第一章

有限元素法的回顧

1-1 引言

在這一章, 我們簡介一維有限元素法的理論. 我們將介紹廣義程序並且說明有限元方程式如何直接由支配方程式 (governing equation, 也就是描述能量的微分方程) 導出, 而不依賴任何古典變分學原理. 兩種方法可以達到此目的: 一個是運用加權餘數法 (Galerkin method). 另一個方法是從整體能量平衡的觀點產生. 但是本篇論文著重於 Galerkin Method, 所以不去討論另一方法.

1-2 什麼是加權餘數法?

加權餘數法是一個求取線性與非線性偏微分方程式近似解的技巧. 加權餘數法牽涉到兩個步驟, 第一步假設一個近似滿足於已知微分方程和邊界條件的一般場變數函數 (不管它是壓力, 溫度, 位移, 應力, 或其它量). 將此近似函數代入原始微分方程式和邊界條件中得到一個殘量 (residual), 這個殘量在整個解空間內必須平均為零.

第二步係求解一所得的方程式 (或方程組), 並且將一般函數型式特殊化, 以得到一特別函數, 此特別函數就是想求得近似解.

我們考慮一典型的問題. 假設我們想求出一求解區域為 D 和邊界為 ∂D 的問題, 其支配方程為

$$L(u) = f \tag{1.1}$$

(L 為橢圓微分算子). u 為場變數, 函數 f 是主變數 x 的已知函數, 而且假設在 ∂D 上有適當的邊界條件. 我們可以利用如下兩個步驟來應用加權餘數法.

首先, 未知正確解 u 以 u_h 估計, 其中 u_h 的函數行為可以用未知參數, 以及所有獨立變數來表示 (所謂獨立變數在數學上我們稱為基底函數), 但基底函數關係為已定義,

而對剩餘的這個變數其函數關係未定義. 因此場變數能以下式表示之

$$u \approx u_h = \sum_{i=1}^{N} u_i \phi_i, \tag{1.2}$$

其中 u_h 是逼近解 (也稱為 trial function), ϕ_i 是基底函數, u_i 是未知參數或是基底函數的未知函數, 相加的上限 N 就是未知 u_i 的個數, N 個 ϕ_i 的選擇方式使得它們能夠滿足總體的邊界條件.

當 u_h 代入方程式 (1.1),滿足 (1.1) 是不可能的, 即

$$L(u_h) - f \neq 0$$

或者

$$L(u_h) - f = R,$$

其中 R 為殘量. 加權餘數的想法是想求取 N 個未知數 u_i 使得整個區域的誤差變小. 將誤差加權平均並要求整個區域的加權平均為零即可完成此一程序. 因此我們選擇 N 個線性獨立的加權函數 W_i 並且要求: 如果

$$\int_D [L(u_h) - f]\, W_i\, dx = \int_D R\, W_i\, dx = 0, \quad i = 1, \ldots, N \tag{1.3}$$

則 $R \approx 0$ 方程式 (1.3) 表示的誤差分佈原理之型式在於加權函數的選擇, 一旦指定加權函數, 方程式代表 N 個方程組. 它將是求解 u_i 的代數方程式. 第二步驟則是求解方程式 (1.3) 的 u_i, 並利用 (1.2) 求得 u 的近似解. 因為加權餘數法和誤差分配原理的選擇有多種可供使用, 如果加權函數與代表 u 的基底函數相同, 我們則稱此法為 Bubnov-Galerkin 法, 不過在這簡稱 Galerkin 法.

Galerkin 法要求,

$$\int_D [L(u_h) - f]\, \phi_i\, dx = 0, \quad i = 1, \ldots, N \tag{1.4}$$

以上的討論我們假設為處理整個區域. 但是因為方程式 (1.1) 對整個區域內任何點都成立, 所以我們把整個區域切割成有限的小區間, 每個區間我們當成一個元素 (element) ,

而整個區域內之任何元素對 (1.1) 式都成立, 因此我們對每一個元素也定義如同方程式 (1.2) 的局部近似函數.

現可用場變數的有限元素表示法, 函數 ϕ_i 轉換成定義於元素內的內插函數 (interpolation function) $\phi_i^{(e)}$. u_i 爲待定之參數, 可能爲場變數或是場變數的節點值. 因此利用 Galerkin Method 我們可以寫下支配元素的方程式爲

$$\int_{D^{(e)}} [L(u_h^{(e)}) - f^{(e)}]\, \phi_i^{(e)}\, dD = 0, \quad i = 1, \ldots, r, \tag{1.5}$$

其中上標 (e) 指明對任一個元素, 而且對整個組合中的單一元素都有如方程式 (1.5) 的一組方程式. 在元素方程式組合成系統方程式之前, 我們要求基底函數 ϕ_i 的選擇必須滿足下列幾項條件:

(1) 基底函數的生成是在元素與元素間, 和元素內選取片段連續函數 (piecewise linear) , 並且要求在元素與元素間必須連續.

(2) 基底函數必須是夠平滑 (smooth) , 並且是在測試函數 (test function) H_0^1 空間內 (ie.$H_0^1(\Omega) = \{u \in H^1(\Omega) | u(0) = u(l) = 0\},\ where\ \Omega = (0, l)$).

1-3 一維 Poisson 方程式

現在我們考慮求取一函數 u 使能滿足

$$-\frac{d^2u}{dx^2} = f(x), \quad 0 < x < l$$

及邊界條件 $u(0) = \alpha$, $u(l) = \beta$. 我們首先用 $u_h(x)$ 來估計未知眞正解 $u(x)$, 而 $u(x)$ 有下列之型式

$$u \approx u_h = \sum_{i=1}^{N} u_i\, \phi_i, \tag{1.6}$$

其中元素基底函數爲 ϕ_i, 我們選擇 ϕ_i 爲一階線性函數, 且對二節點元素 $N = 2$ (i.e. $x_0 = 0$ 與 $x_1 = l$) . 未知節點值爲 $u_i, i = 1, 2$. 我們在元素這一層次不考慮固定邊界條件, 而在組合程序之後才予以考慮. 我們將看到元素方程式的導出過程中會出現自然邊界條件.

運用 Galerkin 準則和方程式 (1.3), 我們得到

$$\int_{x_1}^{x_2} (\frac{d^2 u}{dx^2} + f(x))\, \phi_i(x)\, dx = 0, \quad i = 1, 2, \tag{1.7}$$

其中 x_1 和 x_2 是線元素的端節點之座標. 我們對 u_h 的導數項作分部積分, 得到

$$\phi_i \frac{du_h}{dx}|_{x_1}^{x_2} - \int_{x_1}^{x_2} \frac{du_h}{dx}\frac{d\phi_i}{dx}\, dx + \int_{x_1}^{x_2} f(x)\phi_i\, dx = 0, \quad i = 1, 2. \tag{1.8}$$

利用 u_h 的定義, 我們知道

$$\frac{du_h}{dx} = \sum_{i=1}^{N} \frac{d\phi_i}{dx}\, u_i = [\frac{d\phi_i}{dx}]\{u_i\}^{(e)}$$

其中 $\{u_i\}^{(e)}$ 是未知數 u_i 所形成的行向量. 因此 Galerkin 準則變成

$$\int_{x_1}^{x_2} [\frac{d\phi_j}{dx}]\, \frac{d\phi_i}{dx}\, dx\, \{u_j\}^{(e)} = \phi_i\, \frac{du_h}{dx}|_{x_1}^{x_2} + \int_{x_1}^{x_2} f(x)\, u_h(x)\, dx \quad i = 1, 2. \tag{1.9}$$

方程式右手邊之第一項即爲元素之自然邊界條件. 計算自然邊界條件可以得到

$$\begin{aligned} \phi_i \frac{du_h}{dx}|_{x_1}^{x_2} &= \phi_1(x_2)\frac{du_h}{dx}(x_2) - \phi_1(x_1)\frac{du_h}{dx}(x_1) = -\frac{du_h}{dx}(x_1), \quad i = 1, \\ \phi_i \frac{du_h}{dx}|_{x_1}^{x_2} &= \phi_2(x_2)\frac{du_h}{dx}(x_2) - \phi_2(x_1)\frac{du_h}{dx}(x_1) = \frac{du_h}{dx}(x_2), \quad i = 2. \end{aligned}$$

其中我們用 ϕ_i 的端點値 x_1, x_2. 元素方程式爲

$$\begin{pmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{pmatrix}^{(e)} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}^{(e)} = \begin{pmatrix} -\frac{du_h}{dx}(x_1) \\ \frac{du_h}{dx}(x_2) \end{pmatrix}^{(e)} + \begin{pmatrix} F_1 \\ F_2 \end{pmatrix}^{(e)} \tag{1.10}$$

其中

$$[K]^{(e)} = \int_{x_1}^{x_2} \begin{pmatrix} \frac{d\phi_1}{dx}\frac{d\phi_1}{dx} & \frac{d\phi_1}{dx}\frac{d\phi_2}{dx} \\ \frac{d\phi_2}{dx}\frac{d\phi_1}{dx} & \frac{d\phi_2}{dx}\frac{d\phi_2}{dx} \end{pmatrix} dx,$$

$$[F]^{(e)} = \int_{x_1}^{x_2} \begin{pmatrix} f\phi_1 \\ f\phi_2 \end{pmatrix} dx.$$

這組方程式描述了二點線元素的行爲. 方程式 (1.10) 適用於二節點元素. 對 N 節點元素其步驟是相同的, 但 $i = 1, 2, \ldots, N$. N 節點元素之矩陣包含如方程式 (1.10) 的項, 但有更多的行及列説明其 N 個元素的方程式. 組合元素方程式時須考慮自然邊界條件. 在矩陣組合過程中自然邊界條件 du_h/dx 項在求解區域的所有内節點會互相抵消, 剩下在

點 0 或點 l 的自然邊界條件. 如果端點之 u 值已被指定, 例如 $u(0)=\alpha$, 那麼 du/dx 爲未知, 如果 $du(0)/dx$ 已指定, 那麼 $u(0)$ 爲未知.

這個例子中我們看到分部積分有兩個重要的角色: 第一它使邊界條件產生效果, 第二它將矩陣 $[K]^{(e)}$ 中積分式的最高階導數降爲 1 階. 如此近似函數只須要保有函數值得連續, 而不須要保有函數斜率的連續.

1-4 元素的組合

根據上一節我們所討論的, 每一個元素中的基底都是由線性函數所組成, 任一個元素會有兩個節點, 所以對於每一個元素就會有兩個方程式, 如下:

$$\left.\begin{array}{l} k_{11}^{(e)}u_1^{(e)}+k_{12}^{(e)}u_2^{(e)}=f_1^{(e)} \\ k_{21}^{(e)}u_1^{(e)}+k_{22}^{(e)}u_2^{(e)}=f_2^{(e)} \end{array}\right\} \tag{1.11}$$

(1.11) 的 k 與 (1.10) 的 K 不一樣, 前者以一個元素來看, 後者是以一個基底函數來看. 關於下標 1 和 2 是元素的端節點局部編號, 例如我們找一個元素是介在 6 和 7 之間, 那麼 u_1^e 就代表 u_6, u_2^e 就代表 u_7 我們考慮一個 $N-1$ 個元素, 和 N個節點的區域, 這代表了有 N 個自由度的 N 組方程式, 所以我們可以造出一個 $N\times N$ 的剛度矩陣 (stiffness matrix), 和一個 $N\times 1$ 的載入向量 (load vector).

然而元素矩陣可視爲整個系統的部份矩陣, 對於第一個元素, 它是介在點 x_1 和點 x_2 之間, 根據 (1.11)

$$k_{11}^1u_1+k_{12}^1u_2=f_1^1$$

$$k_{21}^1u_1+k_{22}^1u_2=f_2^1$$

我們將這兩組方程式加入於一個描述整個區域的 $N\times N$ 矩陣中的第一列和第二列.

接下來我們繼續看第二個元素, 它的節點介於第二個元素與第三個元素之間, 根據 (1.11) 我們得到兩個方程式, 分別加在前一個方程組的第二列與第三列, 則兩個元三個節點我們總共有三個方程式,

$$\begin{array}{rcrcl} k_{11}^1u_1+ & & & = & f_1^1 \\ k_{21}^1u_1 & +(k_{22}^1+k_{11}^2)u_2 & +k_{12}^2u_3 & = & f_2^1+f_1^2 \\ & k_{21}^2u_2 & +k_{22}^2u_3 & = & f_2^2 \end{array}$$

如果我們繼續處理整個區域内的 $N-1$ 個元, 我們就會得到 N 個方程式, 如果我們把這 N 個方程式表爲

$$\begin{bmatrix} \tilde{K}_{11} & K_{12} & 0 & 0 & \dots & 0 & 0 \\ K_{21} & K_{22} & K_{23} & 0 & \dots & 0 & 0 \\ 0 & K_{32} & K_{33} & K_{34} & \dots & 0 & 0 \\ 0 & 0 & K_{43} & K_{44} & \dots & 0 & 0 \\ \\ . & . & . & . & \dots & . & . \\ \\ 0 & 0 & 0 & 0 & \dots & K_{N-1,N-1} & K_{N-1,N} \\ 0 & 0 & 0 & 0 & \dots & K_{N-1,N} & \tilde{K}_{NN} \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \\ . \\ . \\ . \\ u_{N-1} \\ u_N \end{bmatrix} = \begin{bmatrix} \tilde{F}_1 \\ F_2 \\ F_3 \\ F_4 \\ . \\ . \\ . \\ F_{N-1} \\ \tilde{F}_N \end{bmatrix} \tag{1.12}$$

而 $N \times N$ 矩陣是由

$$\begin{bmatrix} k_{11}^1 & k_{12}^1 & 0 & 0 & \dots & 0 & 0 \\ k_{21}^1 & k_{22}^1 + k_{11}^2 & k_{12}^2 & 0 & \dots & 0 & 0 \\ 0 & k_{21}^2 & k_{22}^2 + k_{11}^3 & k_{12}^3 & \dots & 0 & 0 \\ 0 & 0 & k_{21}^3 & k_{22}^3 + k_{11}^4 & \dots & 0 & 0 \\ \\ . & . & . & . & \dots & . & . \\ \\ 0 & 0 & 0 & 0 & \dots & k_{22}^{N-2} + k_{11}^{N-1} & k_{12}^{N-1} \\ 0 & 0 & 0 & 0 & \dots & k_{21}^{N-1} & k_{22}^{N-1} \end{bmatrix} \tag{1.13}$$

另外

$$\begin{bmatrix} \tilde{F}_1 \\ F_2 \\ F_3 \\ . \\ . \\ . \\ F_{N-1} \\ \tilde{F}_N \end{bmatrix} = \begin{bmatrix} f_1^1 \\ f_2^1 + f_1^2 \\ f_2^2 + f_1^3 \\ . \\ . \\ . \\ f_2^{N-2} + f_1^{N-1} \\ f_2^{N-1} \end{bmatrix} \tag{1.14}$$

剛度矩陣 (1.12) 是一個稀疏矩陣 (sparse matrix), 並且在本例中是一對稱矩陣 (本論文不討論不對稱問題), 矩陣的係數 $\tilde{K}_{11}$ 和 $\tilde{K}_{NN}$, 及負載向量 (load vactor, 在工程方面代表節點的載重) $\tilde{F}_1$ 和 $\tilde{F}_N$ 還不是一般的形態, 因爲我們還沒考慮邊界條件, 所以它們不是一般式 ($K_{ij} = \int_0^\ell \phi_i\prime \phi_j\prime \, dx$ 和 $F_i = \int_0^\ell \phi_i \, dx, i = 1, 2, \dots, N$), 我們將在下一節討論邊界條件.

1-5 邊界條件

考慮以下兩種邊界條件

1. Dirichlet Boundary Condition:

邊界條件形式爲

$$u(0) = \alpha, \quad u(l) = \beta. \tag{1.15}$$

所以只有 $N-2$ 個未知節點 $u_2, u_3, \ldots, u_{N-1}$ 剩下, 所以 (1.12) 變爲一 $(N-2)\times(N-2)$ 矩陣

$$\begin{bmatrix} K_{22} & K_{23} & 0 & \ldots & 0 & 0 & 0 \\ K_{32} & K_{33} & K_{34} & \ldots & 0 & 0 & 0 \\ 0 & K_{43} & K_{44} & \ldots & 0 & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \ldots & \cdot & \cdot \\ 0 & 0 & 0 & 0 & \ldots & K_{N-1,N-2} & K_{N-1,N-1} \end{bmatrix} \begin{bmatrix} u_2 \\ u_3 \\ u_4 \\ \cdot \\ \cdot \\ \cdot \\ u_{N-1} \end{bmatrix} = \begin{bmatrix} F_2 - K_{21}\alpha \\ F_3 \\ F_4 \\ \cdot \\ \cdot \\ \cdot \\ F_{N-1} - K_{N-1,N}\beta \end{bmatrix} \tag{1.16}$$

和另外兩列表達 x_1 和 x_N 的輔助方程式

$$\left.\begin{aligned} \tilde{K}_{11}\alpha + K_{12}u_2 &= f_1^1 \\ K_{N,N-1}u_{N-1} + \tilde{K}_{NN}\beta &= f_2^{N-1} \end{aligned}\right\} \tag{1.17}$$

這使得剛度矩陣 (1.16) 是一可逆矩陣, 所以 (1.16) 可以解得節點 $u_2, u_3, \ldots, u_{N-1}$. 當然 (1.17) 也提供了有用的資訊, 當 u_2 和 u_{N-1} 被解出來時, 則我們可以得到新的邊界點, 在線性基底方面, 我們可以以此類推算出其他的 u_i 值.

2. Neumann Boundary Condition:

當邊界條件爲

$$u'(0) = \gamma_0, \quad u'(l) = \gamma_l. \tag{1.18}$$

我們稱 (1.18) 爲 Neumann Boundary condition. 而滿足此邊界條件的方程式我們稱爲 Neumann Problem. 通常 Neuman Problem 需要比較明確的支配方程 (governing equation), 在一維的 Poisson 方程式中, 解 u 不唯一, 也就是說假如 u 滿足一維 Poisson 方程式和 (1.18), 則 $u+c_0$ 也會是滿足方程式的解. 因爲 (1.18) 類似於力學系統上的方程式, 常數 c_0 相當於剛體運動 (rigid motion,平移或旋轉), 假如我們想要得到唯一解則這個剛體運動必須被詳細敘述. 此外我們也可以從 (1.12) 得到其它的有限元逼近解, 所以 (1.12) 的解不唯一, 所以矩陣 K 是不可逆的.

根據前面敘述的剛體運動我們必須考慮其它的因素, 我們不能任意指定 f 和 γ_0, γ_l, 它們之間必須是相容的. 然而在 Neumann Problem 的變分形式下, 我們是去找一 $u \in H^1(0,1)$ 使得

$$\int_0^l u'v'dx = \int_0^l f\,v\,dx - \gamma_0\,v(0) + \gamma_l\,v(l), \quad \forall v \in H^1. \tag{1.19}$$

在一維的 Poisson 方程式中, 我們對方程式兩邊同時積分, 可以得到

$$\begin{aligned} -\int_0^l (u')'dx &= -[u'|_0^l] \\ &= u'(0) - u'(l) \\ &= \gamma_0 - \gamma_l \\ &= \int_0^l f(x)dx \end{aligned} \tag{1.20}$$

所以我們得到

$$\int_0^l f(x)dx - \gamma_0 + \gamma_l = 0 \tag{1.21}$$

(1.19) 假如要解得唯一解的話, (1.21) 對 (1.19) 來說是必要 (necessary) 條件. 我們以物理學上的觀點來看 (1.21) 的話, 它就是所謂的整域性能量守衡定律 (global conservation law).

如果我們把 (1.21) 問題離散化, 我們可以得到

$$\tilde{F}_1 + \sum_{i=2}^{N-1} F_i + \tilde{F}_N = 0 \tag{1.22}$$

爲了惕除掉剛體運動, 我們可以指定任意一個 u_h 的節點值 u_j (因爲解是相對解, 而不是絕對解), 使得 $u_j = c_0$ (c_0 是任意值). 我們假設 (1.12) 式中 $u_1 = c_0$, 則我們得到

$$\begin{bmatrix} K_{22} & K_{23} & 0 & \dots & 0 & 0 & 0 \\ K_{32} & K_{33} & K_{34} & \dots & 0 & 0 & 0 \\ 0 & K_{43} & K_{44} & \dots & 0 & 0 & 0 \\ \\ . & . & . & . & \dots & . & . \\ \\ 0 & 0 & 0 & 0 & \dots & K_{N,N-1} & \tilde{K}_{N,N} \end{bmatrix} \begin{bmatrix} u_2 \\ u_3 \\ u_4 \\ . \\ . \\ . \\ u_N \end{bmatrix} = \begin{bmatrix} F_2 - K_{21}c_0 \\ F_3 \\ F_4 \\ . \\ . \\ . \\ f_2^{N-1} + \gamma_l \end{bmatrix} \tag{1.23}$$

和

$$\tilde{K}_{11}c_0 + K_{12}u_2 = f_1^1 - \gamma_0 \tag{1.24}$$

方程式 (1.23) 可以解得唯一解 $u_2, u_3, \dots, u_N$. 通常我們直接令 $c_0 = 0$, 則我們可以由 (1.23) 或 (1.24) 解得 u_2. 值得注意的是 (1.22) 保證 (1.23) 和 (1.24) 相容, 當 (1.22) 成立 , u_2 同樣可以被 (1.24) 所解得.

1-6 矩陣條件數與誤差分析

基本上有限元素法把一微分方程問題轉換爲一解矩陣問題 (ie. $Ku = F$, K 是 stiff matrix ,u 是 u_i 的行向量, F 是 load vactor). 然而直觀上, $\hat{u}$ 爲 $Ku = F$ 之解 u 的一個近似解, 而且剩餘向量 $r = F - K\hat{u}$ 之 r 很小, 則似乎 $\|u - \hat{u}\|$ 也要很小. 通常這是如此, 不過某些實際上時常出現的系統卻沒有這種性質. 如果我們想要得到較小的剩餘向量同時保有準確的近似解, 在一般的情況下, 我們無法依賴系統之幾何來指出何時可能發生問題. 然而,我們可以考慮矩陣 K 之條件數.

定理 **1.1**

假定 $\hat{u}$ 爲 $Ku = F$ 之近似解, K 爲非奇異矩陣, 且 r 爲 $\hat{u}$ 之剩餘向量. 則對任一範數,

$$\|u - \hat{u}\| \le \|K^{-1}\|\|r\|$$

且

$$\frac{\|u - \hat{u}\|}{\|u\|} \le \|K\|\|K^{-1}\|\frac{\|r\|}{\|F\|}, \quad \text{if}\, u \ne 0 \,\text{and}\, F \ne 0. \tag{1.25}$$

定理 1.1 中的不等式蘊涵 $\|K\|\|K^{-1}\|$, 這兩個量提供了剩餘向量與近似解準確度之間的關連指標. 我們對相對誤差 $\|u-\hat{u}\|/\|u\|$ 最感興趣, 並且由不等式 (1.25) 知此相對誤差與相對剩餘 $\|r\|/\|F\|$ 之間的倍率關係是 $\|K\|\|K^{-1}\|$. 此近似解可以使用任何一種便利的範數; 唯一的要求是使用時要完全一致.

定義 1.2

非奇異矩陣 K 相對於範數 $\|\cdot\|$ 之條件數 (condition number) 爲

$$\kappa(K) = \|K\|\|K^{-1}\|,$$

利用這種符號, 定理 1.1 中的不等式成爲

$$\|u-\hat{u}\| \leq \kappa(K)\frac{\|r\|}{\|K\|}$$

與

$$\frac{\|u-\hat{u}\|}{\|u\|} \leq \kappa(K)\frac{\|r\|}{\|F\|}.$$

對於任何一個非奇異矩陣 K 以及任何一種範數 $\|\cdot\|$,

$$1 = \|I\| = \|K\cdot K^{-1}\| \leq \|K\|\|K^{-1}\| = \kappa(K).$$

如果 $\kappa(K)$ 靠近 1, 矩陣 K 就是行爲良好的矩陣, 稱爲條件良好的矩陣 (well-condition matrix); 當 $\kappa(K)$ 遠大於 1 時, 矩陣 K 並非行爲良好的矩陣, 稱爲病態矩陣 (ill-condition matrix). 此處之行爲乃參照於較小的剩餘向量蘊涵一個相應的準確近似解之相對安全性.

理論上, 雖然矩陣的條件數完全取決於此矩陣和其逆矩陣的範數, 但實際上逆矩陣之計算受限於機器誤差, 而與執行的準確度有關. 如果所執行的運算含有 t 位有效準確數字, 則矩陣 K 之近似條件數爲此矩陣之範數乘以利用 t 位數字算術所得到之 K 矩陣的近似逆矩陣範數. 實際上, 此條件數亦與計算 K 之逆矩陣時所用的方法有關.

若我們假設線性系統 $Ku=F$ 之近似解是利用 t 位數字計算以及 Gauss 消去法所求出的, 則我們可証明(參見 Forsythe 與 Moler [1]): 近似解 $\hat{u}$ 之剩餘向量 r 滿足

$$\|r\| \approx 10^{-t}\|K\|\|\hat{u}\|. \tag{1.26}$$

由這個近似式, 我們可以不必求矩陣 K 的逆矩陣, 就可以估計出以 t 位數字計算得到之實際條件數. 事實上, 這個近似式假設 Gauss 消去法中所有的數值運算都是利用 t 位數字算術來執行, 不過求剩餘向量所需的運算都是以雙倍準度 (即 $2t$ 位數字) 算術來執行. 這種手法並不會增加太多計算上的負擔, 而且免除了許多在計算剩餘向量時由於近數相減所失去的準確度.

考慮

$$Ky = r.$$

這組線性系統, 以求原 t 位數字之條件數 $\kappa(K)$ 的近似值. 事實上, $Ky = r$ 之近似解 $\hat{y}$ 滿足

$$\hat{y} \approx K^{-1}r = K^{-1}(F - K\hat{u}) = K^{-1}F - K^{-1}K\hat{u} = u - \hat{u}. \tag{1.27}$$

所以 $\hat{y}$ 爲求原系統之近似解時之誤差估計. 方程式 (1.26) 蘊涵

$$\|\hat{y}\| \approx \|u - \hat{u}\| = \|K^{-1}r\| \le \|K^{-1}\|\|r\| \approx \|K^{-1}\|(10^{-t}\|K\|\|\hat{u}\|) = 10^{-t}\|\hat{u}\|\kappa(K)$$

這給出了利用 Gauss 消去法與方才所描述的 t 爲數字型態之算術求解 $Ku = F$ 這系統時, 所含有的條件數之近似值:

$$\kappa(K) = \frac{\|\hat{y}\|}{\|\hat{u}\|}10^t. \tag{1.28}$$

在 (1.27) 式中, 我們用到了 $\hat{y} \approx u - \hat{u}$ 這個估計, 其中 $\hat{y}$ 爲 $Ky = r$ 這個系統的近似解. 大體上, $\hat{u} + \hat{y}$ 比原近似解 $\hat{u}$ 更近似於 $Ku = F$ 這組線性系統的解. 使用這種假設方法稱爲迭代改良 (iterative refinement 或 iterative improvement).

如果利用 t 位數字算術應用此過程, 而且假使 $\kappa_\infty(K) \approx 10^q$, 則在迭代改良 k 次之後, 解的準確位數大約爲 t 與 $k(t-q)$ 中之較小者. 若此系統條件良好, 一次或兩次迭代就將指出解是準確的. 除非矩陣 K 病態至$\kappa_\infty(K) > 10^t$, 否則在病態系統上還是有可能有極大的改良. 在此情況下, 計算時應利用更多的精度.

一般而言, 所有迭代法的收斂速度都與 $\kappa(K)$ 有關; κ 越靠近 1 (良好), 則收斂速度越快.

第二章

周期性問題與有限元素法

2-1 引言

有限元在一維問題上大部分處理 Dirichlet, Neumann 或是更進一步的 Mixed 問題. 現在我們給定週期性的邊界條件, 測試有限元素法 (內插函數爲一階線性多項式) 在一維週期問題上的表現, 目的在觀察是否和它在 Dirichlet 問題上的表現一致. 我們討論的函數有兩種, 分別爲在邊界上平滑與不平滑的週期函數 (週期函數可視爲兩端相接的函數). 我們將實驗兩者的條件數是否和非周期性有限元的條件數有很大的差異, 另外我們也將討論其平滑度對於逼近解的影響會有多大. 在論文裡我們所用的程式及程式術語皆爲 Matlab 的語法, 在此先說明.

2-2 主要理論

首先, 我們知道當 K 是對應 Dirichlet 問題的剛性矩陣 (stiffness matrix), 它的條件數 (condition number) 理論上應呈 $O(N^2) = O(h^{-2})$ 的成長 (參閱 [2]). 現在我們令討論區間 $\Omega = (0, 1)$, 兩端點 0 和 1. 令 $N = 2^\ell$, 其中 ℓ 是某個正整數. 若邊界條件爲 Dirichlet, 而有限元是分片一階多項式. 讓我們將 Ω 均勻切成 N 等分, 稱節點爲 $x_0 = 0$, $x_1 = h$, …, $x_N = 1$. 令對應 x_i 節點值爲 1 而對應其他 $x_{j\neq i}$ 節點值爲 0 的基底函數爲 $\phi_i(x)$. 那麼在 (2.1) 式中最左右兩端的基底函數, 也就是 $\phi_0(x)$ 和 $\phi_N(x)$, 將被省去 (因爲它們的係數必爲 0). 所以 K 的階數爲 $N-1$.

現在我們只考慮最簡單形式的兩點邊界值問題

$$\begin{aligned} -u'' &= f, \quad x \in \Omega; \\ u(0) &= u(1) = 0. \end{aligned} \tag{2.1}$$

令

$$u_h(x) = \sum_{i=1}^{N-1} u_i \phi_i(x)$$

是爲有限元解. 則矩陣 K 的元素是

$$K_{ij} = \int_0^1 \phi_i' \phi_j' \, dx, \quad 1 \le i, j \le N-1.$$

經過簡單的計算, 我們得到 $K = \frac{1}{h}\mathrm{diag}(-1, 2, -1)$. 我們測驗它的條件數. 爲了配合 Matlab 的 sparse matrix 功能 (它不能直接用 cond 指令,原因我們將在後面解釋), 我們選擇利用 1-norm 的條件數定義

$$\kappa(K) = \|K\|_1 \|K^{-1}\|_1. \tag{2.2}$$

另外, 在解剛度矩陣時, 根據 FRIED 在 [2] 中提到有關條件數與網格分割的大小有一關係存在, 我們發現

$$C_1 h^{-2} \le \kappa(K) \le C_2 h^{-2} \tag{2.3}$$

C_1, C_2是常數,並且

$$\kappa(K) = \frac{\lambda_{max}}{\lambda_{min}}$$

λ_{max} 和 λ_{min} 分別爲矩陣 K 的最大特徵值 (characteristic value) 和最小特徵值, 在一般矩陣分析書中我們可以找到假如矩陣 K 是對稱矩陣 (symetric matrix) 則 (2.3) 式成立. 所以 $\kappa(K)$ 乘以網格大小的次方將會是一定數. 我們在接下來的實驗一併加入討論.

根據上述定義所做的實驗, 所得的結果如下:

N	4	8	16	32	64	128	256	512	1024
$\kappa(A)$	12	40	144	544	2112	8320	33024	131584	525312
$\kappa(A)/N^2$	0.75	0.625	0.5625	0.5312	0.5156	0.5078	0.5039	0.5020	0.5010

我們觀察到, 實驗的結果確實與理論預測一致.

如果將邊界條件改成週期條件, 我們在考慮此問題時, 因爲基底函數 $\phi_0(x)$ 和 $\phi_N(x)$ 不能同時捨去 (因爲週期性的關係, $\phi_0(x)$ 和 $\phi_N(x)$ 兩函數首尾相連, 視爲同一基底函數). 我們令 $\phi_0(x) = \phi_N(x)$, 而取 $\phi_0(x), \phi_1(x), \phi_2(x), \ldots, \phi_{N-1}(x)$ 爲基底函數, 這時候矩陣 K 的階數爲 N, 而基本上矩陣仍然爲 (1.12) 的形式, 只不過右上角和左下角多了幾

個元素. K 的階數是 N, 而且基本上它仍然是 $K=\frac{1}{h}\text{diag}(-1,2,-1)$ 的形態, 只是增加了 $K_{1,N}=-\frac{1}{h}$, 和 $K_{N,1}\ =-\frac{1}{h}$

$$\begin{bmatrix} K_{11} & K_{12} & 0 & 0 & \dots & 0 & K_{1,N} \\ K_{21} & K_{22} & K_{23} & 0 & \dots & 0 & 0 \\ 0 & K_{32} & K_{33} & K_{34} & \dots & 0 & 0 \\ & & & & & & \\ . & . & . & . & \dots & . & . \\ & & & & & & \\ K_{N,1} & 0 & 0 & 0 & \dots & K_{N,N-1} & K_{N,N} \end{bmatrix} \begin{bmatrix} u_0 \\ u_1 \\ u_2 \\ . \\ . \\ . \\ u_{N-1} \end{bmatrix}$$

$$= \begin{bmatrix} \tilde{F}_0 \\ F_1 \\ F_2 \\ . \\ . \\ . \\ \tilde{F}_{N-1} \end{bmatrix} \tag{2.4}$$

但是, 這樣的 K 顯然是不可逆的. 原因是 K 的所有列之和爲 0 向量. 故其列向量爲線性相依 (linear depend) 之向量. 造成這個現像的原因是, 週期性邊界條件本身不能唯一決定一條解曲線. 很顯然的, 如果 $u_1(x)$ 是一個符合 $-u"(x)=f(x)$ 的週期解, 那麼 $u=u_1+c$ 也必然是一個符合 $-u"(x)=f(x)$ 的週期解. 所以矩陣 K 是一個不可逆矩陣 . 所以我們必須附加一個條件, 通常我們加上

$$\int_0^1 u_h\,dx = 1 \tag{2.5}$$

這個條件. 這個條件導出的等式是

$$\sum_{i=1}^{N-1} u_i \int_0^1 \phi_i\,dx = \sum_{i=1}^{N-1} hu_i = 1 \qquad (2.6).$$

(h 是不同基底的寬度) 我們用這條等式代替原聯立方程式的最後一條式子:

$$-\frac{1}{h}u_0 - \frac{1}{h}u_{N-2} + \frac{2}{h}u_{N-1} = (f,\phi_{N-1}) \qquad (2.7).$$

所以我們給定 (2.5) 式加以限制, 然後推導出 (2.6) 式(其中 $N-1$ 點改成 N 點), 並且取代原聯立方程式 (2.7), 所以矩陣 K 階數爲 N, 基本上只改變最後一列, 變成

$\frac{1}{h}(h^2, h^2, \ldots, h^2)$. 很明顯地, 這樣的矩陣非常不平衡. 因爲它的最後一列幾乎是 0 向量 (當網格越切越小時). 我們對這形式的矩陣計算其條件數及條件數與分割點平方的比, 所得的結果如下:

N	4	8	16	32	64	128	256	512	1024
κ	65	257	1025	4097	16385	65537	262145	1048577	4194305
κ/N^2	4.0625	4.0156	4.0039	4.0010	4.0002	4.0001	4.0000	4.0000	4.0000

我們觀察到, 實驗的結果也是與理論預測一致. 但是以絕對值而言, 周期條件下的剛性矩陣, 比 Dirichlet 條件下的要壞很多.

這裡要注意 K 已經不再是對稱矩陣. 當我們計算條件數時直接使用 $cond(K)$ 這個指令的話, 那麼所得的結果將看不出來呈現 $O(N^2)$ 的成長. 這或許是因爲 `Matlab` 以最大與最小矩陣特徵值的比例來當做條件數.

其實我們可以將最後一列的 h^2 移除到右端, 故得到一個如上的矩陣, 但其最後一列是 $\frac{1}{h}(1, 1, \ldots, 1)$. 這樣的矩陣比較平衡得多. 對這種形式的矩陣, 做條件數實驗結果如下:

N	4	8	16	32	64	128	256	512	1024
κ	5	20	80	320	1280	5120	20480	81920	327680
κ/N^2	0.3125	0.3125	0.3125	0.3125	0.3125	0.3125	0.3125	0.3125	0.3125

這是個非常令人高興的實驗結果. 以絕對值而言, 這種比較平衡的矩陣確實得到比較小的條件數.

2-3 誤差分析

在誤差分析方面, 基本上任一函數 u 經過投影到有限元空間中得到一 u_h, 而爲了求解 u_h 我們把它對應到一個矩陣然後去求解 $\hat{u}$, 在這過程當中投影時會造成第一次誤差, 在矩陣計算中又造成另一次誤差. 我們把解出來的 $\hat{u}$ 當作基底函數 $\phi(x)_i$ 的係數得到逼近解 u_h. 另外 $\|u - u_h\|$ 的誤差我們應該考慮在 H^1 中, 但是因爲 H^1 的計算較煩, 所以我們通常只考慮在 L^2 空間的誤差

$$\|u - u_h\| = \left(\int_\Omega |u - u_h|^2 dx\right)^{\frac{1}{2}} \tag{2.8}$$

接著我們做收斂速度的實驗. 我們取 $f(x) = 1$. 將 K 中的 $1/h$ 乘到右端, 那麼右端

向量基本上就是

$$h(f,\phi_i) = h\int_0^1 \phi_i(x)\,dx = h^2.$$

先對 Dirichlet 問題做實驗, 眞解是

$$u(x) = \frac{1}{2}x(1-x).$$

由於有限元的基底函數有內插性質 (interpolatory), 所以, 得到的向量解 u_i 就是 $u_h(x_i)$ 的值 $(1 \le i \le N-1)$. 在這些節點上可以觀察其超收斂性 (superconvergence). 也就是說, 在基底函數等於 1 的節點處 x_i, $u_h(x_i)$ 非常靠近眞解 $u(x_i)$. 事實上, 在這種簡單的狀態, 理論上 $u_h = u_I$; 也就是說, 有限元解也就是插值解. 換句話說, $u_h(x_i) = u(x_i)$. 我們以

$$\|u^h - u\|_s^2 := \sum_{i=1}^{N-1} |u^h(x_i) - u(x_i)|^2$$

來估計 $u_h(x)$ 在節點處的收斂速度. 實驗結果如下:

N	8	16	32	64	128	256	512	1024
$\|u^h - u\|_s$	10^{-16}	10^{-15}	10^{-14}	10^{-14}	10^{-13}	10^{-12}	10^{-11}	10^{-10}

可見這些誤差大概都是由機器造成的. 配合觀察 K 的條件數, 我們知道條件數的一個指數大約相當於計算值的一個失去的有效數字. 所以這樣的實驗結果非常合理.

令 $\|\cdot\|$ 代表 Ω 上的 L^2-norm. 那麼, 超收斂性不能帶到整個空間上. 我們以 1024 個點值來描繪 $u_h(x)$ (分片線性函數) 和 $u(x)$, 然後以簡單的梯形數值積分法來估計 $\|u_h - u\|$. 由實驗結果, 我們可觀察到如理論所估計的 h^2 收斂速度:

N	8	16	32	64	128	256	512
$N^2\|u_h - u\|$	0.0913	0.0913	0.0913	0.0913	0.0913	0.0911	0.0884

其實, 根本不必做這種實驗. 因爲如果 $u_h = u_I$, 那麼 $u_h(x)$ 其實就是 $u(x)$ 的一階插值多項式, 而其 L^2 誤差本該是 $O(h^2)$.

如果 $f(x) = 1$ 但是附加條件變成週期條件和 $\int_0^1 u\,dx = 1$, 則眞解應該是

$$u(x) = -\frac{1}{2}x^2 + \frac{1}{2}x + \frac{11}{12}, \quad x \in [0,1] \text{ 且週期爲 } 1.$$

或是任何的 $u(x-c), 0<c<1$. 注意, 此解函數 $u(x)$ 並非平滑的週期函數. 至於剛性矩陣 K, 我們取上述最後的形式; 也就是最後一個列全是 1. 我們將 K 的共同因數 $1/h$ 乘到右端向量, 故右端向量應是

$$\begin{pmatrix} h(f,\phi_0) \\ \vdots \\ h(f,\phi_{N-2}) \\ 1/h \end{pmatrix} = \begin{pmatrix} h^2 \\ \vdots \\ h^2 \\ 1/h \end{pmatrix}_{N\times 1}.$$

在這個模型之下, u_i 應該是 $u_h(x_{i-1})$, 其中 $1\le i\le N$. 而 $u_h(1)=u_h(0)=\mathtt{u(1)}$. 有趣的是, 我們發現, 這樣做出來的有限元解 $u_h(x)$ 恰好是 $u(x+h)$ 的逼近解. 這並沒有矛盾, 只是我們不太喜歡看到不平滑的部分出現在 Ω 中, 所以做了一點不自然的動作. 我們本該說 `u = [u' u(1)];` 的, 但是我們其實令

```
u = [u(N) u'];
```

因爲如果我們把算出來的 u_i 值打點在眞解 $u(x)$ 的線上, 我們會發現 u_1 是從第一個 step size 開始的, 而非從 0 點開始, 又因爲是週期函數的關係, 所以我們做了上述的處理, 把 u_N 搬到列向量的第一個位子.

此刻我們再用同樣的方法觀察 $\|u^h-u\|_s$ 和 $\|u^h-u\|$.

N	8	16	32	64	128	256	512
$\|u^h-u\|_s$	3.91E-3	1.34E-3	4.67E-4	1.64E-4	5.78E-5	2.04E-5	7.20E-6
$\|u^h-u\|$	5.82E-4	1.46E-4	3.66E-5	9.27E-6	2.44E-6	7.20E-7	2.51E-7
$N^2\|\cdot\|$	3.73E-2	3.73E-2	3.74E-2	3.80E-2	4.00E-2	4.72E-2	6.59E-2

超收斂的現象不見了, L^2 中依稀可見 $O(h^2)$ 的速度, 但是也不甚明確. 我們想要知道, 這是不是因爲 $u(x)$ 不平滑所致呢?

爲了製造一個平滑的週期函數, 我們選 $f(x)=2\pi^2\cos(4\pi x)$. 那麼眞解就是

$$u(x)=1+\frac{1}{8}\cos(4\pi x),\quad x\in[0,1] \text{ 且週期爲 } 1.$$

或是任何的 $u(x-c), 0<c<1$. 利用 `Maple` 的幫助, 我們得到右端向量的嚴格答案. 當 $1\le i\le N-1$ 時,

$$h(f,\phi_i)=f_1(x)|_{x_{i-1}}^{x_i}-x_{i-1}f_2(x)|_{x_{i-1}}^{x_i}+x_{i+1}f_2(x)|_{x_i}^{x_{i+1}}-f_1(x)|_{x_i}^{x_{i+1}};$$

當 $i=0$ 時,

$$h(f,\phi_0)=f_1(x)|_{1-h}^{1}-(1-h)f_2(x)|_{1-h}^{1}+hf_2(x)|_0^h-f_1(x)|_0^h.$$

其中

$$f_1(x)=\frac{1}{8}[\cos(4\pi x)+4\pi x\sin(4\pi x)],$$

$$f_2(x)=\frac{\pi}{2}\sin(4\pi x).$$

依此格式, 我們製造了兩個 `Matlab` 函數以計算 $f_1(x)$ 和 $f_2(x)$, 並且製造 (理論上無誤差的) 右端向量 . 解得了 $u_h(x)$ 之後, 我們發現這時候就沒有平移的現象; 數值解 $u_h(x)$ 看來就是 $u(x)$ 的逼近. 我們仍然觀察 $\|u_h-u\|_s$ 和 $\|u_h-u\|$.

N	8	16	32	64	128	256	512
$\|u^h-u\|_s$	10^{-15}	10^{-15}	10^{-14}	10^{-14}	10^{-13}	10^{-13}	10^{-12}
$\|u_h-u\|$	1.89E-2	4.91E-3	1.24E-3	3.11E-4	7.77E-5	1.94E-5	4.71E-6
$N^2\|\cdot\|$	1.21	1.26	1.27	1.27	1.27	1.27	1.23

此刻又復見超收斂的現象, 而 L^2 的收斂速度也很明顯地是 $O(h^2)$.

對週期函數而言, $u(x)$ 在 $(0,l)$ 內平滑還不夠, 須考慮 $(0-\epsilon,l+\epsilon)$ 內的情況, 因爲這樣才不會因爲邊界上的節點在經過矩陣迭代計算後, 影響整個區域的節點值, 因而造成誤差.

第三章
凌波函數與 Galerkin Method

3-1 引言

在 1-2 節我們提到 u_h 的基底函數可以選擇全域性或是局部性的函數, 本章將介紹新的 Galerkin Method, 一種用自格函數 (scaling function) 爲基底的有限元素法, 我們應用 Daubechies 的三階自格函數爲基底, 它不但具有一般有限元基底的良好性質: 連續性, 可微性與局部性, 並且具有正交性, 與多重解析 (multrisolution) 的能力, 我們根據第二章的方法代入自格函數基底, 比較自格函數與傳統有限元基底在週期性問題上的差異, 我們比較的函數也如同前一章, 分別爲邊界平滑與不平滑函數. 爲求公平與客觀, 我們使用相同的 f (load vector), 看看在逼近的準確度和條件數 (計算複雜度) 上的差別. 爲 3-6 節的 Wavelet-Galerkin Method 做鋪路.

第二節介紹一些我們將使用的自格函數基本性質, 在這我們不去証明它的性質和由來, 有興趣的話請參考一般 Daubechies, Cohen, ...等人的著作, 裡面有詳細說明與証明.

3-2 多重解分析

我們根據 2-2 節所做的一維 Poisson 方程式的步驟, 把眞解 u 投影到以自格函數爲基底的空間上, 並推導剛度矩陣 K 的基本形式.

我們假定區域 $\Omega = (0, 1)$, 同樣的我們只考慮最簡單的邊界值問題

$$\begin{aligned} -u''(x) &= f(x), \quad x \in \Omega, \\ u(0) &= u(1) = 0. \end{aligned} \tag{3.1}$$

令

$$u_h(x) = \sum_{m=0}^{N-1} u_i\, \phi_{j,m} \tag{3.2}$$

是爲有限元解.

上式中的 f 是一個可變動的函數, 在下兩節我們也將代入邊界平滑性函數與邊界不平滑性函數, 與傳統以 shape 爲基底的有限元作比較, 這一節我們將先推導一般式, 在

(3.2) 中的 $\phi_{j,m}(x)$ 我們給定 Daubechies 的自格函數, 其表達意思爲

$$\phi_{j,m}(x) = 2^{\frac{j}{2}}\,\phi(2^j x - m), \quad j, m \in Z$$

j 代表眞解欲作離散化而作投影所在的層數 (leval), m 則是 $\phi(2^j x)$ 的平移量, 這裡我們介紹一下自格函數的性質, 也比較一下兩種基底函數 (shape function 和 scaling function) 的差異性, 設 V_j 是一族 $L^2(R)$ 的子空間且滿足以下條件:

(1) $\cdots \subset V_{-1} \subset V_0 \subset V_1 \subset \cdots$

(2) $\cup_{j=-\infty}^{\infty} V_j$ 在 $L^2(R)$ 中稠密.

(3) $\cap_{j=-\infty}^{\infty} V_j = \{\emptyset\}$

(4) $u(x) \in V_j \Leftrightarrow u(2x) \in V_{j+1}$

(5) $u(x) \in V_j \Leftrightarrow u(x-m) \in V_j$

由以上的五個條件, 我們可知這些空間可以定義一個從 $u \in L^2(R)$ 到 V_j 的投影算子 (othogonal projection) P_j. 從 (2) 的式子可得知:

$$\lim_{j\to\infty} P_j u = u, \quad \text{for } u \in L^2(R).$$

這裡要說明的是, (2) 這個式子 $\cup_{j=-\infty}^{\infty} V_j$ 僅僅稠密在 $L^2(R)$, 不是等於它, 且這收斂是對範數收斂, 不是點收斂. 從 (3) 我們知

$$\lim_{j\to-\infty} P_j u = 0, \quad \text{for } u \in L^2(R).$$

對 (4) 這性質有很好的應用, 使細層與粗層有某種關係, 可經由某種簡單的運算從細層得到粗層的投影.

如果 $\phi \in L^2(R)$, 且滿足 (1) 至 (5) 的條件, 我們稱 ϕ 生成一個多重解分析 (MRA).

從多重解分析條件 (4), (5)中我們看到自格函數的脹縮與平移不變性, 然而, 對於多重解分析基底的要求, 假如我們選擇基底函數 $\phi(t)$ 在 V_0 裡, 則 $\phi(t)$ 將會和它的平移

$\phi(t-k)$ 互相正交(垂直), 並且對於 V_0 來說 $\phi(t)$ 與 $\phi(t-k)$ 會是它的一組單位正交基底, 另外, 當函數 $\phi(t-k)$ 是 V_0 的正交基底時, 它的 rescaled function $\sqrt{2}\phi(2t-k)$ 將是 V_1 的正交基底, 同理 rescaled 到第 j 層時, $\phi(2^j-k)$ 可被 $2^{j/2}$ 單位化爲第 j 層的基底函數.

說到這裡, 我們發現與傳統 shape function 有許多不同的地方, 在網格生成方面, 一般 shape function 的網格大小是人爲切割的, 與函數本身沒有關係, 但是自格函數基底網格的大小卻跟函數本身有很大的關係, 它取決於函數 scale 的大小, 也就是說它可以直接從函數 2^j 中的 j 來控制網格大小, 假設我們用第 4 層來看, 我們把 Ω 切成 16 塊 ($2^4=16$), 並且在函數裡我們就可以直接判斷網格數的多寡. 另外從基底函數的個數方面來看, 在傳統有限元素法中我們先決定切割元素的多寡, 再決定每一元素內的節點數目, 然後再決定用多少階的內插多項式來逼近在元素內的眞解, 一般有限元元素內所含基底個數與所用的函數多項式階數有關, 因爲使用 Lagrange 插值多項式 (基底函數形狀不唯一), 所以個數與元素內分割節點有關, 強調一點: 有限元的元素個數與節點個數不相等.

自格函數是一種可以被自己的漲縮與平移所展開的函數, 也就是一個基底函數就可以展開整個空間, 所以自格函數基底在元素內的個數與所選取的函數階數有關, 而自格函數又與眞解所含的多項式有關, 因爲

$$\int x^m\psi(x)dx=0, \quad m=0,1,\cdots,p-1. \tag{3.3}$$

本式代表凌波函數 $\psi(x)$ 的消失矩 (vanish moment), 其中 p 代表自格函數所用的階數, 因爲階數控制了函數的寬度 (support), 所以當函數平移時, 每一個元素內將有函數寬度的個數通過 (基底函數形狀保持唯一). 與一般有限元最大不同是節點個數與元素個數只差一而已 (也就是節點與所分割區域的關係).

我們更進一步分析 V_j 與其它層的關係, 假設 V_j 與 V_{j+1} 是一組粗細層的關係, 則我們定義 W_j 是V_j 在 V_{j+1} 層的正交補集合 (orthogonal complement) 也就是

$$V_{j+1}=V_j\oplus W_j,$$

則最少存在一組基底 $\psi(x)\in W_0$, 使得 $\psi(x-k), k\in Z$ 仍爲 W_0 的基底, 則函數 $\psi(x)$ 稱

爲凌波函數 (wavelet function), 定義

$$\psi_{j,k} = 2^{\frac{j}{2}}\,\psi(2^j x - k), \quad j,k \in Z$$

我們整理一下, 令 c_k, d_k 是自格函數與凌波函數的自格係數 (scaling coefficient), 則

$$\begin{aligned}\phi_{j-1,k} &= \sum_{k\in Z} c_k\,\phi_{j,k}(x)\\ \psi_{j-1,k} &= \sum_{k\in Z} d_k\,\phi_{j,k}(x)\end{aligned} \tag{3.4}$$

由凌波函數與自格函數所展開的空間中, 我們有兩組正交基底 $\mathcal{B},\mathcal{B}'$

$$\mathcal{B}' = (\phi_{J,k})_k,$$

$$\mathcal{B} = \cup_{j<J}(\psi_{j,k})_k.$$

這一章我們先以 ϕ 來當有限元的基底, 在下一章我們才引進 FWT (fast wavelet transform) 方法把以自格函數造出的剛度矩陣快速轉換爲以凌波函數 ψ 爲主的剛度矩陣, 然後比較其條件數與逼近能力.

3-3 區度上的週期凌波基底

應用凌波方法在閉區間的微分方程, 我們必須建構一個單位正交的基底函數在閉曲線 $\Gamma \in R^2$, 下面敘述的 "強迫週期性 (brutal periodization)" 單位正交基底在 Dorobantu 的論文 [4] 有詳細的介紹.

假設 ϕ 和 ψ 的寬度不超過 Γ 的長度, 我們用第 j_0 層的尺度 (scaled version) 來看它們, 則 ϕ 與 ψ 變爲 $2^{j_0/2}\phi(2^{j_0}x)$ 和 $2^{j_0/2}\psi(2^{j_0}x)$. 方便起見, 我們考慮區間 $I_0 = [0,1]$. 我們只要保持所有函數寬度都包含在 I_0 的這些基底函數, 假如 $\phi_{j,k}$ 這個函數的寬度超過了 I_0 的左邊, 則 $\phi_{j,k}$ 消失的部分剛好與 $\phi_{j,k+2^j}$ 在 I_0 右邊消失的 部分相互吻合, 把這兩函數加在一起和對 I_0 來限制他們的和, 非常簡單就可以看出新的函數一樣和其它基底函數正交, 也依然滿足 two-scale (3.4) 的關係 (註: (3.4) 是快速凌波轉換 (FWT) 的關鍵), 因此, 我們把週期化的 $\phi_{j,k}$ 和 $\psi_{j,k}$ 的函數定義爲 $\tilde{\phi}_{j,k}$, $\tilde{\psi}_{j,k}$ (在 3-5 節中我們會給個週期性的例子).

Theorem 3.1

$\tilde{\phi}_{j,k}$ and $\tilde{\psi}_{j,k}$ defined above form an orthonormal basis in $L^2(I_0)$. For each $j \geq 0$ they generate a sequence of subspace $\tilde{W}_j$ and $\tilde{W}_j$ such that

$$\tilde{V}_{j+1} = \tilde{V}_j + \tilde{W}_j, \qquad \tilde{V}_j \perp \tilde{W}_j, \qquad \overline{\cup_{-\infty}^{\infty} \tilde{V}_j} = L^2(I_0).$$

Let f be the periodic extension to the whole real axis of $\tilde{f} \in L^2(I_0)$, then

$$\begin{aligned} \langle f, \phi_{j,k} \rangle &= \langle \tilde{f}, \tilde{\phi}_{j,k} \rangle, \\ \langle f, \psi_{j,k} \rangle &= \langle \tilde{f}, \tilde{\psi}_{j,k} \rangle, \end{aligned} \tag{3.5}$$

The proof can be found in [4].

(3.4) 告訴我們 $\tilde{f} \in L^2(I_0)$ 的凌波分解可以視爲函數寬度與 I_0 相交的函數 f 的基本凌波分解, 根據上述我們定義一個值爲 $\tilde{\psi}_{j,k}$ 的消失矩 (vanish moment), (實際上 $\tilde{\psi}_{j,k}$ 等於 $\psi_{j,k}$). 從現在起我們不要用 "$\sim$" 這個符號, 我們稱所有的週期化基底爲凌波基底.

空間 $H^s(R)$ 是由無條件基底 $\psi_{j.k}$ 所構成的, 也就是說存在一個函數 $f \in H^s(R)$ 使得 f 可以由凌波函數乘上一組 特徵係數所展出,

Lemma 3.2

Let $f = \sum_{j,k} \langle f, \psi_{j.k} \rangle \psi_{j,k}$. Then $f \in H^s(R)$ if and only if

$$\sum_{j,k} |\langle f, \psi_{j,k} \rangle|^2 (1 + 4^{js}) < +\infty, \quad -r < s < r$$

where r is regularity of MRA .

這個證明我們可以在 Meyer 的書 [5] 中可以找到 . 此外, 當 $\{\psi_{j,k}\}$ 是一個 $H^s(R)$ 的 Riesz 基底, 我們也會有

$$C_1 \sum_{j,k} |\langle f, \psi_{j,k} \rangle|^2 (1 + 4^{js}) \leq \|f\|_{H^s}^2 \leq C_2 \sum_{j,k} |\langle f, \psi_{j,k} \rangle|^2 (1 + 4^{js}), \tag{3.6}$$

同樣的我們可以再以週期性凌波基底爲主的空間 $H^s(I_0)$ 中得到相同的結果.

Theorem 3.3

The characterization of Sobolev space $H^s(I_0)$ by means of wavelet coefficient is

$$\tilde{f} \in H^s(R) \quad iff \quad \sum_{j,k} | < f, \psi_{j,k} > |^2 (1 + 4^{js}) \leq +\infty, \qquad s \geq 0 \tag{3.7}$$

for $s < r$.

3-4 Scaling-Galerkin 法

根據前面敘述的過程, 我們假設函數投影到第 j 層的空間, 則我們把整個區間分割成 2^j 塊區間(元素), 選定點 $x_0, x_1, \cdots, x_{2^j-1}$, 使

$$0 = x_0 < x_1 < x_2 < \cdots < x_{2^j-2} < x_{2^j-1} = 1,$$

對於每一個 $i = 0, 1, 2, \ldots, 2^j - 1$. 令 $h_i = x_{i+1} - x_i$, 根據 3-2 節我們得到一序列基底函數, 分別爲 $\phi_{j,0}, \phi_{j,1}, \ldots, \phi_{j,2^j-1}$.

運用 Galerkin 方法, 我們把 (3.1) 式乘上自格函數 $\phi_{j,n}(x)$ 來作爲方程式的 test function , 並對方程式積分, 得到

$$\int_0^1 (\frac{d^2u}{dx^2} + f(x))\, \phi_{j,n}(x)\, dx = 0, \quad 0 \leq n \leq N-1, \tag{3.8}$$

其中 0 和 1 是區域 Ω 的兩端點. 我們對 u_h 的導數項作分部積分, 得到

$$\phi_{j,n} \frac{du_h}{dx}\Big|_0^1 - \int_0^1 \frac{du_h}{dx} \frac{d\phi_{j,n}}{dx}\, dx + \int_0^1 f(x)\, \phi_{j,n}\, dx = 0, \quad 0 \leq n \leq N-1. \tag{3.9}$$

利用 (3.2) u_h 的定義, 我們知道

$$\frac{du_h}{dx} = \sum_{m=0}^{N-1} \frac{d\phi_{j,m}}{dx}\, u_m = \sum_{m=0}^{N-1} u_m\, 2^{\frac{3j}{2}}\, \phi'(2^j x - m), \tag{3.10}$$

這裡要說明一件事, ϕ 的微分也滿足自格等式,

$$\phi_{j,m}^{(s)}(x) = 2^{\frac{j}{2}} \phi^{(s)}(2^j x - m) \quad j, m \in Z$$

這裡可以參考嚴健彰的論文 [6]. 把 (3.10) 式代入 (3.8) 式, 得到

$$2^{3j}\sum_{m=0}^{N-1}\int_{\Omega^{(e)}}\phi'_{j,m}\,\phi'_{j,n}\,dx\,u_m = 2^{2j}\sum_{m=0}^{N-1}\phi_{j,n}\,\phi'_{j,m}\,u_m|_0^1 + 2^{\frac{j}{2}}\int_{\Omega^{(e)}} f(x)\,\phi_{j,n}\,dx,$$

最後把左右兩邊同除 $2^{j/2}$, 得到

$$2^{\frac{5j}{2}}\sum_{m=0}^{N-1}\int_{\Omega^{(e)}}\phi'_{j,m}\phi'_{j,n}\,dx\,u_m = 2^{\frac{3j}{2}}\sum_{m=0}^{N-1}\phi_{j,n}\,\phi'_{j,m}\,u_m|_0^l + \int_{\Omega^{(e)}} f(x)\,\phi_{j,n}\,dx. \qquad (3.11)$$

方程式 (3.11) 右手邊第一項爲有限元的自然邊界條件, 我們看 $\phi_{j,n}\,\phi'_{j,m}$ 這一項, 因爲週期性的關係(週期性函數可視爲在區間內的函數跨過邊界後在區間另一 邊補上跨過邊界後的函數), 所以跨過 0 和 1 這兩端點的基底函數爲同一個, 這一項在方程式中是互相抵消掉了.

最後我們得到

$$2^{\frac{5j}{2}}\sum_{m=0}^{N-1}\{\int_{\Omega^{(e)}}\phi'_{j,m}\,\phi'_{j,n}\}\,dx\,u_m = \int_{\Omega^{(e)}} f(x)\,\phi_{j,n}\,dx, \quad 0\le n\le N-1. \qquad (3.12)$$

基本上, 如果我們的用的基底函數是 N 項的話, 剛度矩陣 K 表爲

$$k_{m+1,n+1} = \int_{\Omega^{(e)}}\phi'_{j,m}\,\phi'_{j,n}\,dx, \quad 0\le m,n\le N-1.$$

負載向量 F 表爲

$$f_n = \int_{\Omega^{(e)}} f(x)\,\phi_{j,n}\,dx, \quad 0\le n\le N-1$$

$\Omega(e)$ 代表在某一個元素的範圍.

3-5 矩陣元素計算

然而 $\phi(x)$ 在第 j 層的寬度爲

$$\operatorname{supp}\phi(2^j-m) = [\frac{m}{2^j},\frac{m+5}{2^j}],$$

我們必須算出在 Ω 裡的所有基底個數, 因爲第 j 層的基底寬度 $[\frac{m}{2^j},\frac{m+5}{2^j}]$ 覆蓋整個 $\Omega=(0,1)$, 比較兩邊得到

$$(0,1) = [\frac{m}{2^j},\frac{m+5}{2^j}], \quad \text{使得 } 0=\frac{m}{2^j} \text{ 與 } \ell=\frac{m+5}{2^j}.$$

因爲 $\frac{m}{2^j}=0$ 所以 $m=0$, 又 $\frac{m+5}{2^j}=1$ 所以 $m=2^j-5$, 又因三階週期性自格函數有四個跨過邊界的基底函數, 所以 m 的總數爲 2^j (因爲 $m=2^j-5+4+1$). 所以剛度矩陣 K 與 F 表爲

$$K_{mn}=\int_{\frac{m}{2^j}}^{\frac{m+5}{2^j}} \phi'_{j,m}\phi'_{j,n}dx,$$
$$F_n=\int_{\frac{m}{2^j}}^{\frac{m+5}{2^j}} f(x)\phi_{j,n}dx,$$

m,n 皆小於等於 2^j. 對自格函數積分時我們須作變數變換把第 j 層轉換道第 0 層來積分 (因爲不同層的積分值只差了一個常數項), 我們得到

$$\int_{\frac{m}{2^j}}^{\frac{m+5}{2^j}} \phi'(2^jx-m)\phi'(2^jx-n)dx=\frac{1}{2^j}\int_0^5 \phi'(y)\phi'(y+k)dy,$$
$$\int_{\frac{m}{2^j}}^{\frac{m+5}{2^j}} f(x)\phi(2^jx-n)dx=\frac{1}{2^j}\int_0^5 f(\frac{y+m}{2^j})\phi(y+k)dy,$$

上式中 $k=m-n$ 我們代回 (3.8) 式中

$$2^{\frac{5j}{2}}\sum_{m=0}^{N-1}\{\int_0^5\phi'(y)\phi'(y+k)\,dy\}\,u_m=\int_0^5 f(y+k)\,\phi(y)\,dy,\quad 0\le n\le N-1.$$

我們把左式的係數 $2^{\frac{5j}{2}}$ 除到右邊, 得到

$$\sum_{m=0}^{N-1}\{\int_0^5 \phi'(y)\phi'(y+k)dy\}u_m=2^{\frac{-5j}{2}}\int_0^5 f(y+k)\,\phi(y)dy,\quad 0\le n\le N-1. \tag{3.13}$$

在這我們知道 $N=2^j$ 剛度矩陣 K 的元素是

$$k_{m+1,n+1}=\int_0^5\phi'(y)\phi'(y+k)dy,\quad 0\le m,n\le N-1. \tag{3.14}$$

負載向量 F 的元素爲

$$f_{n+1}=2^{\frac{-5j}{2}}\int_0^5 f(y+k)\phi(y)dy,\quad 0\le n\le N-1 \tag{3.15}$$

上式中 $k=m-n$.

由 (3.13) 所給的系統形成一個 $2^j \times 2^j$ 的對稱矩陣系統, 根據前一節矩陣 K 的非零元素表爲

$$
\begin{aligned}
k_{m+1,m+1} &= \int_0^1 \phi'(2^j x - m)\phi'(2^j x - m)dx \\
&= \int_{\frac{m}{2^j}}^{\frac{m+5}{2^j}} \phi'(2^j x - m)\phi'(2^j x - m)dx \\
&= \frac{1}{2^j}\int_0^5 \phi'(y)\phi'(y)dy
\end{aligned}
$$

對於每一個 $m = 0, 1, 2, \ldots, 2^j$-1.

$$
\begin{aligned}
k_{m+1,m+2} &= \int_0^1 \phi'(2^j x - m)\phi'(2^j x - (m+1))dx \\
&= \int_{\frac{m}{2^j}}^{\frac{m+5}{2^j}} \phi'(2^j x - m)\phi'(2^j x - (m+1))dx \\
&= \frac{1}{2^j}\int_0^5 \phi'(y)\phi'(y-1)dy \\
&= \frac{1}{2^j}\int_1^5 \phi'(y)\phi'(y-1)dy \\
&= k_{m+2,m+1}
\end{aligned}
$$

對於每一個 $m = 0, 1, \ldots, 2^j$-2. 同理我們可以得到 $k_{m+1,m+3} = k_{m+3,m+1}$ $(m = 0, 1, 2, \ldots, 2^j$-3), $k_{m+1,m+4} = k_{m+4,m+1}$ $(m = 0, 1, 2, \ldots, 2^j$-4), $k_{m+1,m+5} = k_{m+5,m+1}$ $(m = 0, 1, 2, \ldots, 2^j$-5).

以上這些元素的基底完全都在區域 $\Omega = (0, 1)$ 內, 並沒有跨過邊界的情形出現, 但是因爲週期性的關係, 所有在邊界上會有四個基底函數跨過邊界 (這是因爲 Daubcheies 三階自格函數, 所以寬度 (support) 是 5 $(2P - 1 = 2 \times 3 - 1)$, 這 4 個分別是起始點爲

2^j-4, 2^j-3, 2^j-2, 2^j-1 的基底函數, (給個圖例, 這四張圖是在第四層, ie V_4 層,

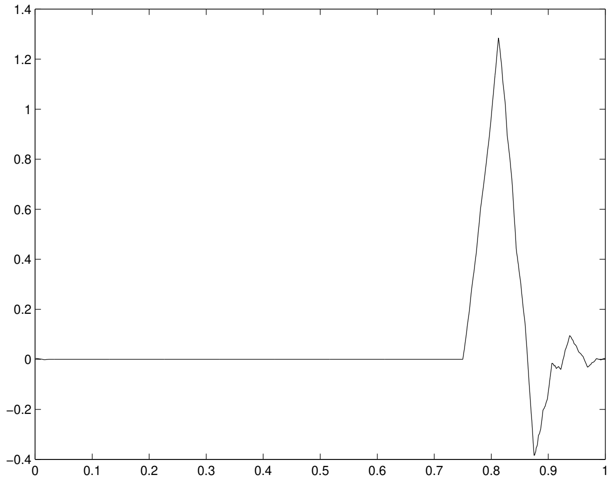

Fig 1.

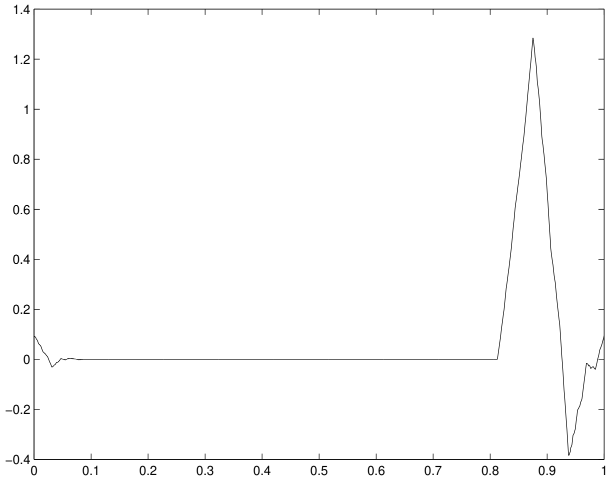

Fig 2.

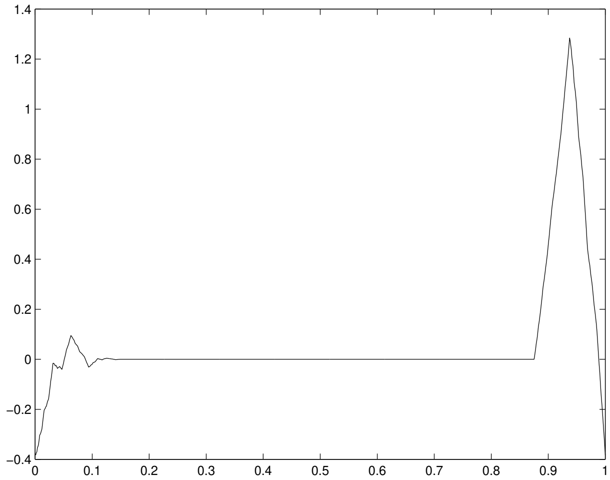

Fig 3.

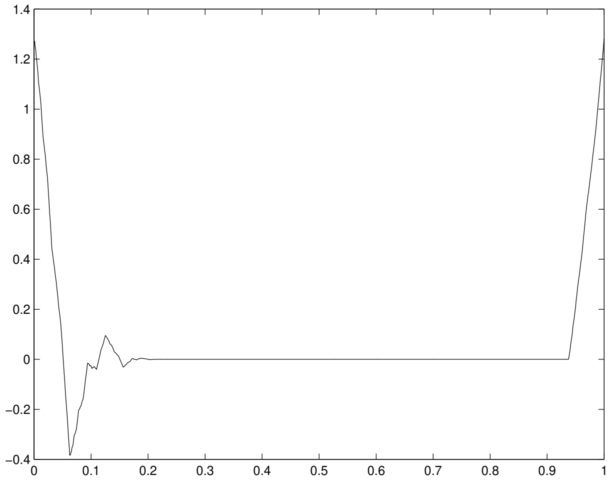

Fig 4.

我們給定 16 個節點, 這些圖看起來不夠平滑, 其實不然, 這是因爲我們在電腦計算時整個區域 Ω 我們才取 1024 點來表示, 而 Ω 又分割成 2^j 塊, 而每一個自格函數又用 5120 點來表示與計算, 但是實際在畫圖時,我們無法用那麼多點, 所以必須從 5120 點中選取適當的點數, 使得在每一層的自格函數基底把函數 u 逼近出來時, 能然保持 1024 點).

我們現在考慮矩陣 K 的元素包含跨過邊界的基底函數時, 矩陣元素 $k_{m+1,n+1}$ 的表示法, 因爲三階自格函數的寬度是 5, 所以 以點 $x = 2^j - 5$ 爲起始點的自格函數末端剛好觸碰到邊界上, 所以從點 $x = 2^j - 4$ 的自格函數開始平移出了 Ω 的區域了, 但是週期

性函數把平移出 Ω 的片段基底函數補在左邊的區域上, 參考圖 Fig1,..,Fig4 四張圖, 假如我們固定第一個自格函數 $\phi(2^j x)$, 另一個函數作平移的動作, 在邊界附近時, 雖然這個基底函數的主體偏移 $\phi(2^j x)$ 很遠了, 但是跨出 Ω 區域的片段函數補在區域的最左邊, 所以能然與原先的基底函數 $\phi(2^j x)$ 有交集, 所以導數積分仍然有值, 這四個基底函數的起始點分別是 $x_{2^j-4}, x_{2^j-3}, x_{2^j-2}, x_{2^j-1}$, 我們現在假設這些點值所代表的基底函數在跨過邊界時並沒有作剪貼(cut and paste)的工作 (ie.剪下跨出邊界的部分函數貼在區間的最左邊) 則這些函數的積分範圍我們假設爲 $[0, 1+\frac{1}{2^j}], [0, 1+\frac{2}{2^j}], [0, 1+\frac{3}{2^j}], [0, 1+\frac{4}{2^j}]$, 而這些積分範圍所造出的矩陣 K 元素, 通常只會出現在矩陣的右上角與左下角, 並且是一個 4×4 的上三角和下三角矩陣 (三角形的大小和所選用的基底函數寬度有關).

所以我們可以找出它們的值分別是

$$
\begin{aligned}
&k_{m+1,m+2^j-3}\\
=&\int_0^{1+\frac{1}{2^j}} \phi'(2^j x)\,\phi'(2^j x-(2^j-4))\,dx\\
=&\int_0^{1} \phi'(2^j x)\,\phi'(2^j x-(2^j-4))\,dx+\int_1^{1+\frac{1}{2^j}} \phi'(2^j x-2^j)\,\phi'(2^j x-(2^j-4))\,dx\\
=&0+\int_1^{1+\frac{1}{2^j}} \phi'(2^j x-2^j)\,\phi'(2^j x-(2^j-4))\,dx\\
=&\frac{1}{2^j}\int_0^1 \phi'(y)\,\phi'(y+4)\,dy\\
=&k_{m+2^j-3,m+1}
\end{aligned}
$$

對於每一個 $m=0,1,2,3.$

$$
\begin{aligned}
&k_{m+1,m+2^j-2}\\
=&\int_0^{1+\frac{2}{2^j}} \phi'(2^j x)\,\phi'(2^j x-(2^j-3))\,dx\\
=&\int_0^{1} \phi'(2^j x)\phi'(2^j x-(2^j-3))\,dx+\int_1^{1+\frac{2}{2^j}} \phi'(2^j x-2^j)\,\phi'(2^j x-(2^j-3))\,dx\\
=&0+\int_1^{1+\frac{2}{2^j}} \phi'(2^j x-2^j)\,\phi'(2^j x-(2^j-3))\,dx\\
=&\frac{1}{2^j}\int_0^2 \phi'(y)\,\phi'(y+3)\,dy\\
=&k_{m+2^j-2,m+1}
\end{aligned}
$$

對於每一個 $m=0,1,2.$

$$
\begin{aligned}
& k_{m+1,m+2^j-1} \\
=& \int_0^{1+\frac{3}{2^j}} \phi'(2^j x)\,\phi'(2^j x-(2^j-2))\,dx \\
=& \int_0^{1} \phi'(2^j x)\,\phi'(2^j x-(2^j-2))\,dx + \int_1^{1+\frac{3}{2^j}} \phi'(2^j x-2^j)\,\phi'(2^j x-(2^j-2))\,dx \\
=& 0+\int_1^{1+\frac{3}{2^j}} \phi'(2^j x-2^j)\,\phi'(2^j x-(2^j-2))\,dx \\
=& \frac{1}{2^j}\int_0^3 \phi'(y)\,\phi'(y+2)\,dy \\
=& k_{m+2^j-1,m+1}
\end{aligned}
$$

對於每一個 $m=0,1,2.$

$$
\begin{aligned}
& k_{m+1,m+2^j} \\
=& \int_0^{1+\frac{4}{2^j}} \phi'(2^j x)\,\phi'(2^j x-(2^j-1))\,dx \\
=& \int_0^{1} \phi'(2^j x-2^j)\,\phi'(2^j x-(2^j-1))\,dx + \int_1^{1+\frac{4}{2^j}} \phi'(2^j x-2^j)\,\phi'(2^j x-(2^j-1))\,dx \\
=& 0+\int_1^{1+\frac{4}{2^j}} \phi'(2^j x-2^j)\,\phi'(2^j x-(2^j-1))\,dx \\
=& \frac{1}{2^j}\int_0^4 \phi'(y)\,\phi'(y+1)\,dy \\
=& k_{m+2^j-3,m+1}
\end{aligned}
$$

對於每一個 $m=0.$ 雖然看似有 9 種積分值要求, 但實際上僅有五種形式, 另外四種是重複的, 分別是

$$
\begin{aligned}
\int_1^5 \phi'(y)\,\phi'(y-1)\,dy &= \int_0^4 \phi'(y)\,\phi'(y+1)\,dy \\
\int_2^5 \phi'(y)\,\phi'(y-2)\,dy &= \int_0^3 \phi'(y)\,\phi'(y+2)\,dy \\
\int_3^5 \phi'(y)\,\phi'(y-3)\,dy &= \int_0^2 \phi'(y)\,\phi'(y+3)\,dy \\
\int_4^5 \phi'(y)\phi'(y-4)dy &= \int_0^1 \phi'(y)\phi'(y+4)dy
\end{aligned}
$$

我們令 $k_{m+1,m+1}=a$, $k_{m+1,m+2}=b$, $k_{m+1,m+3}=c$, $k_{m+1,m+4}=d$, $k_{m+1,m+5}=e$, 對於

所有的 $m=0,1,2,\ldots,2^j$-1. 則我們得到了矩陣

$$K=\begin{pmatrix}
a & b & c & d & e & 0 & 0 & 0 & 0 & \cdots & e & d & c & b\\
b & a & b & c & d & e & 0 & 0 & 0 & \cdots & 0 & e & d & c\\
c & b & a & b & c & d & e & 0 & 0 & \cdots & 0 & 0 & e & d\\
d & c & b & a & b & c & d & e & 0 & \cdots & 0 & 0 & 0 & e\\
e & d & c & b & a & b & c & d & e & \cdots & 0 & 0 & 0 & 0\\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots\\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots\\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots\\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots\\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots\\
e & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & a & b & c & d\\
d & e & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & b & a & b & c\\
c & d & e & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & c & b & a & b\\
b & c & d & e & 0 & 0 & 0 & 0 & 0 & \cdots & d & c & b & a
\end{pmatrix}_{2^j\times 2^j} \tag{3.16}$$

所以我們成功的把以自格函數爲基底的 Poisson 方程式轉爲解一 $Ku=F$ 的矩陣問題, 但是週期化的結果造成一個嚴重的問題, 就是矩陣 K 的每一列的和幾乎爲零 (我們在 2-2 節中討論過), K 是一線性相依的矩陣, 也就是說矩陣 K 是一個 singular 的矩陣, 我們無法正確求得 K^{-1}, 也無法求得 u 的唯一解, 所以我們必須再給定一個唯一解條件, 就是

$$\int_0^1 u_h dx = 1.$$

根據 (3.2) 式我們得到

$$\begin{aligned}
\int_0^1 u_h\,dx &= 2^{\frac{j}{2}} \sum_{m=0}^{N-1} \int_{\frac{m}{2^j}}^{\frac{m+5}{2^j}} \phi(2^j x - m)\,dx\,u_m\\
&= \sum_{m=0}^{N-1} u_m\, 2^{\frac{-j}{2}} \int_0^5 \phi(y)\,dy\\
&= \sum_{m=0}^{N-1} u_m\, 2^{\frac{-j}{2}}\\
&= 1
\end{aligned}$$

所以

$$\sum_{m=0}^{N-1} u_m = 2^{\frac{j}{2}}, \tag{3.17}$$

我們用這一式來取代矩陣 K 的最後一列就可解得唯一解.

做到這裡我們必須強調一點, 就是在考慮積分時所用的方法和 $\phi(x)$ 所取的點數. 在我們第一次做實驗時, 我們對於 $\phi(x)$ 用了 1024 點, 積分法是 Simpson 法, 所得到的積分值用來作爲矩陣的元, 解出來的有限元解誤差極大, 根本沒有規則性也無法達到我們預期的效果, 所以我們在取 $\phi(x)$ 的原始資料時我們取了最大的限度 51202 點, 發現積分出來的值與原來用 1024 點的積分值差異滿大的, 並且再後來的實驗中果然找到了比較有規則性和達到了預先猜想的結果. 用了這麼多的點數當然還是有一些問題的存在, 正如同前面探討邊界問題時所說的, 比如說在畫圖時的解析度方面, 電腦的解析度最大只到 1024 點, 所以我們再把 51202 點畫到 1024 的解析度時, 我們必須做一些捨去, 在這用的方法是隨著我們提高 V_j 的階數 j 時, 我們從 5120 點挑出 5120 除以所切割的網格數目的點數, 來當成 $\phi(2^j x)$ 在第 j 層的點數, 不過這造成另一個問題, 就是在網格切得越來越細的時候, $\phi(2^j x)$ 從原始資料中所選取的點數會越來越少, 到最後細到 $\phi(x)$ 會類似成爲 delta function 的可能, 而最重要的是在逼近能力方面我們無法觀察出 $\phi(x)$ 應該表現的能力. 這裡是我們極需要改進的地方.

我們參考徐敏堂先生的數值積分法 [7], 論文中並不去解一個實際的數值積分計算, 而是把導數的積分轉變爲解矩陣特徵值問題, 得到更爲精確的數值解.

3-6 Wavelet-Galerkin 法

這一節我們使用凌波函數在影像處理方面的方法, 來處理剛度矩陣 K, 我們造出一凌波轉換矩陣 $\mathcal{W}$ 把以 $\phi_{j,k}(x)$ 爲基底的剛度矩陣 K_J 轉換到以 $\phi_{j-1,k}(x)$, $\psi_{j-1,k}(x)$ 爲基底的矩陣 $\mathcal{A}_J$, 因爲

$$\phi(x) = 2^{1/2} \sum_{m=0}^{5} c_m \, \phi(2x - m)$$
$$\psi(x) = 2^{1/2} \sum_{m=0}^{5} d_m \, \phi(2x - m)$$

所以矩陣 K_{J-1} 的元素爲 (在這裡微分後的係數我們不加在式子中, 因爲在 (3.12) 之前

我們已經把爲微後的係數都處理過了)

$$
\begin{aligned}
&\langle\phi' 2^{j-1}x-k), \phi'(2^{j-1}x-\ell)\rangle \\
=&\langle 2^{1/2}\sum_{m=0}^{5} c_m\phi'(2^jx-2k-m), 2^{1/2}\sum_{n=0}^{5} c_n\phi'(2^jx-2\ell-n)\rangle \\
=&2\sum_{m=0}^{5}\sum_{n=0}^{5} c_mc_n\langle\phi'(2^jx-2k-m), \phi'(2^jx-2k-n)\rangle \\
=&2\sum_{s=2k}^{2k+5}\sum_{t=2\ell}^{2\ell+5} c_{s-2k}c_{t-2\ell}\langle\phi'(2^jx-s), \phi'(2^jx-t)\rangle
\end{aligned}
$$

對於所有的 k, ℓ=$0, 1, \ldots, 2^{j-1}$-1,

$$
\begin{aligned}
&\langle\psi' 2^{j-1}x-k), \psi'(2^{j-1}x-\ell)\rangle \\
=&\langle 2^{1/2}\sum_{m=0}^{5} d_m\phi'(2^jx-2k-m), 2^{1/2}\sum_{n=0}^{5} d_n\phi'(2^jx-2\ell-n)\rangle \\
=&2\sum_{m=0}^{5}\sum_{n=0}^{5} d_md_n\langle\phi'(2^jx-2k-m), \phi'(2^jx-2k-n)\rangle \\
=&2\sum_{s=2k}^{2k+5}\sum_{t=2\ell}^{2\ell+5} d_{s-2k}d_{t-2\ell}\langle\phi'(2^jx-s), \phi'(2^jx-t)\rangle \\
=&2(-1)^{s+t-2(k+\ell)}\sum_{s=2k}^{2k+5}\sum_{t=2\ell}^{2\ell+5} c_{2k+5-s}c_{2\ell+5-t}\langle\phi'(2^jx-s), \phi'(2^jx-t)\rangle
\end{aligned}
$$

因爲 $d_k=(-1)^kc_{5-k}$, 並且對於所有的 k, ℓ=$0, 1, \ldots, 2^{j-1}$-1,

$$
\begin{aligned}
&\langle\phi' 2^{j-1}x-k), \psi'(2^{j-1}x-\ell)\rangle \\
=&\langle 2^{1/2}\sum_{m=0}^{5} c_m\phi'(2^jx-2k-m), 2^{1/2}\sum_{n=0}^{5} d_n\phi'(2^jx-2\ell-n)\rangle \\
=&2\sum_{m=0}^{5}\sum_{n=0}^{5} c_md_n\langle\phi'(2^jx-2k-m), \phi'(2^jx-2k-n)\rangle \\
=&2\sum_{s=2k}^{2k+5}\sum_{t=2\ell}^{2\ell+5} c_{s-2k}d_{t-2\ell}\langle\phi'(2^jx-s), \phi'(2^jx-t)\rangle \\
=&2(-1)^{t-2\ell}\sum_{s=2k}^{2k+5}\sum_{t=2\ell}^{2\ell+5} c_{s-2k}c_{2\ell+5-t}\langle\phi'(2^jx-s), \phi'(2^jx-t)\rangle
\end{aligned}
$$

因爲 $d_k=(-1)^kc_{5-k}$, 並且對於所有的 k,ℓ=$0, 1, \ldots, 2^{j-1}$-1.

我們把上面三個式子看成二次式 (bilinear form), 並且造出分解 (decomposition) 矩陣 $\mathcal{W}_J$

$$
\begin{pmatrix}
c_0 & c_1 & c_2 & c_3 & c_4 & c_5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & c_0 & c_1 & c_2 & c_3 & c_4 & c_5 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & c_0 & c_1 & c_2 & c_3 & c_4 & c_5 & 0 & 0 & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
c_4 & c_5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & c_0 & c_1 & c_2 & c_3 \\
c_2 & c_3 & c_4 & c_5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & c_0 & c_1 \\
d_0 & d_1 & d_2 & d_3 & d_4 & d_5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & d_0 & d_1 & d_2 & d_3 & d_4 & d_5 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & d_0 & d_1 & d_2 & d_3 & d_4 & d_5 & 0 & 0 & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
d_4 & d_5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & d_0 & d_1 & d_2 & d_3 \\
d_2 & d_3 & d_4 & d_5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & d_0 & d_1
\end{pmatrix}_{2^j \times 2^j}
\tag{3.18}
$$

並且

$$\frac{1}{2}\mathcal{W}_J\mathcal{W}_J^T = I, \tag{3.19}$$

K_J 與 $\mathcal{A}_J$ 的關係爲

$$\mathcal{A}_J = \mathcal{W}_J K_J \mathcal{W}_J^T.$$

我們把解 $K_J u = F$ 的問題轉換爲

$$\mathcal{A}_J \tilde{u} = \tilde{F}, \tag{3.20}$$

的問題, 則 $\tilde{u} = \mathcal{W}_J u$ 和 $\tilde{F} = \mathcal{W}_J F$, 我們可以利用 (3.20) 式解出 $\tilde{u}$ 並且因爲 (3.19), 所以我們可以很快的得到眞解 u, 隨著我們乘上 $\mathcal{W}_J$ 的次數, 我們可使 K_J 降層的次數增加, 解出降層後的 $\tilde{u}$, 然後在反算求出眞解 u. 在影像處理的技巧裡, 我們常常再降層後處理掉接近零的矩陣元, 以方便在影像壓縮過程中得到較高的壓縮率, 但是我們在解微分方程的過程中, 我們並不能用此方法, 因爲我們需要保留所有的矩陣元素以造回眞實解, 但是在影像方面極小的誤差並不會被我們的眼睛所察覺, 所以並不會影響到視覺上的效果, 但是在解得微分方程中逼近解的過程中卻往往造成大的誤差.

3-7 預條件數處理

這一節我們引進 PCG (preconditioned conjugate gradient) 方法來處理轉換到凌波基底 $\psi(x)$ 爲主的矩陣 $\mathcal{A}_J$, 我們建構一個預條件算子 (preconditioner) 的對角線型式的

矩陣 D_A, 使得預條件處理過後的矩陣之條件數是常數有界的 (也就是與基底函的個數無關). 根據上一節的凌波轉換, 因爲它屬於正交性轉換, 所以 $\mathcal{W}_J$ 並沒有改變矩陣 K_J 的條件數,

$$\kappa(\mathcal{A}_J) = \kappa(K_J) = O(N^2).$$

首先我們假設 $\phi \in V_J$, 則

$$\phi = \sum_{j,k} w_{j,k}\, \psi_{j,k},$$

然後我們得到

$$\mathcal{K}_J = \langle \mathcal{A}(\phi), \phi \rangle = \sum\nolimits_{j,k} \sum\nolimits_{s,t} w_{j,k}\, w_{s,t}\, \langle \mathcal{A}(\psi_{j,k}), \psi_{s,t} \rangle = w^T \mathcal{A}_J\, w \tag{3.21}$$

$w = (w_{j,k})$ 是一向量.

根據有限元中的強制二次逼近 (coercive bilinear form approach) 我們假設 $a(u, v) = \langle \mathcal{A} u, v \rangle$ 是一個在 $H^s(R) \times H^s(R)$ 空間中的橢圓二次式, 則我們得到

$$C_3 \|\phi\|^2_{H^s} \leq \langle \mathcal{A}(\phi), \phi \rangle \leq C_4 \|\phi\|^2_{H^s} \tag{3.22}$$

$C_4 \geq C_3 \geq 0$ 均爲常數, 根據 (3.21), (3.22) 我們得到

$$C_3 \|\phi\|^2_{H^s} \leq w^T \mathcal{A}_J w \leq C_4 \|\phi\|^2_{H^s}$$

代入 (3.7) 式 我們得到

$$C_5 \sum\nolimits_{j,k} \left|2^{js} w_{j,k}\right|^2 \leq w^T \mathcal{A}_J w \leq C_6 \left|2^{js} w_{j,k}\right|^2$$

$C_6 \geq C_5 \geq 0$ 均爲常數,

我們定義一個對角線矩陣 $\mathcal{D}_{\mathcal{A}_J}$ 爲

$$(\mathcal{D}_{\mathcal{A}_J})_{(j,k)(j',k')} = 2^{-j/2} \delta_{j,k} \delta_{j',k'}, \tag{3.23}$$

經過轉換後我們得到對於任一向量 w

$$C_5 \|w\|^2 \leq w^T \mathcal{D}^{-1}_{\mathcal{A}_J} \mathcal{A}_J \mathcal{D}^{-1}_{\mathcal{A}_J} w \leq C_6 \|w\|^2 \tag{3.24}$$

$\|.\|$ 是一 ℓ^2-$norm$. 則我們可以得到條件數是常數收斂 (ie. $O(c)$, c 爲常數)

$$\kappa(\mathcal{D}_{\mathcal{A}_J}^{-1}\mathcal{A}_J\mathcal{D}_{\mathcal{A}_J}^{-1}) = O(1).$$

根據上一節, $\mathcal{A}_J$ 與 K_J 的關係爲

$$\mathcal{A}_J = \mathcal{W}_J K_J \mathcal{W}_J^T,$$

$\mathcal{W}_J$ 是 $\phi(x)$ 基底轉換到 $\psi(x)$ 基底的轉換矩陣, 根據 (3.20)

$$\mathcal{A}_J\tilde{u} = \tilde{F},$$

則 $\tilde{u} = \mathcal{W}_J u$ 和 $\tilde{F} = \mathcal{W}_J F$, 爲求解 $\mathcal{A}_J\tilde{u} = \tilde{F}$ 我們引進 PCG 法的 daigonal preconditioner $\mathcal{D}_{\mathcal{A}_J}^{-1}$. 取代上式 $\mathcal{A}_J\tilde{u} = \tilde{F}$ 我們得到新的預條件處理系統

$$\mathcal{D}_{\mathcal{A}_J}^{-1}\mathcal{A}_J\tilde{u} = \mathcal{D}_{\mathcal{A}_J}^{-1}\tilde{F}.$$

根據前面所述, 本系統的條件數爲

$$\kappa(\mathcal{D}_{\mathcal{A}_J}^{-1}\mathcal{A}_J\mathcal{D}_{\mathcal{A}_J}^{-1}) = O(1),$$

當我們用 PCG 法來求解預條件系統時, 我們得到系統的收斂速度將會是線性的.

第四章
數値實驗

4-1 Scaling-Galerkin 法誤差與條件數估計

爲了比較與傳統有限元的差異, 我們計算與第二章相同的問題, 同樣的我們也是給定週期性函數, 在有限區間邊界上分別爲平滑與不平滑的解函數. 此處以自格函數爲基底的有限元素法來逼近眞實解, 分別計算誤差與條件數的大小, 並且觀察因爲給定 $\int u(x)dx = 1$ 的條件所發生的一些現象.

Example 1: (解函數邊界不平滑, $u \notin H^2_{peri}(R)$)

$$-u''(x) = 1, \qquad x \in (0,1),$$
$$u(0) = u(1) = 0.$$

根據 (3.2) 我們給定

$$u_h(x) = \sum_{m=0}^{N-1} u_i \, \phi_{j,m}$$

爲有限元解. 我們代入 (3.12) 並且得到矩陣 K, 也就是 (3.14) 的矩陣, 但是我們必須用 (3.15) 來取代矩陣 K 的最後一列, 這樣矩陣就可解得唯一解. 再來我們代入 (3.13) 式, 因爲 $f(x)$ 函數是 1, 所以 F 的行向量爲

$$f_{n+1} = 2^{\frac{-5j}{2}}, \quad 0 \le n \le N-1.$$

但是矩陣 K 更改了最後一列, 所以 $f_N = 2^{j/2}$. 根據上述定義所做的實驗, 所得結果如下

N	16	32	64	128	256	512
$\kappa(K)$	237.7	958	3839.6	15366	61471.5	245893.7
$\kappa(K)/N^2$	0.9283	0.9355	0.9374	0.9378	0.9379	0.938

接下來我們看看眞解與逼近解之間的誤差分析

N	16	32	64	128	256	512
$\|u_h - u\|$	2.5918E-2	9.2068E-3	3.2657E-3	1.1569E-3	4.0967E-4	1.4507E-4
$N^2\|u_h - u\|$	6.6351	9.4277	13.3765	18.956	26.8484	38.0311

由上表觀察得知解的誤差並不爲 $O(h^2)$, 而是 $O(h^{1.5})$ (ie. 1.5 階收斂). 我們把眞解與逼近解的圖畫出來做個比較, 左圖 (5,7,9,11) 是第 4 至第 7 層求出來的 u_m 値與眞解 $u(x)$ 的比較, 右圖 (6,8,10,12) 也是第 4 至第 7 層 u_m 乘上自格函數後所造出的 u_h 解與眞實解 $u(x)$ 的比較. 觀察右圖我們發現

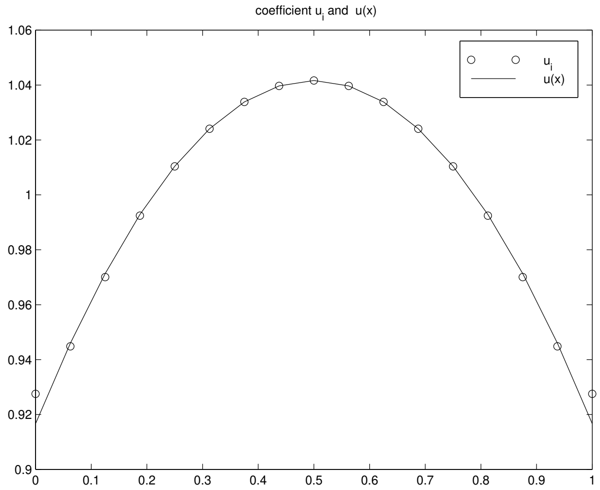

Fig 5.

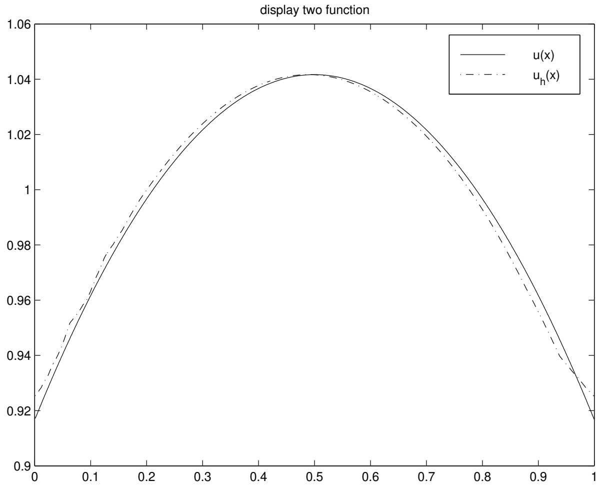

Fig 6.

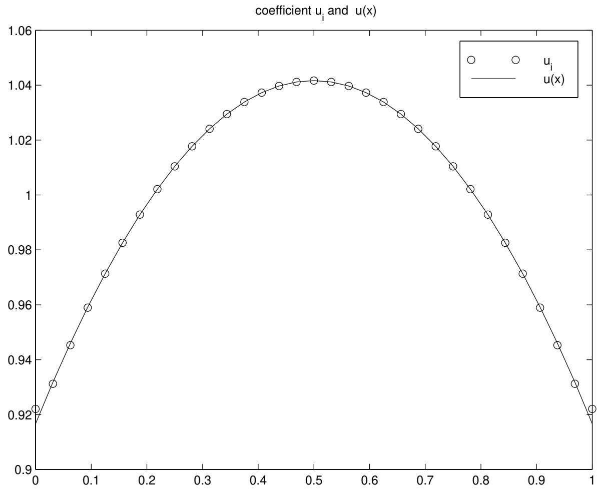

Fig 7.

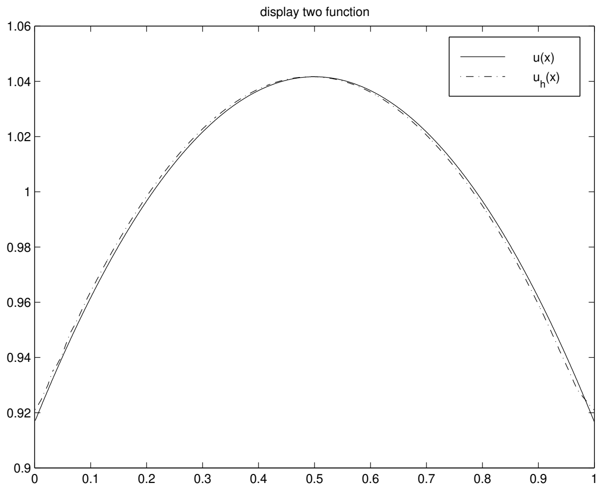

Fig 8.

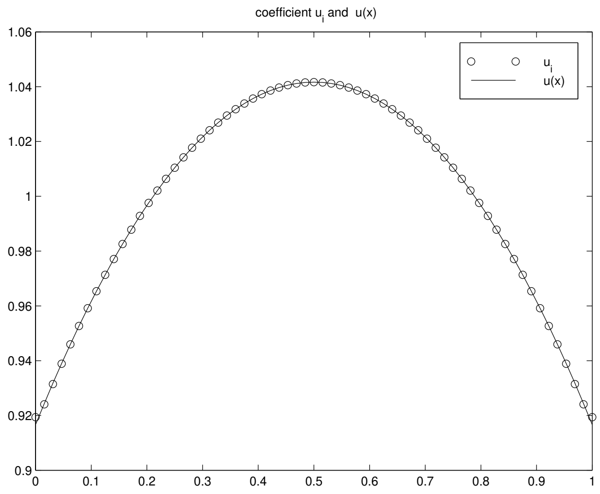

Fig 9.

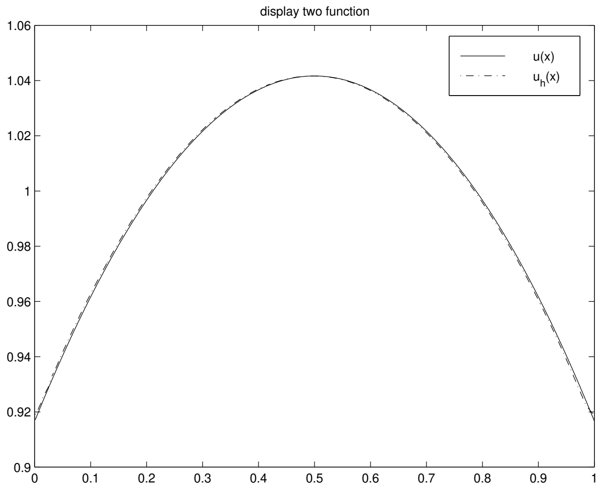

Fig 10.

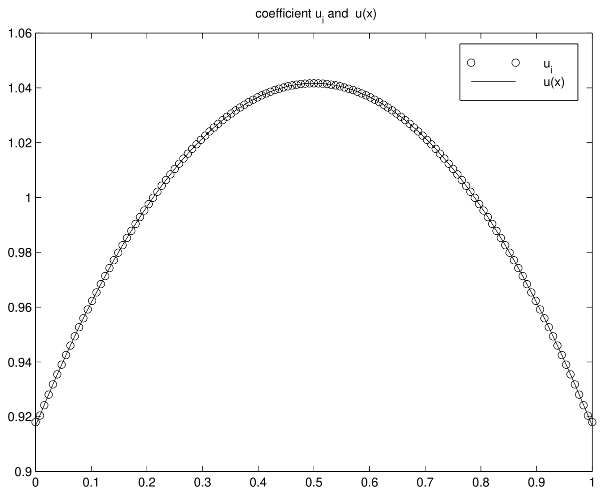

Fig 11.

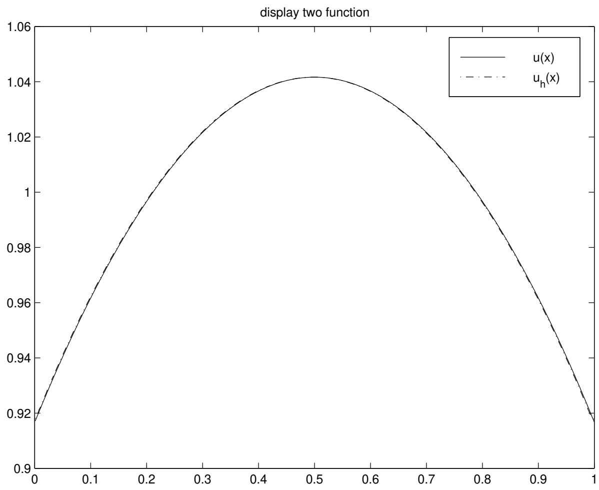

Fig 12.

逼近解係數 u_m 似乎是剛好算在 $u(x)$ 線上, 但是乘上自格函數 $\phi(x)$ 後在第 4 與第 5 層的逼近解有類似平移的現象發生.

我們先來看看 u_m 的值似乎在節點上算得非常準, 這是超收斂的現象. 接下來我們算看看是否用自格函數爲基底的有限元方法與傳統有限元在節點上的逼近能力相當 (傳統有限元解在處理這型問題時, 於節點上具有插值能力, 也就是在處理這型問題時, 於節點上的值就是眞實解的值). 我們同樣以

$$\|u_h - u\|_s^2 := \sum_{i=0}^{N-1} |u_m - u(x_i)|^2$$

來估計 u_h 在節點上的收斂速度. 實驗結果如下

N	16	32	64	128	256	512
$\|u_h - u\|_s$	3.8862E-3	1.3739E-3	4.8577E-4	1.7174E-4	6.0722E-5	2.1468E-5

可是我們從 Fig5, Fig7, Fig9, Fig11, 四張圖來看, 發現在整個區域內的節點上都解得滿準的, 但是在邊界上的 u_m 出現跳躍的情形 . 因此我們想把邊界附近的點去掉, 再看看內部節點的收斂情形. 我們現在去掉左右邊界各三個節點值, 估算誤差得到

N	16	32	64	128	256	512
$\|u_h - u\|_s$	4.8372E-4	1.7102E-4	6.0465E-5	2.1377E-5	7.5581E-6	2.6722E-6

如果去掉左右各四個節點的話, 得到

N	16	32	64	128	256	512
$\|u_h - u\|_s$	1.36E-4	4.8084E-5	1.7E-5	6.0105E-6	2.125E-6	7.5135E-7

我們可以得知離邊界越遠的地方解得越準確. 但是在圖上我們看到有類似超收斂的現象, 這可能是因爲電腦的解析度關係, 所造成視覺上的誤差.

我們觀察係數乘上自格函數後所造出的 u_h 似乎有向左偏移的現象, 在上面的圖我們實際只畫到第 7 層的逼近圖, 第 8, 9 層因爲在圖形上看來幾乎是重疊的, 我們無法觀察誤差現象, 我們只能實際計算誤差才能知道逼近的能力. 我們對於平移的現象做一點小小的實驗. 假設我們解出來的 u_m 差了一個相位 (也就是往左或往右平移一個單位), 我們把整個係數 u_m 的順序左移或是右移一位, 乘上自格函數基底觀察是否會重合. 給定第 5, 6 兩層來觀察 發現代表第 5 層的圖 13 (左移), 圖 15 (右移), 與代表第 6 層的 14 (左移), 圖 16 (右移),

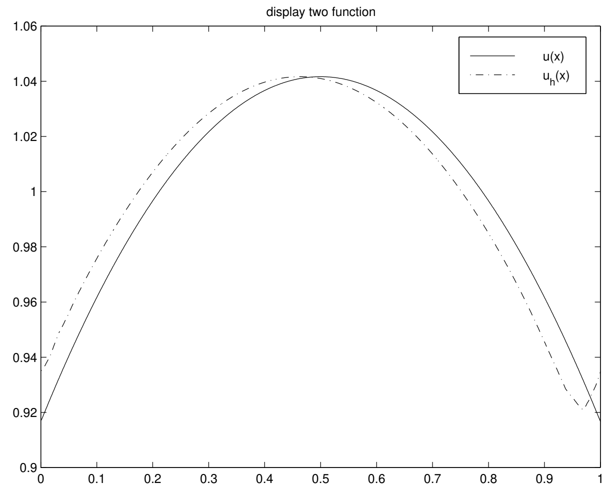

Fig 13.

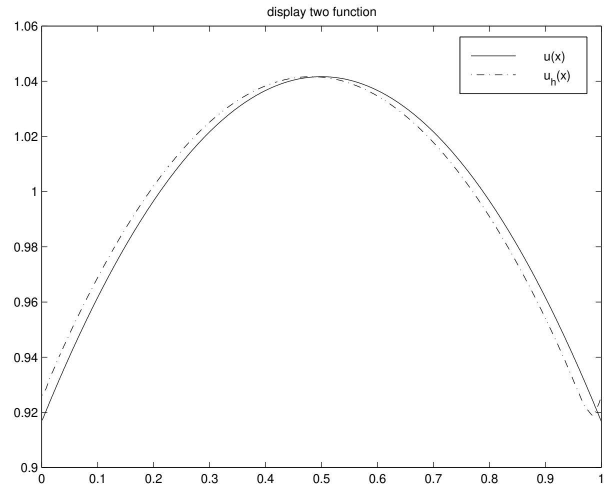

Fig 14.

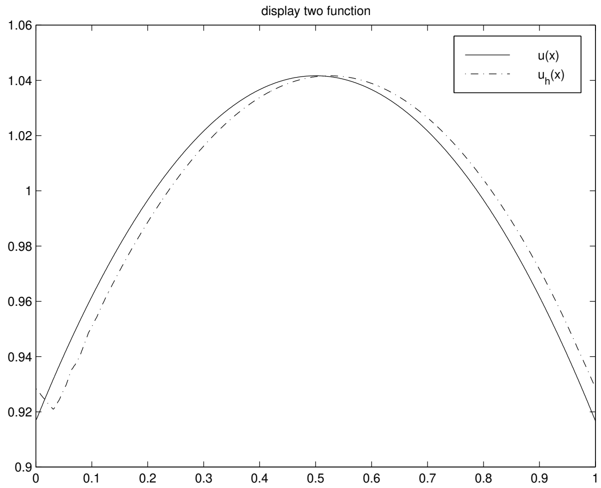

Fig 15.

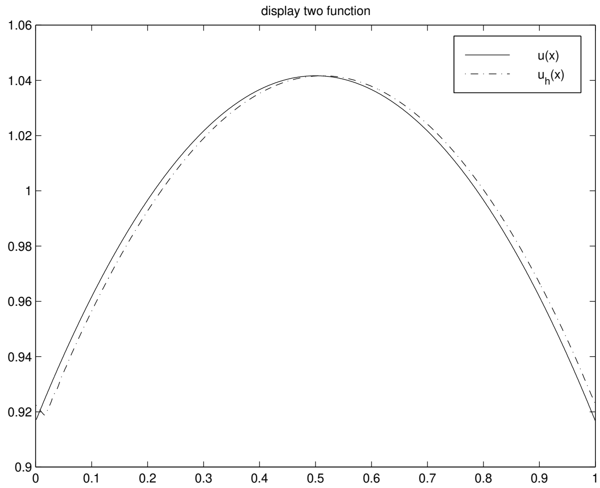

Fig 16.

我們看到並不是因爲係數差了一個相位的關係, 而是我們更改係數的位置時, 平移會同時發生, 但是並不是與眞實解重合. 在傳統有限元方面並未發生這種現象. 所以我們猜想是否是因爲基底的關係, 因爲 Daubechies 自格函數並非一對稱函數 (Fig 17 是 Daubechies 三階自格函數, 圖中橫座標爲描繪圖形的點數, 實際函數寬度是 5).

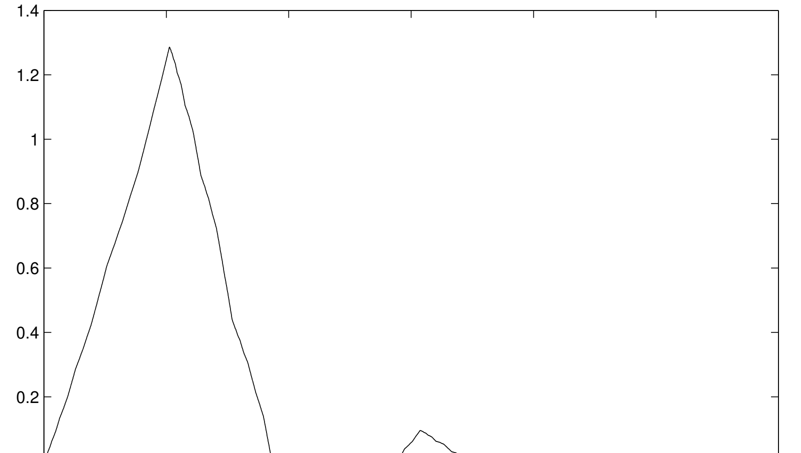

Fig 17.

我們再做一個小小的實驗來驗證我們的假設. 假設一個在區間 $[0,1]$ 的函數 $u = x(1-x)$. 我們把區間 $[0,1]$ 切割成 16 和 32 個節點, 把這些節點代入 $u(x)$ 所製造出來的 $u(x_i)$ $(i = 1, 2, \ldots, 16$ 或 $32)$ 就是函數中的內插值. 把這節點所對應的 $u(x)$ 值乘上第四層和

第五層的自格函數基底, 所得出的結果為

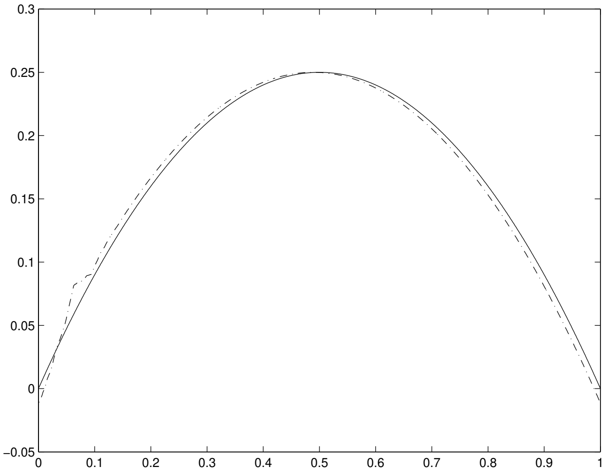

Fig 18.

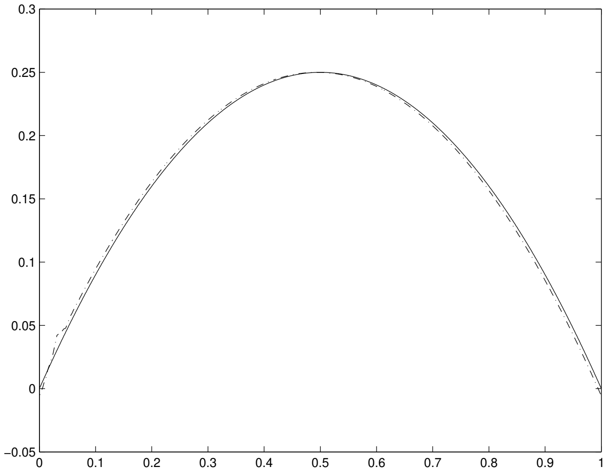

Fig 19.

上面兩張圖驗證了我們的假設, 確定逼近解並沒發生平移, 而是因為自格函數本身的不對稱性所造成的.

在 Dirichlet 條件下的有限元誤差分析中, 對於一階 shape 函數誤差是 $O(h^2)$. 但是我們看到以自格函數為基底的有限元素法並沒有達到這個條件, 而且逼近能力並沒有一階 shape 函數來得準. 在上述的討論之中我們發現誤差 $u(x)$ 可能來自於週期解的不平滑, 如果用 Taylor 展開式來逼近 $u(x)$ 的話, 因為不夠平滑, 所以一階導數不存在, 得知誤差項從一階就開始了. 我們給另外一個例子, 也就是平滑的週期函數, 驗證看看我們假設是否正確.

Example 2.(解函數邊界平滑, $u \in H^2_{peri}(R)$)

$$-u''(x) = 2\pi^2 \cos(4\pi x), \quad x \in (0,1),$$

這樣的 $f(x)$ 在邊界上是平滑的, 所以其導數不會在邊界上造成不連續點. 方程式所造出的矩陣仍然為 (3.13) 式的 K, 並且最後一列都改為 1, F 變為

$$\begin{aligned} f_{m+1} &= \int_{m/2^j}^{m+5/2^j} \phi(2^j x - m) 2\pi^2 \cos(4\pi x) dx \\ &= \frac{2\pi^2}{2^j} \int_0^5 \phi(y) cos(4\pi \frac{y+m}{2^j}) dy \end{aligned}$$

上式中 $0 \le m \le N-1$. 根據上述定義所做的實驗結果, 得到

N	16	32	64	128	256	512
$\kappa(K)$	237.66	958.02	3839.6	15366	61471	245893
$\kappa(K)/N^2$	0.9284	0.9356	0.9374	0.9379	0.938	0.938

可見其條件數仍然是 $O(N^2)$. 在誤差分析方面, 得到

N	16	32	64	128	256	512
$\|u_h - u\|$	4.3848E-3	3.938E-4	4.105E-5	8.2682E-6	3.42E-6	1.6444E-6
$N^2\|u_h - u\|$	1.1125	0.4033	0.1681	0.1354	0.2241	0.431

大約可見 L^2-$norm$ 的收斂速度是 $O(h^2)$.

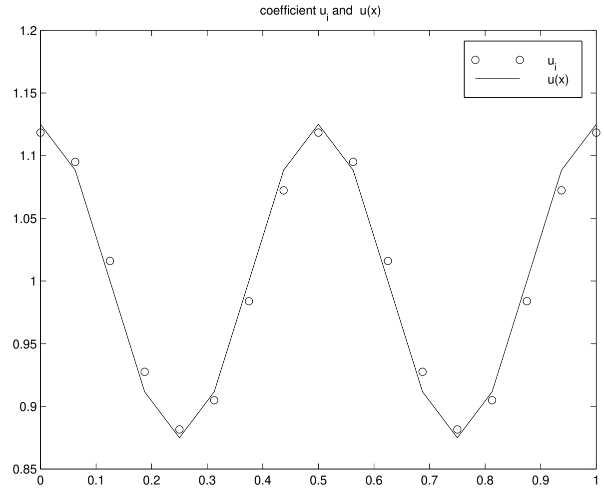

Fig 20.

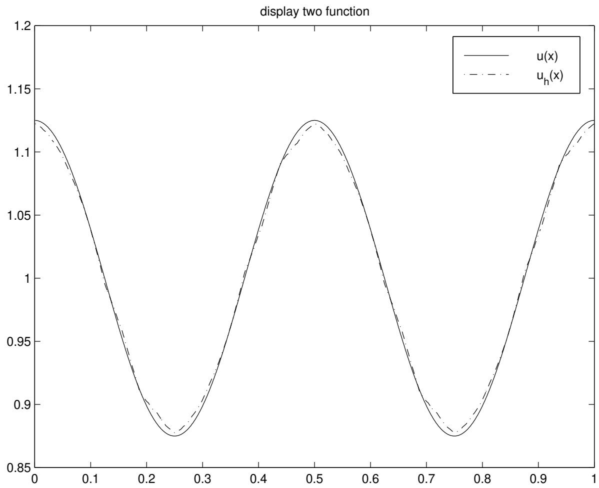

Fig 21.

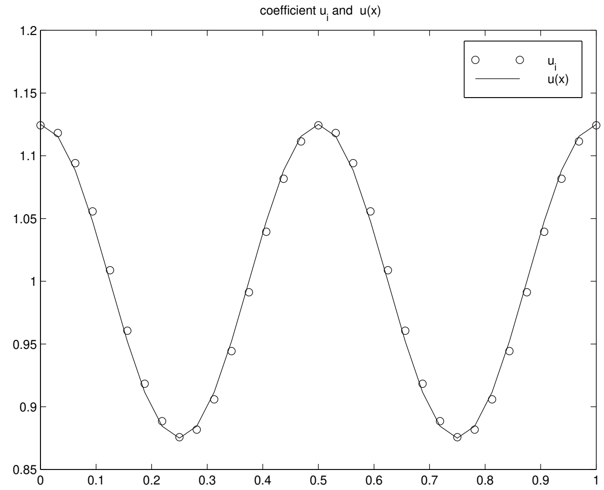

Fig 22.

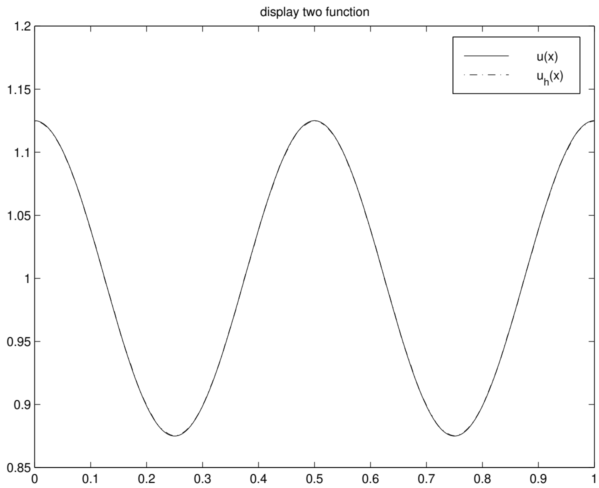

Fig 23.

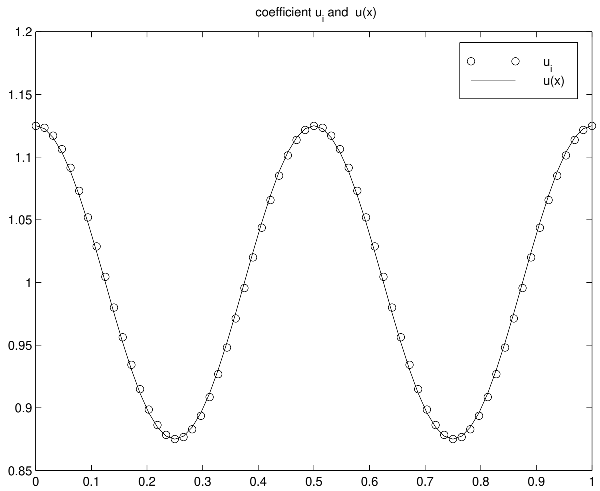

Fig 24.

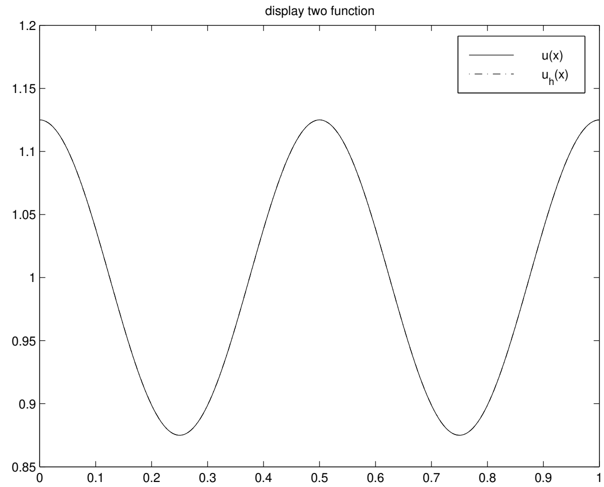

Fig 25.

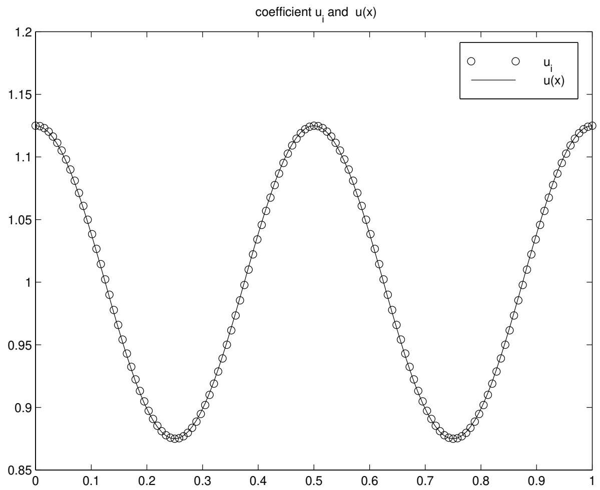

Fig 26.

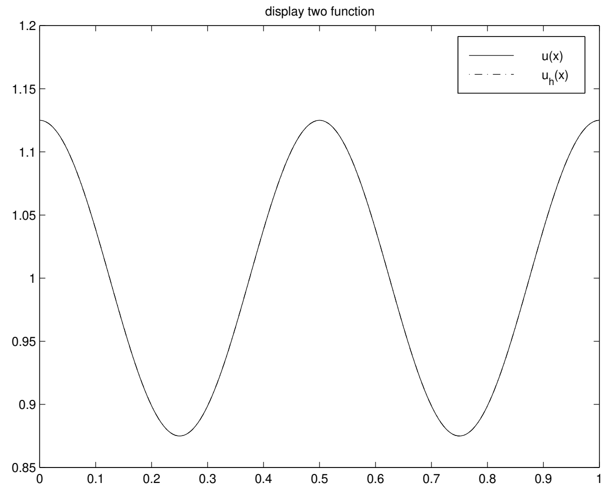

Fig 27.

Fig 20 至 Fig 27 我們給定第 4 層到第 7 層的自格函數, 同樣也是因爲解析度的關係, 並且在第 7 層的逼近解已經與眞實解完全貼合 (這是從圖形上觀察的), 而且係數 u_i 在第 8, 9 層已經太密了, 所以無法看出任何現象. 在邊界平滑的函數逼近能力方面, 顯然比不平滑的函數還要好, 在這方面與我們當初猜想得很近. 逼近的能力與整個週期解的可微程度有關. 在解的逼近能力方面, 對於不平滑解函數的誤差分析, $N = 16$ 點的時候誤差準確度只達到 10^{-2}, 比傳統有限元解的 10^{-4} 還要差很多. 但是在在解曲線平滑度提高的情形下, 同樣在 $N = 16$ 點的情形, 自格函數的逼近能力提高到 10^{-3}, 與傳統有限元解的 10^{-3} 不相上下. 隨著 N 的點數提高, 我們發現自格函數的逼近能力開始超過傳統有限元法, 這是一個令人高興的事.

在這裡我們可以就兩個例題做一個比較與歸納. 在邊界不平滑的條件例子中, 因為傳統有限元在週期性問題裡, 造出剛度矩陣 K 後, 矩陣本身是一個 singular 的矩陣 (因爲解不唯一), 所以我們給定 (2.5) 式後, 確定了所有的係數條件爲 (2.6) 式. 但是邊界的條件 $u(0) = u(l)$ 是給定的條件. 然後開始求解 $Ku = F$, 求解過程採用矩陣迭代消去求得係數矩陣 u, 並且在求解其它係數過程中, 解的的誤差是由矩陣迭代所產生, 是一個分配性的誤差, 而且即使邊界很平滑, 在邊界上並沒有誤差產生. 但是我們用週期性的自格函數基底時, 給定邊界條件後, 無法定出首尾兩項的係數, 但是我們又必須給定因爲週期性所造成的矩陣 K 的解不唯一的條件, 同樣的我們給定 (3.15) 式然後決定 (3.16) 式, 但是我們無法立刻決定首尾兩項係數, 並且邊界上的係數無法準確逼近下, 首尾係數的誤差大小迅速被 (3.16) 式是給吸收了, 連帶的整個區域內的係數都解不準. 在 Fig 5,7,9,11 上的係數超收現象是一個假像, 在 Fig 20,22,24,26 上我們很輕易發現係數並沒想像中解的準確, 但是在乘上自格函數後的逼近能力卻比前者好. 另外我們前面所說的自格函數不對稱性會造成平移的現像, 剛好在第二個例子中得到驗證. 可以說是本來的係數就無法達到超收斂並且有向右偏移的現像, 但是因爲自格函數的不對稱性 (3 階自格函數的最高點在五分之二處, 向左偏移一單位), 把眞解逼近得很準. 在條件數方面, 比較 shape 函數與自格函數之間的條件數, 以自格函數爲基底的矩陣 K 的條件數並沒有比傳統有限元的矩陣 K 來得好, 並且是一個病態矩陣, 降低條件數的計算我們將在下一節討論.

4-2 凌坡轉換與預條件處理計算

這一節我們應用在影像處理的方法來分解矩陣 K_J, 在影像處理時我們把原圖用凌波分解把圖像的高低頻分解出來, 然後再去掉較弱的高頻部分, 使得原圖像保留主要的低頻部分, 以方便壓縮. 在微分方程的凌波分解, 我們並沒有去掉任何的資訊. 因爲影像方面的少量資料遺失並不會造成太大的視覺誤差, 但是在解微分方程方面, 我們必須保留所有分解後的資料, 不然極小的資料遺失將可能會在求解過程中產生大的誤差.

微分方程降層的動作是把矩陣 K_J 的元素的高頻 (大的特徵值) 與低頻 (小的特徵值) 分解出來. 然後再根據預條件數處理方法, 把元素的特徵值做一個調整, 避免矩陣間的特徵值差距過大, 造成較大條件數, 因而使得係數矩陣 u 解不準.

根據前一章所敘述, 我們在改善條件數方面引進 Preconditioner matrix 來降低矩陣 K_J 的條件數並且又能保有原來的逼近能力. 根據 (3.23) 我們把第 J 層的自格函數矩陣 K_J 乘上係數矩陣 $\mathcal{W}$ 以後轉換爲凌波函數基底,

$$\mathcal{W}K_J\mathcal{W}^T = \mathcal{A}_J^{(1)}$$

$\mathcal{A}_J^{(1)}$ 右上角的 1 代表經過 1 次轉換, 但是因爲 $\frac{1}{2}\mathcal{W}\mathcal{W}^T = I$ 所以它的條件數並沒有多大的改變. 我們降層的動作只討論細層經過轉換後往下降 4 層 (i.e. 轉換到粗層), 我們選取 7, 8, 9 三層來實驗. 對於 Example 1. 則我們得到第 7 層爲

$\kappa(K_7)$	$\kappa(\mathcal{A}_7^{(1)})$	$\kappa(\mathcal{A}_7^{(2)})$	$\kappa(\mathcal{A}_7^{(3)})$	$\kappa(\mathcal{A}_7^{(4)})$
15366	17662.8	18199	18196	18338

第 8 層爲

$\kappa(K_8)$	$\kappa(\mathcal{A}_8^{(1)})$	$\kappa(\mathcal{A}_8^{(2)})$	$\kappa(\mathcal{A}_8^{(3)})$	$\kappa(\mathcal{A}_8^{(4)})$
61472	70659	72792	72739	72922

第 9 層爲

$\kappa(K_9)$	$\kappa(\mathcal{A}_9^{(1)})$	$\kappa(\mathcal{A}_9^{(2)})$	$\kappa(\mathcal{A}_9^{(3)})$	$\kappa(\mathcal{A}_9^{(4)})$
186733	282646	291162	290909	291483

根據上面的實驗結果, 降層以後的矩陣條件數並沒有變好.

根據 (3.23) 我們定義一個對角線矩陣 $\mathcal{D}$, 然後做 diagonal rescaling , 使得我們可以得到一個比較平衡的矩陣 (矩陣的最大特徵值與最小特徵值不至於相差太多). 但是如果乘上 Preconditioner matrix $\mathcal{D}^{-1}$ 後, 在第一個例子中, 我們得到第 7 層爲

$\kappa(\mathcal{D}^{-1}\mathcal{A}_7^{(1)}\mathcal{D}^{-1})$	$\kappa(\mathcal{D}^{-1}\mathcal{A}_7^{(2)}\mathcal{D}^{-1})$	$\kappa(\mathcal{D}^{-1}\mathcal{A}_7^{(3)}\mathcal{D}^{-1})$	$\kappa(\mathcal{D}^{-1}\mathcal{A}_7^{(4)}\mathcal{D}^{-1})$
5817	2654	1552	1280

第 8 層爲

$\kappa(\mathcal{D}^{-1}\mathcal{A}_8^{(1)}\mathcal{D}^{-1})$	$\kappa(\mathcal{D}^{-1}\mathcal{A}_8^{(2)}\mathcal{D}^{-1})$	$\kappa(\mathcal{D}^{-1}\mathcal{A}_8^{(3)}\mathcal{D}^{-1})$	$\kappa(\mathcal{D}^{-1}\mathcal{A}_8^{(4)}\mathcal{D}^{-1})$
23264	10587	6115	4793

第 9 層爲

$\kappa(\mathcal{D}^{-1}\mathcal{A}_9^{(1)}\mathcal{D}^{-1})$	$\kappa(\mathcal{D}^{-1}\mathcal{A}_9^{(2)}\mathcal{D}^{-1})$	$\kappa(\mathcal{D}^{-1}\mathcal{A}_9^{(3)}\mathcal{D}^{-1})$	$\kappa(\mathcal{D}^{-1}\mathcal{A}_9^{(4)}\mathcal{D}^{-1})$
93051	42316	24366	18859

我們可以明顯地看到矩陣 K 的條件數降低了許多. 再來我們觀察誤差分析方面, 經過矩陣分解後的逼近能力我們也看 7, 8, 9 三層的值, 在第 7 層爲

$\|u-u_h^1\|^2$	$\|u-u_h^2\|^2$	$\|u-u_h^3\|^2$	$\|u-u_h^4\|^2$
2.2835E-3	4.4693E-3	8.5685E-3	1.57E-2

(註: 上面 $\|u-u_h^1\|^2$ 右上角 2 所代表的是 ℓ^2-$norm$, u_h^1 右上角表示做了第一次的凌波轉換) 我們發現誤差明顯增加, 並且呈兩倍數增長, 我們在看看第 8, 9 層的誤差分析, 在第 8 層得到

$\|u-u_h^1\|^2$	$\|u-u_h^2\|^2$	$\|u-u_h^3\|^2$	$\|u-u_h^4\|^2$
8.1464E-4	1.6127E-3	3.16E-3	6.0599E-3

在第 9 層得到

$\|u-u_h^1\|^2$	$\|u-u_h^2\|^2$	$\|u-u_h^3\|^2$	$\|u-u_h^4\|^2$
2.8943E-4	5.607E-4	1.1407E-3	2.351E-3

誤差的變大跟降層後的 $\phi(x)$ 有很密切的關係, 因爲降至粗層的時候, 相當於我們用了 scale 較大的 $\phi(x)$ 來當基底, 在原來我們沒有降層的動作前的誤差分析中, 我們就已經發現到網格切的越大的有限元素法, 邊界的平滑度對於逼近能力有著決定性的影響. 爲了驗證我們上面所說的想法, 同樣的我們也計算邊界平滑度較高的 Example 2, 看看是否跟我們的猜想一樣.

我們先觀察 Example 2 經過凌波轉換後的條件數變化, 第 7 層爲

$\kappa(K_7)$	$\kappa(\mathcal{A}_7^{(1)})$	$\kappa(\mathcal{A}_7^{(2)})$	$\kappa(\mathcal{A}_7^{(3)})$	$\kappa(\mathcal{A}_7^{(4)})$
15366	17663	18199	18196	18338

第 8 層爲

$\kappa(K_8)$	$\kappa(\mathcal{A}_8^{(1)})$	$\kappa(\mathcal{A}_8^{(2)})$	$\kappa(\mathcal{A}_8^{(3)})$	$\kappa(\mathcal{A}_8^{(4)})$
61472	70659	72791	72738	72921

第 9 層爲

$\kappa(K_9)$	$\kappa(\mathcal{A}_9^{(1)})$	$\kappa(\mathcal{A}_9^{(2)})$	$\kappa(\mathcal{A}_9^{(3)})$	$\kappa(\mathcal{A}_9^{(4)})$
245894	282646	291162	290909	291483

但是如果乘上 Preconditioner matrix $\mathcal{D}^{-1}$ 後, 我們得到第 7 層爲

$\kappa(\mathcal{D}^{-1}\mathcal{A}_7^{(1)}\mathcal{D}^{-1})$	$\kappa(\mathcal{D}^{-1}\mathcal{A}_7^{(2)}\mathcal{D}^{-1})$	$\kappa(\mathcal{D}^{-1}\mathcal{A}_7^{(3)}\mathcal{D}^{-1})$	$\kappa(\mathcal{D}^{-1}\mathcal{A}_7^{(4)}\mathcal{D}^{-1})$
5818	2655	1552	1281

第 8 層爲

$\kappa(\mathcal{D}^{-1}\mathcal{A}_8^{(1)}\mathcal{D}^{-1})$	$\kappa(\mathcal{D}^{-1}\mathcal{A}_8^{(2)}\mathcal{D}^{-1})$	$\kappa(\mathcal{D}^{-1}\mathcal{A}_8^{(3)}\mathcal{D}^{-1})$	$\kappa(\mathcal{D}^{-1}\mathcal{A}_8^{(4)}\mathcal{D}^{-1})$
23264	10587	6115	4794

第 9 層爲

$\kappa(\mathcal{D}^{-1}\mathcal{A}_9^{(1)}\mathcal{D}^{-1})$	$\kappa(\mathcal{D}^{-1}\mathcal{A}_9^{(2)}\mathcal{D}^{-1})$	$\kappa(\mathcal{D}^{-1}\mathcal{A}_9^{(3)}\mathcal{D}^{-1})$	$\kappa(\mathcal{D}^{-1}\mathcal{A}_9^{(4)}\mathcal{D}^{-1})$
93051	42317	24366	18860

同樣的我們看到矩陣 K 的條件數降也低了許多, 再來我們觀察誤差分析方面, 經過矩陣分解後的逼近能力我們也看 7, 8, 9 三層的值, 在第 7 層爲

$\|u-u_h^1\|^2$	$\|u-u_h^2\|^2$	$\|u-u_h^3\|^2$	$\|u-u_h^4\|^2$
8.2682E-6	8.2682E-6	8.2682E-6	8.2682E-6

在第 8 層得到

$\|u-u_h^1\|^2$	$\|u-u_h^2\|^2$	$\|u-u_h^3\|^2$	$\|u-u_h^4\|^2$
3.42E-6	3.42E-6	3.42E-6	3.42E-6

在第 9 層得到

$\|u-u_h^1\|^2$	$\|u-u_h^2\|^2$	$\|u-u_h^3\|^2$	$\|u-u_h^4\|^2$
1.6445E-6	1.6445E-6	1.6445E-6	1.6445E-6

我們觀察到降層求解後重新合成回到原來的層數, 所計算出的逼近解並沒有改變, 所以邊界的平滑度是眞正影響逼近能的眞正原因.

另一方面, 其實在乘上預條件處理矩陣 $\mathcal{D}^{-1}$ 後, 在誤差方面應該會有改變, 因爲條件數變得比較小了, 如果我們用迭代法解矩陣問題, 較小的條件數使得我們可以用很少的次數就可以得到較精確的逼近解. 但是因爲我們用的是 Matlab 來做計算, 事實上還是高斯消去法, 誤差方面並沒有改變 (因爲 $\mathcal{D}^{-1}$ 是對角線矩陣, 所以合成回原來層數時, 乘上 $\mathcal{D}$ 時, 就完全消去, 並沒有發揮它應該有的功能).

4-3 綜合結論

在實驗過程中我們得到三個結論, 第一個結論是, 在轉換基底的實驗過程中我們必須注意一件事, 就是 (3.17) 式的位置. 當我們在計算以自格函數 $\phi(2^j x)$ 算出的矩陣 K_J 時, (3.17) 是矩陣 K_J 造出後, 才加在矩陣的最後一列. 但是轉換到以 $\phi(2^{j-1}x)$ 與

$\psi(2^{j-1}x)$ 的粗層時, (3.17) 這個式子必須是與 $\phi(2^{j-1}x)$ 有關的, 所以這條式子必須是轉換過後才加入矩陣 $\mathcal{A}_J$ 的. 因爲在轉換到粗層後, $\int_0^1 u(x)dx = 1$ 的 $u(x)$ 的基底也必須轉換至粗層, 但是因爲 $\psi(2^{j-1}x)$ 在 (0,1) 區間的積分均爲 0, 以 $\psi(2^{j-1}x)$ 基底函數的係數未影響到其它以 $\phi(2^{j-1}x)$ 爲基底函數的係數, 所以我們必須等到降層後才把 (3.17) 式加至矩陣 $\mathcal{A}_J$內. 可以參考下面四張分解過程圖,

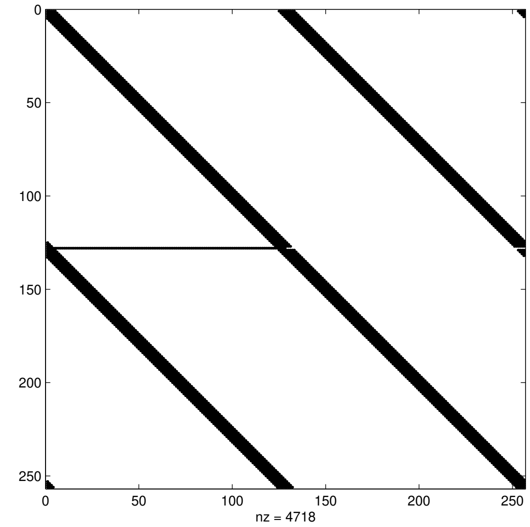

Fig 28.

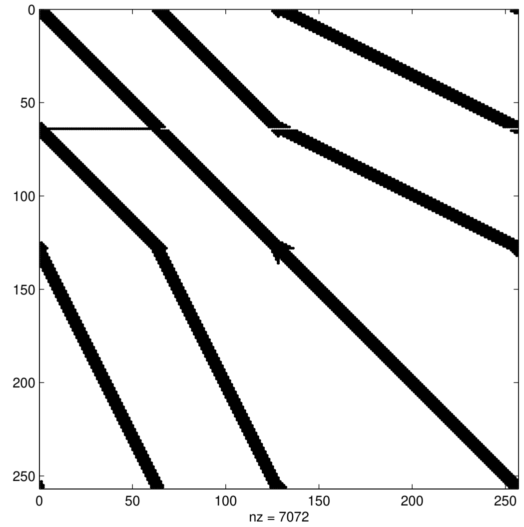

Fig 29.

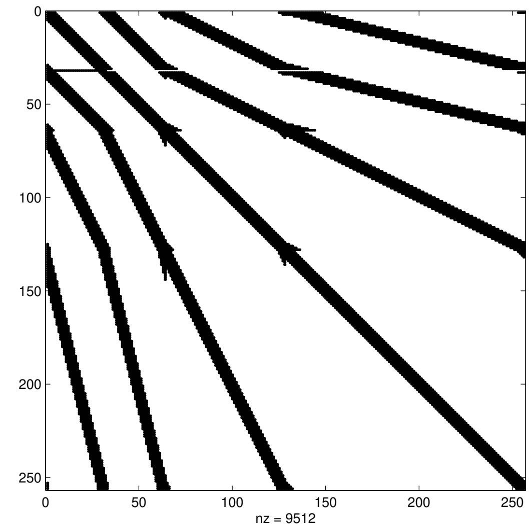

Fig 30.

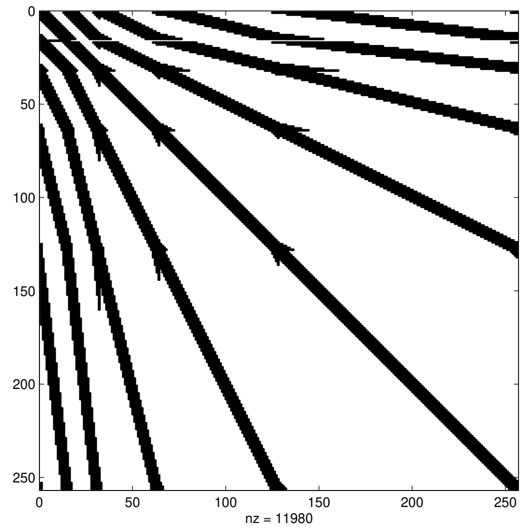

Fig 31.

第二個結論是, 因爲前兩節我們計算的是 $-u"(x) = f(x)$ 的問題, 所以我們只會遇到一個矩陣系統 K, 但是如果題目換成是 $-u"(x) + ku(x) = f(x)$ 的問題, 在建立矩陣時我們又會多出現一個矩陣 M (mass matrix), M 矩陣是由兩個 $\phi(x)$ 相乘後積分得到的,

然而在前一節的預條件處理我們只是針對單一矩陣 K, 如果同時處理 $(K+M)u=F$ 的問題時, 我們想知道預條件處理後的 M 矩陣它的條件數變化. 我們實驗後發現當降 n 層的預條件處理後的 M 條件數變化, 皆是呈現 $4n$ 倍數成長, 這是因爲 M 幾乎是單位矩陣 ($\phi(x)$ 是單位正交基底). 所以在遇到 $-u"(x)+ku(x)=f(x)$ 這類問題時, 我們猜想預條件處理後的條件數變化會是一個先遞減後增加的曲線.

第三個結論是, 我們在處理邊界不平滑的解函數時得知, 以 $\phi(x)$ 當基底時, 邊界的 Taylor 展開式的階數影響到解的逼近能力. 然而我們希望在轉換到 $\psi(x)$ 的基底時, 運用 $\psi(x)$ 函數的消失矩 (vanish moment) 把邊界函數可展開的多項式去掉. 但是因爲邊界的不平滑, 導致邊界的展開式不連續 (週期性函數可視爲無數個有限區間函數相連). 原本應該有的 $\int x\,\psi(x)\,dx=0$, 變成 $\int x\,\tilde{\psi}(x)\,dx\neq 0$ ($\tilde{\psi}(x)$ 是週期性的基底函數). 而這些不爲零的值, 在矩陣求解過程中直接影響逼近解的逼近能力.

Reference

1. Forsythe, G.E.,and C.B.Moler (1967) *Computer solution of linear algebraic systems* , Prentice Hall, Englewood Cliffs,N.J.;148pp. QA297.F57.
2. Eric B. Becker, Grahm F. Carey, and J. Tinsley Oden 'Finite Element An Introduction , Volume 1'. Prentice Hall, Englewood Cliffs,N.J.
3. I.Fried (1972) ' Bounds on the extremal eigenvalues of the F.E stiffness and mass matrices and their special condition number' *Journal of sound and vibration* , 22 (4) 407-478.
4. M.Dorobantu, *Wavelet based algorithms for one-dimension parabolic equation* ,TRITA-NA-9214,(1992).
5. Y. Meyer,Ondelettes et Operateurs 1: Ondelette,Herman,(1990).
6. 嚴健彰, "凌波理論", 中央大學數學系碩士論文, (1995).
7. 徐敏堂, "凌波函數的數值積分法", 中央大學數學系碩士論文, (1995).
8. I.Daubechies, *Ten Lectures on Wavelets*, SIAM, Philadelphia, (1992).
9. J.Douglas Faires, Richadr L. Burden, 'Numerical Methods' , PWS-KENT, FIFTH EDITION.
10. Gordon Erlebacher,M.Yousuff Hussaini, Leland M.Jameson, *Wavelets Theory and Applications*, Oxford University Press (1996).
11. Gilbert Strang, Truong Nguyen, *Wavelets and Filter Banks*, Wellesley-Cambridge Press.
12. G.Strang and G.J.Fix, *An Analysis of the Finite Element Method*, Prentice Hall, Englewood Cliffs, NJ, (1973).
13. A.Cohen and R.D.Ryan, *Wavelets and Multiscale Signal Pricessing*, Chapman & Hall, London, (1995).
14. Eric B. Becker,Grahm F. Carey, and J. Tinsley Oden 'Finite Element Mathematical Aspects, Volume 4'. Prentice Hall, Englewood Cliffs,N.J.
15. Zhang Pingwen, Zhang Yu, *Wavelet Method for Boundary Integral Equations*, Peking University.