東 海 大 學 數 學 研 究 所

碩 士 論 文

再生函數配置法之擾動性及穩定性之研究

The Perturbation and Stability Study of Reproducing Kernel Collocation Method

指導教授：胡馨云 博士

研究生：張智凱

中 華 民 國 九 十 九 年 三 月

Abstract

Solving partial differential equations with strong form collocation and nonlocal approximation functions such as orthogonal polynomials and radial basis functions exhibits exponential convergence rate; however, it yields a full matrix and suffers from ill conditioning. In this work, the local approximation functions, reproducing kernel functions, are used as basis functions. This approach offers algebra convergence rate, but the method is stable like the finite element.

We provide the perturbation and stability analysis of this approach, and the estimation of condition number of the discrete equation is derived. Condition number is used to measure the solution errors resulting from rounding errors, and it plays a critical role in numerical stability. In addition, the new formulas of condition number, called effective condition numbers, are given. Both matrix and right hand side vector of a linear system are taken into consideration in the estimation of condition number, they offer a better measure of conditioning than traditional condition numbers. Numerical results are also presented to validate the mathematical analysis.

摘　要

求解偏微分方程當採用配置法配上非局部函數當基底，例如正交多項式和徑向基函數時得出指數型收斂行為。然而，它產生一個滿矩陣且 condition 過大。此論文中使用配置法配局部基底函數，即再生核心函數，結果為代數收斂行為。此方法類似有限元素法。

本論文主旨為擾動性和穩定性分析，並估計離散方程的 condition number。Condition number 是用來估計解的相對誤差上界，它在數值穩定上扮演一個關鍵的角色。另外，論文中介紹新 condition number 公式，被稱作 effective condition number。此新估計式中線性系統中的矩陣及右端向量兩者都被考量在 condition number 的估計上，它們提供一個比傳統 condition number 更好的測量條件。數值結果也證實了數學分析。

誌　謝

完成了這篇論文，首先我要感謝這幾年來一直指導我的胡馨云教授，除了教導我研究論文的方向和做學問應該有的態度，也不斷給我意見且指出該修正的地方，和撰寫程式的步驟及要點，另外我亦學到研究學問應該有自己的一個明確目標和不斷求新求進步的心態，在遇到研究瓶頸時要更加努力去找出問題癥結所在，培養自己的觀察力和抗壓力，最後完成自己的一篇論文。

此外也要感謝黃皇男教授和袁淵明教授，給我的論文許多的意見和需修正的地方，提醒我一些沒有注意到的地方。最後要感謝家人及朋友默默的支持我，在精神上給我很大的支持，讓我得以完成這篇論文。

符號表

$\psi_I(x)$	再生核心基底函數
$C(x;x-x_I)$	修正函數
x_I	質點
ξ_i	配置點
S	質點集
h	最大的質點間距
N_p	質點數
N_c	配置點數
$\phi_a(x-x_I)$	核心函數
a	核心函數的半徑
n	再生基底函數的階數
δ_{ij}	Kronecker delta
Ω	定義域
Γ	Ω 之邊界
$\mathrm{E}(\bullet)$	連續泛函
$\hat{\mathrm{E}}(\bullet)$	離散泛函
$\|\cdot\|$	範數
A	N_p 乘 N_p 大小的矩陣
$\mathbf{b}$	N_p 維度的已知向量
F	N_c 乘 N_p 大小的矩陣
$\mathbf{r}$	N_c 維度的已知向量
P	投影算子
F^+	F 的偽逆矩陣
$\|v\|_{2,\Omega}$	標準 Sobolev 二範數
$\mathrm{Cond}(\bullet)$	矩陣的 condition number
$\mathrm{Cond\_E}(\bullet)$	矩陣的 effective condition number
β_n	線性系統 $A\mathbf{x}=\mathbf{b}$ 中，矩陣 A 對角化中第 n 個特徵向量與 b 的內積

目錄

一、 前言 ····· 1

二、 再生核心逼近理論 ····· 2

（一）再生核心函數介紹及推導 ····· 2
（二）再生核心函數的高階微分 ····· 4
（三）再生核心函數之基本性質 ····· 6
（四）再生核心函數之反估計式 ····· 10

三、 擾動性及穩定性之探討 ····· 11

（一）函數逼近 ····· 11
（二）偏微分方程解 ····· 14

四、 穩定性之深入探討 ····· 20

（一）新估計方式 ····· 20
（二）數值範例 ····· 21
（三）Effective condition number 之推導 ····· 24

五、 結論 ····· 30

參考文獻 ····· 31

附錄 A ····· 32

附錄 B ····· 35

第一章 前言

近二十年來，無網格法[1,2]已成為計算力學界的新趨勢，無網格法不需建立網格，僅需一組質點，以此組質點建構出 local 或 global 的基底函數。基底的建構快速且簡單，而且搭配調適法來解移動不連續問題、衝擊波問題或多尺度問題比傳統網格法更有效率。

前面提到無網格法的基底可為 local 型態也可為 global 型態，各有優缺點。當我們選用 global 正交函數[3]或徑向基函數[4,5]作為基底，可得到指數型的收斂行為，但穩定性就比較差，condition number 指數成長。若選用 local 核心函數[6]則可得到代數型的收斂行為，其穩定性就比較好，condition number 之呈現是代數成長。

在此論文中選用 local 函數基底，即再生核心基底函數，來探討穩定性。先作擾動分析，再觀察 condition number 的變化。我們利用泛函分析方法得出傳統 condition number 的界，數值範例之結果與理論分析相穩合。而收斂性及複雜性已探討過，請參見相關論文[7]。

本文第二章介紹再生核心函數及其導函數、基本性質和反估計式。第三章針對方陣型系統（函數逼近）及非方陣型系統（求解微分方程解）作擾動分析，並估計傳統 condition number 的上界，第四章介紹新的 condition numbers，有別於傳統 condition numbers，具更客觀更好的穩定分析。最後，第五章為結論。

第二章 再生核心逼近理論

第一節 再生核心函數介紹及推導

假設連續函數 $f(x)$ 可由下面的近似函數 $f^h(x)$ 逼近

$$f^h(x)=\sum_{I=1}^{Np} d_I\psi_I(x) \tag{2.1}$$

此近似函數是由一組局部的基底函數線性組合而成，這組基底函數由下列質點集

$$S=\{x_I\}_{I=1}^{Np}=\{x_1,x_2,\ldots,x_{Np}\}$$

為中心建構而成，最大的質點間距是 h ，函數形式如下[8]

$$\psi_I(x)=H^T(0)M^{-1}(x)H(x-x_I)\phi_a(x-x_I) \tag{2.2}$$

此種基底稱之為『再生核心基底函數』，其中矩陣及向量為

$$M(x)=\sum_{I=1}^{Np} H(x-x_I)H^T(x-x_I)\phi_a(x-x_I) \tag{2.3}$$

$$H^T(x-x_I)=[1,x-x_I,(x-x_I)^2,\ldots,(x-x_I)^n] \tag{2.4}$$

$$H^T(0)=[1,0,\ldots,0] \tag{2.5}$$

函數 $\phi_a(x-x_I)$ 稱為核心函數（kernel function）可選用三次 B-spline

$$\phi_a(z)=\begin{cases}\frac{2}{3}-4z^2+4z^3, & 0\le z<\frac{1}{2}\\ \frac{4}{3}-4z+4z^2-\frac{4}{3}z^3, & \frac{1}{2}\le z<1\\ 0, & z\ge 1\end{cases},\qquad z=\frac{|x-x_I|}{a} \tag{2.6}$$

或五次 B-spline

$$\phi_a(z)=\begin{cases}\frac{11}{20}-\frac{9z^2}{2}+\frac{81z^4}{4}-\frac{81z^5}{4}, & 0\le z<\frac{1}{3}\\ \frac{17}{40}+\frac{15z}{8}-\frac{63z^2}{4}+\frac{135z^3}{4}-\frac{243z^4}{8}+\frac{81z^5}{8}, & \frac{1}{3}\le z<\frac{2}{3}\\ \frac{81}{40}-\frac{81z}{8}+\frac{81z^2}{4}-\frac{81z^3}{4}+\frac{81z^4}{8}-\frac{81z^5}{40}, & \frac{2}{3}\le z<1\\ 0, & z\ge 1\end{cases},\ z=\frac{|x-x_I|}{a} \tag{2.7}$$

其中 a 表示核心函數的半徑。此半徑可隨不同位置改變，可令 x_I 位置之核心函數的半徑為 a_I，此組核心函數的半徑大小不可太過懸殊（稱之為擬一致分佈）。

接下來將介紹如何推導出形式如(2.2)的再生核心基底函數。

假設基底函數由修正函數 $C(x;x-x_I)$ 及核心函數 $\phi_a(x-x_I)$ 組合而成

$$\psi_I(x)=C(x;x-x_I)\phi_a(x-x_I),\qquad x_I\in S \tag{2.8}$$

修正函數再經由多項式函數建構出來，可進一步表示成向量內積形式如下

$$C(x;x-x_I)=\sum_{i=0}^{n}(x-x_I)^i b_i(x)=:H^T(x-x_I)b(x) \tag{2.9}$$

其中 $b_i(x)$ 為係數，稍後可推導出。

基底函數(2.8)滿足下列再生性（reproducing conditions）

$$\sum_{I=1}^{Np}\psi_I(x)\,x_I^i=x^i,\ \ i=0,1,\ldots,n \tag{2.10}$$

或寫成

$$\sum_{I=1}^{Np}C(x;x-x_I)\,\phi_a(x-x_I)\,x_I^i=x^i,\ \ i=0,1,\ldots,n \tag{2.11}$$

由數學歸納法可證明(2.11)式是等價於下列式子

$$\sum_{I=1}^{Np}C(x;x-x_I)\,\phi_a(x-x_I)\,(x-x_I)^i=\delta_{i0},\ \ i=0,1,\ldots,n \tag{2.12}$$

其中 δ_{ij} 是 Kronecker delta。

可進一步將(2.12)寫成向量形式，如下

$$\sum_{I=1}^{Np}C(x;x-x_I)\phi_a(x-x_I)H(x-x_I)=H(0) \tag{2.13}$$

若將修正函數(2.9)放入式(2.13)中，可得到

$$\sum_{I=1}^{Np}H(x-x_I)\,H^T(x-x_I)\phi_a(x-x_I)\,b(x)=H(0) \tag{2.14}$$

簡記為矩陣形式

$$M(x)b(x)=H(0) \tag{2.15}$$

其中 $M(x)$ 矩陣為

$$M(x)=\begin{bmatrix} \sum_{I=1}^{Np}\phi_a(x-x_I) & \sum_{I=1}^{Np}(x-x_I)\phi_a(x-x_I) & \cdots & \sum_{I=1}^{Np}(x-x_I)^n\phi_a(x-x_I) \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ \sum_{I=1}^{Np}(x-x_I)^n\phi_a(x-x_I) & \sum_{I=1}^{Np}(x-x_I)^{n+1}\phi_a(x-x_I) & \cdots & \sum_{I=1}^{Np}(x-x_I)^{2n}\phi_a(x-x_I) \end{bmatrix}$$

其規模為 $(n+1)\times(n+1)$，通常 n 取 2 或 3，因此反矩陣很容易被計算出來。

從式(2.15)，我們得到係數向量，如下

$$b(x)=M^{-1}(x)H(0) \tag{2.16}$$

再將此結果代入修正函數(2.9)式中，得到

$$\begin{aligned} C(x;x-x_I) &= H^T(x-x_I)M^{-1}(x)H(0) \\ &= H^T(0)M^{-1}(x)H(x-x_I) \end{aligned} \tag{2.17}$$

因此，逼近函數(2.1)變成

$$\begin{aligned} f^h(x) &= \sum_{I=1}^{Np} d_I\, C(x;x-x_I)\phi_a(x-x_I) \\ &= \sum_{I=1}^{Np} d_I\, H^T(x-x_I)b(x)\,\phi_a(x-x_I) \\ &= \sum_{I=1}^{Np} d_I\, H^T(x-x_I)M^{-1}(x)H(0)\,\phi_a(x-x_I) \\ &= \sum_{I=1}^{Np} d_I\, H^T(0)M^{-1}(x)H(x-x_I)\phi_a(x-x_I) \\ &=: \sum_{I=1}^{Np} d_I\,\psi_I(x) \end{aligned}$$

其中基底函數定義如下

$$\psi_I(x)=H^T(0)M^{-1}(x)H(x-x_I)\,\phi_a(x-x_I) \tag{2.18}$$

其中心點為 x_I，核心函數半徑為 a 之局部函數。

第二節 再生核心函數的高階微分

前一節已介紹核心基底函數由修正函數及核心函數組成

$$\psi_I(x) = C(x; x - x_I)\phi_a(x - x_I) \tag{2.19}$$

所以，再生核心基底函數其一階及二階微分可根據乘法原理得到

$$\psi_{I,x}(x) = C_{,x}(x; x - x_I)\phi_a(x - x_I) + C(x; x - x_I)\phi_{a,x}(x - x_I) = \psi'_I(x) \tag{2.20}$$

$$\begin{aligned}\psi_{I,xx}(x) = {} & C_{,xx}(x; x - x_I)\phi_a(x - x_I) + 2C_{,x}(x; x - x_I)\phi_{a,x}(x - x_I) \\ & + C(x; x - x_I)\phi_{a,xx}(x - x_I) = \psi''_I(x)\end{aligned} \tag{2.21}$$

修正函數及其導函數分別為

$$C(x; x - x_I) = H^T(0)M^{-1}(x)H(x - x_I) \tag{2.22}$$

$$C_{,x}(x; x - x_I) = H^T(0)M^{-1}{}_{,x}(x)H(x - x_I) + H^T(0)M^{-1}(x)H_{,x}(x - x_I) \tag{2.23}$$

$$\begin{aligned}C_{,xx}(x; x - x_I) = {} & H^T(0)M^{-1}{}_{,xx}(x)H(x - x_I) + 2H^T(0)M^{-1}{}_{,x}(x)H_{,x}(x - x_I) \\ & + H^T(0)M^{-1}(x)H_{,xx}(x - x_I)\end{aligned} \tag{2.24}$$

其中 $M(x)$ 矩陣及其導函數如下

$$M(x) = \sum_{I=1}^{NP} H(x - x_I)H^T(x - x_I)\phi_a(x - x_I) \tag{2.25}$$

$$\begin{aligned}M_{,x}(x) = \sum_{I=1}^{NP} \{ & H_{,x}(x - x_I)H^T(x - x_I)\phi_a(x - x_I) + H(x - x_I)H_{,x}{}^T(x - x_I)\phi_a(x - x_I) \\ & + H(x - x_I)H^T(x - x_I)\phi_{a,x}(x - x_I)\}\end{aligned} \tag{2.26}$$

$$\begin{aligned}M_{,xx}(x) = \sum_{I=1}^{NP} \{ & H_{,xx}(x - x_I)H^T(x - x_I)\phi_a(x - x_I) + 2H_{,x}(x - x_I)H_{,x}{}^T(x - x_I)\phi_a(x - x_I) \\ & + 2H_{,x}(x - x_I)H^T(x - x_I)\phi_{a,x}(x - x_I) + H(x - x_I)H_{,xx}{}^T(x - x_I)\phi_a(x - x_I) \\ & + 2H(x - x_I)H_{,x}{}^T(x - x_I)\phi_{a,x}(x - x_I) + H(x - x_I)H^T(x - x_I)\phi_{a,xx}(x - x_I)\}\end{aligned} \tag{2.27}$$

$M(x)$ 矩陣之反矩陣的高階微分，即(2.23)及(2.24)式中之 $M^{-1}{}_{,x}(x)$ $M^{-1}{}_{,xx}(x)$ ，可利用隱微分技巧推導出來。因為

$$M(x)M^{-1}(x) = \boldsymbol{I} \tag{2.28}$$

其中 $\boldsymbol{I}$ 是單位矩陣。我們在兩端取微分可得

$$\frac{d}{dx}(M(x)M^{-1}(x)) = \frac{d}{dx}\boldsymbol{I} \tag{2.29}$$

再根據乘法原理

$$M_{,x}(x)M^{-1}(x)+M(x)M^{-1}{}_{,x}=\boldsymbol{0} \tag{2.30}$$

其中$\boldsymbol{0}$為零矩陣。移項整理可得

$$M^{-1}{}_{,x}=-M^{-1}(x)M_{,x}(x)M^{-1}(x) \tag{2.31}$$

再進一步求二階微分

$$\frac{d}{dx}(M^{-1}{}_{,x})=\frac{d}{dx}\{-M^{-1}(x)M_{,x}(x)M^{-1}(x)\} \tag{2.32}$$

同樣使用乘法原理

$$M^{-1}{}_{,xx}=-\{M^{-1}{}_{,x}(x)M_{,x}(x)M^{-1}(x)+M^{-1}(x)M_{,xx}(x)M^{-1}(x)+M^{-1}(x)M_{,x}(x)M^{-1}{}_{,x}(x)\} \tag{2.33}$$

整理可得

$$\begin{aligned} M^{-1}{}_{,xx} &= -M^{-1}(x)\{M_{,xx}(x)M^{-1}(x)+2M_{,x}(x)M^{-1}{}_{,x}(x)\} \\ &= M^{-1}(x)\{2M_{,x}M^{-1}(x)M_{,x}-M_{,xx}(x)\}M^{-1}(x) \end{aligned} \tag{2.34}$$

導函數(2.31)及(2.34)在建立基底之導函數(2.20)及(2.21)時需用到。

第三節 再生核心函數之基本性質

接下來看再生核心函數之基本性質。考慮一維定義域$\Omega=\{x\mid 0<1<x\}$，在$\overline{\Omega}$中取N_p個質點，此組質點可為等距點集亦可為不等距質點。若不等距，其原則是點分佈不可太過懸殊（擬一致分布怖）。選用三階 B-spline 如(2.6)當核心函數，核心函數半徑將隨其基底的階數變化。在第一節已介紹過核心基底是透過再生性推導出來的，在此我們先檢視此項性質。此組基底函數本身滿足

$$\sum_{I=1}^{N_p}\psi_I(x)x_I{}^k=x^k\ ,\quad k=0,1,\ldots,n \tag{2.35}$$

而其一階導函數滿足

$$\sum_{I=1}^{N_p}\psi_I{}'(x)x_I{}^k=kx^{k-1}\ ,\quad k=0,1,\ldots,n \tag{2.36}$$

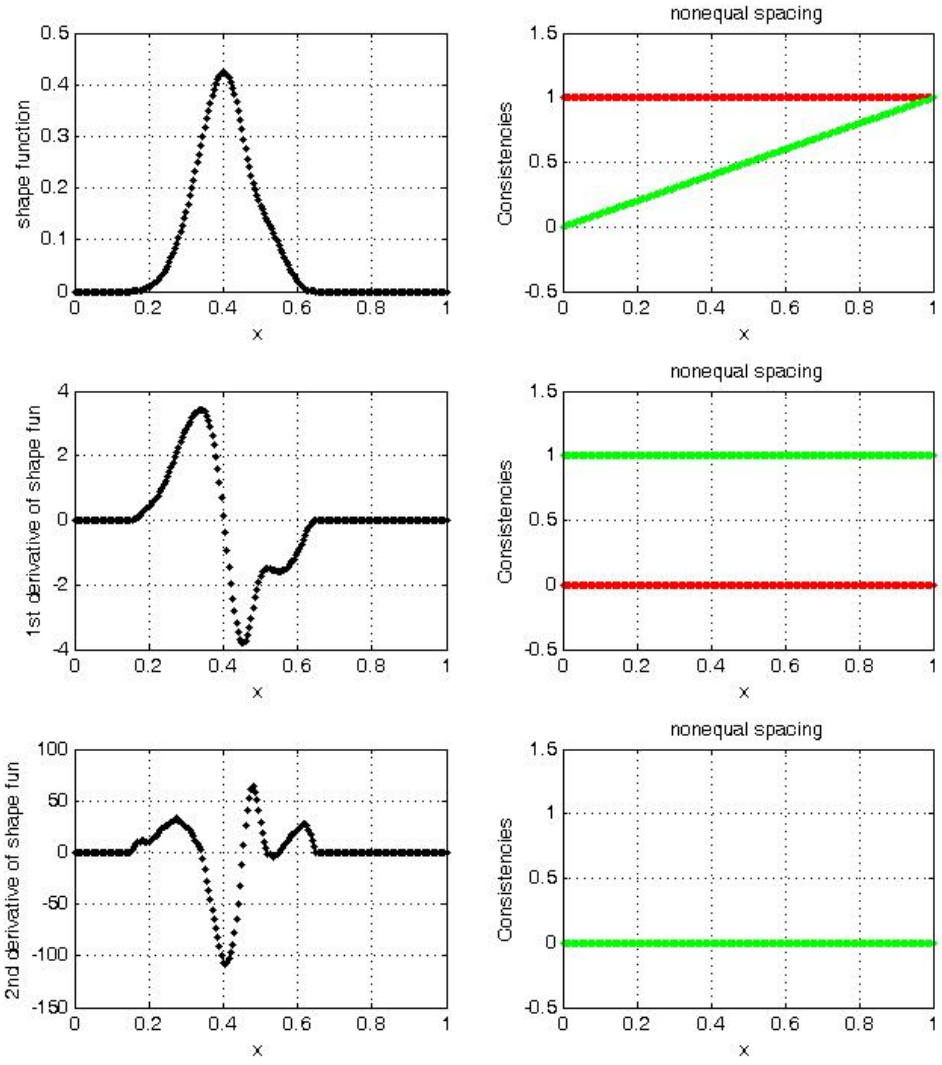
二階導函數滿足

$$\sum_{I=1}^{N_p}\psi_I''(x)x_I^{\,k}=k(k-1)x^{k-2},\quad k=0,1,\ldots,n \tag{2.37}$$

下面圖一裡所顯示的是選用一組『不等距質點』所得到的結果。此組非等距的質點集為

$$S=\{x_I\}_{I=1}^{N_p}=\{0.0,0.08,0.22,0.34,0.4,0.48,0.64,0.69,0.82,0.93,1.0\} \tag{2.38}$$

共有 11 個點，$N_p=11$。圖左由上而下依序為第五個基底函數$\psi_5(x)$及其第一階第二階導函數；而圖右則表示此組再生核心基底具有再生性，其中紅色線代表 (2.35) - (2.37) 式中 $k=0$ 的情況，而綠色線則代表 $k=1$ 的情況。

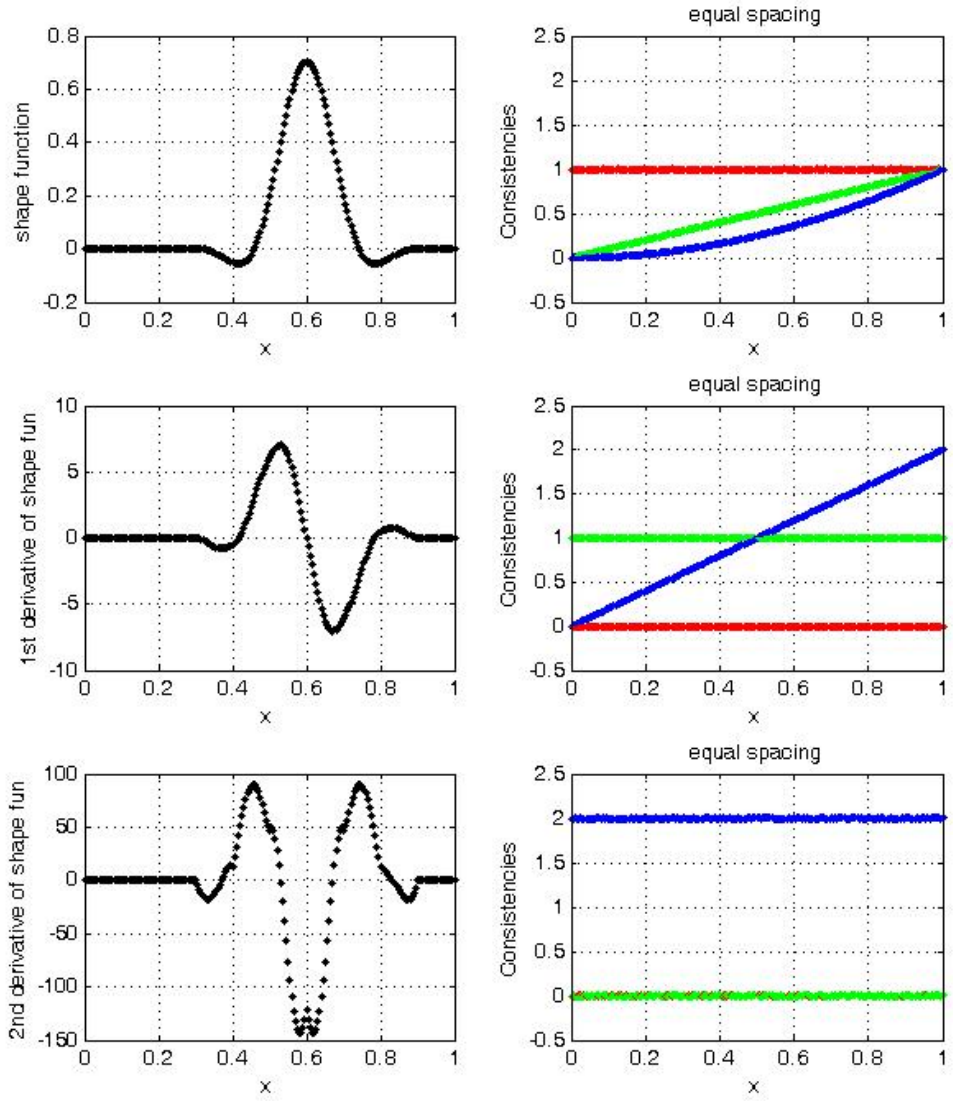


圖一：一階再生核心函數基底函數和導函數（$a=0.25$）及再生性

圖二中則選用一組『等距質點』所產生的結果，質點集為

$$S=\{x_I\}_{I=1}^{N_p}=\left\{0.0+\frac{I-1}{10}\right\}_{I=1}^{11} \tag{2.39}$$

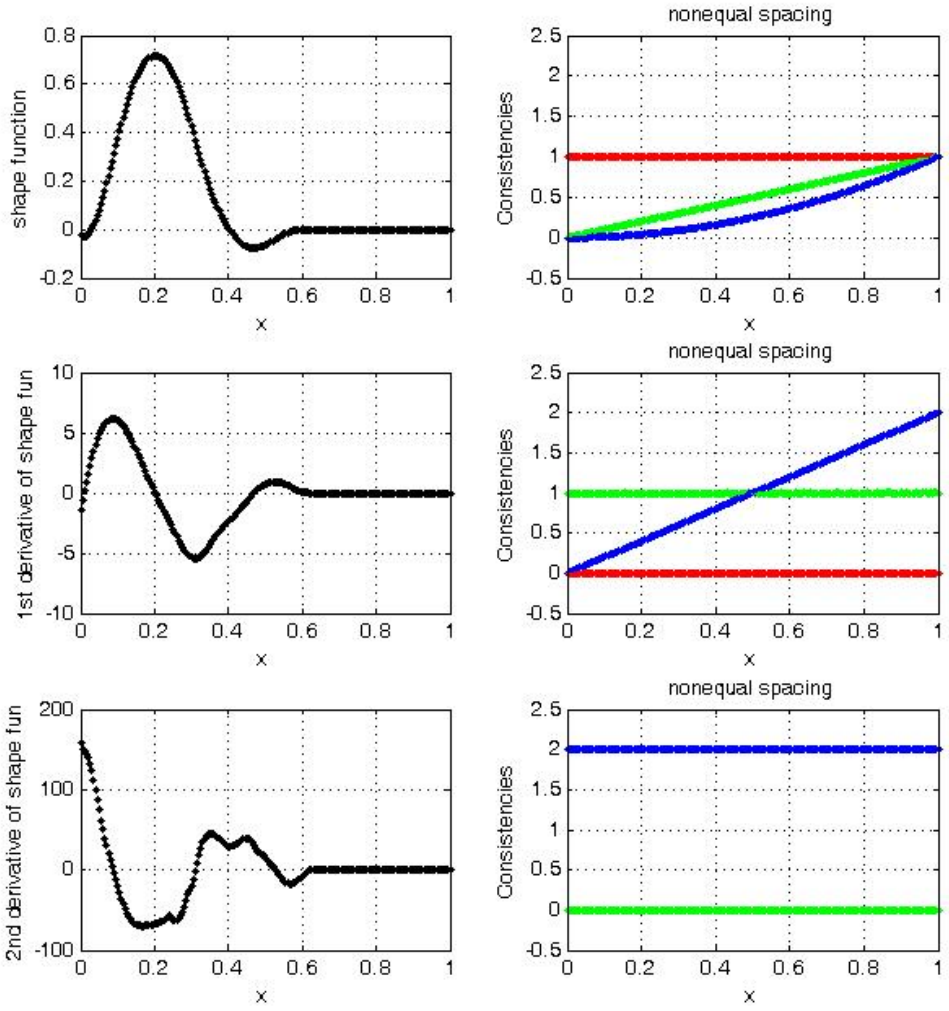
其中考慮二階基底，$n=2$。下圖左依序代表第 7 個基底函數$\psi_7(x)$及導函數的輪廓；圖右表示此組基底具有再生性，紅色線表示$k=0$的情況，綠色線代表$k=1$，而藍色線代表$k=2$。



圖二：二階再生核心基底函數和導函數（$a=0.3$）及再生性

接下來，圖三是也是由不等距質點(2.38)，所產生之結果。與圖二相同皆採用二階基底，但核心函數半徑大一些，由 $a=0.3$ 變成 $a=0.4$ 。

由圖二及圖三可觀察出，等距質點集所形成的再生核心基底函數是對稱的，而非等距質點集產生的基底是較不規則的。



圖三：二階再生核心基底函數和導函數（ $a=0.4$ ）及再生性

更進一步可由圖一至圖三看出此基底函數不具 Kronecker delta 性質，即

$$\psi_i\left(x_j\right)\neq\delta_{ij} \tag{2.40}$$

當 $i=j$ ，從圖中可見 $\psi_i\left(x_i\right)\neq 1$ ，值介於 0.4 至 0.8 間未達 1。

另外核心函數另具有二個性質如下述[9]：

(i) 再生基底函數的可微性是依賴於核心函數的平滑性。

(ii) 核心函數的半徑不可太小，否則將導致矩陣 M 變得奇異，其反矩陣不存在，因此無法建構出一組適合的基底來。若核心函數的半徑過大也不好，會使得離散系統的穩定性變差，此部分，稍後會再作詳細的說明。

第四節 再生核心函數之反估計式

此章最後來探討基底函數的高階微分之反估計。假設函數 $v(x)$ 是由一組再生核心基底函數線性組合而成

$$v(x)=\sum_{I=1}^{Np} b_I \psi_I(x), \quad \text{V=span}\left\{\psi_1,\psi_2,\cdots,\psi_{N_p}\right\} \tag{2.41}$$

其高階微分具有下列不等關係。

引理 1 假設質點集為擬一致分佈[10]，則函數 $v(x)$ 具有下列關係

$$\|v\|_{\ell,\Omega} \le C_1 a^{-\ell} n^{2\ell} \|v\|_{0,\Omega,} \quad , \ell = 1,2,3,\ldots \tag{2.42}$$

$$\|v\|_{\ell,\Gamma} \le C_2 a^{-\ell} n^{2\ell} \|v\|_{0,\Gamma,} \quad , \ell = 1,2,3,\ldots \tag{2.43}$$

$$\|v_x\|_{\ell,\Gamma} \le C_3 a^{-(\ell+1)} n^{2(\ell+1)} \|v\|_{1,\Omega,} \quad , \ell = 1,2,\cdots \tag{2.44}$$

其中 Γ 表示 Ω 之邊界，而 C_1, C_2 和 C_3 是與 a, n 及 ℓ 無關之常數。$\|\cdot\|$ 是標準 Sobolev 範數。在後面的收斂性及穩定性分析中，將需引用此引理去做詳細估計。

第三章 擾動性及穩定性之探討

第一節 函數逼近

此節我們將利用前一章所介紹的基底來近似已知函數，想了解利用再生核心函數作函數逼近所形成之線性系統的穩定性。

我們考慮一已知函數 $f(x)=\sin \pi x$ ， $x\in[0,1]$，近似函數定義如下

$$f^h(x)=\sum_{I=1}^{N_p} d_I \psi_I(x)，0<x<1 \tag{3.1}$$

核心函數 $\phi_a(x-x_I)$ 選用三階 B-spline。利用配置法（collocation methods）求出係數 d_I ，其中需再定義一組配置點（collocation points），可同於前述的質點集，亦可選用不同的點集。我們稱此組點為配置點集

$$E=\{\xi_i\}_{i=1}^{N_p} ， \xi_i\in[0,1] \tag{3.2}$$

我們強制具有下述插值條件：

$$f^h(\xi_i)=f(\xi_i)， \forall i \tag{3.3}$$

因而得到一線性系統

$$\mathbf{Ax}=\mathbf{b} \tag{3.4}$$

其中矩陣及向量元素分別為

$$[A]_{ij}=\psi_j(\xi_i)， \ [\mathbf{b}]_i=f(\xi_i)，\forall i,j=1,2,\cdots,N_p \tag{3.5}$$

在系統(3.4)中的向量 $\mathbf{x}$ 即為未知係數 d_I ，可利用高斯消去法解此線性系統。但若考慮較多的配置點時，例如使用 $2\times N_p$ 或 $4\times N_p$ 個配置點時，則線性系統(3.4)將變成一最小二乘系統，此時則需選用 QR 分解法或 SVD 分解法解此系統。

理論上若被近似函數的 $n+1$ 階導函數是可積，可證明具有下列收斂行為[10]

$$\left\| f(x)-f^h(x)\right\|_{\ell,\Omega}\le ca^{n+1-\ell}\left|f(x)\right|_{n+1,\Omega} \tag{3.6}$$

其中 $\ell\ge 0$ ， a 代表核心函數的半徑。當 ℓ=0 時上述之估計式變成

$$\left\| f(x)-f^h(x)\right\|_{L^2,\Omega}\le ca^{n+1}\left|f(x)\right|_{n+1,\Omega} \tag{3.7}$$

其中 $n=1,2,3,\ldots$。

表一是誤差結果，其中採用等距質點集，並且配置點集的數目與位置相同於質點集，而核心函數半徑 $a=(n+1)\cdot h$，其中 $h=1/(N_p-1)$。前面提過，核心基底函數的半徑需隨階數增加而增大，如此一來，M 矩陣才不至於變成奇異矩陣，基底函數可順利建構出來。隨 n 增加收斂行為越好。

$n=1$		$n=2$		$n=3$	
N_p	$\left\|f-f^h\right\|_{L^2}$	N_p	$\left\|f-f^h\right\|_{L^2}$	N_p	$\left\|f-f^h\right\|_{L^2}$
6	7.6×10^{-3}	6	4.70×10^{-3}	6	4.12×10^{-4}
11	1.9×10^{-3}	11	4.34×10^{-4}	11	5.00×10^{-6}
21	4.9×10^{-4}	21	3.99×10^{-5}	21	1.47×10^{-7}

表一：使用一至三階基底函數做函數逼近所產生之誤差

由擾動理論可知當線性系統(3.4)的左右邊有擾動時，其解的相對誤差與矩陣 A 的 condition number 有密切關係：系統解的相對誤差是矩陣 condition number 乘上右端向量的相對誤差。藉由觀察 condition number，即可知此系統的穩定性。

探討穩定性共有兩種情況，分述如下（詳細證明請參閱附錄 A）。

第一種擾動狀況：僅系統右邊向量有微擾。假設 $\hat{\mathbf{x}}=\mathbf{x}+\Delta\mathbf{x}$ 是下列系統

$$\mathbf{A}\hat{\mathbf{x}}=\mathbf{b}+\Delta\mathbf{b} \tag{3.8}$$

的解，則可以證明得到下列估計式 [11]

$$\frac{\|\hat{\mathbf{x}}-\mathbf{x}\|}{\|\mathbf{x}\|}=\frac{\|\Delta\mathbf{x}\|}{\|\mathbf{x}\|}\le \mathrm{Cond(A)}\frac{\|\Delta\mathbf{b}\|}{\|\mathbf{b}\|} \tag{3.9}$$

其中 Cond(**A**)為矩陣 **A** 的 condition number。傳統定義是矩陣 **A** 最大特徵值除上最小特徵值。

第二種擾動狀況：系統左邊矩陣及右邊向量皆有微擾。假設 $\tilde{\mathbf{x}}=\mathbf{x}+\delta\mathbf{x}$ 是下列系統

$$\left(\mathbf{A}+\delta\mathbf{A}\right)\tilde{\mathbf{x}}=\mathbf{b}+\delta\mathbf{b} \tag{3.10}$$

的解，則可證明得出 [12]

$$\frac{\left\|\tilde{\mathbf{x}}-\mathbf{x}\right\|}{\left\|\mathbf{x}\right\|}=\frac{\left\|\delta\mathbf{x}\right\|}{\left\|\mathbf{x}\right\|}\le\frac{\mathrm{Cond}(\mathbf{A})}{1-\mathrm{Cond}(\mathbf{A})\dfrac{\left\|\delta\mathbf{A}\right\|}{\left\|\mathbf{A}\right\|}}\left\{\frac{\left\|\delta\mathbf{A}\right\|}{\left\|\mathbf{A}\right\|}+\frac{\left\|\delta\mathbf{b}\right\|}{\left\|\mathbf{b}\right\|}\right\} \tag{3.11}$$

附註：當 $\delta\mathbf{A}=0$ 時，則(3.11)式中的 $\left\|\delta\mathbf{A}\right\|/\left\|\mathbf{A}\right\|$ 項為零，因此，(3.11)式的右端最後是等於(3.9)式的估計。相應於表一，我們選用相同的質點集及配置點集，所計算出的 condition numbers 列於下面表中。

$n=1$		$n=2$		$n=3$	
N_p	Cond(**A**)	N_p	Cond(**A**)	N_p	Cond(**A**)
6	1.53	6	1.45	6	2.68
11	1.58	11	1.58	11	6.54
21	1.60	21	1.62	21	8.77

表二：一至三階基底函數所形成矩陣 **A** 之 condition numbers

值得注意的是當再生核心基底函數階增加時，由 $n=1$ 至 $n=3$，系統的規模並不像有限元素法一樣增大。不同階所形成的矩陣 **A**，其實都一樣大，差別在於形成矩陣 **A** 中的元素所花時間不同。從表二可觀察出，當 $n=1$ 由變到 $n=3$ 時 condition numbers 並沒有太劇烈的增加。而下列表三是使用 “傳統多項式基底” 逼近後所得到的 condition numbers。

N_p	Cond(**A**)
6	4.92×10^{3}
11	1.16×10^{8}
21	8.06×10^{16}

表三：傳統多項式基底所形成矩陣 **A** 之 condition numbers.

兩相比較之下，採用 local 的核心基底做函數逼近所得之結果，比較起 global 的基底算是相當穩定的。

第二節 偏微分方程解

接下來我們將再生核心基底函數配合配置法（collocation method）求解微分方程。考慮一簡單卜松問題如下

$$-\Delta u = f \qquad \text{in}\,\Omega \tag{3.12}$$
$$u_{\upsilon} = q_1 \qquad \text{on}\,\Gamma_N \tag{3.13}$$
$$u_{\upsilon} + \beta u = q_2 \qquad \text{on}\,\Gamma_R \tag{3.14}$$

其中$\partial\Omega = \Gamma_N \cup \Gamma_R$且$\Gamma_N \cap \Gamma_R = \phi$，且$\beta > 0$，而$\upsilon$代表外法向量。由配置法[3]得到的最優解 u^R滿足下列泛函問題

$$\hat{E}\left(u^R\right) = \min_{v \in V} \hat{E}(v), \quad \text{V=span}\left\{\psi_1, \psi_2, \cdots, \psi_{N_p}\right\} \tag{3.15}$$

其中離散泛函$\hat{E}(\cdot)$定義如下：

$$\hat{E}(v) = \frac{1}{2}\left\{ \hat{\int}_{\Omega} \left(\Delta v + f\right)^2 d\Omega + \hat{\int}_{\Gamma_N} \left(v_{\upsilon} - q_1\right)^2 d\ell + \hat{\int}_{\Gamma_R} \left(v_{\upsilon} + \beta v - q_2\right)^2 d\ell \right\} \tag{3.16}$$

其中$\hat{\int}$代表數值積分，而近似解v定義如前述

$$v = \sum_{I=1}^{N_p} a_I \psi_I(x) \tag{3.17}$$

『泛函極小化問題』可等價於下面『非方陣型線性系統之求解問題』

$$-\sqrt{\alpha_J} \sum_{I=1}^{N_p} \Delta\psi_I(\xi_J) a_I = \sqrt{\alpha_J} f(\xi_J)，\forall \xi_J \in \Omega \tag{3.18}$$
$$\sqrt{\alpha_J{}^N} \sum_{I=1}^{N_p} \psi_{I,\upsilon}(\xi_J) a_I = \sqrt{\alpha_J{}^N} q_1(\xi_J)，\forall \xi_J \in \Gamma_N \tag{3.19}$$
$$\sqrt{\alpha_J{}^R} \sum_{I=1}^{N_p} \left\{\psi_{I,\upsilon} + \beta\psi_I\right\}(\xi_J) a_I = \sqrt{\alpha_J{}^R} q_2(\xi_J)，\forall \xi_J \in \Gamma_R \tag{3.20}$$

其中配置點ξ_J在定義域內及邊上均勻分佈，共計有N_c個點。我們將(3.18) - (3.20)式改寫為一線性系統如下

$$\mathbf{F}\,\mathbf{y} = \mathbf{r} \tag{3.21}$$

其中**F**是一 $N_c \times N_p$ 的矩陣，$N_c > N_p$，**r**是一 $N_c \times 1$ 的向量，而 N_c 是配置點總數目。於此，我們選用較多的配置點原因是在證明收斂性時必要的條件 [3]。

我們先來看收斂性，需定義範數如下

$$\|v\|_H = \left\{ \|v\|_{1,\Omega}^2 + \|\Delta v\|_{0,\Omega}^2 + \|v_\upsilon\|_{0,\Gamma_N}^2 + \|v_\upsilon + \beta v\|_{0,\Gamma_R}^2 \right\}^{\frac{1}{2}} \tag{3.22}$$

可證明得到最優誤差估計如下[7, 10]

$$\left\| u - u^R \right\|_H \le \min_{v \in V} \| u - v \|_H \le C a^{n-1} |u|_{n+1,\Omega} \tag{3.23}$$

由上面的估計式可得知：當使用再生核心基底函數用來求解微分方程時，基底的階必須至少二階，$n \ge 2$，才有收斂的行為。

我們考慮一維卜松方程，藉以了解此方法的收斂性質

$$\begin{aligned} &-u''(x) = \pi^2 \sin \pi x, \quad 0 < x < 1 \\ &u(0) = 0 \\ &u(1) = 0 \end{aligned}$$

僅觀察 Sobolev 一範數的誤差變化情形，結果列表如下：

$n=1$		$n=2$		$n=3$	
N_p	$\left\|u-u^R\right\|_{1,\Omega}$	N_p	$\left\|u-u^R\right\|_{1,\Omega}$	N_p	$\left\|u-u^R\right\|_{1,\Omega}$
6	6.59×10^{-1}	6	1.08×10^{-1}	6	8.80×10^{-3}
11	6.32×10^{-1}	11	1.99×10^{-2}	11	2.80×10^{-4}
21	6.27×10^{-1}	21	3.50×10^{-3}	21	1.61×10^{-5}

表四：一至三階基底函數所產生收斂行為

從表四中可看出，的確在一階，$n=1$，時是不收斂的。更多收斂結果參閱論文[7]。

接下來觀察穩定性，在此方面我們同樣得先了解擾動分析。共分二種情況分

述如下（詳細證明請參閱附錄 B）。

第一種擾動狀況：僅系統右邊向量有微擾。假設 $\hat{\mathbf{y}} = \mathbf{y} + \Delta\mathbf{y}$ 是下列非方陣系統

$$\mathbf{F}\hat{\mathbf{y}} = \mathbf{r} + \Delta\mathbf{r} \tag{3.24}$$

的解，則可得到誤差估計如下

$$\frac{\|\hat{\mathbf{y}} - \mathbf{y}\|}{\|\mathbf{y}\|} = \frac{\|\Delta\mathbf{y}\|}{\|\mathbf{y}\|} \le \mathrm{Cond}(\mathbf{F})\frac{\|\mathrm{P}\Delta\mathbf{r}\|}{\|\mathrm{P}\mathbf{r}\|} \tag{3.25}$$

其中 Cond(**F**) 為矩陣 **F** 的 condition number 。傳統定義為矩陣 **F** 最大奇異值除以最小奇異值。此外，P 代表投影算子，其定義為 $\mathrm{P} = \mathbf{F}\left(\mathbf{F}^T\mathbf{F}\right)^{-1}\mathbf{F}^T = \mathbf{F}\mathbf{F}^+$ ，而 $\mathbf{F}^+$ 是 **F** 的偽逆矩陣[12]。

第二種擾動狀況：在系統左邊矩陣及右邊向量皆有微擾。假設 $\tilde{\mathbf{y}} = \mathbf{y} + \delta\mathbf{y}$ 是下列系統

$$\left(\mathbf{F} + \delta\mathbf{F}\right)\tilde{\mathbf{y}} = \mathbf{r} + \delta\mathbf{r} \tag{3.26}$$

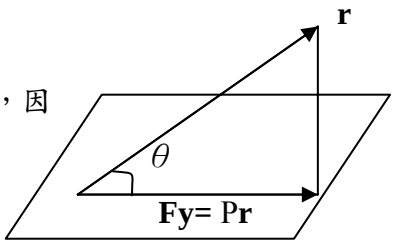
的解，可證明出

$$\frac{\|\tilde{\mathbf{y}} - \mathbf{y}\|}{\|\mathbf{y}\|} = \frac{\|\delta\mathbf{y}\|}{\|\mathbf{y}\|} \le \frac{\varepsilon_F + \varepsilon_F \cdot \varepsilon_{\mathbf{r}} + \bar{\varepsilon}_{\mathbf{r}} \cdot \mathrm{Cond}(\mathbf{F})}{1 - \left(\varepsilon_F + {\varepsilon_F}^2 + \varepsilon_F \cdot \mathrm{Cond}(\mathbf{F})\right)} \cdot \sec(\theta) \tag{3.27}$$

其中

$$\varepsilon_{\mathbf{F}} = \frac{\|\delta\mathbf{F}\|}{\|\mathbf{F}\|}，\varepsilon_{\mathbf{r}} = \frac{\|\delta\mathbf{r}\|}{\|\mathbf{r}\|}，\bar{\varepsilon}_{\mathbf{r}} = \frac{\|\mathrm{P}\delta\mathbf{r}\|}{\|\mathbf{r}\|}，\cos(\theta) = \frac{\|\mathbf{F}\mathbf{y}\|}{\|\mathbf{r}\|} = \frac{\|\mathrm{P}\mathbf{r}\|}{\|\mathbf{r}\|} \tag{3.28}$$

角度 θ 是向量 **r** 及 F **y** 的夾角。



附註: 當 δ **F**=0 時，則(3.27)中的 $\varepsilon_F = \|\delta\mathbf{F}\|/\|\mathbf{F}\| = 0$ ，因此，(3.27)式的右端變成

$$\frac{\|\tilde{\mathbf{y}} - \mathbf{y}\|}{\|\mathbf{y}\|} = \frac{\|\delta\mathbf{y}\|}{\|\mathbf{y}\|} \le \bar{\varepsilon}_{\mathbf{r}} \cdot \mathrm{Cond}(\mathbf{F}) \cdot \sec(\theta)$$

$$=\text{Cond}(\mathbf{F})\ \frac{\|\mathrm{P}\delta\mathbf{r}\|}{\cancel{\|\mathbf{r}\|}}\cdot\frac{\cancel{\|\mathbf{r}\|}}{\|\mathrm{P}\mathbf{r}\|}=\text{Cond}(\mathbf{F})\frac{\|\mathrm{P}\delta\mathbf{r}\|}{\|\mathrm{P}\mathbf{r}\|} \tag{3.29}$$

等同於(3.25)式的估計。

前面提到用再生核心函數求解方程其階至少需為二階，$n\geq 2$，請參考表四。因此，我們檢試穩定性時，僅考慮 $n=2$ 及 $n=3$。下列表五是系統(3.21)矩陣 F 的 condition numbers。

$n=2$		$n=3$	
N_p	Cond(**F**)	N_p	Cond(**F**)
6	359.2	6	404.1
11	1991.3	11	2063.7
21	11085.4	21	11401.1

表五：二及三階基底所形成矩陣**F**之 condition numbers。

從表五可觀察出 condition numbers 隨質點數 N_p 增加及階數 n 增加而增大。

接下來我們將針對 condition numbers 之增加情況，給出一明確的估計式。在前面已提及偏微分方程的最優解 u^R 是透過泛函極小的方式得到，它同時等於下列離散的變分形式

$$\hat{B}\left(u^R,v\right)=\hat{F}(v)$$

其中雙線性及線性形式定義如下

$$\hat{B}(u,v)=\hat{\int}_{\Omega}\Delta u\Delta v d\Omega+\hat{\int}_{\Gamma_N}u_v v_v d\ell+\hat{\int}_{\Gamma_R}\left(u_v+\beta v\right)d\ell \tag{3.30}$$

$$\hat{F}\left(v\right)=-\hat{\int}_{\Omega}f\Delta v d\Omega+\hat{\int}_{\Gamma_N}q_1 v_\upsilon d\ell+\hat{\int}_{\Gamma_R}q_2\left(v_\upsilon+\beta v\right)d\ell \tag{3.31}$$

我們需再定義一個新的範數如下

$$\overline{\|v\|}_E^2=\hat{B}(v,v) \tag{3.32}$$

此離散範數可進一步表示為

$$\overline{\|v\|}_E^2=\mathbf{y}^T\mathbf{F}^T\mathbf{F}\,\mathbf{y}=:\mathbf{y}^T\mathbf{G}\,\mathbf{y} \tag{3.33}$$

其中矩陣 F 及向量 $\mathbf{y}$ 就是在線性方程組(3.21)中的矩陣及向量。

我們可得到下列兩組關係式[10]

$$c_1\|v\|_E \le \overline{\|v\|}_E \le c_2\|v\|_E \tag{3.34}$$

$$c_3\|v\|_{0,\Omega} \le \|v\|_E \le c_4\|v\|_{2,\Omega} \tag{3.35}$$

其中 $\|v\|_E$ 代表相應 $\overline{\|v\|}_E$ 連續型態的範數，其定義如下

$$\|v\|_E^2 = B(v,v) \tag{3.36}$$

而 $B(\cdot,\cdot)$ 是指不具積分近似的雙線型式，在(3.35)式中的範數 $\|\cdot\|_{2,\Omega}$ 則代表標準 Sobolev 二範數。

根據 Rayleigh-Ritz 定理可得到矩陣 G 之最大及最小特徵值的界分別為

$$\lambda\min(\mathbf{G}) = \min\frac{\mathbf{y}^T\mathbf{G}\mathbf{y}}{\mathbf{y}^T\mathbf{y}} \ge \min\frac{c_1\|v\|_E^2}{\mathbf{y}^T\mathbf{y}} \ge \min\frac{c_3\|v\|_{0,\Omega}^2}{\mathbf{y}^T\mathbf{y}} \tag{3.37}$$

$$\lambda\max(\mathbf{G}) = \max\frac{\mathbf{y}^T\mathbf{G}\mathbf{y}}{\mathbf{y}^T\mathbf{y}} \le \min\frac{c_2\|v\|_E^2}{\mathbf{y}^T\mathbf{y}} \le \max\frac{c_4\|v\|_{2,\Omega}^2}{\mathbf{y}^T\mathbf{y}} \tag{3.38}$$

接下來再使用第二章第四節所提到的反估計式，可進一步得到 G 矩陣的 condition number 的界如下

$$\mathrm{Cond}(\mathbf{G}) = \frac{\lambda\max(\mathbf{G})}{\lambda\min(\mathbf{G})} \le \frac{c_3\|v\|_{2,\Omega}^2}{c_4\|v\|_{0,\Omega}^2} \le \frac{c_5\kappa a^{-4}n^8\|v\|_{0,\Omega}^2}{c_4\|v\|_{0,\Omega}^2} \le C\kappa a^{-4}n^8 \tag{3.39}$$

最後，可得到 F 矩陣的 condition number 的上界為

$$\mathrm{Cond}(\mathbf{F}) = \{\mathrm{Cond}(\mathbf{G})\}^{1/2} \le \widetilde{C}a^{-2}n^4 \tag{3.40}$$

其中常數 $\widetilde{C}$ 與覆蓋數 κ 有關。

所以，矩陣 $\mathbf{F}$ 的 condition number 與再生核心基底的半徑有直接關聯。而此半徑又與質點距離成正比關係，$a = (n+1)\cdot h$。因此，更清楚來說 condition number 與質點距成平方反比的關係

$$\mathrm{Cond}(\mathbf{F}) \le C \frac{n^4}{(n+1)^2} \cdot h^{-2} \tag{3.41}$$

質點距與質點數具下列關係

$$h \approx \mathrm{O}\left(N_p^{-1}\right) \tag{3.42}$$

進一步可得

$$\mathrm{Cond}(\mathbf{F}) \approx \mathrm{O}\left(n^2 N_p^2\right) \tag{3.43}$$

所以，condition number 將隨質點數增加而呈平方倍數增大。

我們回頭檢視表五，由呈現的數據可看出質點數 N_p 加倍時 condition number 約增加 5 倍左右。此外，condition numbers 與再生核心基底函數的階數 n 也有些許關聯。

當 N_p 數大時（$N_p = 21$），階數 n 由 $n = 2$ 改變至 $n = 3$ 時 condition numbers 變化不大，所以使用高階基底函數能有較好收斂行為，請參考表四，且 condition number 與低階基底相比也不會過大。建議可考慮使用三階基底。

第四章 穩定性之深入探討

第一節 新穩定性之估計方式

我們在第三章已探討過線性系統的擾動理論。其中 condition number 是扮演相當重要的角色，傳統 condition number 定義如下

『方陣型』線性系統：$A\mathbf{x}=\mathbf{b}$，其中 $A\in\mathbf{R}^{n\times n}$，$\mathbf{b}\in\mathbf{R}^{n}$，定義為

$$\mathrm{Cond}(\mathbf{A})=\frac{\lambda_{\max}}{\lambda_{\min}} \tag{4.1}$$

『非方陣型』線性系統：$F\mathbf{y}=\mathbf{r}$，其中 $F\in\mathbf{R}^{m\times n}$，$\mathbf{r}\in\mathbf{R}^{m}$，$m>n$，定義為

$$\mathrm{Cond}(\mathbf{F})=\frac{\sigma_{\max}}{\sigma_{\min}} \tag{4.2}$$

然而有一些新的研究[13,14] 指出有更好的擾動理論。因而衍生出新的 condition number 定義，稱之為 effective condition number，它與線性系統右端向量有關，其估計式如下

$$\mathrm{Cond\_E}(\mathbf{A})=\frac{\|\mathbf{b}\|}{\sqrt{\dfrac{\|\mathbf{b}\|^{2}-\beta_n^{2}}{Cond(\mathbf{A})^{2}}+\beta_n^{2}}} \tag{4.3}$$

其中 β_n 是矩陣 A 對角化中第 n 個特徵向量與 $\mathbf{b}$ 的內積，也就是 $\mathbf{A}=U\Lambda U^{T}$，$U=(\mathbf{u}_1,\mathbf{u}_2,\cdots\mathbf{u}_n)$，其中

$$\beta_n=\mathbf{u}_n^{T}\cdot\mathbf{b} \tag{4.4}$$

由(4.3)中可看出當 $\beta_n=0$ 時，Cond_E(**A**)就回歸到傳統 Cond(**A**)。在一般 $\beta_n\neq 0$ 之情形下，可知

$$\mathrm{Cond\_E}(\mathbf{A})<\mathrm{Cond}(\mathbf{A}) \tag{4.5}$$

詳細推導及証明我們放在本章第三節中。

同樣地，對於非方陣型線性系統，我們亦可得相似的 effective condition number，定義如下

$$\text{Cond\_E}(\mathbf{F}) = \frac{\|\mathbf{r}\|}{\sqrt{\frac{\|\mathbf{r}\|^2 - \beta_n^2}{\text{Cond}(\mathbf{F})} + \beta_n^2}} \tag{4.6}$$

其中 β_n 是矩陣 F 奇異分解中，將 U 第 n 個向量與系統右端 $\mathbf{r}$ 向量之投影量 P$\mathbf{r}$ 作內積，也就是 $\mathbf{F} = U\Sigma V^T$ ，$U = (\mathbf{u}_1, \mathbf{u}_2, \cdots \mathbf{u}_m)$，其中

$$\beta_n = \mathbf{u}_n^T \cdot P\mathbf{r} \tag{4.7}$$

而投影算子定義如下

$$\text{P} = \mathbf{F}\left(\mathbf{F}^T\mathbf{F}\right)^{-1}\mathbf{F}^T = \mathbf{F}\mathbf{F}^+ \tag{4.8}$$

其中 $\mathbf{F}^+$ 為 $\mathbf{F}$ 之偽逆矩陣。同樣地，可觀察出當 $\beta_n = 0$ 時，Cond_E($\mathbf{F}$) 會等於 Cond($\mathbf{F}$)。在一般情形 $\beta_n \neq 0$，新的 conditon number 比傳統的小一些

$$\text{Cond\_E}(\mathbf{F}) < \text{Cond}(\mathbf{F}) \tag{4.9}$$

詳細推導過程放在本章第三節中。

第二節 數值範例

考慮一維卜松方程

$$-u''(x) = \pi^2 \sin \pi x, \quad 0 < x < 1 \tag{4.10}$$

$$u(0) = 0 \tag{4.11}$$

$$u(1) = 0 \tag{4.12}$$

於此，我們同時計算出“新的 condition number”與“傳統 condition number”，並作比較。下列表六是選用二階核心基底函數（$n = 2$）的結果，而表七則是採用三階基底函數（$n = 3$）的結果。

N_p	$h=1/(N_p-1)$	Cond(**F**)	Cond_E(**F**)	$\sigma_1=\sigma_{\max}$	$\sigma_n=\sigma_{\min}$	β_n
6	1/5	359.2	44.3	207.3	0.577	-0.699
7	1/6	563.9	46.2	301.3	0.534	-0.737
9	1/8	1147.4	50.4	540.8	0.471	0.782
11	1/10	1991.3	54.7	848.9	0.426	-0.807
13	1/12	3125.8	58.7	1225.9	0.392	-0.824
17	1/16	6373.6	66.2	2185.9	0.343	0.843
21	1/20	11085.4	73.0	3420.7	0.309	-0.855

表六：新的與傳統的 condition numbers 之比較（$n=2$）

N_p	$h=1/(N_p-1)$	Cond(**F**)	Cond_E(**F**)	σ_1	σ_n	β_n
6	1/5	404.1	42.5	233.2	0.577	0.730
7	1/6	598.2	45.0	319.7	0.534	0.757
9	1/8	1194.4	49.9	562.9	0.471	0.791
11	1/10	2063.7	54.3	879.8	0.426	0.812
13	1/12	3229.7	58.5	1225.9	0.392	0.827
17	1/16	6565.2	66.1	2251.6	0.343	-0.845
21	1/20	11401.1	72.9	3518.1	0.309	0.856

表七：新的與傳統的 condition numbers 之比較（$n=3$）

可從表六及七看出 β_n 值並不小，$0.7<|\beta_n|<0.9$，其中 β_n 即是向量 $\mathbf{u}_n$ 與系統右

端向量 **r** 的投影量內積之結果。所以，Cond_E(**F**)比 Cond(**F**)小許多。假若 $\beta_n \to 0$ 時，則 Cond_E(**F**)值將與 Cond(**F**)的值差不多。針對不同系統（來自於不同微分方程）β_n 值會不同，所以 condition number 有不同程度的改進。

考慮另一卜松方程如下：

$$u''(x) = e^x, \quad 0 < x < 1 \tag{4.13}$$

$$u(0) = 1 \tag{4.14}$$

$$u(1) = e \tag{4.15}$$

與前一範例相比較，因問題解不同，所以線性系統左端矩陣相同，但右端向量不同。表八及表九分別是選用二階及三階基底函數的計算出之結果。由表八及表九可看出此範例的 effective condition number 比上一個例子小許多，而其 β_n 值比上一個例子大些，因為 $2.4 < |\beta_n| < 2.5$。同樣地，當 $\beta_n = 0$ 或 $\beta_n \to 0$ 時，Cond_E(F) 將等於 Cond(F)。

N_p	$h = 1/(N_p - 1)$	Cond(**F**)	Cond_E(**F**)	σ_1	σ_n	β_n
6	1/5	359.2	3.34	207.3	0.577	-2.468
7	1/6	563.9	3.65	301.3	0.534	2.462
9	1/8	1147.4	4.20	540.8	0.471	2.454
11	1/10	1991.3	4.69	848.9	0.426	-2.449
13	1/12	3125.8	5.13	1225.9	0.392	-2.446
17	1/16	6373.6	5.92	2185.9	0.343	2.442
21	1/20	11085.4	6.61	3420.7	0.309	-2.440

表八：新與傳統 condition numbers 之比較（$n = 2$）

N_p	$h=1/(N_p-1)$	Cond(**F**)	Cond_E(**F**)	σ_1	σ_n	β_n
6	1/5	404.1	3.34	233.2	0.577	2.469
7	1/6	598.2	3.65	319.7	0.534	2.462
9	1/8	1194.4	4.20	562.9	0.471	2.454
11	1/10	2063.7	4.69	879.8	0.426	2.450
13	1/12	3229.7	5.13	1225.9	0.392	-2.446
17	1/16	6565.2	5.91	2251.6	0.343	-2.442
21	1/20	11401.1	6.61	3518.1	0.309	2.440

表九：新與傳統 condition numbers 之比較（$n=3$）

從表六至表九可觀察出：effective condition number 的增長是相當緩慢地。可估計出大約是下列成長的速率

$$\text{Cond\_E}(\mathbf{F}) \approx \text{O}\left(N_p^{\frac{1}{2}} \right) \tag{4.16}$$

與傳統估計相比

$$\text{O}\left(n^2 N_p^{\,2}\right) \approx \text{Cond}(\mathbf{F}) > \text{Cond\_E}(\mathbf{F}) \approx \text{O}\left(N_p^{\frac{1}{2}} \right) \tag{4.17}$$

兩數值結果與理論分析相吻合。

第三節 Effective condition numbers 之推導

首先針對『方陣型』線性系統作相對誤差界之推導。

定理 1： 考慮一滿秩的方陣型線性系統，$A\mathbf{x}=\mathbf{b}$，其中 $A\in R^{n\times n}$，$\mathbf{x}\in R^n$, $\mathbf{b}\in R^n$。

矩陣 A 可對角化，其特徵向量及特徵值之關係為：$\lambda_1 \geq \lambda_2 \geq \cdots \lambda_n$，$A\mathbf{u}_i = \lambda_i \mathbf{u}_i$，$\forall i$。我們假設 $\hat{\mathbf{x}} = \mathbf{x} + \Delta\mathbf{x}$ 滿足

$$\mathbf{A}(\mathbf{x} + \Delta\mathbf{x}) = \mathbf{b} + \Delta\mathbf{b} \text{ ，} \Delta\mathbf{x} \in R^n \text{ ，} \Delta\mathbf{b} \in R^n \tag{4.16}$$

則存在相對誤差

$$\frac{\|\Delta\mathbf{x}\|}{\|\mathbf{x}\|} \leq \text{Cond\_E}(\mathbf{A}) \times \frac{\|\Delta\mathbf{b}\|}{\|\mathbf{b}\|} \tag{4.17}$$

其中

$$\text{Cond\_E}(\mathbf{A}) = \frac{\|\mathbf{b}\|}{\sqrt{\frac{\|\mathbf{b}\|^2 - \beta_n^{\,2}}{\text{Cond}(A)^2} + \beta_n^{\,2}}}, \qquad \beta_n = \mathbf{u}_n^{\,T}\mathbf{b} \tag{4.18}$$。

證明：

假設 A 為滿秩對稱矩陣，則可對角化，$A = U\Lambda U^T$ ，其中 $\Lambda = \text{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)$ ，$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n > 0$ ，且 $U = (\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_n)$ ， $\mathbf{u}_i^{\,T}\mathbf{u}_j = \delta_{ij}$ 。 因為 $A\mathbf{u}_i = \lambda_i \mathbf{u}_i$ ，所以可得

$$A^{-1}\mathbf{u}_i = \frac{1}{\lambda_i}\mathbf{u}_i \text{，} \forall i \tag{4.19}$$

令 $\mathbf{b} = \sum_{i=1}^{n} \beta_i \mathbf{u}_i$ ，則 $\beta_i = \mathbf{u}_i^{\,T}\mathbf{b}$ 且

$$\|\mathbf{b}\| = \sqrt{\sum_{i=1}^{n} \beta_i^{\,2}} \tag{4.20}$$

我們再令 $\Delta\mathbf{b} = \sum_{i=1}^{n} \alpha_i \mathbf{u}_i$ ，則 $\alpha_i = \mathbf{u}_i^{\,T}\Delta\mathbf{b}$ ，所以

$$\|\Delta\mathbf{b}\| = \sqrt{\sum_{i=1}^{n} \alpha_i^{\,2}} \,. \tag{4.21}$$

藉由 (4.19)，我們可得

$$\mathbf{x} = A^{-1}\mathbf{b} = A^{-1}\sum_{i=1}^{n} \beta_i \mathbf{u}_i = \sum_{i=1}^{n} \beta_i A^{-1}\mathbf{u}_i = \sum_{i=1}^{n} \beta_i \frac{1}{\lambda_i}\mathbf{u}_i = \sum_{i=1}^{n} \left(\frac{\beta_i}{\lambda_i}\right) \mathbf{u}_i \tag{4.22}$$

因此

$$\|\mathbf{x}\| = \sqrt{\sum_{i=1}^{n}\left(\frac{\beta_i}{\lambda_i}\right)^2}\ . \tag{4.23}$$

相似地，我們可得

$$\|\Delta\mathbf{x}\| = \sqrt{\sum_{i=1}^{n}\left(\frac{\alpha_i}{\lambda_i}\right)^2}\ . \tag{4.24}$$

結合 (4.24)及(4.21)，我們可推得

$$\|\Delta\mathbf{x}\|^2 = \sum_{i=1}^{n}\frac{\alpha_i^{\,2}}{\lambda_i^{\,2}} \le \frac{\sum_{i=1}^{n}\alpha_i^{\,2}}{\lambda_n^{\,2}} = \frac{\|\Delta\mathbf{b}\|^2}{\lambda_n^{\,2}}$$

兩端取方根

$$\|\Delta\mathbf{x}\| \le \frac{\|\Delta\mathbf{b}\|}{\lambda_n} \tag{4.25}$$

再利用 (4.23)，我們得到

$$\|\mathbf{x}\|^2 = \sum_{i=1}^{n}\frac{\beta_i^{\,2}}{\lambda_i^{\,2}} = \sum_{i=1}^{n-1}\frac{\beta_i^{\,2}}{\lambda_i^{\,2}} + \frac{\beta_n^{\,2}}{\lambda_n^{\,2}} \ge \frac{\sum_{i=1}^{n-1}\beta_i^{\,2}}{\lambda_1^{\,2}} + \frac{\beta_n^{\,2}}{\lambda_n^{\,2}} = \frac{1}{\lambda_n^{\,2}}\left\{\frac{\|\mathbf{b}\|^2 - \beta_n^{\,2}}{\left(\lambda_1/\lambda_n\right)^2} + \beta_n^{\,2}\right\} \tag{4.26}$$

兩端取方根

$$\|\mathbf{x}\| \ge \frac{1}{\lambda_n}\sqrt{\frac{\|\mathbf{b}\|^2 - \beta_n^{\,2}}{\mathrm{Cond(A)}^2} + \beta_n^{\,2}} \tag{4.27}$$

結合 (4.25) 和 (4.27) ，

$$\frac{\|\Delta\mathbf{x}\|}{\|\mathbf{x}\|} \le \frac{\|\Delta\mathbf{b}\|}{\lambda_n}\cdot\frac{\lambda_n}{\sqrt{\frac{\|\mathbf{b}\|^2 - \beta_n^{\,2}}{\mathrm{Cond(A)}^2} + \beta_n^{\,2}}}$$

$$= \frac{\|\mathbf{b}\|}{\sqrt{\frac{\|\mathbf{b}\|^2 - \beta_n^{\,2}}{\mathrm{Cond(A)}^2} + \beta_n^{\,2}}}\cdot\frac{\|\Delta\mathbf{b}\|}{\|\mathbf{b}\|} =: Cond\_E(A)\times\frac{\|\Delta\mathbf{b}\|}{\|\mathbf{b}\|} \tag{4.28}$$

其中 Cond(A) $= \lambda_1 / \lambda_n$ 。

接下來針對『非方陣型』線性系統做推導。

定理 2： 考慮一滿秩非方陣線性系統 $A\mathbf{y}=\mathbf{b}$，其 $A\in R^{m\times n},\ m\geq n$，$\mathbf{y}\in R^n$，$\mathbf{b}\in R^m$。矩陣 A 可作奇異值分解，$A=U\Sigma V^T$，它的奇異值與左特徵向量及右特徵向量之關係為 $A\mathbf{v}_i=\sigma_i\mathbf{u}_i,\forall i$，$\sigma_1\geq\sigma_2\cdots\geq\sigma_n>0$。我們假設 $\hat{\mathbf{y}}=\mathbf{y}+\Delta\mathbf{y}$ 滿足

$$\mathbf{A}\left(\mathbf{y}+\Delta\mathbf{y}\right)=\mathbf{b}+\Delta\mathbf{b},\quad \Delta\mathbf{y}\in R^n,\quad \Delta\mathbf{b}\in R^n \tag{4.29}$$

則存在相對誤差

$$\frac{\left\|\Delta\mathbf{y}\right\|}{\left\|\mathbf{y}\right\|}\leq \text{Cond\_E}(\mathbf{A})\times\frac{\left\|P\Delta\mathbf{b}\right\|}{\left\|P\mathbf{b}\right\|} \tag{4.30}$$

其中

$$\text{Cond\_E}(\mathbf{A})=\frac{\left\|\mathbf{b}\right\|}{\sqrt{\dfrac{\left\|\mathbf{b}\right\|^2-\beta_n{}^2}{\text{Cond}^2(\text{A})}+\beta_n{}^2}},\quad \beta_n=\mathbf{u}_n{}^T\mathbf{b} \tag{4.31}$$

且 $\mathbf{Ay}=P\mathbf{b}$，P 為正交投影算子。

證明：

設矩陣 A 為滿秩，則具有 SVD 分解，$A=U\Sigma V^T$，其中 $U=\left(\mathbf{u}_1,\mathbf{u}_2,\cdots,\mathbf{u}_m\right)$，且

$$V=\left(\mathbf{v}_1,\mathbf{v}_2,\cdots,\mathbf{v}_n\right),\quad \Sigma=\begin{bmatrix}\sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_n \\ & & 0 & \end{bmatrix},\quad \sigma_1\geq\sigma_2\geq\cdots\geq\sigma_n>0$$

$\mathbf{u}_i{}^T\mathbf{u}_j=\delta_{ij}$，$\mathbf{v}_i{}^T\mathbf{v}_j=\delta_{ij}$。進一步可得 $AV=U\Sigma$，即 $A\mathbf{v}_i=\sigma_i\mathbf{u}_i,\forall i$，或記作

$$A^+\mathbf{u}_i=\frac{1}{\sigma_i}\mathbf{v}_i,\forall i \tag{4.32}$$

令 $P\mathbf{b}=\sum_{i=1}^{n}\beta_i\mathbf{u}_i$，則 $\beta_i=\mathbf{u}_i{}^TP\mathbf{b}$ 且

$$\|P\mathbf{b}\| = \sqrt{\sum_{i=1}^{n} \beta_i^2} \tag{4.33}$$

我們再令 $P\Delta\mathbf{b} = \sum_{i=1}^{n} \alpha_i \mathbf{u}_i$ ，則 $\alpha_i = \mathbf{u}_i^T P\Delta\mathbf{b}$ ，可得

$$\|P\Delta\mathbf{b}\| = \sqrt{\sum_{i=1}^{n} \alpha_i^2} \tag{4.34}$$

從 (4.32)，且因為 $A^+(I-P)=0$ 我們可得

$$\mathbf{y} = A^+\mathbf{b} = A^+P\mathbf{b} = A^+\sum_{i=1}^{n}\beta_i\mathbf{u}_i = \sum_{i=1}^{n}\beta_i A^+\mathbf{u}_i = \sum_{i=1}^{n}\beta_i\frac{1}{\sigma_i}\mathbf{v}_i = \sum_{i=1}^{n}\left(\frac{\beta_i}{\sigma_i}\right)\mathbf{v}_i \tag{4.35}$$

因此

$$\|\mathbf{y}\| = \sqrt{\sum_{i=1}^{n}\left(\frac{\beta_i}{\sigma_i}\right)^2} \tag{4.36}$$

相似地，可得

$$\|\Delta\mathbf{y}\| = \sqrt{\sum_{i=1}^{n}\left(\frac{\alpha_i}{\sigma_i}\right)^2} \tag{4.37}$$

結合(4.34)及(4.37)，我們可得

$$\|\Delta\mathbf{y}\|^2 = \sum_{i=1}^{n}\frac{\alpha_i^2}{\sigma_i^2} \le \frac{\sum_{i=1}^{n}\alpha_i^2}{\sigma_n^2} = \frac{\|P\Delta\mathbf{b}\|^2}{\sigma_n^2} \tag{4.38}$$

兩端取方根

$$\|\Delta\mathbf{y}\| \le \frac{\|P\Delta\mathbf{b}\|}{\sigma_n} \tag{4.39}$$

從 (4.36) 式，我們進一步可得

$$\|\mathbf{y}\|^2 = \sum_{i=1}^{n}\frac{\beta_i^2}{\sigma_i^2} = \sum_{i=1}^{n-1}\frac{\beta_i^2}{\sigma_i^2} + \frac{\beta_n^2}{\sigma_n^2} \ge \frac{\sum_{i=1}^{n-1}\beta_i^2}{\sigma_1^2} + \frac{\beta_n^2}{\sigma_n^2} = \frac{1}{\sigma_n^2}\left\{\frac{\|P\mathbf{b}\|^2 - \beta_n^2}{(\sigma_1/\sigma_n)^2} + \beta_n^2\right\} \tag{4.40}$$

因此，得到

$$\|\mathbf{y}\| \ge \frac{1}{\sigma_n}\sqrt{\frac{\|\mathrm{Pb}\|^2 - \beta_n^2}{\mathrm{Cond(A)}^2} + \beta_n^2} \tag{4.41}$$

結合 (4.39) 及 (4.41)

$$\frac{\|\Delta \mathbf{y}\|}{\|\mathbf{y}\|} \le \frac{\|\mathrm{P}\Delta \mathbf{b}\|}{\sigma_{\mathrm{n}}} \cdot \frac{\sigma_{\mathrm{n}}}{\sqrt{\frac{\|\mathrm{P}\mathbf{b}\|^2 - \beta_{\mathrm{n}}{}^2}{\mathrm{Cond}^2(\mathrm{A})} + \beta_{\mathrm{n}}{}^2}}$$

$$= \frac{\|\mathrm{P}\mathbf{b}\|}{\sqrt{\frac{\|\mathrm{P}\mathbf{b}\|^2 - \beta_{\mathrm{n}}{}^2}{\mathrm{Cond}(\mathrm{A})^2} + \beta_{\mathrm{n}}{}^2}} \cdot \frac{\|\mathrm{P}\Delta \mathbf{b}\|}{\|\mathrm{P}\mathbf{b}\|} =: \mathrm{Cond\_E}(\mathrm{A}) \times \frac{\|\mathrm{P}\Delta \mathbf{b}\|}{\|\mathrm{P}\mathbf{b}\|} \qquad (4.42)$$

其中 $\mathrm{Cond}(\mathrm{A}) = \sigma_1 / \sigma_n$ 。

第五章 結論

本論文主要研究再生函數配置法的穩定性，但先探討擾動性：系統解的相對誤差是矩陣 condition number 乘上右端向量的相對誤差。

由第三章分析之結果可知利用再生函數配置法求偏微分方程所得到之線性系統 $\mathbf{Fy}=\mathbf{r}$ ，傳統之 condition number 與再生函數的階數 n 及再生核心基底函數 N_p 成平方正比關係

$$\mathrm{Cond}(\mathbf{F})=\mathrm{O}\left(n^2 {N_p}^2\right) \tag{5.1}$$

又可寫成

$$\mathrm{Cond}(\mathbf{F})\approx \mathrm{O}\left(n^2 h^{-2}\right), \qquad h={N_p}^{-1} \tag{5.2}$$

此結果相似於有限元素法及有限差分法，是相當穩定的方法。

而在第四章我們探討較新式的 condition number，稱為 effective condition number，其實這 effective condition number 比起傳統 condition number 是更精確地描述穩定性，因為它與線性系統右端項 $\mathbf{r}$ 密切相關。

因此，effective condition number 是比較客觀的一個指標。當一個方程的邊界條件有更改時，系統 $F\mathbf{y}=\mathbf{r}$ 的右端 $\mathbf{r}$ 就會有些許更改，β_n 就有些不同，因此系統的擾動也會有些許不同。只要 $\beta_n\neq 0$ ，新的 condition number 一定比傳統的 condition number 來的小。因此，系統解的相對誤差並比想像中的好。

參考文獻

[1] W. K. Liu, T. Belytschko and J. T. Oden（Eds.）, Special Issue on Meshless Methods, *Comput. Methods Appl. Mech. Engrg.,* Vol. 139, 1996.

[2] J. S. Chen and W. K. Liu（Eds.）, Special Issue on Meshfree Methods: Recent Advances and New Applications, *Comput. Methods Appl. Mech. Engrg.,* Vol. 193, 2004.

[3] Z. C. Li, T. T. Lu, H. Y. Hu and A. H.-D. Cheng, **Trefftz and Collocation Methods**, WITpress, Southampton, UK, 2008.

[4] E. J. Kansa, Multiqudrics -A scattered data approximation scheme with applications to computational fluid dynamics Ⅰ&Ⅱ, *Computer Math. Applic*, Vol.19, pp. 127-161, 1990.

[5] R. Schaback, Error estimates and condition numbers for radial basis function interpolation, *Advances in Computational Mathematics*, Vol. 3, pp. 251-264, 1995.

[6] W. K. Liu, S. Jun and Y. F. Zhang, Reprodicing kernel particle methods, *Int. J. Numer. Methods Fluids*, Vol. 20, pp. 1081-1106, 1995.

[7] H. Y. Hu, C. K. Lai and J. S. Chen, A study on convergence and complexity of reproducing kernel collocation method, *Interaction Multiscale Mech.,* Vol.2, No.3, pp. 295-319, 2009

[8] J. S. Chen, C. T. Wu. S. Yoon and Y. You, A stabilized conforming nodal integration for Galerkin meshfree method, *Int. J. Numer. Methods Eng.,* Vol. 50, pp. 435-466, 2001.

[9] W. Han and X. Meng, Error analysis of the reproducing kernel particle method, *Comput. Methods Appl. Mech. Engrg.,* Vol. 190, pp. 6157-6181, 2001.

[10] H. Y. Hu, J. S. Chen and W. Hu, Error analysis of collocation method based on reproducing kernel approximation, *Numer. Methods Partial Differential Eq.,* to appear, 2010.

[11] G. H. Golub and C. F. Van Loan, **Matrix Computations**, The Johns Hopkins University Press, Baltimore and London, 1996.

[12] J. W. Demmel, **Applied Numerical Linear Algebra**, SIAM, Philadelphia, PA, 1997.

[13] F. C. Chan and D. E. Foulser, Effectively well-conditioned linear systems, *SIAM J. Stat. Comput.*. Vol. 9, pp. 963-969, 1988.

[14] Z. C. Li and H. T. Huang, Effective condition number for numerical partial differential equations, *Numer. Linear Algebra Appl.,* Vol. 15, pp. 575-594, 2008.

附錄 A　方陣型線性系統

考慮 $A\mathbf{x}=\mathbf{b}$ ，$A\in R^{n\times n}$ ，是一滿秩方陣型矩陣，$\mathbf{x}\in R^n$，$\mathbf{b}\in R^n$ ，其特徵值分別為 $\lambda_1\geq\lambda_2\geq\cdots\lambda_n$ ，特徵函數 $\mathbf{u}_1,\mathbf{u}_2,\cdots,\mathbf{u}_n$，並滿足 $A\mathbf{u}_i=\lambda_i\mathbf{u}_i$，$\forall i$

<u>第一種情形</u>：　$A(\mathbf{x}+\Delta\mathbf{x})=\mathbf{b}+\Delta\mathbf{b}$ ，其中 $\Delta\mathbf{x}\in R^n$ ，$\Delta\mathbf{b}\in R^n$ 則存在誤差邊界

$$\frac{\|\Delta\mathbf{x}\|}{\|\mathbf{x}\|}\leq \mathrm{Cond(A)}\frac{\|\Delta\mathbf{b}\|}{\|\mathbf{b}\|} \tag{A.1}$$

其中

$$\mathrm{Cond(A)}=\frac{\lambda_{\max}}{\lambda_{\min}} \tag{A.2}$$

<u>第二種情形</u>：　$(A+\delta A)(\mathbf{x}+\delta\mathbf{x})=\mathbf{b}+\delta\mathbf{b}$ ，$\delta x\in R^n$，　$\delta b\in R^n$ 。其中 $\delta A\in R^{m\times n}$, 則存在誤差邊界

$$\frac{\|\delta\ \ \mathbf{x}\|}{\|\mathbf{x}\|}\leq\frac{\mathrm{Cond(A)}}{1-\mathrm{Cond(A)}\dfrac{\|\delta\mathrm{A}\|}{\|\mathrm{A}\|}}\left\{\frac{\|\delta\mathrm{A}\|}{\|\mathrm{A}\|}+\frac{\|\delta\ \ \mathbf{b}\|}{\|\mathbf{b}\|}\right\}$$

(A.3)

其中 Cond(A)的定義如 (A.2)。

證明一：

因為 $A\mathbf{x}+A\Delta\mathbf{x}=A(\mathbf{x}+\Delta\mathbf{x})=\mathbf{b}+\Delta\mathbf{b}$ 且 $A\mathbf{x}=\mathbf{b}$ 。因此，$A\Delta\mathbf{x}=\Delta\mathbf{b}$ 或 $\Delta\mathbf{x}=A^{-1}\Delta\mathbf{b}$ 。我們可得

$$\|\Delta\mathbf{x}\|\leq\|A^{-1}\|\|\Delta\mathbf{b}\| \tag{A.4}$$

另外，我們得到

$$\|\mathbf{b}\|\leq\|A\|\|\mathbf{x}\| \tag{A.5}$$

將(A.5)整理可得

$$\frac{1}{\|\mathbf{x}\|}\leq\frac{\|A\|}{\|\mathbf{b}\|} \tag{A.6}$$

結合(A.4)和(A.6) 則有下列估計

$$\frac{\|\Delta\mathbf{x}\|}{\|\mathbf{x}\|} \le \|A\|\|A^{-1}\|\frac{\|\Delta\mathbf{b}\|}{\|\mathbf{b}\|} =: \mathrm{Cond(A)}\frac{\|\Delta\mathbf{b}\|}{\|\mathbf{b}\|} \tag{A.7}$$

其中 $\|A\| = \|A\|_2$。因為 $\|A\| = \sqrt{\rho(A^T A)} = \max_i \lambda_i = \lambda_1$，且 $\|A^{-1}\| = \max_i \frac{1}{\lambda_i} = \frac{1}{\min_i \lambda_i} = \frac{1}{\lambda_n}$

所以 $\mathrm{Cond(A)} = \frac{\lambda_1}{\lambda_n}$ 。

證明二：

因為 $(A+\delta A)(\mathbf{x}+\delta\mathbf{x}) = \mathbf{b}+\delta\mathbf{b}$ 且 $A\mathbf{x} = \mathbf{b}$ 。因此，我們得 $(A+\delta A)\delta\mathbf{x} = \delta\mathbf{b} - \delta A\mathbf{x}$

或寫作

$$\delta\mathbf{x} = (A+\delta A)^{-1}(\delta\mathbf{b} - \delta A\mathbf{x}) \tag{A.8}$$

進一步可得

$$\|\delta\mathbf{x}\| \le \|(A+\delta A)^{-1}\|(\|\delta\mathbf{b}\| + \|\delta A\|\|\mathbf{x}\|) \tag{A.9}$$

其中

$$\|(A+\delta A)^{-1}\| = \|(I + A^{-1}\delta A)^{-1}A^{-1}\| \le \|(I + A^{-1}\delta A)^{-1}\|\|A\|^{-1}$$

$$\le \frac{\|A^{-1}\|}{1-\|A^{-1}\delta A\|} \le \frac{\|A^{-1}\|}{1-\|A^{-1}\|\|\delta A\|} \tag{A.10}$$

將(A.10)放入(A.9)，可得

$$\|\delta\mathbf{x}\| \le \frac{\|A^{-1}\|}{1-\|A^{-1}\|\|\delta A\|}\{\|\delta\mathbf{b}\| + \|\delta A\|\|\mathbf{x}\|\}$$

$$= \frac{\|A^{-1}\|\|A\|}{1-\|A^{-1}\|\|\delta A\|}\left\{\frac{\|\delta\mathbf{b}\|}{\|A\|} + \frac{\|\delta A\|\|\mathbf{x}\|}{\|A\|}\right\} \tag{A.11}$$

此外，我們可得

$$\frac{1}{\|\mathbf{x}\|} \le \frac{\|A\|}{\|\mathbf{b}\|} \tag{A.12}$$

整理可得

$$\frac{1}{\|A\|\|\mathbf{x}\|} \le \frac{1}{\|\mathbf{b}\|} \tag{A.13}$$

結合(A.11) - (A.12)及 (A.13)，我們可得

$$\frac{\|\delta\mathbf{x}\|}{\|\mathbf{x}\|} \le \frac{\|A^{-1}\|\|A\|}{1-\|A^{-1}\|\|A\|\frac{\|\delta A\|}{\|A\|}}\left\{\frac{\|\delta\mathbf{b}\|}{\|A\|\|\mathbf{x}\|}+\frac{\|\delta A\|}{\|A\|}\right\}$$

$$\le \frac{\mathrm{Cond(A)}}{1-\mathrm{Cond(A)}\frac{\|\delta\mathrm{A}\|}{\|\mathrm{A}\|}}\left\{\frac{\|\delta\mathbf{b}\|}{\|\mathbf{b}\|}+\frac{\|\delta\mathrm{A}\|}{\|\mathrm{A}\|}\right\} \tag{A.14}$$

附錄 B 非方陣型線性系統

考慮 $A\mathbf{y}=\mathbf{b}$ 其中 $A\in R^{m\times n},\ m\geq n$ ，是一滿秩非方陣型矩陣，$\mathbf{y}\in R^{n},\mathbf{b}\in R^{m}$ 。將 A 作奇異質分解，$A=U\Sigma V^{T}$ ，$A\mathbf{v}_i=\sigma_i\mathbf{u}_i$ ，其奇異值關係為 $\sigma_1\geq\sigma_2\cdots\geq\sigma_n>0,\forall i$

<u>第一種情形</u>： $A(\mathbf{y}+\Delta\mathbf{y})=\mathbf{b}+\Delta\mathbf{b}$ ，其中 $\Delta\mathbf{y}\in R^{n},\Delta\mathbf{b}\in R^{m}$ 。則存在誤差界

$$\frac{\|\Delta\mathbf{y}\|}{\|\mathbf{y}\|}\leq \mathrm{Cond(A)}\frac{\|P\Delta\mathbf{b}\|}{\|P\mathbf{b}\|} \tag{B.1}$$

其中

$$\mathrm{Cond(A)}=\frac{\sigma_1}{\sigma_n} \tag{B.2}$$

且 P 為正交投影，$P=A\left(A^{T}A\right)^{-1}A^{T}$ ，$\beta_n=\mathbf{u}_n{}^{T}\mathbf{b}$ ，$A\mathbf{y}=P\mathbf{b}$ 。

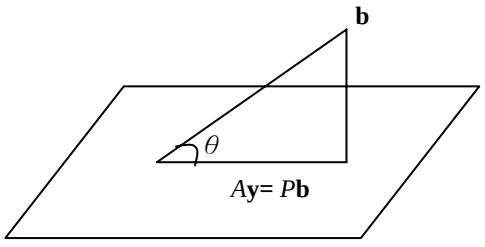
<u>第二種情形</u>： $(A+\delta A)(\mathbf{y}+\delta\mathbf{y})=\mathbf{b}+\delta\mathbf{b}$ ，其中 $\delta A\in R^{m\times n},\ \delta\mathbf{y}\in R^{n},\ \ \delta\mathbf{b}\in R^{m}$ 。則存在誤差界如下

$$\frac{\|\delta\mathbf{y}\|}{\|\mathbf{y}\|}\leq\frac{\varepsilon_{\mathrm{A}}+\varepsilon_{\mathrm{A}}\cdot\varepsilon_{\mathrm{b}}+\overline{\varepsilon_{\mathrm{b}}}\cdot\mathrm{Cond(A)}}{1-\left(\varepsilon_{\mathrm{A}}+\varepsilon_{\mathrm{A}}{}^{2}+\varepsilon_{\mathrm{A}}\cdot\mathrm{Cond(A)}\right)}\cdot\sec(\theta) \tag{B.3}$$

其中

$$\varepsilon_A=\frac{\|\delta A\|}{\|A\|},\varepsilon_b=\frac{\|\delta\mathbf{b}\|}{\|\mathbf{b}\|},\overline{\varepsilon_b}=\frac{\|P\delta\mathbf{b}\|}{\|\mathbf{b}\|},\ \cos(\theta)=\frac{\|A\mathbf{y}\|}{\|\mathbf{b}\|}=\frac{\|P\mathbf{b}\|}{\|\mathbf{b}\|} \tag{B.4}$$

角度 θ 為 向量 **b** 和 $A\mathbf{y}$ 之間的角度。



令 $A\in R^{m\times n},m\geq n$ ，$m\times n$ 矩陣為滿秩，則 $A\mathbf{y}=\mathbf{b}$ 有唯一的最小二乘解，其滿

足 $A^TA\mathbf{y}=A^T\mathbf{b}$。

最小二乘系統具有下列特性：

(i) 我們表示 $A^+\equiv\left(A^TA\right)^{-1}A^T$，則此解可表示為 $\mathbf{y}=A^+\mathbf{b}$。A^+ 稱為 A 的偽逆矩陣。假如 $m=n$，則 A 是非奇異，且 $A^+=\left(A^TA\right)^{-1}A^T=A^{-1}\left(A^T\right)^{-1}A^T=A^{-1}$。

(ii) 令 $A=U\Sigma V^T$，為 A 的 SVD，其中

$$\Sigma=\begin{bmatrix}\sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_n \\ \hline & & 0 & \end{bmatrix},\ \sigma_1\geq\sigma_2\geq\cdots\geq\sigma_n>0$$

因此

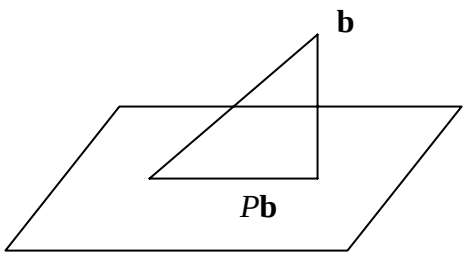
$$A^+=\left(A^TA\right)^{-1}A^T=\left[\left(V\Sigma^TU^T\right)\left(U\Sigma V^T\right)\right]^{-1}V\Sigma^TU^T=\left[V\Sigma^T\Sigma V^T\right]^{-1}V\Sigma^TU^T$$

$$=V\begin{bmatrix}\sigma_1^{\,2} & & & \\ & \sigma_2^{\,2} & & \\ & & \ddots & \\ & & & \sigma_n^{\,2}\end{bmatrix}^{-1}V^{-1}V\Sigma^TU^T$$

$$=V\left(\Sigma^T\Sigma\right)^{-1}\Sigma^TU^T=:V\Sigma^+U^T=V\left[\begin{array}{ccc|c}\frac{1}{\sigma_1} & & & \\ & \ddots & & 0 \\ & & \frac{1}{\sigma_n} & \end{array}\right]U^T$$

(iii) 令 $P=AA^+$ 為 $R(A)$ 之上的正交投影，則 $A^+(I-P)=0$ 且 $A\mathbf{y}=P\mathbf{b}$。

$$A^+(I-P)=A^+-A^+P=A^+-A^+AA^+=A^+-(A^TA)^{-1}A^TAA^+=A^+-IA^+=A^+-A^+=0$$



證明三：

假設 $\mathbf{y}$ 和 $\hat{\mathbf{y}} = \mathbf{y} + \Delta\mathbf{y}$ 分別為 $A\mathbf{y} = \mathbf{b}$ 和 $A\hat{\mathbf{y}} = \mathbf{b} + \Delta\mathbf{b}$ 的最小二乘解。則我們有

$$\begin{aligned}\Delta\mathbf{y} = \hat{\mathbf{y}} - \mathbf{y} &= A^{+}(\mathbf{b} + \Delta\mathbf{b}) - A^{+}\mathbf{b} = A^{+}\Delta\mathbf{b} \\ &= A^{+}(P\Delta\mathbf{b} + (I - P)\Delta\mathbf{b}) \\ &= A^{+}P\Delta\mathbf{b} + A^{+}(I - P)\Delta\mathbf{b} = A^{+}P\Delta\mathbf{b}\end{aligned} \tag{B.5}$$

因此

$$\|\Delta\mathbf{y}\| \le \|A^{+}\| \, \|P\Delta\mathbf{b}\| \tag{B.6}$$

此外， 因 $A\mathbf{y} = P\mathbf{b}$

$$\|P\mathbf{b}\| \le \|A\|\|\mathbf{y}\| \tag{B.7}$$

整理可得

$$\frac{1}{\|\mathbf{y}\|} \le \frac{\|A\|}{\|P\mathbf{b}\|} \tag{B.8}$$

結合(B.6)和(B.8)，則得到

$$\frac{\|\Delta\mathbf{y}\|}{\|\mathbf{y}\|} \le \|A\|\|A^{+}\|\frac{\|P\Delta\mathbf{b}\|}{\|P\mathbf{b}\|} =: \mathrm{Cond(A)}\frac{\|P\Delta\mathbf{b}\|}{\|P\mathbf{b}\|} \tag{B.9}$$

因為 $\|A\| = \|A\|_2 = \|U\Sigma V^T\|_2 = \|\Sigma V^T\|_2 = \|\Sigma\|_2 = \sigma_1$， $\|A^{+}\| = \|U\Sigma^{+}V^T\|_2 = \|\Sigma^{+}\|_2 = 1/\sigma_n$ 。

所以， $\mathrm{Cond(A)} = \dfrac{\sigma_1}{\sigma_n}$ 。

證明四：

假設 $\mathbf{y}$ 和 $\tilde{\mathbf{y}} = \mathbf{y} + \delta\mathbf{y}$ 分別為 $A\mathbf{y} = \mathbf{b}$ 和 $(A + \delta A)\tilde{\mathbf{y}} = \mathbf{b} + \delta\mathbf{b}$ 的最小二乘解。則我們有

$$\delta\mathbf{y} = \tilde{\mathbf{y}} - \mathbf{y} = (A + \delta A)^{+}(\mathbf{b} + \delta\mathbf{b}) - A^{+}\mathbf{b}$$

$$= \left((A + \delta A)^T (A + \delta A)\right)^{-1}(A + \delta A)^T(\mathbf{b} + \delta\mathbf{b}) - A^{+}\mathbf{b}$$

$$= \left(A^T A + \delta A^T A + A^T \delta A + \delta A^T \delta A\right)^{-1} \left(A^T + \delta A^T\right)\left(\mathbf{b} + \delta \mathbf{b}\right) - A^+ \mathbf{b}$$

$$= \left[I + \left(A^T A\right)^{-1} \delta A^T A + \left(A^T A\right)^{-1} A^T \delta A + \left(A^T A\right)^{-1} \delta A^T \delta A\right]^{-1} \left(A^T A\right)^{-1} \left(A^T + \delta A^T\right)\left(\mathbf{b} + \delta \mathbf{b}\right) - A^+ \mathbf{b}$$

然而 $A^+ = \left(A^T A\right)^{-1} A^T$ ，我們將 $\left(A^T A\right)^{-1} \delta A^T$ 記作 A^* 。因此得到

$$\delta \mathbf{y} = \left(I + A^* A + A^+ \delta A + A^* \delta A\right)^{-1} \left(A^+ + A^*\right)\left(\mathbf{b} + \delta \mathbf{b}\right) - A^+ \mathbf{b}$$

$$\leq \left(I + A^* A + A^+ \delta A + A^* \delta A\right)^{-1} \left(A^+ \delta \mathbf{b} + A^* \mathbf{b} + A^* \delta \mathbf{b}\right)$$

$$= \left(I + A^* A + A^+ \delta A + A^* \delta A\right)^{-1} \left(A^+ P \delta \mathbf{b} + A^* \mathbf{b} + A^* \delta \mathbf{b}\right) \qquad \text{(B.10)}$$

因為

$$\left\| \left(I + A^* A + A^+ \delta A + A^* \delta A\right)^{-1} \right\| \leq \frac{1}{1 - \left\| A^* A + A^+ \delta A + A^* \delta A \right\|}$$

$$\leq \frac{1}{1 - \left(\left\| A^* \right\| \left\| A \right\| + \left\| A^+ \right\| \left\| \delta A \right\| + \left\| A^* \right\| \left\| \delta A \right\| \right)} \qquad \text{(B.11)}$$

且

$$\left\| A^* \right\| \leq \frac{\left\| \delta A \right\|}{\left\| A \right\|^2} \qquad \text{(B.12)}$$

(B.12)整理可得

$$\left\| A^* \right\| \left\| A \right\| \leq \frac{\left\| \delta A \right\|}{\left\| A \right\|} \qquad \text{(B.13)}$$

此外

$$\frac{1}{\left\| \mathbf{y} \right\|} \leq \frac{\left\| A \right\|}{\left\| P\mathbf{b} \right\|} \qquad \text{(B.14)}$$

結合 (B.10)至(B.14)，可得

$$\frac{\left\| \delta \quad \mathbf{y} \right\|}{\left\| \mathbf{y} \right\|} \leq \frac{\left\| A^+ \right\| \left\| P\delta \quad \mathbf{b} \right\| + \left\| A^* \right\| \left\| \mathbf{b} \right\| + \left\| A^* \right\| \left\| \delta \quad \mathbf{b} \right\|}{1 - \left(\left\| A^* \right\| \left\| A \right\| + \left\| A^+ \right\| \left\| \delta A \right\| + \left\| A^* \right\| \left\| \delta A \right\| \right)} \cdot \frac{\left\| A \right\|}{\left\| P\mathbf{b} \right\|} \qquad \text{(B.15)}$$

其中

$$\varepsilon_A = \frac{\|\delta A\|}{\|A\|}, \varepsilon_b = \frac{\|\delta \mathbf{b}\|}{\|\mathbf{b}\|}, \bar{\varepsilon}_b = \frac{\|P\delta \mathbf{b}\|}{\|\mathbf{b}\|}, \ \cos(\theta) = \frac{\|P\mathbf{b}\|}{\|\mathbf{b}\|}$$

因此，(B.15) 變為

$$\frac{\|\delta \mathbf{y}\|}{\|\mathbf{y}\|} \le \frac{\mathrm{Cond(A)} \cdot \bar{\varepsilon}_\mathrm{b} + \varepsilon_\mathrm{A} + \varepsilon_\mathrm{A} \cdot \varepsilon_\mathrm{b}}{1 - \left(\varepsilon_\mathrm{A} + \varepsilon_\mathrm{A} \cdot \mathrm{Cond(A)} + {\varepsilon_\mathrm{A}}^2\right)} \frac{\|\mathbf{b}\|}{\|\mathrm{P}\mathbf{b}\|}$$

$$\le \frac{\varepsilon_\mathrm{A} + \varepsilon_\mathrm{A} \cdot \varepsilon_\mathrm{b} + \bar{\varepsilon}_\mathrm{b} \cdot \mathrm{Cond(A)}}{1 - \left(\varepsilon_\mathrm{A} + \varepsilon_\mathrm{A} \cdot \mathrm{Cond(A)} + {\varepsilon_\mathrm{A}}^2\right)} \cdot \sec\theta \qquad \text{(B.16)}$$

其中

$$\mathrm{Cond(A)} = \frac{\sigma_1}{\sigma_n}$$

附註：當 $\delta A = 0$ ，則 $\varepsilon_A = 0$ ，我們可得

$$\frac{\|\delta \mathbf{y}\|}{\|\mathbf{y}\|} \le \frac{\bar{\varepsilon}_\mathrm{b} \cdot \mathrm{Cond(A)}}{1 - 0} \cdot \frac{\|\mathbf{b}\|}{\|\mathrm{P}\mathbf{b}\|}$$

$$= \mathrm{Cond(A)} \cdot \frac{\|P\delta \mathbf{b}\|}{\|\mathbf{b}\|} \cdot \frac{\|\mathbf{b}\|}{\|P\mathbf{b}\|}$$

$$= \mathrm{Cond(A)} \ \frac{\|P\delta \mathbf{b}\|}{\|P\mathbf{b}\|}$$

相當於(B.1)。