-x^2)^n dx}=0,$$

$$\lim_{n\to\infty}\frac{\int_0^\delta(1-x^2)^n dx}{\int_0^1(1-x^2)^n dx}=1$$

證明: 我們作個簡單的估計

$$\int_\delta^1(1-x^2)^n dx<(1-\delta^2)^n(1-\delta)<(1-\delta^2)^n$$

$$\int_0^1(1-x^2)^n dx > \int_0^1 (1-x)^n dx = \frac{1}{n+1},$$

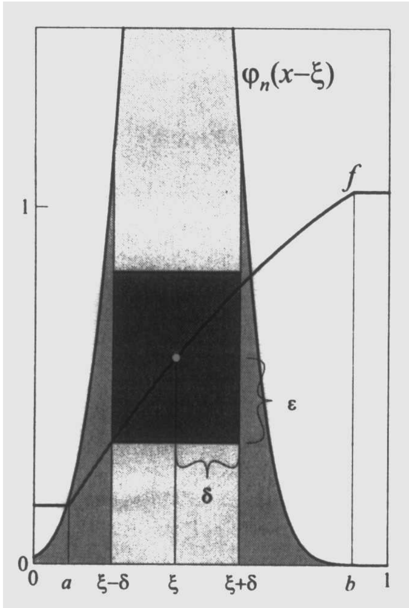
但 $\lim_{n\to\infty}(n+1)(1-\delta^2)^n = 0$ 故得證。

3.2 褶積convolution之引進

如果作個平移考慮函數 $\phi_n(x-\xi)$, $\forall \xi \in [0,1]$, 則由 ϕ_n 之特質可知此時只是將山峰由 $x=0$ 的位置平移至 $x=\xi$。所以乘積 $f(x)\cdot\phi_n(x-\xi)$ 差不多就是 $f(\xi)$ 乘上山峰的高度再利用 (3.2) 平均一下 我們可以猜測新的函數

$$P_n(\xi) \equiv \int_0^1 f(x)\phi_n(x-\xi)dx = c_n\int_0^1 f(x)[1-(x-\xi)^2]^n dx \quad (3.6)$$

與 $f(\xi)$ 是非常接近。這就是 Landau 的想法, 也是 Weierstrass 的想法。這兒所引進的就是褶積 (convolution) 的概念; 在數學分析中是非常重要的工具。如果將因式 $[1-(x-\xi)^2]^n$ 用二項式定理展開, 其實它是 ξ 的 $2n$ 次多項式; 所以可用多項式 $P_n(\xi)$ 來逼近連續函數 $f(\xi)$。



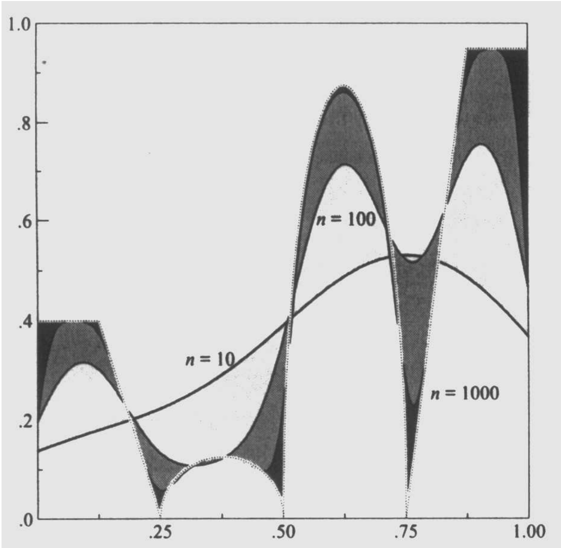
定理證明: 我們所需要的是作估計; 看 $P_n(\xi)$ 與 $f(\xi)$ 兩者間的誤差; 利用三角不等式

$$\begin{aligned}|P_n(\xi)-f(\xi)| &\le \Big|\int_0^1 f(x)\phi_n(x-\xi)dx - \int_{\xi-\delta}^{\xi-\delta} f(x)\phi_n(x-\xi)dx\Big| \\ &+\Big|\int_{\xi-\delta}^{\xi+\delta} f(x)\phi_n(x-\xi)dx - \int_{\xi-\delta}^{\xi+\delta} f(\xi)\phi_n(x-\xi)dx\Big| \\ &+\Big|f(\xi)\int_{\xi-\delta}^{\xi+\delta}\phi_n(x-\xi)dx - f(\xi)\Big| \quad (3.7)\end{aligned}$$

正如 Bernstein 的證明方法, 我們還是要用到一致有界與一致連續: 若$|f(x)| < M$ 則 (3.7) 右邊的第一式小於$M\epsilon$; 同理第三式也小於ϵ。至於第二式則由一致連續性可結論小於ϵ 故當n很大時,

$$|P_n(\xi)-f(\xi)| \le (2M+1)\epsilon$$

證明完畢。



褶積的意義

3.2 定義: f、g 爲 $\mathbf{R}$ 上的兩個函數則其褶積 (convolution) $f*g$ 定義爲 (如果積分存在)

$$f*g(x) \equiv \int f(x-\xi)g(\xi)d\xi \qquad (3.8)$$

利用變數變換與積分順序互換可容易得到褶積的基本代數運算:

3.3 定理: f、g、h 爲 $\mathbf{R}$ 上的三個函數則其褶積具有底下之關係:

(分配律) $f*(ag+bh)=a(f*g)+b(f*h)$

(交換律) $f*g=g*f$

(結合律) $f*(g*h)=(f*g)*h \qquad (3.9)$

關於褶積的意義可以如此看: 將 (3.8) 視爲 Riemann 和

$$\int f(x-\xi)g(\xi)d\xi \approx \sum f(x-\xi_j)g(\xi_j)\triangle\xi_j \qquad (3.10)$$

其中 $f(x-\xi_j)$ 是沿 x 軸平移 ξ_j 的量因此 (3.10) 表示所有f之平移量的線性組合(係數等於 $g(\xi_j)\triangle\xi_j$)。

褶積的另一個意義是—平均: 函數f在區間 $[a,b]$ 之平均值是

$$\bar{f}=\frac{1}{b-a}\int_a^b f(\xi)d\xi$$

如果有權函數 (weight funtion), 則

$$\bar{f}=\frac{1}{\int_a^b w(\xi)d\xi}\int_a^b f(\xi)w(\xi)d\xi$$

現在假設 $g \geq 0$ 且 $\int g(\xi)d\xi = 1$ 並且刻意將 $f*g$ 視爲 $\int f(\xi)g(x-\xi)d\xi$, 因此 $f*g(x)$ 表示權函數爲 $w(\xi)=g(x-\xi)$ 的平均。例如

$$g(x)=\begin{cases}\frac{1}{2a}, |x|<a\\ 0 \quad \text{其它}\end{cases}$$

則

$$f*g(x)=\frac{1}{2a}\int_{x-a}^{x+a} f(\xi)d\xi$$

我們可以把褶積視爲 —— 函數的乘積, 但是這個乘法有個最大的缺憾, 就是沒有單位元素即找不到函數 g 滿足 $f*g=f$, 這個問題的解決則有賴於 δ 函數。

3.3 Dirac Delta 函數

Landau 所引用的函數 $\phi_n(x)$, 當 $n \to \infty$ 時會變得很特殊, 與傳統上所理解的函數有些 (事實上是極大!) 的差異。用現代的語言來說就是 Dirac δ 函數; 它被介紹到科學的領域是起源於英國理論物理學家 Dirac 所研究的量子力學, 在那兒他有系統用 δ 函數作爲工具並獲得相當大的成果。

δ 函數並不是傳統上的函數, 但我們並不能因此說它不存在。數學不能否定已知的

事實, 因此解決之道就是回到數學的基礎——函數是什麼? 對函數有更新、更高層次的認識是當務之急; 爲著與傳統的函數有所區別我們將這種新的函數稱爲廣義函數 (generalized function)。

廣義函數的數學理論基礎是後來由蘇聯數學家 S. L. Sobolev(1936) 與法國數學家 L. Schwartz(1950-1951) 所建立, 其中 L. Schwartz 還因這項工作獲得第二屆 Fields 獎 (1950)。爾後經過多人的努力, 目前廣義函數己成爲研究數學、物理甚至工程的必備工具。我們正式給定 δ 函數的定義 (這是 Dirac 的貢獻)。

3.4 定義: δ 函數定義爲

$$\delta(x)=\begin{cases}+\infty, & x=0\\ 0 & x\neq 0,\end{cases}\quad \int_{-\infty}^{\infty}\delta(x)dx=1 \tag{3.11}$$

但從常理與情感上而言, 這樣子的函數怎能稱之爲函數呢? 就好比我們要測量物質在某一點的密度; 這怎麼可能呢? 我們唯一可能測量的是在某一點附近的平均密度。這就是整個問題的核心——平均。實際上廣義函數正是透過積分(也就是平均) 來定義。

考慮函數族

$$\phi_n(x)=\begin{cases}\frac{n}{2}, & |x|<\frac{1}{n},\\ 0, & 其它\end{cases} \tag{3.12}$$

則顯然

$$\int_{-\infty}^{\infty}\phi_n(x)dx=1 \tag{3.13}$$

$$\delta(x)=\lim_{n\to\infty}\phi_n(x)=\begin{cases}+\infty, & x=0;\\ 0, & x\neq 0.\end{cases} \tag{3.14}$$

但 (3.14) 是很有爭議的, 如果對它積分

$$\int_{-\infty}^{\infty}\lim_{n\to\infty}\phi_n(x)dx=0$$
$$\neq 1=\lim_{n\to\infty}\int_{-\infty}^{\infty}\phi_n(x)dx$$

顯然逐點收斂 (pointwise convergence) 在此事實上有極大的困難, 克服的方法就是引進弱收斂 (weak convergence) 的概念。

3.5 引理: 對任意的連續函數 $\psi(x)$ 下列等式成立

$$\lim_{n\to\infty}\int_{-\infty}^{\infty}\phi_n(x)\psi(x)dx=\psi(0) \tag{3.15}$$

證明: 直接相減並利用 $\int_{-\infty}^{\infty}\phi_n(x)dx=1$。

$$\Big|\int_{-\infty}^{\infty}\phi_n(x)\psi(x)dx-\psi(0)\Big|$$
$$=\Big|\int_{\frac{-1}{n}}^{\frac{1}{n}}\phi_n(x)(\psi(x)-\psi(0))dx\Big|$$
$$\leq\frac{n}{2}\int_{\frac{-1}{n}}^{\frac{1}{n}}|\psi(x)-\psi(0)|dx$$

$\forall\epsilon>0$ 取$|x|<\frac{1}{n}\leq\delta$, 使得 $|\psi(x)-\psi(0)|<\epsilon$ 則

$$\Big|\int_{-\infty}^{\infty}\phi_n(x)\psi(x)dx-\psi(0)\Big|\leq\epsilon$$

證明完畢。

這個引理告訴我們逐點看時 $\{\phi_n(x)\}$ 的極限是有問題但藉由積分 (平均) 就可以克服。典型的例子就是 $\{\sin nx\}$ 逐點收斂有問題, 但由 Riemann-Lebesgue 引理可得

$$\int_{-\pi}^{\pi}\sin nx f(x)dx\to 0\ n\to\infty$$
$$\forall f\in C[-\pi,\pi]$$

雖然逐點而言極限不存在但弱極限 (weak limit) 卻是存在的 (請參考 [3]), 換句話說弱

收斂 (weak convergence) 是研究許多數學物理的自然語言; 如果我們想對大自然有更深入的認識則第一步需改變傳統上對函數的看法。有了 (3.15) 再加上弱收斂的概念, 我們可以重新解釋 (3.14)

3.6 定義: $\phi_n(x)$ 弱收斂到 $\delta(x)$ 記爲

$$\phi_n(x) \xrightarrow{w} \delta(x), \qquad n \to \infty$$

若且唯若

$$\int_{-\infty}^{\infty} \phi_n(x)\psi(x)dx \to \psi(0) \equiv (\delta, \psi), \quad \forall \psi \in C_c(\mathbf{R}) \quad (3.16)$$

如果取$\psi = 1$ 則

$$(\delta, 1) = \int_{-\infty}^{\infty} \delta(x)dx = 1$$

(3.16) 也可以這麼看 (由$\phi_n(x)$之定義)

$$\lim_{n\to\infty} \int_{-\infty}^{x} \phi_n(\xi)d\xi = \begin{cases} 0, & x < 0 \\ 1, & x > 0 \end{cases}$$

這告訴我們 δ 函數可視爲 Heaviside 函數 $H(x)$ 之微分

$$\delta(x) = \frac{d}{dx}H(x), \ H(x) = \begin{cases} 0, & x < 0 \\ 1, & x > 0, \end{cases}$$

因此 (3.16) 可另外表示爲 Stieltjes 積分

$$\int_{-\infty}^{\infty} \psi(x)\delta(x)dx = \int_{-\infty}^{\infty} \psi(x)dH(x) = \psi(0) \quad (3.17)$$

另外由 (3.16) 再加上平移可得

$$f(x) = \int_{-\infty}^{\infty} f(\xi)\delta(x-\xi)d\xi = f * \delta(x) \quad (3.18)$$

所以 δ 函數是乘法單位元素 (在褶積的意義下)。取特殊的 $f(x) = e^{-iwx}$、e^{-xt} 則

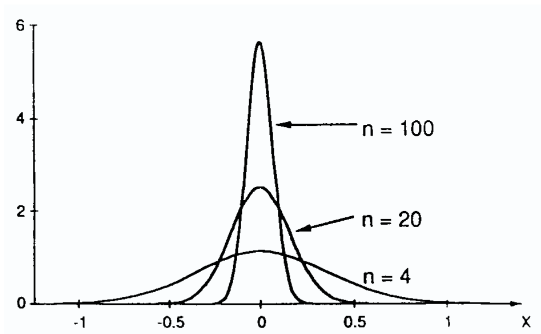
$$\int_{-\infty}^{\infty} e^{-iw\xi}\delta(x-\xi)dx = e^{-iwx},$$
$$\int_{-\infty}^{\infty} e^{-\xi t}\delta(x-\xi)dx = e^{-xt}$$

令 $x = 0$ 則 (3.19) 告訴我們δ函數的 Fourier 變換與 Laplace 變換等於1;

$$\mathcal{L}[\delta(x)] = \mathcal{F}[\delta(x)] = 1$$

這再一次印證 δ 函數所扮演的角色是乘法單位元素。

註: 以褶積爲乘法的代數結構是泛函分析一重要分枝, 稱爲算子理論, 其中最著名的有 Banach 代數 、C^* 代數與 Von Neumann 代數。



最常見的 Dirac 序列是常態分配 (normal distribution):

$$\phi_n(x) = \sqrt{\frac{n}{\pi}}e^{-nx^2}, \ \int_{-\infty}^{\infty} \phi_n(x)dx = 1$$

因此廣義函數論也稱爲荷佈理論 (distribution theory)。由 $\phi_n(x)$ 之特質可知當n很大時幾乎都集中在0點而外部爲0, 所以可對任意連續函數 $\psi(x)$ 在0點附近可視爲常數

$$\lim_{n\to\infty} \int_{-\infty}^{\infty} \phi_n(x)\psi(x)dx$$

$$\approx \int_{-\infty}^{\infty} \phi_n(x)\psi(0)dx$$
$$= \psi(0)\int_{-\infty}^{\infty} \phi_n(x)dx = \psi(0)$$

3.7 引理: $\phi_n(x) \xrightarrow{w} \delta(x)$。

證明: 這相當等於

$$\int_{-\infty}^{\infty} \delta(x)\psi(x)dx$$
$$\equiv \lim_{n\to\infty}\int_{-\infty}^{\infty} \phi_n(x)\psi(x)dx = \psi(0)$$

直接相減並利用均值定理 (mean value theorem)

$$\Big|\int_{-\infty}^{\infty}\sqrt{\frac{n}{\pi}}e^{-nx^2}\psi(x)dx - \psi(0)\Big|$$
$$= \Big|\int_{-\infty}^{\infty}\sqrt{\frac{n}{\pi}}e^{-nx^2}(\psi(x)-\psi(0))dx\Big|$$
$$\le |\psi'(x)|\Big|\int_{-\infty}^{\infty}\sqrt{\frac{n}{\pi}}e^{-nx^2}xdx\Big|$$
$$\le ||\psi'||\frac{1}{\sqrt{n\pi}} \to 0, \qquad n\to\infty$$

讀者有興趣可嘗試: 下列均為 Dirac 序列

$$\frac{1}{\pi x}\sin nx, \quad \frac{1}{\pi}\frac{n}{n^2x^2+1}$$

通常可由一個函數來建構 Dirac 序列, 為此我們有必要引進一個重要觀念。

3.8 定義 (mollifier): 已知函數 $\phi > 0$, $\phi \in C^\infty(\mathbf{R})$ 且

$$\phi(x) = 0,\ |x| > 1;\ \int_{-\infty}^{\infty}\phi(x)dx = 1 \tag{3.19}$$

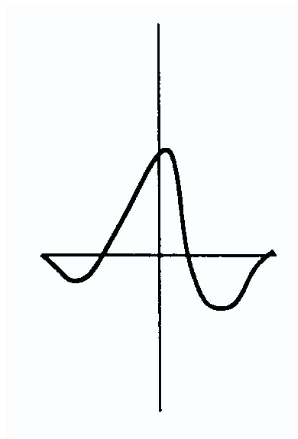
則稱 ϕ 爲緩和劑 (mollifier)。

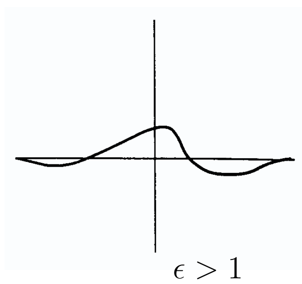
我們還可以作一下尺度 (scaling) 變換: 如果ϕ是可積函數 ($\phi \in L^1(\mathbf{R})$) 令

$$\phi_\epsilon(x) = \frac{1}{\epsilon}\phi(\frac{x}{\epsilon}) \tag{3.20}$$

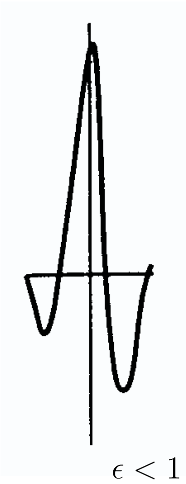
這個新函數的幾何意義是什麼呢? 當$\epsilon < 1$ 會壓縮 (compressing), $\epsilon > 1$ 則伸張 (stretching), 而係數 $\frac{1}{\epsilon}$ 的作用則是當壓縮時 ($\epsilon < 1$) 也就是變瘦了, 但還要維持一定的面積因此身高必須高一點; 同理若伸張時 ($\epsilon > 1$) 變胖了所以要長的矮一點。(如圖所示) 直接計算可得

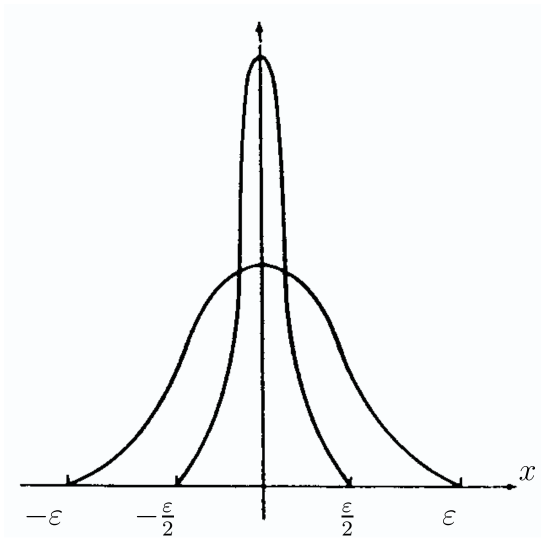
$$\int \phi_\epsilon(x)dx = \int \phi(\frac{x}{\epsilon})d(\frac{x}{\epsilon})$$
$$= \int \phi(y)dy, \quad \phi_\epsilon \xrightarrow{w} \delta \tag{3.21}$$





$\epsilon > 1$





3.9 定理: $\{\phi_\epsilon\}$ 是 Dirac 序列。

證明: 方法與引理同。由變換 $x \to \frac{x}{\epsilon}$ 可得

$$\int_{-\infty}^{\infty} \phi_\epsilon(x)dx = 1$$
$$\lim_{\epsilon\to 0}\int_{|x|>M} \phi_\epsilon(x)dx = 0 \quad \forall M > 0$$
$$\lim_{\epsilon\to 0}\int_{|x|<M} \phi_\epsilon(x)dx = 1 \quad \forall M > 0$$

另外仿照3.7引理的證明可得

$$\lim_{\epsilon\to 0}\int_{-\infty}^{\infty} \phi_\epsilon(x)\psi(x)dx = \psi(0), \quad \forall\psi \in C(\mathbf{R})$$

註: 由於這個原因我們也稱 $\{\phi_\epsilon\}$ 是近似單位元素 (approximate identity)。

最典型的 mollifier 是

$$\phi(x) = \begin{cases} c\exp\left(\frac{1}{x^2-1}\right), & |x| \le 1 \\ 0, & |x| > 1. \end{cases} \tag{3.22}$$

則由此 mollifier 可得 Dirac 序列

$$\phi_\epsilon = \begin{cases} C_\epsilon \exp\left(-\frac{\epsilon^2}{\epsilon^2-x^2}\right) & |x| \le \epsilon. \\ 0, & |x| > \epsilon. \end{cases} \tag{3.23}$$

其中 C_ϵ 滿足

$$C_\epsilon \epsilon \int_{-1}^{1} \exp\left(-\frac{1}{1-\xi^2}\right)d\xi = 1$$

其他常見的 Dirac 序列如下:

(1) Poisson 核 (Poisson kernel)

$$\phi(x) = \frac{1}{\pi(1+x^2)},\ \phi_y(x) = \frac{y}{\pi(x^2+y^2)}$$

Poisson 核在勢論 (potential theory) 有極重要之應用, 它出現在上半平面的 Dirichlet 問題, 此時的解是一調和函數 (harmonic function), 可藉由 Poisson 核與邊界值的褶積表示出來, 為著保證當

$y \to 0^+$ 時此調和函數解會趨近邊界值所以不難想像底下之等式

$$\lim_{y\to 0^+} \phi_y(x) = \delta, \quad \epsilon = y$$

(2) Gauss–Weierstrass 核 (Gauss–Weierstrass kernel)

$$\phi(x) = \frac{e^{-\frac{x^2}{4}}}{\sqrt{4\pi}},\ \phi_t(x) = \frac{e^{-\frac{x^2}{4t}}}{\sqrt{4\pi t}}$$

Gauss-Weierstrass 核是熱傳導方程的基本解 (fundamental solution), 在機率論有重要之應用, 爲著保證當 $t \to 0^+$ 時解會趨近初始値所以

$$\lim_{t\to 0^+} \phi_t(x) = \delta(x), \quad \epsilon = \sqrt{t}$$

(3)

$$\phi(t) = H(t)\frac{e^{-\frac{1}{4t}}}{\sqrt{4\pi}t^{\frac{3}{4}}},$$

$$\phi_x(t) = H(t)\frac{xe^{-\frac{x^2}{4t}}}{\sqrt{4\pi}t^{\frac{3}{4}}},$$

$$H:\ \text{Heaviside 函數}$$

$$\lim_{x\to 0^+} \phi_x(t) = \delta(t), \qquad \epsilon = x^2$$

(4) Fejer 核 (Fejer kernel)

$$\phi(x) = \frac{\sin^2 x}{\pi x^2}, \quad \phi_R(x) = \frac{\sin^2 Rx}{\pi R x^2}$$

$$\lim_{R\to\infty} \phi_R(x) = \delta(x), \quad \epsilon = \frac{1}{R}$$

Fejer核 (Fejer kernel) 是由匈牙利數學家 Fejer 在研究 Fourier 級數時, 以算術平均得其收斂性而來的。

不用 mollifier 也可以得到 Dirac 序列:

(1) Dirichlet 核 (Dirichlet kernel)

$$\phi(x) = \frac{1}{\pi x}\sin x,$$

$$\lim_{n\to\infty} \phi_n(x) = \lim_{n\to\infty} \frac{\sin nx}{\pi x} = \delta(x)$$

Dirichlet核的出現是在於處理 Fourier 級數的收斂性問題。

(2) Landau 的證明方法就是用這個 Dirac 序列

$$\phi_n(x) = \begin{cases} (1-x^2)^n/I_n, & |x| < 1. \\ 0, & |x| \ge 1 \end{cases}$$

$$\lim_{n\to\infty} \phi_n(x) = \delta(x),$$

$$I_n = \int_{-1}^{1} (1-x^2)^n dx$$

這個序列與 Legendre 多項式$P_n(x) = \frac{1}{2^n n!}\frac{d^n}{dx^n}(x^2-1)^n$ 有極密切之關係。

(3) Poisson 核 (Poisson kernel)

$$\phi_r(\theta) = \begin{cases} \frac{1}{2\pi}\left(\frac{1-r^2}{1+r^2-2r\cos\theta}\right) & |\theta| \le \pi \\ 0 & |\theta| > \pi. \end{cases}$$

這個形式的 Poisson 核出現在單位圓的 Dirichlet 問題, 爲著保證此調和函數解在單位圓會趨近邊界值, 所以

$$\lim_{r\to 1^-} \phi_r(\theta) = \delta$$

4、Weierstrass的方法:

Weierstrass 的證明方法不像 Bernstein 或 Landau 那麼直接, 中間多了一道手續—— 平滑化 (regularization)。f爲任意可積函數, 則其平滑化爲

$$f_\epsilon(x) = f * \varphi_\epsilon(x) \tag{4.1}$$

φ 爲一 mollifier,φ_ϵ 如 (3.20)。

4.1 引理: f 爲有界區間 $[a,b]$ 上的連續函數, 則 f_ϵ 一致收斂到 f

$$\lim_{\epsilon\to 0} f_\epsilon(x)=\lim_{\epsilon\to 0} f*\varphi_\epsilon(x)=f(x), \quad \forall x\in[a,b] \qquad (4.2)$$

證明: 由定義

$$f_\epsilon(x)=\frac{1}{\epsilon}\int_{|x-y|<\epsilon}\varphi(\frac{x-y}{\epsilon})f(y)dy$$
$$=\int_{|z|\le 1}\varphi(z)f(x-\epsilon z)dz \qquad (z=\frac{x-y}{\epsilon})$$

但 $[a,b]$ 是緊緻區間,f是一致連續顯然

$$\sup_x|f(x)-f_\epsilon(x)|$$
$$\le\sup_x\int_{|z|\le 1}\varphi(z)|f(x)-f(x-\epsilon z)|dz$$
$$\le\sup_x\sup_{|z|\le 1}|f(x)-f(x-\epsilon z)|\to 0 \ \epsilon\to 0$$

故f_ϵ一致收斂到f。

註: 4.1 引理在分析中扮演著非常重要的角色, 甚至 f 只是可積函數 (L^1) 或者 p 次方可積函數 (L^p) 都是成立, 有興趣的讀者可參閱 [5]、[8]。

4.2 定理 (Weierstrass): $f\in C[a,b]$, 則 $\forall\delta>0,\exists$ 多項式 P 滿足

$$\sup_{a\le x\le b}|f(x)-P(x)| \equiv \max_{a\le x\le b}|f(x)-P(x)|<\delta \qquad (4.3)$$

證明: Weierstrass選取的是

$$\varphi(x)=\frac{1}{\sqrt{\pi}}e^{-x^2}$$

直接微分, φ 具有底下之性質

$$\varphi^{(k)}(x)=q_k(x)e^{-x^2}, \quad q_k \text{ 爲 } k \text{ 次多項式}$$
$$|\varphi^k(x)|\le C_k e^{-|x|}$$

由4.1引理知

$$f_\epsilon=f*\varphi_\epsilon\to f \quad \text{一致收斂}$$

$$\sup_{a\le x\le b}\Big|f(x)-\frac{1}{\epsilon\sqrt{\pi}}\int_a^b e^{-(x-y)^2/\epsilon^2}f(y)dy\Big|<\frac{\delta}{2} \qquad (4.4)$$

另一方面由於 $[a,b]$ 是緊緻集, 故 e^{-x^2} 的 Taylor 級數 $\sum_{n=0}^\infty\frac{(-1)^n x^{2n}}{n!}$ 一致收斂到 e^{-x^2}, 換句話說, 存在多項式 $P(x)$ 滿足

$$\sup_{a\le x\le b}|f_\epsilon(x)-P(x)|\le\frac{\delta}{2} \qquad (4.5)$$

由 (4.4)、(4.5) 可得

$$\sup_{a\le x\le b}|f(x)-P(x)|<\delta$$

證明完畢。

Weierstrass 逼近定理的意義

1. 分析:

我們在高等微積分研究的主題是連續函數, 所用到的收斂是均勻收斂或一致收斂 (uniform convergence)。換句話說, 我們所探討的空間是

$$(C[a,b],||\cdot||) \quad ||f||\equiv\max_{x\in[a,b]}|f(x)|$$

令 $P[a,b]$ 代表所有多項式所形成之空間, 則 Weierstrass 逼近定理告訴我們任意的連續函數都能以多項式近似之, 而且可以達到任意要求的精度。這情形正如有理數在實數中一樣, 對於任意的實數 $r\in\mathbf{R}$ 永遠可以找到有理數 $q\in\mathbf{Q}$ 使得 $|r-q|<\epsilon$。同理

如果f是任意的連續函數 ($f \in C[a,b]$), 總是存在多項式 $p_n(x) \in P[a,b]$ 滿足 $||f-p_n|| < \epsilon$。所以 Weierstrass 定理告訴我們多項式函數在連續函數內是稠密的 (dense), 換個角度而言 (用分析的語言)$(C[a,b], \|\cdot\|)$ 是 $(P[a,b], \|\cdot\|)$ 的完備化 (completion)

$$\begin{array}{ccc} (\mathbf{Q}, |\cdot|) & \longrightarrow & (\mathbf{R}, |\cdot|) \\ (P[a,b], \|\cdot\|) & \longrightarrow & (C[a,b], \|\cdot\|) \end{array}$$

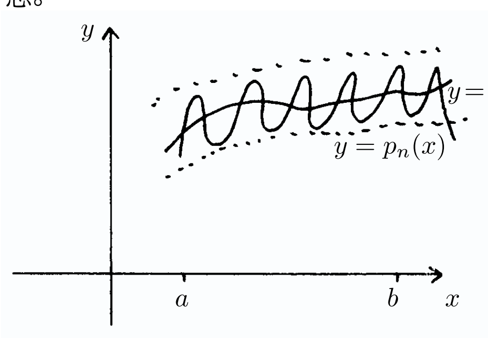
2. 幾何:

Weierstrass 逼近定理從圖形上來看是相當明確的, 已知任意連續函數 f 與任意正數 ϵ, 我們可以取一寬度爲2ϵ的帶狀區域使得 f 通過該區域之中線 (你可以想像成一個口徑爲 2ϵ 的水管將 f 包在裡面), 我們恆可以取得次數夠高之多項式

$$p_n(x) = a_0 + a_1 x + \cdots + a_n x^n$$

使得 $p_n(x)$ 之圖形全部落在帶狀區域之內。

在這裡値得一提的是帶狀區域之 ϵ 對於任意點 $x \in [a,b]$ 都是一樣的。因此也不難想像 Weierstrass 是如何發現均勻收斂的觀念。



附記: 中研院數學所劉豐哲教授於1987年在數學傳播十一卷二期 p.44-47 之著作:「弱大數法則, Bernstein 多項式及 Weierstrass 逼近定理」。對於 Weierstrass 定理在機率上的意義有詳盡的討論, 有興趣的讀者可參閱該文。

參考資料:

1. 數學之內容方法及意義 (I,II,III), 徐氏基金會出版 (1970)。
2. 近代分析之研究, 徐氏基金會出版 (1971)。
3. 林琦焜, Riemann-Lebesgue 引理與弱收斂, 數學傳播季刊, 81, pp.17-28, (1997)。
4. Howard Eves, *An Introduction to the History of Mathematics*, (1976).
5. Gerald B. Folland, *Fourier Analysis and Its Applications*, Wadsworth & Brooks/Cole Mathematics Series, Pacific Grove, California, (1992).
6. E. Hairer and G. Wanner, *Analysis by Its History*, UTM : Reading in Mathematics, Springer-Verlag (1995).
7. Jerrold E. Marsden, *Elementary Classical Analysis*, Marcel Dekker, New York (1973).
8. R. Wheeden and A. Zygmund, *Measure and Integral*, Marcel Dekker, New York (1977).

—本文作者任教於國立成功大學數學系暨應用數學研究所—