1977

Finite element solutions to boundary value problems

Chew, Kok-Thai

FINITE ELEMENT SOLUTIONS TO BOUNDARY VALUE PROBLEMS

A thesis submitted to
Lakehead University
in partial fulfillment of the requirements
for the degree of
Master of Science

by

Kok-Thai Chew

1977

ACKNOWLEDGMENT

I wish to thank my supervisor, Professor P. O. Frederickson, for his advice and encouragement during the preparation of this thesis.

ABSTRACT

The finite element solution of certain two-point boundary value problems is discussed.

In order to obtain more accuracy than the linear finite element method can give, an order-h^4 *global superconvergence* technique is studied. This technique, which uses a quasi-inverse of the Rayleigh-Ritz-Galerkin (finite element) method, is motivated by the papers of C. de Boor and G. J. Fix [14] and P. O. Frederickson [25]. The Peano kernel theorem is generalized and used to approximate the rate of convergence of the global superconvergence.

Following Sard's theory on *best quadrature formulae* [50], with some generalization, several quadrature formulae are derived. These quadrature formulae are shown to be *consistent*, and have some advantages over those obtained by Herbold, Schultz and Varga [34].

For solution of large linear systems which result from the finite element method, LU decomposition (Gaussian Elimination Method) is fast and accurate. However, when it comes to a singular or a nearly singular system, LU decomposition fails. The algorithm FAPIN developed by P. O. Frederickson for 2-dimensional systems is able to solve singular systems as we demonstrate.

We found FAPIN will work more efficiently in 1-dimensional case if we replace the DB_q approximate inverse C, developed by Benson [3], with other approximate inverses.

For the sake of verifying the theory, appropriate numerical experiments are carried out.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

1.1. Two-Point Boundary Value Problems

Two-point boundary value problems (abbrev. TPBVP) associated with ordinary differential equations mostly arise in physics and engineering problems. For these problems conditions are specified at the two ends of an interval and a solution to the ordinary differential equation is sought to satisfy the boundary conditions. For example the vertical deflection $y(x)$ of a transverse loaded string with two ends fixed satisfies the ordinary differential equation $-y''(x)=f(x)$ with the boundary conditions $y(0)=a$, $y(1)=b$. Numerous analytical techniques for solving TPBVP have been developed. The characteristic of an analytical method is that it expresses particular solutions of ordinary differential equations in terms of series or integrals involving elementary or special functions. However there are restrictions on analytical methods. For example, the TPBVP $-y''(x)=e^{x^2}$, $y(0)=y(1)=0$ has the solution $y(x)=-\int_0^x \int_0^\zeta e^{\xi^2} d\xi d\zeta$. This integral can not be expressed in closed form in terms of known functions, thus numerical quadrature is necessary to approximate the $y(x)$.. By numerical methods we mean methods to approximate the solution of TPBVP without any assistance of an analytical solution. Numerical methods provide practical procedures for approximating the solutions of a very general class of TPBVP. In parallel with the development of the modern computer, numerical methods are becoming increasingly important. Many numerical methods have been developed, for example finite

difference methods which have been investigated in detail in [24], and shooting methods for which a detailed investigation can be found in [46]. [38] and [39] cover some other numerical methods. [53] is devoted to the numerical methods which are under current research activities. However, this thesis is dedicated to finite element methods.

1.2. Basic Notations

Let $u^{[n]}$ denotes the n^{th} derivatives of u. Let

$$(1.2.1) \qquad H^n[0,1] = \{ u \mid u^{[n]} \in L_2[0,1] \} \quad \text{for} \quad n \geq 0$$

Define

$$(1.2.2) \qquad (f,g)_n = \sum_{i=0}^{n} (f^{[i]},g^{[i]})$$

then $||f|| = (f,f)_n^{\frac{1}{2}}$ is a (Sobolev) norm on $H^n[0,1]$ and $(H^n[0,1] , ||\cdot||_n)$ is known as a *Sobolev space* (cf. [2], [61]). From (1.2.1) and (1.2.1), we have

$$(1.2.3) \qquad ||\cdot||_{n_1} \leq ||\cdot||_{n_2} \quad \text{for} \quad n_1 \leq n_2$$

hence it is clear that $H^{n_1}[0,1] \supset H^{n_2}[0,1]$, i.e. the embedding of $H^{n_2}[0,1]$ into $H^{n_1}[0,1]$ is continuous (cf. [2], pp.21).

We shall denote by

$$(1.2.4) \qquad H_0^n[0,1] = \{ u \in H^n[0,1] \mid u(0) = u(1) = 0 \} .$$

Obviously it is a subspace of $H^n[0,1]$.

Define (cf. [2] [56])

$$(1.2.5) \qquad H^{-1}[0,1] = \{ u \mid \int_0^x u(t)dt \in H^0[0,1], \quad 0 \le x \le 1 \}$$

with norm

$$(1.2.6) \qquad ||f||_{-1} = \max_{v \in H_0^1[0,1]} \frac{\left| \int_0^1 f(x)v(x)dx \right|}{||v||_1}$$

We observe that the Dirac δ-function, δ_x, is an element of H^{-1} if $x \in (0,1)$

Let $C^n[0,1]$ be the set of all real-valued functions which have continuous derivatives of order at least n in $[0,1]$, where n is a non-negative integer.

Let $\Pi : 0 = x_0 < x_1 < \cdots\cdots < x_n = 1$ be a partition on $[0,1]$. Let $I_i = [x_{i-1}, x_i]$, $h_i = x_i - x_{i-1}$, for $i = 1,\cdots,n$, and $h = \max_i h_i$. If $h_i \equiv h$ for all i, we shall denote by Π_h the regular partition on $[0,1]$ with regular mesh $h = \frac{1}{n}$.

We shall denote by $S_\pi^{n,k}$ the space of spline functions (definition is given in 1.9.) defined on Π.

Let $E = [a,b] \subset [0,1]$. Denote by $\mathcal{P}^m(E)$ the set of polynomials of degree m defined on E and $\mathcal{P}_0^m(E) = \{ p \in \mathcal{P}^m(E) \mid p(a) = p(b) = 0 \}$.

The truncated power function is defined as

$$x_+^m = \begin{cases} x^m & x > 0 \\ 0 & x \leq 0 \end{cases}$$

For $m = 0$, this is the well known Heaviside function.

1.3. Problem Formulation

Consider the differential equation

(1.3.1) $$-(p(x)u'(x))' + q(x)u(x) = f(x) \qquad x \in [0,1]$$

Define the differential operator $L : H^1[0,1] \to H^{-1}[0,1]$ (cf. [8], [56]) by

(1.3.2) $$Lu = -(pu')' + qu$$

The problem is : given a $f \in H^{-1}[0,1]$, we are asked to find a $u \in H^1[0,1]$ such that

(1.3.3) $$Lu = f$$

with the boundary conditions :

(1.3.4) $$u(0) = g_0 , \quad u(1) = g_1 .$$

To ensure that the equation (1.3.2) - (1.3.4) has a solution, we assume that ([8], [37], [56], [59])

(1.3.5) $\quad p(x) \in C^0[0,1], \quad q(x) \in C^0[0,1]$

and

(1.3.6) $\quad p(x) \geq p_{min} > 0, \quad q(x) \geq 0$

Let $H_g^1[0,1] = \{ u \in H^1[0,1] \mid u(0)=g_0, \ u(1)=g_1 \}$; it is the space of admissible functions. It is not a vector space, but an *affine* space since $u_1, u_2 \in H_g^1[0,1]$ imply $u_1 - u_2 \in H_0^1[0,1]$.
With the assumptions (1.3.5) - (1.3.6), L is a one-to-one continuous linear operator from $H_g^1[0,1]$ to $H^{-1}[0,1]$. Thus for each $f \in H^{-1}[0,1]$, the BVP (1.3.2) - (1.3.4) has a unique solution $u \in H_g^1[0,1]$ ([56]). Moreover, there exists a constant ρ_1 , independent of f, such that

(1.3.7) $\quad ||u||_1 \leq \rho_1 ||f||_{-1}$

If $f \in H^0[0,1]$, then $u \in H_g^2[0,1]$ and there exists ρ_2 such that

(1.3.8) $\quad ||u||_2 \leq \rho_2 ||f||_0$

(with ρ_2 also independent of f).

Lemma 1.3.1 (i) L is positive definite on $H_0^1[0,1]$,
i.e. $(Lv_0, v_0) > 0$ for all non-zero $v_0 \in H_0^1[0,1]$.

(ii) L is symmetric, i.e. $(Lu, v_0) = (Lv_0, u)$
for $u \in H_g^1[0,1]$, $v_0 \in H_0^1[0,1]$
(cf. [37], [40], [56]).

Proof: (i) If $v_0 \in H_0^1[0,1]$, $p(x) > 0$ and $q(x) \geq 0$, then

$$(Lv_0, v_0) = \int_0^1 \{-(p(x)v_0'(x))' + q(x)v_0(x)\}v_0(x)\, dx$$

$$= \int_0^1 \{p(x)(v_0'(x))^2 + q(x)(v_0(x))^2\}\, dx \quad > 0$$

(ii) Let $u \in H_g^1[0,1]$, $v_0 \in H_0^1[0,1]$, then

$$(Lu, v_0) = \int_0^1 -p(x)u'(x)v_0'(x) + q(x)u(x)v_0(x)\, dx$$

$$= (Lv_0, u)$$

1.4. The Variational Formulation of The Problem

Let $\Phi : H_g^1[0,1] \to R$ be a quadratic functional defined by (cf. [8], [9], [40], [56])

$$\Phi(v_0) = \int_0^1 \{p(x)(v_0'(x))^2 + q(x)(v_0(x))^2 - 2f(x)v_0(x)\}\, dx \tag{1.4.1}$$

For any fixed $w \in H_g^1[0,1]$, for any ε and $v_0 \in H_0^1[0,1]$

$$\Phi(w+\varepsilon v_0) = \Phi(w) + 2\varepsilon\{\int_0^1 (-(p(x)w'(x))' + q(x)w(x) - f(x))\, v_0(x)dx\}$$
$$+ \varepsilon^2 \int_0^1 \{p(x)(v_0'(x))^2 + q(x)(v_0(x))^2\}\, dx \tag{1.4.2}$$

Define the first variation ([37], [56])

$$\delta\Phi(w,v_0) = \lim_{\varepsilon\to 0} \frac{\Phi(w+\varepsilon v_0) - \Phi(w)}{\varepsilon}$$

(1.4.3)

$$= 2\int_0^1 (-(p(x)w'(x))' + q(x)w(x) - f(x))v_0(x)\,dx$$

Then (1.4.2) can be written as

$$\Phi(w+\varepsilon v_0) = \Phi(w) + \varepsilon\cdot\delta\Phi(w,v_0)$$

(1.4.4)

$$+ \varepsilon^2 \int_0^1 (p(x)(v_0'(x))^2 + q(x)(v_0(x))^2)\,dx$$

From (1.4.4), we have the following important notes:

(cf. [8], [37], [40], [56])

(i) For each $w \in H_g^1[0,1]$

$$\delta\Phi(w,v_0) = 0 \quad \text{iff} \quad (Lw,v_0) = (f,v_0)$$

$$\forall\ v_0\in H_0^1[0,1] \qquad \forall\ v_0\in H_0^1[0,1]$$

The right-hand side is known as the Galerkin weak form.

(ii) If $w \in H_g^1[0,1]$ has the property that

$$\delta\Phi(w,v_0) = 0 \quad \text{for all} \quad v_0 \in H_0^1[0,1]$$

then

$$\Phi(w) < \Phi(w + \varepsilon v_0) \quad \text{for any} \quad \varepsilon \neq 0,$$

$$\text{and non-zero} \quad v_0 \in H_0^1[0,1]$$

The reverse is true.

In other words, the element w in $H_g^1[0,1]$ which minimizes the quadratic functional Φ over $H_g^1[0,1]$ is the unique element at which the first variation of Φ is zero.

(iii) If $w \in H_g^1[0,1]$ such that $\delta\Phi(w,v_0) = 0 \quad \forall\, v_0 \in H_0^1[0,1]$ then w is the solution to (1.3.2) - (1.3.4). The reverse is true.

Thus, we have the following theorem:

Theorem 1.4.1

(i) u^*, the unique solution to (1.3.2) - (1.3.4), strictly minimizes $\Phi(v)$ over the admissible space $H_g^1[0,1]$.

(ii) The first variation of $\Phi(v)$ at u^* vanishes.

(cf. [8], [37], [40], [56])

1.5. The Energy Norm

Since L in (1.3.3) is a positive definite and symmetric linear operator on $H_0^1[0,1]$. we can define a new inner product $a(u,v)$ on $H_0^1[0,1]$ by (cf. [2], [40], [56], [59])

$$(1.5.1) \qquad a(u,v) = (Lu,v) \qquad \text{for all } u,v \in H_0^1[0,1]$$

The proof that $a(u,v)$ is an inner product is straight forward by the definition of an inner product and the properties of L.

Following [56], we shall call $a(u,v)$ the *energy* inner product on $H_0^1[0,1]$.

Define a norm $||\cdot||_A$ on $H_0^1[0,1]$ by

(1.5.2) $$||u||_A = [a(u,u)]^{\frac{1}{2}} \quad \text{for all} \quad u \in H_0^1[0,1].$$

We shall refer it as the *energy* norm.

Lemma 1.5.1

The norm $||\cdot||_A$ is equivalent to $||\cdot||_1$ on $H_0^1[0,1]$, i.e. there exists constants $\rho_3, \rho_4 \neq 0$, such that

(1.5.3) $$\rho_3||v||_1 \leq ||v||_A \leq \rho_4||v||_1$$

Proof: (cf. [56], pp.42), let $v_0 \in H_0^1[0,1]$

$$\begin{aligned} ||v_0||_A^2 &= a(v_0,v_0) \\ &= \int_0^1 p(x)(v_0'(x))^2 + q(x)v_0^2(x)\ dx \\ &\leq \max[p(x),q(x)] \int_0^1 (v_0'(x))^2 + (v_0(x))^2)\ dx \\ &= \max[p(x),q(x)]\ ||v_0||_1 \end{aligned}$$

On the other hand, if $v_0(0) = 0$, then we have

$$v_0(x_0) = \int_0^{x_0} v_0'(z)\, dz$$

By the Schwarz inequality

$$\begin{aligned} |\, v_0(x_0)\,|^2 &\le \Big(\int_0^{x_0} 1^2\, dz\Big)\Big(\int_0^{x_0} |v_0'(z)|^2\, dz\Big) \\ &\le \Big(\int_0^{1} 1\, dz\Big)\Big(\int_0^{x_0} |v_0'(z)|^2\, dz\Big) \\ &\le \int_0^{x_0} |v_0'(z)|^2\, dz \\ &\le \int_0^{1} |v_0'(z)|^2\, dz \end{aligned}$$

Integrating w.r.t. x_0 over the interval $[0,1]$, we have

$$\int_0^1 (v_0(x))^2\, dx \le \int_0^1 (v_0'(z))^2\, dz \tag{1.5.4}$$

and

$$\begin{aligned} ||v_0||_A^2 &= \int_0^1 p(x)(v_0'(x))^2 + q(x)(v_0(x))^2\, dx \\ &\ge p_{min} \int_0^1 (v_0'(x))^2\, dx \\ &\ge \frac{p_{min}}{2} \int_0^1 (v_0'(x))^2 + (v_0(x))^2\, dx \\ &\frac{p_{min}}{2}\, ||v_0||_1^2 \end{aligned}$$

This completes the proof.

1.6. The Rayleigh-Ritz-Galerkin (RRG) method

From theorem 1.4.1 we know that the solution of (1.3.2)-(1.3.4) is equivalent to the minimization of $\Phi(v)$ in (1.4.1). Thus instead of solving (1.3.2)-(1.3.4) directly, we could, alternatively, concentrate on the following problem ([37], [40], [56]):

$$
(1.6.1)\quad \begin{cases} \text{Given} \quad f \in H^{-1}[0,1] \\ \text{find the function} \quad u^* \in H_g^1[0,1] \qquad \text{s.t.} \\ \Phi(u^*) \le \Phi(v) \quad \text{for all} \quad v \in H_g^1[0,1] \\ \text{where} \quad \Phi \quad \text{is defined in (1.4.1)} \end{cases}
$$

For this kind of problem, a simple yet efficient method was proposed by W. Ritz in [49] in 1908. Ritz's method has been used widely in applied mechanics ([10],[35], [44]). In 1915, B. G. Galerkin ([23], [30]) proposed a method in solving BVP. It is well known that Ritz's method is a special case of Galerkin's method ([8], [37],[40]). For the self-adjoint elliptic problems the two methods are equivalent.

Ritz's method is to approximate u^* by a u_h^* from an affine subspace $S_{h,g}$ in the sense that $\Phi(u_h^*)$ is minimum over $S_{h,g}$ ([4], [37], [40]). More precisely, to approximate u^* by a sequence of more accurate solutions $u_{h_n}^* \in S_{h_n,g}$ such that

(1.6.2) $$\Phi(u^*_{h_1}) \geq \Phi(u^*_{h_2}) \geq \cdots\cdots$$

and

(1.6.3) $$\lim_{n\to\infty} \Phi(u^*_{h_n}) = \Phi(u^*)$$

where for fixed n

(1.6.4) $$\Phi(u^*_{h_n}) \leq \Phi(v_{h_n}) \qquad \forall\, v_{h_n} \in S_{h_n,g}$$

The Galerkin method is to approximate u^* by any $u^*_h \in S_{h_n,g}$ which satisfies

(1.6.5) $$(Lu^*_h, v_{h_n}) = (f, v_{h_n}) \qquad \forall\, v_{h_n} \in S_{h_n,0}$$

More precisely, to approximate u^* by a sequence of more accurate solution $u^*_{h_n} \in S_{h_n,g}$ with the properties

(1.6.6) $$S_{h_n,g} \subset S_{h_m,g} \qquad \text{if } n < m$$

and

(1.6.7) $$(Lu^*_{h_k}, v_{h_k}) = (f, u_{h_k}) \qquad \forall\, v_{h_k} \in S_{h_k,0}$$

(cf. [37], [40]).

The proof that (1.6.4) and (1.6.6) are equivalent is similar to the proof of Theorem 1.4.1.

Let $S_{h,g}$ be a closed affine subspace and let $u_h^* \in S_{h,g}$ be the RRG solution to u^* in the sense that u_h^* minimizes $\Phi(v_h)$ over $S_{h,g}$. The following theorem concerning the minimization property of the error $e = u^* - u_h^*$ (note that $e \in H_0^1[0,1]$) in the energy norm is equivalent to Theorem 1.1 of [56, pp.39].

Theorem 1.6.1

(i) u_h^* minimizes $a(u^*-v_h,u^*-v_h)$ over $S_{h,g}$, i.e.

$$(1.6.8) \quad a(u^*-u_h^*,u^*-u_h^*) = \min_{v_h \in S_{h,g}} a(u^*-v_h,u^*-v_h)$$

(ii) $$L(u^*-u_h^*,v_{h,0}) = 0 \qquad \forall\, v_{h,0} \in S_{h,0}$$

(iii) $$L(u_h^*,v_{h,0}) = (f,v_{h,0}) \qquad \forall\, v_{h,0} \in S_{h,0}$$

In particular, if $S_{h,g} = H_g^1[0,1]$, then

$$L(u^*,v_0) = (f,v_0) \qquad \forall\, v_0 \in H_0^1[0,1]$$

Theorem 1.6.1 is fundamental to the Ritz theory. The Ritz method provides us with an idea that we could approximate u^* from a close affine subspace $S_{h,g}$. The problem we are now facing is: "how do we construct the Ritz solution (equivalently, the Galerkin solution)?" This is essential because for a method to be practical, it has to be constructive.

The construction of the Ritz solution is based on the choice of an affine subspace $S_{h,g}$. Since the difference of any elements in $S_{h,g}$ is in $S_{h,0}$, it is clear that $S_{h,g}$ is a shift of $S_{h,0}$; we have $S_{h,g} = g + S_{h,0}$, where $g \in S_{h,g}$.

Suppose that the dimension of $S_{h,0}$ is m-1, then it has a basis $\{\phi_i\}_{i=1}^{m-1}$ such that every element $u_{h,0} \in S_{h,0}$ can be represented in the form $u_{h,0} = \sum_{i=1}^{m-1} a_i\phi_i$. Then $u_{h,g} \in S_{h,g}$ can be written as $u_{h,0} = g + \sum_{i=1}^{m-1} a_i\phi_i$. In the finite element method, the *trial* functions ϕ_i are piecewise polynomials (spline functions, cf. 1.9). We shall call ϕ_i, $i = 1,\cdots,m-1$, *finite elements* when they are taken to be spline functions.

The basic steps in the finite element method are:

(i) The conversion of the operational form of the problem to the variational form as we have discussed in section 1.4. .

(ii) The construction of the spline trial functions. This is the main subject of section 1.8. .

(iii) The computations of the stiffness matrix and the solution of the discrete large linear system. This will be discussed in Chapter 2 and Chapter 5 respectively.

The convergence rate of the RRG (finite element) approximation is the main topic of Chapter 4.

The next section is devoted to the Peano Kernel Theorem. We find that it is very useful in the error analysis.

1.7. The Peano Kernel Theorem

The Peano Kernel Theorem was due to G. Peano in his paper [43] in 1914. It is a very useful tool in the evaluation of the error functional either in interpolation or quadrature ([5], [12], [49], [57]).

Theorem 1.7.1 (Peano)

Suppose the linear functional

$$E : C^n[0,1] \to R \tag{1.7.1}$$

has the property that

$$E(p) = 0 \quad \forall\, p \in \mathcal{P}^m, \quad m < n \tag{1.7.2}$$

Then there exists a function K_{m+1} such that, for $f \in C^{m+1}[0,1]$,

$$E(f) = \int_0^1 K_{m+1}(t) f^{[m+1]}(t)\, dt \tag{1.7.3}$$

Furthermore, the Peano Kernel K_{m+1} is of the form

$$K_{m+1}(t) = \frac{1}{m!} E_x (x-t)_+^m \tag{1.7.4}$$

(The subscript x is used to indicate that $E_x(x-t)_+^m$ is a function of x).

Proof: Suppose that E is defined on $L_1[0,1]$, then by the Riesz representation theorem there exists a function $K_0 \in L_\infty[0,1]$ such that

$$E(f) = \int_0^1 K_0(t)f(t)\,dt \tag{1.7.5}$$

If $f \in C^1[0,1]$ and $\mathcal{P}^0 \in \eta(E)$ (the null set of E) then

$$E(f) = [-K_1(t)f(t)]_0^1 + \int_0^1 K_1(t)f'(t)\,dt \tag{1.7.6}$$

where

$$K_1(t) = -\int_0^t K_0(s)\,ds \tag{1.7.7}$$

Let $f = c$ (non-zero constant), then $f' = 0$ and $k_1(0) = 0$, these imply that $0 = E(c) = K_1(t)\cdot c$ hence $K_1(1) = 0$. Thus (1.7.6) becomes

$$E(f) = \int_0^1 K_1(t)f'(t)\,dt \tag{1.7.8}$$

By induction, assume that there is a K_{k+1}, $k < m$, such that

$$E(f) = \int_0^1 K_{k+1}(t)f^{[k+1]}(t)\,dt \tag{1.7.9}$$

Let

$$f(x) = \frac{x^{k+1}}{(k+1)!} \tag{1.7.10}$$

then

$$E(f) = 0, \quad f^{[k+1]}(x) = 1, \text{ and } f^{[k+2]}(x) = 0 \tag{1.7.11}$$

Define

$$K_{k+2}(t) = -\int_0^1 K_{k+1}(s)\, ds \tag{1.7.12}$$

then we have

$$E(f) = [-K_{k+2}(t)f^{[k+1]}(t)]_0^1 + \int_0^1 K_{k+2}(t)f^{[k+2]}(t)\, dt \tag{1.7.13}$$

From (1.7.10) and (1.7.11), we have $0 = K_{k+2}(1)\cdot 1 + 0$

thus

$$E(f) = \int_0^1 K_{k+2}(t)f^{[k+2]}(t)\, dt \tag{1.7.14}$$

Hence, there is a K_{m+1} such that

$$E(f) = \int_0^1 K_{m+1}(t)f^{[m+1]}(t)\, dt$$

To evaluate $K_{m+1}(t)$, let us consider a function $f_t \in C^{m-1}$ such that $f_t^{[m]} \in L_2[0,1]$ and $f_t^{[m+1]}(\tau) = \delta_t(\tau)$,
then from (1.7.3)

$$E(f_t) = K_{m+1}(t) \tag{1.7.15}$$

By integrating $f_t^{[m+1]}(\tau)$ $m+1$ times, we have

$$f_t(\tau) = \frac{(\tau-t)_+^m}{m!} \tag{1.7.16}$$

Thus

$$K_{m+1}(t) = \frac{1}{m!} E_\tau (\tau-t)_+^m$$

This completes the proof.

1.8. The Peano-Sard Kernel Theorem

A. Sard in [48] and [49] generalized the Peano Kernel Theorem and developed the theory of best approximation especially on the topic of best quadrature formulae.

Let $T : H^k[0,1] \to A \subset H^k[0,1]$ be a bounded linear operator such that

$$Tu = u \qquad \text{for all } u \in \mathcal{P}^m \subset A \tag{1.8.1}$$

then T can be viewed as an approximation of the identity mapping I.

Define the error function

$$E = I - T$$

From (1.8.1) we have

$E(\mathcal{P}^m) = 0$. For such an approximation T, we say that it is exact for polynomial of degree m or call it a *m-exact* approximation to I .

Theorem 1.8.1 (Peano-Sard)

Let $E : H^k[0,1] \to H^k[0,1]$ be a bounded linear function such that $E(\mathcal{P}^m) = 0$, then there exists $K_{m+1} : [0,1]\times[0,1] \to R$ such that

$$E(w) = \int_0^1 K_{m+1}(\cdot,t)w^{[m+1]}(t)\, dt \tag{1.8.2}$$

Furthermore,

$$K_{m+1}(\cdot,t) = \frac{1}{m!} E_\tau((\tau-t)_+^m) \tag{1.8.3}$$

Proof :

For each $k \geq 0$; consider for each $x \in [0,1]$, the proof is similar to the proof of Theorem 1.7.1 .

Lemma 1.8.1 If $K_0 \in S^{n,k}$ and $K_0(t) = \frac{d^m}{dt^m} K_m(t)$, then $K_m \in S^{n+m,k+m}$ (cf. (1.9.)).

Proof :

$$S^{n',k'} = \left\{ u \in C^{k'} \mid \exists\ \Pi \text{ s.t. } u|_{[x_{i-1},x_i)} \in \mathcal{P}^{n'}[x_{i-1},x_i) \right\}$$

so if $u \in S^{n',k'}$, then $\frac{du}{dt} \in C^{k'-1}$ $k' > -1$

and $$\frac{du}{dt}\Big|_{[x_{i-1},x_i)} \in \mathcal{P}^{n'-1}[x_{i-1},x_i)$$

We shall make use of the Peano-Sard Kernel Theorem in Chapter 2 to investigate the errors of the quadratures; and the applications of the Peano Kernel Theorem will be discussed in detail in Chapter 4. Now we go on to discuss spline functions and spaces of spline functions which will be used in the thesis.

1.9. Spline Functions

A draftsman's spline is a mechanical tool, consisting of a strip of wood or some flexible material, used by draftsman to draw a smooth curve to pass through specified points, called knots.

The idea of a mathematical spline came from the draftsman's spline. The term *spline function*, first used by Schoenberg ([50]) in 1946, is intended to suggest that the graph of such a function is similar to a curve drawn by the draftsman's spline which is approximately a cubic spline function. However, a generalization of this idea leads to the following definition:

Definition 1.9.1 A *spline function* (of degree m and q times differentiable, $0 \le q < m$) is a function s which satisfies the the following properties:

$$\text{(i)} \quad s|_{[x_{i-1},x_i]} \in \mathcal{P}^m[x_{i-1},x_i] \quad i=1,\cdots,n$$

(1.9.1)

$$\text{(ii)} \quad s \in C^q[0,1]$$

(cf. [1], [33], [50])

We shall denote by $S_\pi^{m,q}[0,1]$ the class of spline functions defined on Π. If Π is of regular mesh h, we shall denote $S_\pi^{m,q}[0,1]$ by $S_h^{m,q}[0,1]$.

Remarks

(i) When $q = m-1$, the definition is the same as that of Schoenberg in [50].

(ii) The above definition extends for $q = -1$. A spline $s \in S_h^{m,-1}$ is a piecewise polynomial with discontinuities at the knots.

(iii) It follows from the definition that $s^{(q+1)} \in L_2[0,1]$ and $s^{(q)}$ is absolutely continuous. Thus $s \in H^{q+1}[0,1]$.

1.10. Approximation by Splines

1.10a. Introduction

Polynomials have long been used to approximate functions, partly because they are simple and can be easily handled. However, evaluation of a high degree polynomial will not be that

simple; to interpolate a function f with a polynomial p of degree n at m points, $m \leq n+1$, we need to evaluate $n+1$ unknowns which are the coefficients of p. Due to the accuracy of the digital computer, for a fairly large n, the round off errors by the computer will be of considerable significance; also, when one fits a high degree polynomial to a large number of data points, the result is often rather undulated. There is now evidence that in many circumstances a spline function is a more adaptable approximating function than a polynomial involving a comparable number of parameters. It has been shown that a variety of problems of best approximation turn out to have as a solution spline function ([13], [51], [52]). Many other properties such as "minimum curvature" ([36]) and "smoothest interpolation" ([13], [32], [36], [51]) have been widely investigated. Spline functions have been used widely in smoothing datas ([45]), approximation of linear functions ([52]) and solving differential equations ([31], [59]). The theory of finite element method is one of the many successes of the spline functions in the application on solving boundary value problems.

In the next two sections, we shall discuss the spline spaces $S_h^{1,0}$ and $S_h^{3,2}$.

1.10.b. The Spline Spaces

1.10b.1 The space $S_h^{1,0}$

An $s \in S_h^{1,0}$ is a spline function which is continuous over [0,1] and reduces to a linear function in each interval $[x_{i-1},x_i]$, $i = 1,\cdots,n$.

The dimension of $S_h^{1,0}$ is n+1 . Thus $S_h^{1,0}$ has a basis $\{\phi_i\}_{i=0}^n$ of n+1 elements.

A basis $\{\phi_i\}_{i=0}^n$, shown in fig. 1.10.1 , is defined as

$$(1.10.1)\quad \phi_i(x) = \begin{cases} t & x \in [x_{i-1},x_i] \quad t = \dfrac{x-x_{i-1}}{h} \\ 1-t & x \in [x_i,x_{i+1}] \quad t = \dfrac{x_{i+1}-x}{h} \\ 0 & \text{otherwise} \end{cases}$$

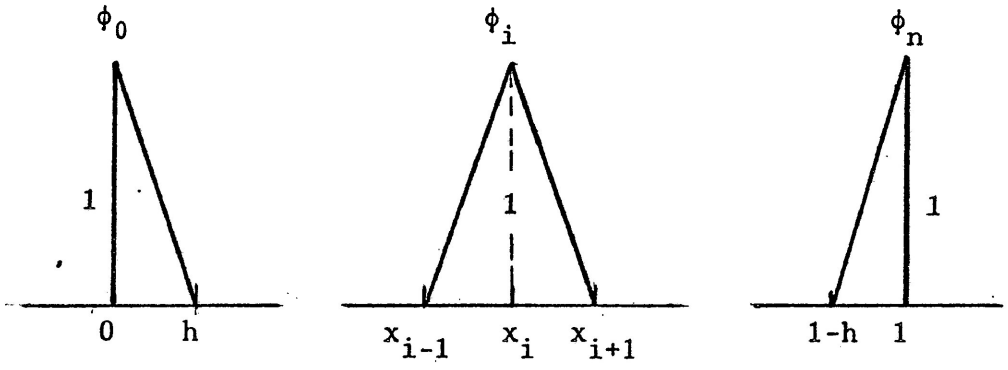

Fig. 1.10.1

This space of functions has been used for a long time as interpolation functions.

1.10b.2 The space $S_h^{3,2}$

An $s \in S_h^{3,2}$ is a cubic spline with continuous second derivative. This function has an important property that it is analogous to a curve drawn by the draftsman's spline. ([1], [50], [33]).

The dimension of $S_h^{3,2}$ is n+3. To construct a basis for $S_h^{3,2}$, we introduce four additional knots x_{-2}, x_{-1}, x_{n+1} and x_{n+2} such that $x_{-2} < x_{-1} < x_0 = 0$ and $1 = x_n < x_{n+1} < x_{n+2}$.

Define the functions $\{B_i\}_{i=1}^{n+1}$, called basis B-splines ([11], [17], [18], [51]) by

$$
\begin{aligned}
&\text{(i)} && B_i \in S_h^{3,2} \\
&\text{(ii)} && B_i \text{ is identical to zero outside } (x_{i-2}, x_{i+1}) \\
&\text{(iii)} && B_i(x_{i-2}) = B_i(x_{i+2}) = 0 \\
& && B_i(x_{i-1}) = B_i(x_{i+1}) = \tfrac{1}{4} \\
& && B_i(x_i) = 1
\end{aligned}
\tag{1.10.2}
$$

By the constraints in (iii), we have

$$
B_i(x) = \begin{cases}
\dfrac{t^3}{4} & x \in [x_{i-2}, x_{i-1}] \\
-\dfrac{1}{4}(3t^3 - 3t^2 - 3t - 1) & x \in [x_{i-1}, x_i] \\
1 - \dfrac{3}{2}t^2 + \dfrac{3}{4}t^3 & x \in [x_i, x_{i+1}] \\
\dfrac{(1-t)^3}{4} & x \in [x_{i+1}, x_{i+2}] \\
0 & \text{otherwise}
\end{cases}
\tag{1.10.3}
$$

where t, $0 \le t \le 1$, is a local variable defined by

$$(1.10.4) \qquad t = \frac{x - x_{k-1}}{x_k - x_{k-1}} \quad \text{for} \quad x \in [x_{k-1}, x_k]$$

The graph of $B_i(x)$ is shown in fig. 1.10.2

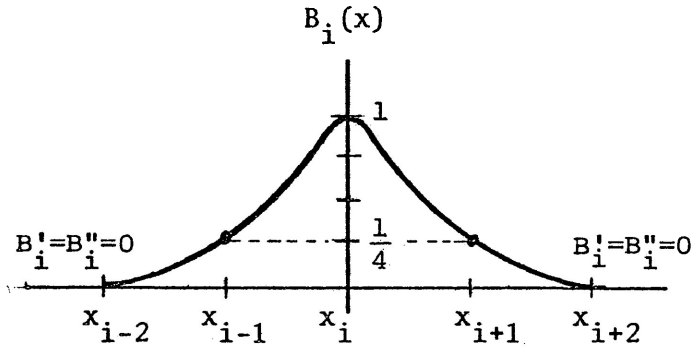

Fig. 1.10.2

1.10c. Quasi-Interpolation

Quasi-interpolation was first introduced by C. de Boor and G. Fix in [14] and was generalized by P.O. Frederickson in [25]. The problem can be stated as follows : for each $f \in C^k[0,1]$, let $F_\pi f : C^{k+1}[0,1] \to S_\pi^{k,k-1}$ such that $F_\pi f$, which called a *quasi-interpolant* of f, has the following properties :

(i). $F_\pi f$ is local in the sense that its value at a point x depends only on the values of f in a uniformly small neighbourhood of x .

(ii) F_π reproduces polynomials; $F_\pi(p) = p \quad \forall\, p \in \mathcal{P}^k$.

(iii) $F_\pi f - f = O(h^{k+1})$.

The explicit forms of $F_\pi f$ were given in [14] which are written in the linear combinations of the k-degree B-spline (A B-spline basis for $S^{k,k-1}$). As indicated in [25], quasi-interpolations have two strong advantages over interpolations. The first of these is ease of computation and the second advantage is that strong error estimates,very nearly sharp, are easy to obtain for almost any norm. The idea of (ii) is a source of motivation for the works in Chapter 3.

CHAPTER 2

THE FINITE ELEMENT SOLUTION

2.1. The Discrete Linear Systems due to the Piecewise Linear Approximation

We have shown in section 1.5 that the exact solution to (1.3.2) - (1.3.4) is equivalent to the finding of a $u^* \in H_g^{-1}[0,1]$ which minimizes $\Phi(v)$ over $H_g^1[0,1]$. The RRG approach in section 1.6 is to approximate u^* by $u_{h,g}^*$ from an affine subspace $S_{h,g}$ such that $\Phi(u_{h,g}^*)$ is minimum over $S_{h,g}$. In this section, we shall construct $u_{h,g}^*$ by using the affine subspace $S_{h,g}^{1,0}$. A set of trial functions is taken to be $\{\phi_i\}_{i=0}^n$ which has been defined in 1.10b. . From section 1.6, an element $u_{h,g} \in S_{h,g}^{1,0}$ can be written as $u_{h,g} = g + u_{h,0}$, where $g \in S_{h,g}^{1,0}$ and $u_{h,0} \in S_{h,0}^{1,0}$; since $g \in S_{h,g}^{1,0}$, it can be taken as

$$g = g_0\phi_0(x) + g_1\phi_n(x)$$

Thus

$$u_{h,g} = g_0\phi_0(x) + \sum_{i=1}^{n-1} a_i\phi_i(x) + g_1\phi_n(x) \tag{2.1.1}$$

The coefficients a_i are to be determined so that $\Phi(u_{h,g})$ is minimum. From (1.4.1), we have

$$\Phi(u_{h,g}) = \int_0^1 \{p(x)(\sum_{i=0}^n a_i\phi_i'(x))^2 + q(x)(\sum_{i=0}^n a_i\phi_i(x))^2 - 2(\sum_{i=0}^n a_i\phi_i(x))f(x)\}\,dx$$

$$= \sum_{k=0}^n \sum_{j=0}^n a_k a_j A_{k,j} + \sum_{k=0}^n \sum_{j=0}^n a_k a_j B_{k,j} - 2\sum_{j=0}^n a_j f_j$$

where $a_0 = g_0$, $a_n = g_1$ and

$$(2.1.3)\qquad \begin{cases} A_{k,j} = \int_0^1 p(x)\phi_k'(x)\phi_j'(x)\,dx \\ B_{k,j} = \int_0^1 q(x)\phi_k(x)\phi_j(x)\,dx \\ f_j = \int_0^1 f(x)\phi_j(x)\,dx \end{cases} \qquad k,j = 0,\cdots,n$$

(Note that $A_{k,j} = A_{j,k}$, $B_{k,j} = B_{j,k}$)

Thus (2.1.2) can be written as

$$(2.1.4)\qquad \Phi(u_{h,g}) = \sum_{k=0}^{n} a_k \Big(\sum_{j=0}^{n} a_j A_{k,j} + \sum_{j=0}^{n} a_j B_{k,j} - 2f_k \Big)$$

The variables are $a_1,\cdots,a_{n-1}$. Thus

$$(2.1.5)\qquad \Phi(u_{h,g}) = \Phi(a_1,\cdots,a_{n-1})$$

To determine a_i , $i = 1,\cdots,n-1$, such that $\Phi(u_{h,g})$ is minimum, we solve $\frac{\partial \Phi}{\partial a_i} = 0$, and obtain a system of linear equations :

$$(2.1.6)\qquad \sum_{k=0}^{n} a_k (A_{k,i}+B_{k,i}) = f_i \qquad i = 1,\cdots,n-1$$

Let $M_{k,j} = A_{k,j} + B_{k,j}$, then (2.1.6) becomes

$$(2.1.7)\qquad \sum_{k=0}^{n} a_k M_{k,i} = f_i \qquad i = 1,\cdots,n-1$$

Since $a_0 = g_0$ and $a_n = g_1$ are fixed, (2.1.7) is a linear system of $n - 1$ unknowns. It is

$$(2.1.8)\qquad \sum_{k=1}^{n-1} a_k M_{k,i} = f_i - g_0 M_{0,i} - g_n M_{n,i}$$

$$= F_i \qquad i = 1,\cdots,n-1$$

Let

$$(2.1.9)\qquad \begin{cases} M = (M_{k,i})^T \ (= (M_{i,k})) & 1 \le k,i \le n-1 \\ M_0 = (M_{0,1},\cdots,M_{0,n-1})^T \\ M_n = (M_{n,1},\cdots,M_{n,n-1})^T \\ a = (a_1,\cdots,a_{n-1})^T \\ f = (f_1,\cdots,f_{n-1})^T \\ F = (F_1,\cdots,F_{n-1})^T \\ \quad = [\, f - g_0 M_0 - g_n M_n \,]^T \end{cases}$$

Then (2.1.8) can be written as

$$(2.1.10)\qquad Ma = F$$

If the boundary conditions (1.3.4) were homogeneous, i.e. $g_0 = g_1 = 0$, then the linear system would be

$$(2.1.11)\qquad Ma = f$$

The matrix M is positive definite since for all a, there exists a $u \in H_0^1[0,1]$ such that $a^T Ma = (Lu,u) > 0$. It follows from (2.1.3) that M is symmetric.

The *finite element (RRG) solution* has been constructed if the

linear system (2.1.10) (or equivalently (2.1.11) for homogeneous boundary conditions) has been solved. The solving of (2.1.10) (or (2.1.11)) will be discussed in Chapter 5.

We shall evaluated M for constant p and q. From (2.1.3), we have

$$(2.1.12) \qquad \begin{cases} A_{i,j} = p(\phi_i', \phi_j') \\ B_{i,j} = q(\phi_i, \phi_j) \end{cases}$$

After some calculations, we have

$$(2.1.13) \qquad A_{i,j} = \frac{p}{h} \begin{cases} -1 & |j-i| = 1 \\ 2 & j = i \\ 0 & \text{otherwise} \end{cases}$$

$$(2.1.14) \qquad B_{i,j} = \frac{qh}{6} \begin{cases} 1 & |j-1| = 1 \\ 4 & j = i \\ 0 & \text{otherwise} \end{cases}$$

By using (2.1.13) and (2.1.14), the matrix M is of the form

$$M = \frac{p}{h} \begin{bmatrix} 2 & -1 & 0 & 0 & & \\ -1 & 2 & -1 & 0 & & \\ 0 & -1 & 2 & -1 & & \\ & & & & -1 & \\ & & & & -1 & 2 \end{bmatrix} + \frac{qh}{6} \begin{bmatrix} 4 & 1 & 0 & 0 & & \\ 1 & 4 & 1 & 0 & & \\ 0 & 1 & 4 & 1 & & \\ & & & & & 1 \\ & & & & 1 & 4 \end{bmatrix}$$

For the right-hand side vector F, we have

$$F_1 = f_1 - \left(-\frac{p}{h} + \frac{qh}{6}\right)g_0$$

$$F_i = f_i \qquad i = 2,\cdots,n-2$$

$$F_{n-1} = f_{n-1} - \left(-\frac{p}{h} + \frac{qh}{6}\right)g_n$$

For general $p(x)$ and $q(x)$, M is either obtained by exact integration or by quadrature formulae which we shall discuss in the subsequent sections.

2.2. Best Quadrature Formulae

2.2a. Introduction

In (2.1.3), there are integrations namely

$I_i(f) = \int_0^1 f(x)\phi_i(x)\,dx$, $I'_{i,j}(p) = \int_0^1 p(x)\phi'_i(x)\phi'_j(x)\,dx$ and $I_{i,j}(q) = \int_0^1 q(x)\phi_i(x)\phi_j(x)\,dx$. Beside performing the actual integrations, we could approximate these integrations numerically. Especially, when the analytical solution is not possible to obtain, then an approximation will be become necessary. There are several ways to approximate these integrations. For example, to approximate $I_i(f)$, we could interpolate f by a spline s and then integrate $I_i(s)$ exactly ([8],[34],[56]); or we could approximate $I_i(f)$ by a quadrature rule H_i of the form $H_i(f) = \sum_{j=0}^{N} a_j f(\xi_j)\phi_i(\xi_j)$. However, in the subsequent sections, we shall extend Sard's approach on

quadrature formulae to obtain *best quadratures* for a more general type of integrations, namely $T(g) = \int_0^1 g(x)w(x)\,dx$, where $w(x)$ is a weight function.

Let

$$T(g) = \int_0^1 g(x)w(x)\,dx \tag{2.2.1}$$

Consider a discrete approximation Q to T, or quadrature formulae, of the form

$$Q(g) = \sum_{i=0}^{N} a_i g(\xi_i) \tag{2.2.2}$$

If $w(x) \equiv 1$, $x \in [0,1]$, then (2.2.1) - (2.2.2) reduces to the case which considered in detail by A. Sard in [48] and [49]. Assume that Q is *m-exact* (exact for $\mathcal{P}^m$). Define the *error functional* $E : H^{m+1}[0,1] \to R$ by

$$E(g) = T(g) - Q(g) \tag{2.2.3}$$

From the assumption, $E(\mathcal{P}^m) = 0$. Thus we could apply the Sard kernel theorem that there exists a function K_{m+1} such that

$$E(g) = \int_0^1 K_{m+1}(t)g^{[m+1]}(t)\,dt \tag{2.2.4}$$

and

$$K_{m+1}(t) = \frac{1}{m!} E_x((x-t)_+^m) \tag{2.2.5}$$

$$= \frac{1}{m!}\left[T(x-t)_+^m - \sum_{i=0}^{N} a_i(\xi_i-t)_+^m \right]$$

Then by applying the Hölder inequality on (2.2.4), we have,

for $p,q \geq 1$ with $\frac{1}{p} + \frac{1}{q} = 1$,

$$(2.2.6) \qquad |E(g)| \leq \left\{\int_0^1 |K_{m+1}(t)|^q dt\right\}^{\frac{1}{q}} \left\{\int_0^1 |g^{[m+1]}(t)|^p dt\right\}^{\frac{1}{p}}$$

$$= ||K_{m+1}||_{L_q} \cdot ||g^{[m+1]}||_{L_p}$$

Equality holds for the function g with the property

$$(2.2.7) \qquad g^{[m+1]}(x) = \operatorname{sgn}(K_{m+1}(x)) |K_{m+1}(x)|^{\frac{q}{p}} \qquad \text{(a.e. for } p \geq 1)$$

In particular, there exists a function $g^* \in H^{m+1}[0,1]$ such that

$$(2.2.8) \qquad (g^{*[m+1]})^p = K_{m+1}^q$$

Thus from (2.2.6) and (2.2.7)

$$(2.2.9) \qquad |E(g^*)| = ||K_{m+1}||_{L_q} \cdot ||g^{*[m+1]}||_{L_p}$$

Define

$$(2.2.10) \qquad |||g|||_p = ||g^{[m+1]}||_{L_p}$$

Note that $|||g|||_p = 0$ if $g \in \mathcal{P}^m$, so we do not distinguish f and g if $f - g \in \mathcal{P}^m$, then $|||\cdot|||_p$ is a norm on $H^{m+1}[0,1]/\mathcal{P}^m$. (2.2.6) can he written as

$$(2.2.11) \qquad |E(g)| \leq ||K_{m+1}||_{L_q} \cdot |||g|||_p$$

Hence if $E : H^{m+1} \to R$ annihilates polynomials of degree m, then the L_q-norm of E is given by

$$||E||_q = ||K_{m+1}||_{L_q} \tag{2.2.12}$$

In particular,

(i) If $p = \infty$, $q = 1$, we have

$$\begin{aligned} |E(g)| &\leq ||K_{m+1}||_{L_1} \cdot |||g|||_\infty \\ ||E||_1 &= ||K_{m+1}||_{L_1} \end{aligned} \tag{2.2.13}$$

(ii) If $p = 2$, $q = 2$, we have

$$\begin{aligned} |E(g)| &\leq ||K_{m+1}||_{L_2} \cdot |||g|||_2 \\ ||E||_2 &= ||K_{m+1}||_{L_2} \end{aligned} \tag{2.2.14}$$

Different quadratures have different Kernel functions for the error functionals. We shall denote by E_Q if the dependency of E on Q is to be emphasized. Following [48] and [49], we have the following definition:

Definition 2.2.1 Let $\mathcal{Q}(N,m,\{\xi_i\}_{i=1}^N)$ (abbrev. $\mathcal{Q}$) be the class of *m*-exact quadrature Q to T of the form (2.2.2). $Q^b \in \mathcal{Q}$ is called the *best* quadrature (w.r.t. $\mathcal{Q}$) if

$$||E_{Q^b}||_2 = \min_{Q \in \mathcal{Q}} ||E_Q||_2 \tag{2.2.15}$$

The subsequent sections are devoted to the evaluation of the best quadratures for I_i, $I'_{i,j}$ and $I_{i,j}$

2.2b. Best Quadrature for $\int_0^1 f(x)\phi_i(x)\,dx$

Let $I_i(f) = \int_0^1 f(x)\phi_i(x)\,dx$, $1 \le i \le n-1$, where ϕ_i , $1 \le i \le n-1$, is defined in (1.10.1). Since $\phi_k(x) = \phi(x-kh)$, where

$$\phi(x) = \begin{cases} t & x \in [-h,0] \quad t = \frac{x+h}{h} \\ 1-t & x \in [0,h] \quad t = \frac{h-x}{h} \end{cases}$$

we only need to evaluate one quadrature formula for $I_0(f) = \int_{-h}^{h} f(x)\phi(x)\,dx$; this quadrature applies to I_i , $\forall$ i .

Notation :

(i) $Q_0(N,m,\{\xi_i\}_{i=1}^N)$: the class of m-exact quadrature Q_0 of I_0 of the form (2.2.2).

(ii) Q_0^b : the *best* quadrature formula w.r.t. $Q_0(N,m,\{\xi_i\}_{i=1}^N)$.

The following is a list of quadrature formulae which are *best* in each class of quadratures. The bounds for the error functionals are computed via the Peano-Sard kernel theorem.

The best quadrature in the class $Q_0(1,1,\{0\})$ is

$$Q_0^b(f) = h \cdot f(0)$$

If $f \in H^2[0,1]$, then

$$|E(f)| \le 0.0833 \cdot h^3 \cdot \|f''\|_{L_\infty[0,1]}$$

and

$$|E(f)| \le 0.0891 \cdot h^{2.5} \cdot \|f''\|_{L_2[0,1]}$$

The best quadrature in the class $Q_0(2,1,\{-\frac{h}{2},\frac{h}{2}\})$ is

$$Q_0^b(f) = \frac{h}{2}\cdot f(-\frac{h}{2}) + \frac{h}{2}\cdot f(\frac{h}{2})$$

If $f \in H^2[0,1]$, then

$$|E(f)| \le 0.0542\cdot h^3\cdot ||f''||_{L_\infty[0,1]}$$

and

$$|E(f)| \le 0.0570\cdot h^{2.5}\cdot ||f''||_{L_2[0,1]}$$

The best quadrature in the class $Q_0(2,3,\{-\frac{h}{\sqrt{6}},\frac{h}{\sqrt{6}}\})$ is

$$Q_0^b(f) = \frac{h}{2}\cdot f(-\frac{h}{\sqrt{6}}) + \frac{h}{2}\cdot f(\frac{h}{\sqrt{6}})$$

If $f \in H^2[0,1]$, then

$$|E(f)| \le 0.0282\cdot h^3\cdot ||f''||_{L_\infty[0,1]}$$

and

$$|E(f)| \le 0.0267\cdot h^{2.5}\cdot ||f''||_{L_2[0,1]}$$

If $f \in H^4[0,1]$, then

$$|E(f)| \le 0.00162\cdot h^5\cdot ||f^{[4]}||_{L_\infty[0,1]}$$

and

$$|E(f)| \le 0.00178\cdot h^{4.5}\cdot ||f^{[4]}||_{L_2[0,1]}$$

This is an example of a *Gaussian type best quadrature* formula . It shows that there exist two-point best quadratures which are exact for $\mathcal{P}^3$. An disadvantage of this quadrature is that we have to place the weights at $\pm h/\sqrt{6}$. Another way to obtain higher order quadratures

is to place the weights at more points.

The best quadrature in the class $Q_0(3,3,\{-h,0,h\})$ is

$$Q_0^b(f) = \frac{h}{12}\cdot f(-h) + \frac{10h}{12}\cdot f(0) + \frac{h}{12}\cdot f(h)$$

If $f \in H^2[0,1]$, then

$$|E(f)| \leq 0.0409\cdot h^3\cdot ||f''||_{L_\infty[0,1]}$$

and

$$|E(f)| \leq 0.0381\cdot h^{2.5}\cdot ||f''||_{L_2[0,1]}$$

If $f \in H^4[0,1]$, then

$$|E(f)| \leq 0.00417\cdot h^5\cdot ||f^{[4]}||_{L_\infty[0,1]}$$

and

$$|E(f)| \leq 0.00404\cdot h^{4.5}\cdot ||f^{[4]}||_{L_2[0,1]}$$

Herbold, Schultz and Varga [34] obtained the same quadrature by interpolating f by a quadratic (Lagrange) polynomial p and then integrating $I_0(p)$ exactly.

The best quadrature in the class $Q_0(3,3,\{-\frac{h}{2},0,\frac{h}{2}\})$ is

$$Q^b(f) = \frac{h}{3}\cdot f(-\tfrac{h}{2}) + \frac{h}{3}\cdot f(0) + \frac{h}{3}\cdot f(\tfrac{h}{2})$$

If $f \in H^2[0,1]$, then

$$|E(f)| \leq 0.0151\cdot h^3\cdot ||f''||_{L_\infty[0,1]}$$

and

$$|E(f)| \leq 0.0135\cdot h^{2.5}\cdot ||f''||_{L_2[0,1]}$$

If $f \in H^4[0,1]$, then

$$|E(f)| \leq 0.00104 \cdot h^5 \cdot ||f^{[4]}||_{L_\infty[0,1]}$$

$$|E(f)| \leq 0.00107 \cdot h^{2.5} \cdot ||f^{[4]}||_{L_2[0,1]}$$

2.2c. Best Quadrature for $\int_0^1 p(x)\phi_i'(x)\phi_j'(x)\,dx$

In (2.2.2), if we take $w(x) = \phi_i'(x)\phi_j'(x)$, then we have $I'_{i,j}(p) = \int_0^1 p(x)\phi_i'(x)\phi_j'(x)\,dx$.

We need only to consider two cases : (i) $|j-i| = 1$ and (ii) $j = i$, since $I'_{i,j}(p) = 0 \quad \forall\, p \in H^k[0,1]$, $\forall\, k \geq 0$ if $|j-i| > 1$.

2.2c.1. The Case $|j-i| = 1$, $1 \leq i,j \leq n-1$

For $j = i-1$, we have

$$I'_{i,i-1}(p) = -\frac{1}{h^2}\int_{(i-1)h}^{ih} p(x)\,dx \tag{2.2.16}$$

This case has been considered in detail in Sard [49].

The quadrature for $I'_{1,0}$ is valid for $I'_{i,i-1}$, $i = 2, \cdots, n-1$ (cf. 2.2b.) , and for the case when $j = i+1$, since $I'_{i-1,i} = I'_{i,i-1}$, the same quadrature applies, so we shall evaluate the quadrature for $I'_{1,0}$

Notation :

(i) $Q'_{1,0}(N,m,\{\xi_i\}_{i=1}^N)$: the class of m-exact quadratures $Q'_{1,0}$ of the form (2.2.2) of $I'_{1,0}$.

(ii) $Q'^{b}_{1,0}$: the *best* quadrature w.r.t. $Q'_{1,0}(N,m,\{\xi_i\}_{i=1}^N)$.

The best quadrature in $Q'_{1,0}(3,3,\{0,\frac{h}{2},h\})$ is the Simpson's rule :

$$Q'_{1,0}(p) = -\frac{1}{6h}\cdot p(0) - \frac{4}{6h}\cdot p(\tfrac{h}{2}) - \frac{1}{6h}\cdot p(h)$$

If $f \in H^2[0,1]$, then

$$|E(f)| \le 0.0123\cdot h\cdot ||f''||_{L_\infty[0,1]}$$

and

$$|E(f)| \le 0.0152\cdot h^{0.5}\cdot ||f''||_{L_2[0,1]}$$

If $f \in H^4[0,1]$, then

$$|E(f)| \le 0.000347\cdot h^3\cdot ||f^{[4]}||_{L_\infty[0,1]}$$

and

$$|E(f)| \le 0.000464\cdot h^{2.5}\cdot ||f^{[4]}||_{L_2[0,1]}$$

2.2c.2. The Case $j = i$, $1 \le i,j \le n-1$

For $j = i$, we have

$$I'_{i,i} = \frac{1}{h^2}\int_{(i-1)h}^{(i+1)h} p(x)\,dx$$

For notational simplicity, we shall evaluate the best quadratures by considering

$$I'_{0,0}(p) = \frac{1}{h^2}\int_{-h}^{h} p(x)\, dx$$

(This case can be found in Sard [49]).

Notation :

(i) $Q'_{0,0}(N,m,\{\xi_i\}_{i=1}^N)$: the class of m-exact quadrature $Q'_{0,0}$ of the form (2.2.2) of $I'_{0,0}$.

(ii) $Q'^{b}_{0,0}$: the *best* quadrature w.r.t. $Q'_{0,0}(N,m,\{\xi_i\}_{i=1}^N)$.

The best quadrature in $Q'_{0,0}(3,3,\{-h,0,h\})$ is the Simpson's rule :

$$Q'_{0,0}(p) = \frac{1}{3h}p(-h) + \frac{4}{3h}p(0) + \frac{1}{3h}p(h)$$

If $f \in H^2[0,1]$, then

$$|E(p)| \le 0.0987\cdot h\cdot ||p''||_{L_\infty[0,1]}$$

and

$$|E(p)| \le 0.0861\cdot h^{0.5}\cdot ||p''||_{L_2[0,1]}$$

If $f \in H^4[0,1]$, then

$$|E(p)| \le 0.0111\cdot h^3\cdot ||p^{[4]}||_{L_\infty[0,1]}$$

and

$$|E(p)| \le 0.0105\cdot h^{2.5}\cdot ||p^{[4]}||_{L_2[0,1]}$$

2.2d. Best Quadrature for $\int_0^1 q(x)\phi_i(x)\phi_j(x)\,dx$

If we take $w(x) = \phi_i(x)\phi_j(x)$ in (2.2.1), then we have $I_{i,j}(q) = \int_0^1 q(x)\phi_i(x)\phi_j(x)\,dx$; we need only to consider two cases : (i) $|j-i| = 1$ and (ii) $j = i$.

2.2d.1 The Case $|j-i| = 1$, $1 \le i,j \le n-1$

In this case , we need only to evaluate the quadratures for $I_{1,0}$, which is

$$I_{1,0}(q) = \frac{1}{h^2}\int_0^h q(x)\cdot(h-x)\cdot x\,dx$$

Notation:

(i) $Q_{1,0}(N,m,\{\xi_i\}_{i=1}^N)$ (abbrev. $Q_{1,0}$) : the class of quadratures m-exact $Q_{1,0}$ of the form (2.2.2) of $I_{1,0}$.

(ii) $Q_{1,0}^b$: the *best* quadrature w.r.t. $Q_{1,0}(N,m,\{\xi_i\}_{i=1}^N)$.

The following is a list of *best* quadratures in each class $Q_{1,0}$:

The best quadrature in the class $Q_{1,0}(1,1,\{\frac{h}{2}\})$ is

$$Q_{1,0}^b(q) = \frac{h}{6}\cdot q(\tfrac{h}{2})$$

If $q \in H^2[0,1]$, then

$$|E(q)| \leq 0.00417 \cdot h^2 \cdot ||q''||_{L_\infty[0,1]}$$

and

$$|E(q)| \leq 0.00616 \cdot h^{1.5} \cdot ||q''||_{L_2[0,1]}$$

The best quadrature in the class $Q_{1,0}(3,3,\{0,\frac{h}{2},h\})$ is

$$Q^b_{1,0}(q) = \frac{h}{60} \cdot q(0) + \frac{8h}{60} \cdot q(\frac{h}{2}) + \frac{h}{60} \cdot q(h)$$

If $q \in H^2[0,1]$, then

$$|E(q)| \leq 0.00189 \cdot h^3 \cdot ||q''||_{L_\infty[0,1]}$$

and

$$|E(q)| \leq 0.00246 \cdot h^{2.5} \cdot ||q''||_{L_2[0,1]}$$

If $q \in H^4[0,1]$, then

$$|E(q)| \leq 0.0000496 \cdot h^5 \cdot ||q^{[4]}||_{L_\infty[0,1]}$$

and

$$|E(q)| \leq 0.0000676 \cdot h^{4.5} \cdot ||q^{[4]}||_{L_2[0,1]}$$

2.2d.2. The Case $j = i$, $1 \leq i,j \leq n-1$

As before, for notational simplicity, we shall consider

$$I_{0,0}(q) = \frac{1}{h} \int_{-h}^{0} q(x) \cdot (x+h)\ dx + \frac{1}{h} \int_{0}^{h} q(x) \cdot (h-x)\ dx$$

Notation :

(i) $Q_{0,0}(N,m,\{\xi_i\}_{i=1}^N)$ (abbrev. $Q_{0,0}$) : the class of m-exact quadrature $Q_{0,0}$ of the form (2.2.2) of $I_{0,0}$.

(ii) $Q_{0,0}^b$: the *best* quadrature in $Q_{0,0}(N,m,\{\xi_i\}_{i=1}^N)$.

The following is a list of best quadratures in each class $Q_{0,0}$.

The best quadrature in the class $Q_{0,0}(1,1,\{0\})$ is

$$Q_{0,0}^b(q) = \frac{2h}{3}\cdot q(0)$$

If $q \in H^2[0,1]$, then

$$|E(q)| \le 0.0334\cdot h^3\cdot ||q''||_{L_\infty[0,1]}$$

and

$$|E(q)| \le 0.0323\cdot h^{2.5}\cdot ||q''||_{L_2[0,1]}$$

The best quadrature in the class $Q_{0,0}(3,3,\{-\frac{h}{2},0,\frac{h}{2}\})$ is

$$Q_{0,0}^b(q) = \frac{2h}{15}\cdot q(-\frac{h}{2}) + \frac{6h}{15}\cdot q(0) + \frac{2h}{15}\cdot q(\frac{h}{2})$$

If $q \in H^2[0,1]$, then

$$|E(q)| \le 0.0223\cdot h^3\cdot ||q''||_{L_\infty[0,1]}$$

and

$$|E(q)| \le 0.0246\cdot h^{2.5}\cdot ||q''||_{L_2[0,1]}$$

If $q \in H^4[0,1]$, then

$$|E(q)| \leq 0.0000993 \cdot h^5 \cdot \|q^{[4]}\|_{L_\infty[0,1]}$$

and

$$|E(q)| \leq 0.0001050 \cdot h^{4.5} \cdot \|q^{[4]}\|_{L_2[0,1]}$$

The best quadrature in the class $Q_{0,0}(3,3,\{-h,0,h\})$ is

$$Q^b_{0,0}(q) = \frac{h}{30} \cdot q(-h) + \frac{18h}{30} \cdot q(0) + \frac{h}{30} \cdot q(h)$$

If $q \in H^2[0,1]$, then

$$|E(q)| \leq 0.0219 \cdot h^3 \cdot \|q''\|_{L_\infty[0,1]}$$

and

$$|E(q)| \leq 0.0208 \cdot h^{2.5} \cdot \|q''\|_{L_2[0,1]}$$

If $q \in H^4[0,1]$, then

$$|E(q)| \leq 0.00199 \cdot h^5 \cdot \|q^{[4]}\|_{L_\infty[0,1]}$$

and

$$|E(q)| \leq 0.00196 \cdot h^{4.5} \cdot \|q^{[4]}\|_{L_2[0,1]}$$

We shall end this Chapter by concluding that, as far as we know, the best quadrature formulae in Section 2.2b, except the best quadrature in $Q_0(3,3,\{-h,0,h\})$, and Section 2.2d are new.

CHAPTER 3

SUPERCONVERGENCE

3.1. Introduction

In [16], C. de Boor and B. Swartz showed that in solving a certain BVP by a certain collocation method, the error at the knots of the spline being employed is of much higher order than it can be either uniformly or in L_2. J. Douglas and T. Dupont in [19] discovered the knot higher order phenomenon does occur when they approximated certain TPBVP by the Galerkin method using the space $S_\pi^{m,0}$; they used the word "*Superconvergence*" to represent this knot higher order phenomenon. In [20], they demonstrated several methods to obtain higher order convergence and extended the meaning of superconvergence to include methods of obtaining higher order convergence. The characteristic of their methods in [20] is *local* in the sense that superconvergence results at a certain small number of subintervals. However, motivated by [20], together with [14] and [25], we shall introduce a method called "*global superconvergence via local quasi-inverse*" in section 3.3 to obtain higher order convergence for the solution of (1.3.3) with homogeneous (1.3.4).

For the sake of simplicity, in this chapter, we shall restrict our discussion on the solution of (1.3.3) with homogeneous (1.3.4).

3.2. The Superconvergence Phenomenon at the Knots

The solution of (1.3.3) with homogeneous (1.3.4) by the

Ritz method has been shown in section 1.5 and 1.6 to be equivalent to the Galerkin solution. In this section, we shall modify the proof in [21] to show the superconvergence phenomenon at the knots for the RRG solution of (1.3.3) with homogeneous (1.3.4).

Let u be the true solution of (1.3.3) with homogenous (1.3.4). Let $u_h \in S_{h,0}^{m,0}$ be the RRG solution to u in the sense that

$$(3.2.1) \quad (pu_h',v_h') + (qu_h,v_h) = (f,v_h) \quad \forall\, v_h \in S_{h,0}^{m,0}$$

Let $\zeta = u-u_h$, then

$$(3.2.2) \quad (p\zeta',v_h') + (q\zeta,v_h) = 0 \quad \forall\, v_h \in S_{h,0}^{m,0}$$

Let $G(x,\xi)$ be the Green's function for (1.3.3) with homogeneous (1.3.4), i.e.

$$(3.2.3) \qquad \begin{aligned} u(x) &= (f,G(x,\cdot)) \\ &= (Lu,G(x,\cdot)) \\ &= (pu',\frac{\partial G}{\partial \xi}(x,\cdot)) + (qu,G(x,\cdot)) \end{aligned}$$

for sufficiently smooth u. In particular, the representation holds for $u \in H_0^1[0,1]$, so that it can be applied to ζ . Thus

$$(3.2.4) \quad \begin{aligned} \zeta(x_i) &= (p\zeta',\frac{\partial G}{\partial \xi}(x_i,\cdot)) + (q\zeta,G(x_i,\cdot)) \\ &= (p\zeta',\frac{\partial G}{\partial \xi}(x_i,\cdot)-v_h') + (q\zeta,G(x_i,\cdot)-v_h) \end{aligned}$$

for all $v_h \in S_{h,0}^{m,0}$ and

$$(3.2.5) \qquad |\zeta(x_i)| \le C \cdot \|\zeta\|_1 \cdot \inf_{v_h \in S^{m,0}_{h,0}} \|G(x_i,\cdot)-v_h\|_1$$

where $C = C(p,q)$

If $p \in H^{m+2}[0,1]$ and $q \in H^{m+1}[0,1]$, then

$$(3.2.6) \qquad G(x_i,\cdot) \in H^{m+1}([0,x_i]) \cap H^{m+1}([x_i,1])$$

and

$$(3.2.7) \qquad \|G(x_i,\cdot)\|_{H^{m+1}[0,x_i]} + \|G(x_i,\cdot)\|_{H^{m+1}[x_i,1]} \le C_1(p,q)$$

Since the functions in $S^{m,0}_{h,0}$ are not required to be differentiable at $x = x_i$, it is clear (by using Lagrange interpolation at each interval) that

$$(3.2.8) \qquad \inf_{v_h \in S^{m,0}_{h,0}} \|G(x,\cdot)-v_h\|_1 \le C_2(p,q)\cdot h^m$$

It is known that ([8],[20],[21],[41])

$$(3.2.9) \qquad \|\zeta\|_0 + h\cdot\|\zeta\|_1 \le C_3 \cdot \|u\|_{k+1}\cdot h^{k+1} \qquad 0 \le k \le m$$

From (3.2.5),(3.2.8) and (3.2.9)

$$(3.2.10) \qquad |\zeta(x_i)| \le C\ (p,q)\cdot \|u\|_{k+1}\cdot h^{m+k} \qquad 0 \le k \le m$$

Thus we have proved the following theorem:

Theorem 3.2.1 If the solution of (1.3.3) with homogeneous (1.3.4) is such that $u \in H^m_0[0,1]$, then the knot estimate (3.2.10) is valid, $i = 1,\cdots, n-1$.

Remarks:

(i) If the coefficients $p(x)$ and $q(x)$ are not smooth enough for (3.2.7) to hold but are such that, for some $j \in [0,m]$, $G(x_i,\cdot) \in H^{j+1}([0,x_i]) \cap H^{j+1}([x_i,1])$, then

$$|\zeta(x_i)| < C_5(p,q)\cdot \|u\|_{k+1}\cdot h^{j+k} \qquad 0 \le k \le m \tag{3.2.11}$$

(ii) In particular, if $k = m$, then we have

$$|\zeta(x_i)| \le C_6(p,q)\cdot \|u\|_{m+1}\cdot h^{2m}$$

for $i = 1, \cdots, n-1$.

(iii) It is worthwhile to mention that Wheeler in [60] made use of (3.2.10) in showing that

$$\|\zeta\|_{L_\infty[0,1]} \le C_7(p,q)\cdot \|u\|_{W^{\infty,m+1}[0,1]}\cdot h^{m+1} \tag{3.2.12}$$

where $$\|f\|_{W^{p,k}[0,1]} = \sum_{i=0}^{k} \|f^{[i]}\|_{L^p[0,1]}$$

$$W^{p,k}[0,1] = \left\{ u \in C^{k-1}[0,1] \;\middle|\; \begin{array}{l} u^{[k-1]} \text{ is absolutely} \\ \text{continuous and } u^{[k]} \in L_p[0,1] \end{array} \right\}$$

(see a similar result by Douglas, Dupont and Wahbin in [22]).

(iv) In particular, if $m = 1$, i.e. $u_h \in S^{1,0}_{h,0}$, then we have

$$|\zeta(x_i)| \le C_8(p,q)\cdot \|u\|_2 \cdot h^2 \qquad 1 \le i \le n-1$$

$$\|\zeta\|_{L_\infty[0,1]} \le C_9(p,q)\cdot \|u\|_{W^{\infty,2}[0,1]}\cdot h^2$$

and $$\|\zeta\|_0 + h\|\zeta\|_1 \le C_{10}\cdot \|u\|_k \cdot h^k \qquad k = 1,2$$

(v) If $p = 1$ and $q = 0$, the Green's function $G(x,\cdot) \in S^{1,0}_{h,0} \subset S^{m,0}_{h,0}$, hence $\inf_{v_h \in S^{m,0}_{h,0}} \|G(x,\cdot)-v_h\|_1 = 0$, which implies, (3.2.5), $\zeta(x_i) = 0$, $1 \le i \le n-1$.

3.3. Global Superconvergence via Local Quasi-Inverse

In this section, we shall introduce a constructive method to obtain a "*global superconvergence*" solution to (1.3.3) with homogeneous (1.3.4) in the sense that the global error is of higher order then the RRG solution in either the energy norm or H_0 norm.

The outline of the method is as follows :

Firstly, we solve (1.3.3) - homogeneous (1.3.4) by RRG (finite element) method using the space $S^{1,0}_{h,0}$ to obtain the RRG approximation u_h . We shall write the RRG method in an operational form by R_h . Thus $R_h : H^1_0[0,1] \to S^{1,0}_{h,0}$ such that $u_h = R_h(u)$. Note that u_h is a linear function of u .

As soon as we know u_h , it is nice to know the inverse of R_h , then we can compute u exactly. But, unfortunately, this is impossible in general. However, motivated by C. de Boor and G. Fix in [14] and P.O. Frederickson in [25], we know that we could make use of an approximate inverse (for more about an approximate inverse of a linear operator, please refer to Benson [3]) of R_h to obtain a better approximate solution s of u . An easy type of an approximate inverse to compute is what we shall call a *quasi-inverse* Q of R_h ; thus we define $s = Q \cdot R_h(u)$.

In section 4.5 , we will obtain error estimates for the error operator $E = I - Q \cdot R_h$.

Definition 3.3.1 ([14,pp.19], [25,pp.159])

An approximation operator $H : A \subset H^k[0,1] \to B \subset H^k[0,1]$ is called a *local approximation operator* if $(Hf)(x)$ is independent of the function f outside a certain neighborhood of the point x .

To be more precise, we require a compact set K such that for any f in A and any x in $[0,1]$, $[f(y) = 0 \quad \forall\, y \in x+K \Rightarrow (Hf)(x) = 0]$.

Definition 3.3.2

Let $\psi : A \subset H^k[0,1] \to B \subset H^k[0,1]$ be a linear operator. $Q : B \to C \subset H^k[0,1]$, a linear operator, is a *r-exact quasi-inverse*

of ψ if (i) Q is a local approximation operator and (ii) $Q\cdot\psi(p) = p \quad \forall\ p \in \mathcal{P}^r[0,1]$.

In this thesis, we shall consider in particular, $A = H_0^1[0,1]$, $B = S_{h,0}^{1,0}$, $C = S_{h,0}^{3,2}$, $r=3$ and $\psi = R_h$. i.e. we shall make use of a 3-exact quasi-inverse of the RRG method to obtain a superconvergence approximation $s \in S_{h,0}^{3,2}$ to u.

Let us go on to the construction of the superconvergence solution. From (1.6.9),

$$(3.3.1) \qquad a(u-u_h,w_h) = 0 \quad \forall\ w_h \in S_{h,0}^{1,0}$$

$$\Rightarrow\ a(u-u_h,\phi_j) = 0 \quad j = 1,\cdots,n-1$$

$$\Rightarrow\ a(u,\phi_j) = a(u_h,\phi_j) = \sum_{i=1}^{n-1} a_i\cdot a(\phi_i,\phi_j)$$

where $u_h = \sum_{i=1}^{n-1} a_i\phi_i$ with a_i which have been determined earlier by R_h.

$Q(u_h)$ is in $S_h^{3,2}$, it can be written as $Q(u_h) = \sum_{k=-1}^{n+1} b_k B_k$, where $\{B_k\}_{k=-1}^{n+1}$ is a basis of $S_h^{3,2}$ (cf. 1.10b.2). We want $Q\cdot R_h$ to reproduce p for $p \in \mathcal{P}_0^3$. By the *local* property of B_k in $[x_{k-2},x_{k+2}]$ (B_k is identical to zero outside (x_{k-2},x_{k+2})), we could have Q to be a local approximation operator by defining

$$(3.3.2) \qquad b_k = \alpha_k u_{h_{k-1}} + \beta_k u_{h_k} + \alpha_k u_{h_{k+1}}$$

where $u_{h_i} = u_h(x_i)$. Note that

$$(3.3.3) \quad Q(u_h)(x_j) = \frac{1}{4}b_{j-1} + b_j + \frac{1}{4}b_{j+1} \quad (= a_j \text{ if } u \in \mathcal{P}_0^3)$$

for $j = 0,\cdots,n-1$

$\mathcal{P}_0^3$ is of dimension 2, and $\{u^{00}(x) = x(1-x)\ ,\ u^{01}(x) = x^2(1-x)\}$ form a basis for $\mathcal{P}_0^3$. $Q\cdot R_h$ will reproduce $p \in \mathcal{P}_0^3$ if α_k and β_k are determined through :

$$(3.3.4) \quad \begin{cases} Q\cdot R_h(u^{00})(x_k) = u^{00}(x_k) \\ \\ Q\cdot R_h(u^{01})(x_k) = u^{01}(x_k) \end{cases}$$

where

$$(3.3.5) \quad \begin{cases} u^{00}(x_k) = kh(1-kh) \\ \\ u^{01}(x_k) = k^2h^2(1-kh) \end{cases}$$

After we have obtained the RRG solutions $\{a_k^{00}\}$ and $\{a_k^{01}\}$, $1 \le k \le n-1$, the next step is to evaluate b_k^{0t}, $-1 \le k \le n+1$, $t = 0,1$, and $\{\alpha_k,\beta_k\}_{k=-1}^{n+1}$. These determine $\{b_k\}_{k=-1}^{n+1}$, and hence s is obtained. From (3.3.3), (3.3.4) and (3.3.5), we have

$$(3.3.6) \quad \begin{cases} Qu_h^{00}(x_k) = \frac{1}{4}b_{k-1}^{00} + b_k^{00} + \frac{1}{4}b_{k+1}^{00} = kh(1-kh) \\ \\ Qu_h^{01}(x_k) = \frac{1}{4}b_{k-1}^{01} + b_k^{01} + \frac{1}{4}b_{k+1}^{01} = k^2h^2(1-kh) \end{cases}$$

A solution of (3.3.6) is

$$(3.3.7)\qquad \begin{cases} b_k^{00} = \frac{2}{3}kh(1-kh) + \frac{2h^2}{9} \\ b_k^{01} = \frac{2}{3}k^2h^2(1-kh) + \frac{2}{3}kh^3 - \frac{2}{9}h^2 \end{cases} \qquad -1 \le k \le n+1$$

From (3.3.2), $\{\alpha_k,\beta_k\}_{k=1}^{n-1}$ are obtained by solving the linear systems :

$$(3.3.8)\qquad \alpha_k u^{0t}_{h_{k-1}} + \beta_k u^{0t}_{h_k} + \alpha_k u^{0t}_{h_{k+1}} = b_k^{0t} \qquad t = 0,1$$

For $k = -1,0,n,n+1$, we need extrapolations as follows :

$$(3.3.9)\qquad \gamma_k u^{0t}_{h_0} + \beta_k u^{0t}_{h_1} + \alpha_k u^{0t}_{h_2} = b_k^{0t} \qquad k = -1,0 \quad t = 0,1$$

$$(3.3.10)\qquad \alpha_k u^{0t}_{h_{n-2}} + \beta_k u^{0t}_{h_{n-1}} + \gamma_k u^{0t}_{h_n} = b_k^{0t} \qquad k = n,n+1 \quad t = 0,1$$

Note that $u^{00}_{h_0} = u^{00}_{h_n} = u^{01}_{h_0} = u^{01}_{h_n} = 0$, so we need only to evaluate $\{\alpha_k,\beta_k\}$ for $k = -1,0,n,n+1$. The solutions of (3.3.8) - (3.3.10) can be easily obtained by Gaussian elimination method.

For p a constant and $q = 0$, $\{\alpha_k,\beta_k\}_{k=-1}^{n+1}$ can be evaluated by pen and paper calculation. The evaluation is as follows :

From (2.2.12) we have $a(\phi_i,\phi_j) = \frac{p}{h}A$, where A is defined in (2.2.13) . The RRG solution for u^{00} and u^{01} can be derived from (3.3.1) with A :

(i) <u>For $u^{00}(x) = x(1-x)$</u>

From (3.3.1) , $a(u^{00},\phi_j) = \frac{p}{h}[-a^{00}_{j-1} + 2a^{00}_j - a^{00}_{j+1}]$, but from

the definition, $a(u^{00},\phi_j) = 2ph$, $j = 1,\cdots,n-1$. The solution is

$$a_k^{00} = kh(1-kh) \qquad 1 \le k \le n-1 \tag{3.3.11}$$

This fits Remark (v) in 2.2 .

(ii) For $u^{01}(x) = x^2(1-x)$

Similar to (i), we have $-2ph + 6ph^2j^2 = \frac{p}{h}[-a_{j-1}^{01} + 2a_j^{01} + a_{j+1}^{01}]$. The solution is

$$a_k^{01} = k^2h^2(1-kh) \qquad 1 \le k \le n-1 \tag{3.3.12}$$

This fits Remark (v) in 2.2 too.

With (3.3.11) and (3.3.12), the solutions of (3.3.7) - (3.3.10) by Gaussian elimination method are :

$$\alpha_{-1} = \frac{1}{9}\left(\frac{5+2h}{1-2h}\right) \qquad \beta_{-1} = -\frac{8}{9}\left(\frac{2+h}{1-h}\right) \qquad \gamma_{-1} = 0$$

$$\alpha_0 = -\frac{1}{9}\left(\frac{1+h}{1-2h}\right) \qquad \beta_0 = \frac{2}{9}\left(\frac{1+2h}{1-h}\right) \qquad \gamma_0 = 0$$

$$\alpha_k = -\frac{1}{9} \qquad \beta_k = \frac{8}{9} \qquad 2 \le k \le n-1$$

$$\alpha_n = \alpha_0 \qquad \beta_n = \beta_0 \qquad \gamma_n = \gamma_0$$

$$\alpha_{n+1} = \alpha_{-1} \qquad \beta_{n+1} = \beta_{-1} \qquad \gamma_{n+1} = \gamma_{-1}$$

Remarks :

(i) For general $p(x)$ and $q(x)$, the cofficients $\{\alpha_k,\beta_k\}_{k=-1}^{n+1}$ can be solved numerically in a similar way by Gaussian elimination method.

(ii) For p constant and $q = 0$, the extrapolation for $\{\alpha_k,\beta_k\}$, $k = -1,0,n,n+1$, can be evaluated in an alternate way as follows :

For $k = -1,0$, we let

$$(3.3.13) \qquad b_k^{0t} = \alpha_k u_{h_0}^{0t} + \beta_k u_{h_1}^{0t} + \gamma_k u_{h_2}^{0t} + \delta_k u_{h_3}^{0t} \qquad t = 0,1$$

and for $k = n,n+1$, we let

$$(3.3.14) \qquad b_k^{0t} = \delta_k u_{h_{n-3}}^{0t} + \gamma_k u_{h_{n-2}}^{0t} + \beta_k u_{h_{n-1}}^{0t} + \alpha_k u_{h_n}^{0t} \qquad t = 0,1$$

Then from (3.3.3) - (3.3.7) , we have

$$(3.3.15) \qquad \frac{1}{4}b_{k-1}^{0t} + b_k^{0t} + \frac{1}{4}b_{k+1}^{0t} = u_{h_k}^{0t} \qquad \begin{matrix} k = -1,0,n,n+1 \\ t = 0,1 \end{matrix}$$

By expanding both sides of (3.3.15) in terms of order of h and then identifying the coefficients of the terms in order of h, we will have, for each k, t, a linear system. After solving these linear systems, we obtain

$$\alpha_{-1} = \alpha_{n+1} = 0 \qquad \alpha_0 = \alpha_n = 0$$

$$\beta_{-1} = \beta_{n+1} = -\frac{28}{9} \qquad \beta_0 = \beta_n = \frac{5}{9}$$

$$\gamma_{-1} = \gamma_{n+1} = \frac{17}{9} \qquad \gamma_0 = \gamma_n = -\frac{4}{9}$$

$$\delta_{-1} = \delta_{n+1} = -\frac{4}{9} \qquad \delta_0 = \delta_n = \frac{1}{9}$$

The next chapter is on the error analysis, in which we shall extend the Peano kernel theorem and make use of it to estimate the convergence rate of the global superconvergence.

CHAPTER 4

ERROR ANALYSIS

In this chapter, we shall discuss the errors and obtain error bounds on different norms for the RRG approximation and the global superconvergence. We shall employ the Peano Kernel Theorem as a tool in the error analysis especially in 4.1 when we make use of the linear interpolant $u_I(x)$ of $u(x)$ to obtain the bounds for $\|u-u_h\|_A$ as well as the bounds for $\|u-u_h\|_{L_\infty[0,1]}$. We shall, in 4.3, extend the Peano Kernel Theorem to apply on the BVP so that we could make use of the theorem to obtain the error bounds for the global superconvergence approximations. Finally, in 4.5, the effects of quadrature rules on the RRG approximation will be discussed in detail.

4.1. The Applications of the Peano Kernel Theorem on Interpolation

In the section, we discuss the applications of the Peano kernel theorem on the evaluation of the error bounds for the interpolation of functions. These applications will be used extensively in 4.2 , 4.4 , and 4.5 . A more general discussion can be found in [5].

Let $f \in H^{m+1}[0,1]$. Let $s \in S_h^{m,k}$ be an interpolant of f at the knots $\{x_i\}_{i=0}^n$.

Assume that s is exact for $\mathcal{P}^m$.

Let $E(\xi,\cdot) : H^{m+1}[0,1] \to R$ be the error functional, when f is interpolated by s, at each point $\xi \in [0,1]$. i.e.

(4.1.1) $$E(\xi,f) = f(\xi) - s(\xi)$$

Then by the Peano kernel theorem, we have

(4.1.2) $$E(\xi,f) = \int_0^1 K_{m+1}(\xi,t) f^{[m+1]}(t)\, dt$$

where $K_{m+1}(\xi,t)$ is the Peano kernel function which has been shown in section 1.7 to have the explicit form :

(4.1.3) $$K_{m+1}(\xi,t) = \frac{1}{m!} E_x(\xi,(x-t)_+^m)$$

Let, for each $f \in H^{m+1}[0,1]$,

(4.1.4) $$e(\xi) = E(\xi,f)$$

Then by Hölder's inequality, we have

(4.1.5) $$|e(\xi)| \le \|f^{(m+1)}\|_{L_p} \cdot \left[\int_0^1 |K_{m+1}(\xi,t)|^q\, dt\right]^{1/q}$$

If we define

(4.1.6) $$k_q(\xi) = \int_0^1 |K_{m+1}(\xi,t)|^q\, dt$$

then (4.1.5) becomes

(4.1.7) $$|e(\xi)| \le \|f^{(m+1)}\|_{L_p} \cdot [k_q(\xi)]^{1/q}$$

Since, in (4.1.7), $\|f^{(m+1)}\|_{L_p}$ is a constant, we could easily obtain

(4.1.8) $$\|e\|_{L_s} \le \|f^{(m+1)}\|_{L_p} \cdot k_{q,s}$$

where

$$(4.1.9) \qquad k_{q,s} = \left(\int_0^1 \left(\int_0^1 |K_{m+1}(\xi,t)|^q \, dt \right)^{s/q} ds \right)^{1/s}$$

In particular,

(i) $p = \infty$, $q = 1$, $s = \infty$

$$(4.1.10) \qquad \begin{cases} \|e\|_{L_\infty} \le \|f^{(m+1)}\|_{L_\infty} \cdot k_{1,\infty} \\ k_{1,\infty} = \sup_{0\le\xi\le1} \int_0^1 |K_{m+1}(\xi,t)| \, dt \end{cases}$$

(ii) $p = 1$, $q = \infty$, $s = \infty$

$$(4.1.11) \qquad \begin{cases} \|e\|_{L_\infty} \le \|f^{(m+1)}\|_{L_1} \cdot k_{\infty,\infty} \\ k_{\infty,\infty} = \sup_{0\le\xi\le1} \sup_{0\le t\le1} |K_{m+1}(\xi,t)| \end{cases}$$

(iii) $p = 2$, $q = 2$, $s = 2$

$$(4.1.12) \qquad \begin{cases} \|e\|_{L_2} \le \|f^{(m+1)}\|_{L_2} \cdot k_{2,2} \\ k_{2,2} = \left\{ \int_0^1 \int_0^1 |K_{m+1}(\xi,t)|^2 \, dt \, d\xi \right\}^{1/2} \end{cases}$$

4.2. The Errors in the Interpolation by $S_h^{1,0}$

Theorem 1.6.1 tells us that the energy norm of the RRG error is minimum over the space $S_{h,g}^{1,0}$ if the RRG approximation is from $S_{h,g}^{1,0}$. i.e. $\|u-u_h\|_A \le \|u-v_h\|_A \quad \forall\, v_h \in S_{h,g}^{1,0}$. Generally, the bound for $\|u-u_h\|_A$ is difficult to obtain directly,

but from the fact that $\|u-u_h\|_A$ is minimum over $S^{1,0}_{h,g}$, we could suitably choose a $u_I \in S^{1,0}_{h,g}$ such that the bounds of $\|u-u_I\|_A$ can be found easily and hence we could make use of these bounds for $\|u-u_I\|_A$ to be the bounds of $\|u-u_h\|_A$. Such a u_I to be chosen is the linear interpolant of u, i.e. the piecewise linear function which agrees with u at the knots $\{x_i\}_{i=0}^n$ Note that u_I can be written as

$$u_I(x) = \sum_{j=0}^{n} u(x_j)\phi_j(x)$$

Obviously, u_I is exact for $\mathcal{P}^1$. Hence by the Peano kernel theorem, the error functional $E(\xi,\cdot)$, for fixed $\xi \in [0,1]$, can be written as, for $u \in H^2[0,1]$,

$$(4.2.1) \qquad E(\xi,u) = u(\xi) - \sum_{j=0}^{n} u(x_j)\phi_j(\xi) = \int_0^1 K_2(\xi,t)u''(t)\,dt$$

where

$$(4.2.2) \qquad K_2(\xi,t) = E_x(\xi,(x-t)_+)$$

Suppose that $\xi \in [x_{k-1},x_k]$, for some $k \in \{1,\cdots,n\}$, then the Peano kernel $K_2(\xi,\cdot)$ is of the form:

$$(4.2.3) \qquad K_2(\xi,t) = (\xi-t)_+ - [(x_{k-1}-t)_+^m\phi_{k-1}(\xi)+(x_k-t)_+^m\phi_k(\xi)]$$

where

$$(4.2.4) \qquad \begin{cases} \phi_{k-1}(\xi) = \dfrac{x_k-\xi}{h} \\ \phi_k(\xi) = \dfrac{\xi-x_{k-1}}{h} \end{cases}$$

We have, from (4.2.3) and (4.2.4),

$$(4.2.5)\qquad K_2(\xi,t) = \begin{cases} 0 & t \le x_{k-1} \quad \text{or} \quad t \ge x_k \\ (\xi-t)-(x_k-t)\phi_k(\xi) & x_{k-1} \le t \le \xi \\ -(x_k-t)\phi_k(\xi) & \xi \le t \le x_k \end{cases}$$

The graph of $K_2(\xi,t)$ is

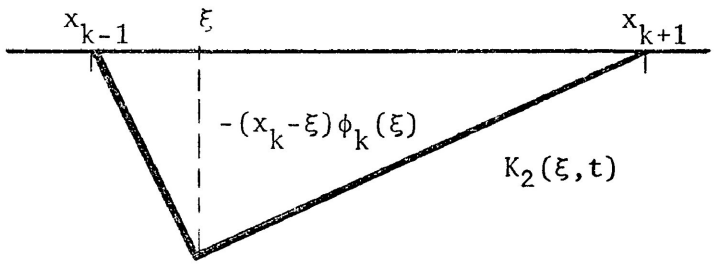

Fig. 4.2.1

and

$$(4.2.6)\qquad \|K_2(\xi,\cdot)\|_{L_1[0,1]} = \frac{1}{2}(\xi-x_{k-1})(x_k-\xi) \le \frac{1}{8}h^2$$

If $u \in H^2[0,1]$, and u'' is bounded, then

$$|E(\xi,u)| \le \|K_2(\xi,\cdot)\|_{L_1[0,1]} \cdot \|u''\|_{L_\infty[0,1]}$$

and

$$(4.2.7)\qquad \|E(\cdot,u)\|_{L_\infty[0,1]} \le \frac{1}{8}h^2 \|u''\|_{L_\infty[0,1]}$$

The constant $\frac{1}{8}$ is the best constant, since there exist functions such that the equality holds in (4.2.7) (cf. [56, pp.44]).

To establish bounds for $u'-u_I'$, we let, at each point $\xi \in [0,1]$,

$$E(\xi,u') = u'(\xi) - u_I'(\xi) \tag{4.2.8}$$

we see that the right hand side of (4.2.8) is exactly the quantity $\frac{d}{dx} E(x,u)\big|_{x=\xi}$ (we shall denote it by $E'(\xi,u)$)

It is clear that u_I' is exact for p' if $p \in \mathcal{P}^1$. By the Peano kernel theorem and similar way of working as before, we have, if $u \in H^2$ and for $\xi \in [x_{k-1},x_k]$,

$$E'(\xi,u) = \int_0^1 K_2^*(\xi,t)u''(t)\,dt \tag{4.2.9}$$

$$K_2^*(\xi,t) = \begin{cases} 0 & t \le x_{k-1} \\ \frac{t-x_{k-1}}{h} & x_{k-1} \le t \le \xi \\ -\frac{x_k-t}{h} & \xi \le t \le x_k \\ 0 & t \ge x_k \end{cases} \tag{4.2.10}$$

The graph of K_2^* is shown in fig. 4.2.2

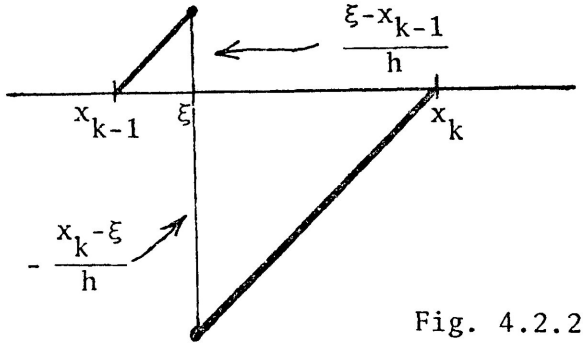

Fig. 4.2.2

From (4.2.9) , we have

$$(4.2.11)\qquad \|K_2(\xi,\cdot)\|_{L_1[0,1]} = \int_{x_{k-1}}^{\xi} \left(\frac{t-x_{k-1}}{h}\right)dt + \int_{\xi}^{x_k} \frac{x_k-t}{h}dt \le \frac{h}{2}$$

Then

$$(4.2.12)\qquad |E'(\xi,u)| \le \frac{h}{2}\|u''\|_{L_\infty[0,1]}$$

and

$$(4.2.13)\qquad \|E'(\cdot,u)\|_{L_\infty[0,1]} \le \frac{h}{2}\|u''\|_{L_\infty[0,1]}$$

The constant $\frac{1}{2}$ in (4.2.13) is an improvement of the constant in [8] in which $C = 1$. However [56] has the same result.

Thus we have proved the following theorem :

Theorem 4.2.1 If $u \in H^2[0,1]$ and u'' is bounded, then

$$(4.2.14)\qquad \|u-u_I\|_{L_\infty[0,1]} \le \frac{1}{8}h^2\|u''\|_{L_\infty[0,1]}$$

$$(4.2.15)\qquad \|u'-u_I'\|_{L_\infty[0,1]} \le \frac{1}{2}h\|u''\|_{L_\infty[0,1]}$$

For the bounds of the errors $u-u_I$ and $u'-u_I'$ in L_2-norm , we have the following theorem :

Theorem 4.2.2 If $u \in H^2[0,1]$, then

$$\|u-u_I\|_{L_2[0,1]} \le \frac{1}{3\sqrt{10}} h^2 \|u''\|_{L_2[0,1]} \tag{4.2.16}$$

$$\|u'-u_I'\|_{L_2[0,1]} \le \frac{1}{\sqrt{6}} h \|u''\|_{L_2[0,1]} \tag{4.2.17}$$

Proof :

From (4.1.12) , we compute, for $\xi \in [x_{k-1},x_k]$,

$$k_2(\xi) = \frac{1}{3h} (x_k-\xi)^2(\xi-x_{k-1})^2 \tag{4.2.18}$$

and

$$k_{2,2} = \left[\int_0^1 k_2(\xi)\, d\xi\right]^{1/2} \tag{4.2.19}$$

$$= \frac{1}{3\sqrt{10}} h^2$$

Thus, if $u \in H^2[0,1]$,

$$\|u-u_I\|_2 \le k_{2,2} \|u''\|_{L_2[0,1]}$$

$$= \frac{1}{3\sqrt{10}} h^2 \|u''\|_{L_2[0,1]}$$

Similary, we have

$$\|u'-u_I'\|_{L_2[0,1]} \le \|u''\|_{L_2} \cdot k_{2,2}^*$$

where

$$k_{2,2}^* = \left[\int_0^1 k_2^*(\xi)\, d\xi\right]^{\frac{1}{2}}$$

with

$$(4.2.20) \qquad k_2^*(\xi) = \int_0^1 |K_2^*(\xi,t)|^2 \, dt$$

$$= \frac{1}{3h^2}(\xi - x_{k-1})^3 + \frac{1}{3h^2}(x_k - \xi)^3$$

Then

$$k_{2,2}^* = \frac{h}{\sqrt{6}}$$

This completes the proof.

For the H^1 and energy error bounds, we have :

Theorem 4.2.3 If $u \in H_g^2[0,1]$, then

$$(4.2.21) \qquad \|u-u_I\|_1^2 \le (\frac{1}{6} + \frac{h^2}{90}) \cdot \|u''\|_0^2 \cdot h^2$$

$$(4.2.22) \qquad \|u-u_I\|_A^2 \le (\frac{1}{6} p_{max} + \frac{h^2}{90} q_{max}) \cdot \|u''\|_0^2 \cdot h^2$$

Proof : (4.2.21) follows from Theorem 4.2.2.

For (4.2.22) , we have :

$$\|u-u_I\|_A^2 = \int_0^1 p(u'-u_I')^2 + q(u-u_I)^2 \, dx$$

$$\le (p_{max} \cdot \frac{h^2}{6} + q_{max} \cdot \frac{h^4}{90}) \cdot \|u''\|_0^2$$

Theorem 4.2.3 will be used to establish error bounds for the RRG solution.

4.3. The Errors in the RRG Solution

Since the energy norm of the RRG error is minimum over $S_{h,g}^{1,0}$, thus from theorem 4.2.3, we have the following theorem :

Theorem 4.3.1 For all $f \in H^0[0,1]$,

$$(4.3.1) \qquad \|u-u_h\|_A \leq \left(\frac{p_{max}}{6} + \frac{q_{max}}{90} h^2\right)^{\frac{1}{2}} h \cdot \|u''\|_0$$

$$\leq \rho_2 \left(\frac{p_{max}}{6} + \frac{q_{max}}{90} h^2\right)^{\frac{1}{2}} h \cdot \|f\|_0$$

Proof : If $f \in H^0[0,1]$, then $u \in H^2[0,1]$; and from (1.3.8), we have

$$\|u''\|_0 \leq \rho_2 \|f\|_0$$

and from Theorem 1.6.1 and Theorem 4.2.3, we have

$$\|u-u_h\|_A \leq \|u-u_I\|_A$$

$$\leq \left(\frac{p_{max}}{6} + \frac{q_{max}}{90} h^2\right)^{\frac{1}{2}} h \cdot \|u''\|_0$$

$$\leq \rho_2 \left(\frac{p_{max}}{6} + \frac{q_{max}}{90} h^2\right)^{\frac{1}{2}} h \cdot \|f\|_0$$

Theorem 4.3.2 If $u \in H^2[0,1]$, then

$$(4.3.2) \qquad \|u-u_h\|_0 \le \rho_2 \cdot (\frac{1}{6} p_{max} + \frac{h^2}{90} q_{max}) h^2 \|u''\|_0$$

$$\le \rho_2^2 (\frac{1}{6} p_{max} + \frac{h^2}{90} q_{max}) h^2 \|f\|_0$$

Proof : (cf. [56], pp.49]) .

However, for a self-adjoint second order ODE (1.3.2 - 1.3.4), a theorem in ([60], pp. 914) could be applied.

Theorem 4.3.3 [Wheeler]

$\exists$ a constant C such that if $u \in W^{2,\infty}[0,1]$,

then

$$(4.3.3) \qquad \|u-u_h\|_0 \le Ch^2 \|u''\|_{W^{0,\infty}}$$

(cf. Section 3.2)

From theorem 4.3.1 to theorem 4.3.3, we have shown that, the order of accuracy in the error $u-u_h$ in different norms are: $\|\cdot\|_A$: $O(h)$, $\|\cdot\|_0$: $O(h^2)$ and $\|\cdot\|_{L_\infty[0,1]}$: $O(h^2)$.

4.4. A Generalization of the Peano Kernel Theorem

By a generalization of the Peano kernel theorem, we are able to establish error bounds for the Global Superconvergence.

Theorem 4.4.1 (*Generalized Peano Kernel Theorem*)

Suppose $E : H_0^k[0,1] \to R$ has the property that

(4.4.1) $E(u) = 0$ if $u(t) = t(1-t)p(t)$ $\forall\ p \in \mathcal{P}^m$

Then there exists a K_{m+1} such that

(4.4.2) $$E(u) = \int_0^1 K_{m+1}(t)u^{(m+1)}(t)\ dt$$

where

(4.4.3) $$u^{(m+1)} = \frac{d^{m+1}}{dt^{m+1}} \frac{u(t)}{t(1-t)}$$

Furthermore, the generalized Peano kernel has the form :

(4.4.4) $$K_{m+1}(t) = E(u_t)$$

where

(4.4.5) $$u_t(\tau) = \frac{(\tau-t)_+^m}{m!}\tau(1-\tau)$$

Proof :

By Riesz's representation theorem, there exists a $\mathring{K}$ such that

(4.4.6) $$E(u) = \int_0^1 \mathring{K}(t)u(t)\ dt$$

Let $K_0 = t(1-t)\mathring{K}$ and $u^{(0)} = \frac{u(t)}{t(1-t)}$, then (4.4.6) can be written as

$$(4.4.7) \qquad E(u) = \int_0^1 K_0(t)u^{(0)}(t)\,dt$$

Define $$K_1(t) = -\int_0^t K_0(\tau)\,d\tau$$

Then since $$u^{(1)}(t) = \frac{d}{dt}u^{(0)}(t)$$

we have

$$(4.4.8) \qquad E(u) = \left[-K_1(t)u^{(0)}(t)\right]_0^1 + \int_0^1 K_1(t)u^{(1)}(t)\,dt$$

Take $u(t) = t(1-t)$, then $u^{(0)} = 1$, $u^{(1)} = 0$, and from (4.4.1), $E(u) = 0$; these with $K_1(0) = 0$ give us $K_1(1) = 0$. Thus

$$(4.4.9) \qquad E(u) = \int_0^1 K_1(t)u^{(1)}(t)\,dt$$

Assume that we have, for some $1 \le s \le m$,

$$(4.4.10) \qquad E(u) = \int_0^1 K_s(t)u^{(s)}(t)\,dt$$

Define $$K_{s+1}(t) = -\int_0^t K_s(\tau)\,d\tau$$

Then since $$u^{(s+1)}(t) = \frac{d}{dt}u^{(s)}(t)$$

and by integration by parts, we have

$$E(u) = \left[-K_{s+1}(t)u^{(s)}(t)\right]_0^1 + \int_0^1 K_{s+1}(t)u^{(s+1)}(t)\,dt$$

Take $u(t) = t(1-t)\frac{t^s}{s!}$, then from (4.4.1), $E(u) = 0$. We also have $u^{(s+1)}(t) = 0$ and $K_{s+1}(0) = 0$, these give $K_{s+1}(1) = 0$ and hence

$$E(u) = \int_0^1 K_{s+1}(t)u^{(s+1)}(t)\,dt$$

The function $K_{s+1}(t)$, $1 \le s \le m$, is a generalized Peano kernel function.

To evaluate K_{s+1} , we want

$$u^{(s+1)}(\tau) = \delta_t(\tau) \tag{4.4.11}$$

so we take $u = u_t$, where u_t is defined as

$$u_t(\tau) = \frac{(\tau-t)_+^s}{s!}\tau(1-\tau) \tag{4.4.12}$$

Then

$$K_{s+1}(t) = E(u_t) \tag{4.4.13}$$

This completes the proof.

We shall, in the following Lemma, establish conditions on u to ensure the existence of $u^{(s+1)}$. We shall define

$$C_0^k[0,1] = \{u \in C^k[0,1] \mid u(0) = u(1) = 0\} .$$

Lemma 4.4.1

(i) If $u \in C_0^1[0,1]$, then $u^{(0)} \in C^0[0,1]$.

(ii) If $u \in C_0^2[0,1]$, then $u^{(1)} \in C^0[0,1]$.

(iii) If $u \in C_0^3[0,1]$, then $u^{(2)} \in C^0[0,1]$.

Proof :

It is clear that (i) - (iii) are valid for $t \in (0,1)$. We

shall show that the derivatives exist at $t = 0,1$ for the three cases.

(i) If $u \in C^1[0,1]$, then we have

$$u(t) = tu'(0) + t\cdot o(1)$$

Let $$\frac{1}{t(1-t)} = \frac{1}{t}(1 + t + t\cdot o(1))$$

Then $$u^{(0)}(t) = \frac{u(t)}{t(1-t)} = u'(0) + o(1)$$

Thus we have $u^{(0)}(0) = u'(0)$.

Similarly, we have $u^{(0)}(1) = -u'(1)$.

(ii) If $u \in C^2[0,1]$, then we have

$$u(t) = tu'(0) + \frac{t^2}{2}u''(0) + t^2\cdot o(1)$$

Let $$\frac{1}{t(1-t)} = \frac{1}{t}(1 + t + \frac{t^2}{2} + t^2\cdot o(1))$$

Then we have $$u^{(0)}(t) = (1+t)u'(0) + \frac{t}{2}u''(0) + t\cdot o(1)$$

Thus $$u^{(1)}(0) = \frac{1}{2}u''(0) + u'(0)$$

Similarly we have $u^{(1)}(1) = \frac{1}{2}u''(1) - u'(1)$.

(iii) If $u \in C^3[0,1]$, then we have

$$u(t) = tu'(0) + \frac{t^2}{2}u''(0) + \frac{t^3}{6}u'''(0) + t^3\cdot o(1)$$

Let $$\frac{1}{t(1-t)} = \frac{1}{t}(1 + t + \frac{t^2}{2}u''(0) + \frac{t^3}{6}u'''(0) + t^3\cdot o(1))$$

Then we have $$u^{(0)}(t) = u'(0)(1+t+t^2) + u''(0)\frac{t(1-t)}{2} + u'''(0)\frac{t^2}{6} + t^2\cdot o(1)$$

and hence $$u^{(2)}(0) = 2u'(0) + u''(0) + \frac{1}{3}u'''(0)$$.

Similarly $$u^{(2)}(1) = -2u'(1) + u''(1) - \frac{1}{3}u'''(1).$$

4.5. The Rate of Convergence of the Global Superconvergence

The global superconvergence $S = Q \cdot R_h$ has the property

$$S(p) = p \qquad \forall\, p \in \mathcal{P}_0^{m+2} \qquad m \geq 0 \tag{4.5.1}$$

Each $p \in \mathcal{P}_0^{m+2}$ can be written as

$$p(t) = t(1-t)p^*(t) \qquad p^* \in \mathcal{P}^m \tag{4.5.2}$$

From Section 4.4 , for each $\xi \in [0,1]$, the error functional $E(\xi,\cdot) : H_0^{m+2}[0,1] \to R$, $E(\xi,u) = u(\xi)-s(\xi)$, has a representation form

$$E(\xi,u) = \int_0^1 K_{m+1}(\xi,t)u^{(m+1)}(t)\, dt \tag{4.5.3}$$

Similar to Section 2.2, we define, for each $u \in H_0^{m+2}[0,1]$,

$$e(\xi) = E(\xi,u) \tag{4.5.4}$$

Following Section 4.1, we have, for $p,q \geq 1$, s.t. $\frac{1}{p} + \frac{1}{q} = 1$,

$$\|e\|_{L_s[0,1]} \leq \|u^{(m+1)}\|_{L_p[0,1]} \cdot k_{q,s} \tag{4.5.5}$$

where

$$k_{q,s} = \left(\int_0^1 \left(\int_0^1 |K_{m+1}(\xi,t)|^q\, dt\right)^{s/q} d\xi\right)^{1/s} \tag{4.5.6}$$

From Section 4.4, the kernel function $K_{m+1}(\xi,t)$ is of the form $K_{m+1}(\xi,t) = E(\xi,u_t)$ where $u_t(\tau) = \frac{(\tau-t)_+^m}{m!}\tau(1-\tau)$. The analytical solution of K_{m+1} is rather difficult to obtain. However, in Section 6.2 , we shall evaluate, in particular, $K_2(\xi,t)$ numerically

and then compute the quantities $k_{\infty,\infty}$ and $k_{1,\infty}$ to obtain the error bounds

$$\begin{cases} \|e\|_{L_\infty[0,1]} \leq \|u^{(2)}\|_{L_1[0,1]} \cdot k_{\infty,\infty} \\ \\ \|e\|_{L_\infty[0,1]} \leq \|u^{(2)}\|_{L_\infty[0,1]} \cdot k_{1,\infty} \end{cases}$$

Thus the *rate of convergence* of the global superconvergence can be approximated.

4.6. The Effects of Quadrature Errors on the RRG Solution

In this section, we shall discuss the effects of quadrature errors on the RRG solution of (1.3.3) , for simplicity, with homogeneous (1.3.4) .

From Section 2.1 , the RRG solution u_h is obtained by solving the linear system

(4.6.1) $$Ma = f$$

Our discussion will be concentrated on the following cases:

(i) When f is approximated by the quadrature schemes in 2.2b alone.

(ii) When M is approximated by the quadrature schemes in 2.2c and 2.2d alone.

(iii) When both f and M are approximated by the quadrature schemes in 2.2b and (2.2c - 2.2d) respectively.

For the case when f alone is approximated by quadrature rules, then we have a new linear system

$$M\hat{a} = \hat{f} \tag{4.6.2}$$

and the solution of (4.6.2) gives us an approximation

$$\hat{u}_h = \sum_{i=1}^{n-1} \hat{a}_i \phi_i \tag{4.6.3}$$

We shall discuss the choice of quadrature schemes for a given sequence of spline subspaces $\{S^{1,0}_{h_i,0}\}_{i=1}^{\infty}$ of $H^1_0[0,1]$ so that the theoretical approximations $\{u_{h_i}\}_{i=1}^{\infty}$, determined successively from (4.6.1), and the approximations $\{\hat{u}_{h_i}\}_{i=1}^{\infty}$, determined successively from (4.6.3), have a general order of accuracy.

Let

$$I_i(f) = f_i = \int_0^1 f(x)\phi_i(x)\,dx \qquad 1 \le i \le n-1 . \tag{4.6.4}$$

We associate with the subspace $S^{1,0}_{h,0}$ a set of quadrature rules $\{Q_i\}_{i=1}^{n-1}$ which is to approximate $\{I_i\}_{i=1}^{n-1}$.

Let

$$\hat{f}_i = Q_i(f) = \sum_{j=0}^{n'} c_j f(x'_j) \tag{4.6.5}$$

as the approximation of f_i in (4.6.4).

It can be easily verified that the quadratic form $y^T My$ can be

expressed in terms of the norm $\|\cdot\|_A$ as follow

$$y^T M y = \| \sum_{i=1}^{n-1} y_i \phi_i \|_A^2 \tag{4.6.6}$$

Subtracting (4.6.2) from (4.6.1), we have

$$M(a-\hat{a}) = f - \hat{f} \tag{4.6.7}$$

Then by multiplying both sides by $(a - \hat{a})^T$ and using (4.6.6) , we have

$$(a-\hat{a})^T M(a-\hat{a}) = \|u_h-\hat{u}_h\|_A^2 = (a-\hat{a})^T(f-\hat{f}) \tag{4.6.8}$$

Up to this point the discussion is similar to [34] ; what is different is the form of the quadrature rules. In [34] Herbold, Schultz and Varga considered the quadrature rule H_i for I_i of the form

$$H_i(f) = \sum_{j=0}^{n'} \sigma_i(x'_j) \tag{4.6.9}$$

where $\sigma_i = f\phi_i$

i.e. their quadrature rules take on the values of $f\phi_i$ at the selected points. The following works will show the quadrature rules which we are employing have the following advantages over the quadrature rules of the form (4.6.9) .

(i) The bound for $\|u_h-\hat{u}_h\|_A$ is easy to obtain.

(ii) The order of accuracy for the bound of $\|u_h-\hat{u}_h\|_A$ is higher.

Let

$$(4.6.10)\qquad \begin{cases} e = a - \hat{a} & i = 1,\cdots,n-1 \\ E_i = I_i - Q_i & i = 1,\cdots,n-1 \\ E = (E_1,\cdots,E_{n-1}) \end{cases}$$

Then we have, from (4.6.4) and (4.6.5),

$$(4.6.11)\qquad E_i(f) = \int_0^1 f(x)\phi_i(x)\,dx - \sum_{j=0}^{n'} c_j f(x_j')$$

and the last quantity in (4.6.8) can be written as

$$(4.6.12)\qquad \begin{aligned} (a-\hat{a})^T(f-\hat{f}) &= \sum_{i=1}^{n-1} (a_i-\hat{a}_i)(I_i(f)-Q_i(f)) \\ &= \sum_{i=1}^{n-1} e_i\, E_i(f) \\ &= \sum_{i=1}^{n-1} E_i(fe_i) \\ &= E\cdot fe \end{aligned}$$

Thus (4.6.8) is of the form

$$(4.6.13)\qquad e^TMe = \|u_h-\hat{u}_h\|_A^2 = E\cdot fe$$

This equation will be used to obtain bounds for $\|u_h-\hat{u}_h\|_A$.

As we have done in Section 2.3, given a $f \in H^m[0,1]$, we could select a quadrature formula of the form (4.6.5) such that the quadrature errors of $f_i-\hat{f}_i$ satisfying

$$(4.6.14)\quad |I_i(f)-Q_i(f)| \le K_{r,i}\cdot h^{m_r}\cdot \|f^{[m]}\|_{L_r[x_{i-1},x_{i+1}]}$$

where $K_{r,i}$ is independent of h.

In particular, we have, from 2.2b, $m_\infty = m + 1$ and $m_2 = m + \frac{1}{2}$.

Theorem 4.6.1

Let $f \in H^m[0,1]$, $u_h, \hat{u}_h \in S^{1,0}_{h,0}$. If for $m \ge 1$, the linear functionals Q_i defined in (4.6.5) satisfies (4.6.14), then we have, for $q \ge 2$,

$$(4.6.15)\quad \|u_h-\hat{u}_h\|_A \le 2^{1/q}\cdot K_q\cdot \|f^{[m]}\|_{L_q[0,1]}\cdot h^{m_{q+1/q-1/2}}$$

In particular, we have

$$(4.6.16)\quad \|u_h-\hat{u}_h\|_A \le K_\infty\cdot \|f^{[m]}\|_{L_\infty[0,1]}\cdot h^{m+1/2}$$

$$(4.6.17)\quad \|u_h-\hat{u}_h\|_A \le 2^{1/2}\cdot K_2\cdot \|f^{[m]}\|_{L_2[0,1]}\cdot h^{m+1/2}$$

Proof :

By applying the Hölder's inequality on (4.6.12), we have, for $p,q \ge 1$ with $\frac{1}{p} + \frac{1}{q} = 1$,

$$(4.6.18)\quad \begin{aligned} E\cdot fe &= \sum_{i=1}^{n-1} E_i\cdot fe_i \\ &= \sum_{i=1}^{n-1} e_i\cdot E_i f \\ &= \Big(\sum_{i=1}^{n-1} |e_i|^p\Big)^{1/p}\Big(\sum_{i=1}^{n-1} |E_i f|^q\Big)^{1/q} \\ &\le \|e\|_{\ell_p}\cdot\Big(\sum_{i=1}^{n-1} [h^{m_q}\cdot K_{q,i}\cdot \|f^{[m]}\|_{L_q[x_{i-1},x_{i+1}]}]^q\Big)^{1/q} \end{aligned}$$

Since Q_i , $1 \le i \le n-1$, are identical, (cf. 2.2b) , we have $K_{qi} = K_q$, $1 \le i \le n-1$ and

$$
\begin{aligned}
&\sum_{i=1}^{n-1} \left[K_{q,i}^{q} \cdot \| f^{[m]} \|^{q}_{L_q[x_{i-1},x_{i+1}]} \right. \\
&= K_q^q \cdot \sum_{i=1}^{n-1} \int_{x_{i-1}}^{x_{i+1}} |f^{(m)}(x)|^q \, dx \qquad (4.6.19) \\
&\le 2K_q^q \cdot \| f^{[m]} \|^q_{L_q[0,1]}
\end{aligned}
$$

Thus (4.6.18) becomes

$$
E \cdot fe \le \| e \|_{\ell_p} \cdot h^{m_q} \cdot 2^{1/q} \cdot K_q \cdot \| f^{[m]} \|_{L_q[0,1]} \qquad (4.6.20)
$$

Define

$$
\| e \|_M^2 = e^T M e \qquad (4.6.21)
$$

Since M is positive definite, $\| \cdot \|_M$ is a norm on R^{n-1} and is equivalent to $\| \cdot \|_{\ell_p}$ (Young [63], pp.27). In particular, for $p \le 2$, we have

$$
\| e \|_{\ell_p} \le \| e \|_M \cdot h^{1/q-1/2} \qquad (4.6.22)
$$

Then from (4.6.13), (4.6.20), (4.6.21) and (4.6.22), we have

$$
\| u_h - \hat{u}_h \|_A \le 2^{1/q} \cdot K_q \cdot \| f^{[m]} \|_{L_q[0,1]} \cdot h^{m_q+1/q-1/2}
$$

This completes the proof.

Herbold, Schultz and Varga [34,pp.253] obtained $\|u-u_h\|_A \leq K \cdot h^{m-1}$, $m \geq 1$; compare this with (4.6.16), (4.6.16) has an order $O(h^{1.5})$ higher.

We see from (4.6.15) that if we have a sequence $\{S^{1,0}_{h_i,0}\}_{i=1}^{\infty}$ of finite dimensional subspaces of $H_0^1[0,1]$ such that $\lim_{i\to\infty} h_i = 0$ then if m , dependent only on f , satisfies $m \geq 0$, we evidently have

$$\lim_{i\to\infty} \|u_h-\hat{u}_h\|_A = 0$$

This means that the quadrature errors, introduced by computing $\hat{u}_h$ rather than u_h , tends to zero with i . These errors may or may not be small relative to $\|u-u_h\|_A$. Following [34], we have the following definition :

Definition 4.6.1

The choice of quadrature rules in (4.6.5) is said to be *consistent* in the norm $\|\cdot\|_N$ if the order of $\|u-\hat{u}_h\|_N$ is of the same order with $\|u-u_h\|_N$. i.e. $\exists\, K_4$ if

$$\|u-u_h\|_N \leq K_3 \cdot h^{\ell} \tag{4.6.23}$$

then

$$\|u_h-\hat{u}_h\|_N \leq K_4 \cdot h^{\ell} \tag{4.6.24}$$

where K_3, K_4 and ℓ are positive constants, which are independent of h. With the triangle inequality, the bounds of (4.6.23) for the norm $\|\cdot\|_A$ and the result of Theorem 4.6.1, it follows that

$$(4.6.25)\quad \|u-\hat{u}_h\|_A \le \|u-u_h\|_A + \|u_h-\hat{u}_h\|_A$$
$$\le K_5\cdot h^{\ell} + 2^{1/q}\cdot K_q\cdot \|f^{[m]}\|_{L_q[0,1]}\cdot h^{m_q+1/q-1/2}$$

It follows that $m_q + 1/q - 1/2 \ge \ell$ gives a consistent choice of quadratures in (4.6.15) in the norm $\|\cdot\|_A$ which preserves the asymptotic accuracy of (4.6.23) in this norm.

As we notice from Theorem 4.3.1, for sufficiently small h, K_5 is independent of h, and $\ell = 1$, thus $m_\infty \ge \frac{3}{2}$, equivalently $m \ge \frac{1}{2}$, will give a consistent choice of quadrature in (4.6.16) in the norm $\|\cdot\|_A$ Thus an 1-exact quadrature in Section 2.2b is good for a consistent quadrature.

Now we consider the case when we apply quadrature rules on M and compute f exactly. In this case, we have a different linear system

$$(4.6.26)\qquad \tilde{M}\,\tilde{a} = f$$

The solution of (4.6.26) will give us an approximation

$$(4.6.27)\qquad \tilde{u}_h = \sum_{i=1}^{n-1} \tilde{a}_i\,\phi_i$$

Let

$$(4.6.28)\qquad \begin{cases} \tilde{e} = a - \tilde{a} \\ \delta M = M - \tilde{M} \end{cases}$$

then from (4.6.1), (4.6.26) and (4.6.28), we have

(4.6.29) $$(M - \delta M)(a - \tilde{e}) = Ma$$

After neglecting the term $\delta M \cdot \tilde{e}$ in (4.6.29), we have

$$(\delta M)a + M\tilde{e} = 0$$

Thus

(4.6.30) $$\tilde{e} = -M^{-1}(\delta M)a$$

and for $q \geq 2$,

(4.6.31) $$\|\tilde{e}\|_{\ell_q} \leq \|M^{-1}\|_s \cdot \|\delta M\|_s \cdot \|a\|_{\ell_2} \cdot h^{1/q-1/2}$$

where $\|\cdot\|_s$ is the spectral norm. For symmetric and positive definite M,

(4.6.32) $$\|M\|_s = \lambda_{max}$$

where λ_{max} is the largest eigenvalue of M.

If $p \in H^{m_1}[0,1]$ and $q \in H^{m_2}[0,1]$, we could select quadratures (cf. Section 2.2c and 2.2d) such that

(4.6.33) $$\begin{cases} (\delta M)_{i-1,i} \leq K_6 h^{m_1 q} + K_7 h^{m_2 q} \\ (\delta M)_{i,i} \leq K_8 h^{m_1 q} + K_9 h^{m_2 q} \\ (\delta M)_{i,i+1} \leq K_6 h^{m_1 q} + K_7 h^{m_2 q} \end{cases}$$

where K_6, K_7, K_8 and K_9 are independent of h. Note that $(\delta M)_{i,j} = 0$ if $|j-i| > 1$. Thus there exists a constant K_{10} which is

independent of h, for sufficiently small h, such that

$$(4.6.34) \qquad (\delta M)_{i,j} \le K_{10} \cdot h^{\min(m_{1q}, m_{2q})}$$

Then, there is a K_{11}, which is independent of h for sufficiently small h, such that

$$(4.6.35) \qquad \|\delta M\|_s \le K_{11} \cdot h^{\min(m_{1q}, m_{2q})}$$

By multiplying M^{-1} on both sides of (4.6.1), we have

$$a = M^{-1} f$$

and

$$(4.6.36) \qquad \|a\|_{\ell_2} \le \|M^{-1}\|_s \|f\|_{\ell_2}$$

From (4.6.35) and (4.6.36), (4.6.31) becomes

$$(4.6.37) \qquad \|\tilde{e}\|_{\ell_q} \le \|M^{-1}\|_s^2 \cdot \|f\|_{\ell_2} \cdot K_{11} \cdot h^{\min(m_{1q}, m_{2q}) + 1/q - 1/2}$$

To evaluate $\|M^{-1}\|_s$, we shall consider first the case when p and q are both constants. For this case, M is given in Section 2.1. The eigenvalues and the corresponding eigenvectors of M is

$$(4.6.38) \begin{cases} u_k(i) = \sin \pi \dfrac{i}{n} & 1 \le k \le n-1 \\ \\ \lambda_k = 2 \cdot \dfrac{p}{h}\left[1 - \cos \dfrac{\pi k}{n}\right] + \dfrac{2qh}{6}\left[\cos \dfrac{\pi k}{n} + 2\right] \end{cases}$$

where $1 \le i \le n-1$ denote the i^{th} component of u_k.

For $h < \sqrt{6p/q}$, we have

$$\lambda_{min} = \frac{2p}{h} + \frac{4qh}{6} + (-\frac{2p}{h} + \frac{2qh}{6})\cos\frac{\pi}{n}$$

(4.6.39)

$$\lambda_{max} = \frac{2p}{h} + \frac{4qh}{6} + (-\frac{2p}{h} + \frac{2qh}{6})\cos\frac{(n-1)\pi}{n}$$

For large n , equivalently small h , we have

$$\lambda_{min} \simeq p\pi^2 h - \frac{\pi^2 q}{6} h^3 \simeq p\pi^2 h$$

(4.6.40)

$$\lambda_{max} \simeq \frac{4p}{h} + h[\frac{2q}{6} - p\pi^2] + h^3[\frac{q\pi^2}{6}] \simeq \frac{4p}{h}$$

The maximum eigenvalue for M^{-1} is $\frac{1}{\lambda_{min}}$, thus we have

$$\|M^{-1}\|_s = \frac{1}{\lambda_{min}} \simeq \frac{1}{p\pi^2} h^{-1} \tag{4.6.41}$$

For general $p(x)$ and $q(x)$, by the Sturm comparison theorem [6, pp.290], the minimum eigenvalue of L in (1.3.3) is greater that the minimum eigenvalue of $L' = p_{min}u'' + (\lambda - q_{min})u'$, thus we have

$$\lambda_{min} \geq p_{min}\pi^2 h \tag{4.6.42}$$

and

$$\|M^{-1}\|_s \leq \frac{1}{p_{min}\pi^2} h^{-1} \tag{4.6.43}$$

By substituting (4.6.43) into (4.6.37), we have

$$\|\tilde{e}\|_{\ell q} \leq \frac{K_{11}}{p_{min}^2 \pi^4} \|f\|_{\ell_2} \cdot h^{\min(m_{1q}, m_{2q}) - 5/2 + 1/q} \tag{4.6.44}$$

From the definition of f_i , we easily obtain

$$(4.6.45) \qquad \|f\|_{\ell_2} \leq \frac{2}{\sqrt{3}} \cdot h^{1/2} \cdot \|f\|_{L_2[0,1]}$$

Substituting (4.6.45) into (4.6.44), we have

$$(4.6.46) \qquad \|\tilde{e}\|_{\ell_q} \leq \frac{2K_{11}}{p_{min}^2 \pi^4} \cdot \frac{1}{\sqrt{3}} \cdot h^{\min(m_{1q}, m_{2q})+1/q-2} \cdot \|f\|_{L_2[0,1]}$$

Up to this step, it is not hard to obtain bounds for $\|u_h - \tilde{u}_h\|_A$. We observe that, for $p,q \geq 1$ with $\frac{1}{p} + \frac{1}{q} = 1$,

$$(4.6.47) \qquad \|u_h - \tilde{u}_h\|_0^2 = \int_0^1 (u_h - \tilde{u}_h)^2(x)\, dx$$

$$= \int_0^1 \left[\sum_{i=1}^{n-1} \tilde{e}_i \phi_i(x)\right]^2 dx$$

$$\leq \int_0^1 \|\tilde{e}\|_{\ell_q}^2 \, \|\phi_i(x)\|_{\ell_p}^2 \, dx$$

$$= \|\tilde{e}\|_{\ell_q}^2 \cdot \left(\int_0^1 \left[\sum_{i=1}^{n-1} |\phi_i(x)|^p \right]^{2/p} dx \right)$$

$$\leq \|\tilde{e}\|_{\ell_q}^2 \cdot h^{-2/p}$$

$$(4.6.48)\quad \|u_h'-\tilde{u}_h'\|_0^2 = \int_0^1 (u_h'-\hat{u}_h')^2(x)\,dx$$

$$= \int_0^1 \Big[\sum_{i=1}^{n-1} \tilde{e}_i \phi_i'(x)\Big]^2 dx$$

$$\le \int_0^1 \|\tilde{e}\|_{\ell_q}^2 \cdot \|\phi_i'(x)\|_{\ell_p}^2 dx$$

$$= \|\tilde{e}\|_{\ell_q}^2 \cdot \Big(\int_0^1 \Big[\sum_{i=1}^{n-1} |\phi_i'(x)|^p\Big]^{2/p} dx\Big)$$

$$\le \|\tilde{e}\|_{\ell_q}^2 \cdot \Big(\frac{2}{h}\Big)^2$$

By using (4.6.47) and (4.6.48), we have

$$(4.6.49)\qquad \|u_h-\tilde{u}_h\|_A \le (h^{-2/p} + 4h^{-2})^{1/2} \cdot \|\tilde{e}\|_{\ell_q}$$

Then from (4.6.46), we have

$$(4.6.50)\qquad \|u_h-\tilde{u}_h\|_A \le \frac{2K_{11}}{p_{\min}^2 \pi^4} \cdot \frac{1}{\sqrt{3}} \cdot (h^{-2/p} + 4h^{-2})^{1/2}$$

$$\cdot h^{\min(m_{1q}, m_{2q})-2+1/q} \cdot \|f\|_{L_2[0,1]}$$

In particular,

(i) when $p = 1$ and $q = \infty$, we have

$$m_{1,\infty} = m_1 + 1\ ,\ m_{2,\infty} = m_2 + 1$$

and

$$(4.6.51)\qquad \|u_h-\tilde{u}_h\|_A \le \frac{2K_{11}}{p_{min}^2\pi^4}\cdot\left(\frac{5}{3}\right)^{1/2}\cdot h^{\min(m_1,m_2)-2}\cdot\|f\|_{L_2[0,1]}$$

(ii) when $p = 2$ and $q = 2$, we have

$$m_{1,2} = m_1 + \frac{1}{2}\ ,\quad m_{2,2} = m_2 + \frac{1}{2}$$

and

$$(4.6.52)\qquad \|u_h-\tilde{u}_h\|_A \le \frac{2K_{11}}{p_{min}^2\pi^4}\cdot\frac{1}{\sqrt{3}}\cdot(h+4)^{1/2}\cdot h^{\min(m_1,m_2)-2}\cdot\|f\|_{L_2[0,1]}$$

Consider (4.6.51), it shows that $\min(m_1,m_2)-2 \ge \ell$ will give a consistent choice of quadratures in the norm $\|\cdot\|_N$ in (4.6.23). For $\ell = 1$ (w.r.t. $\|\cdot\|_A$) we need $\min(m_1,m_2) \ge 3$. This shows that the quadratures in 2.2c and 2.2d are consistent in the energy norm $\|\cdot\|_A$.

Now we shall consider the case when M and f are both approximated by quadratures. In this case, the linear system is

$$(4.6.53)\qquad \tilde{M}\overset{*}{a} = \hat{f}$$

Let

$$(4.6.54)\qquad \begin{cases} \delta M = M - \tilde{M} \\ \overset{*}{e} = a - \overset{*}{a} \\ \delta f = f - \hat{f} \end{cases}$$

Then from (4.6.53) and (4.6.54), we have

$$(4.6.55)\qquad (M - \delta M)(a - \overset{*}{e}) = f - \delta f$$

By expanding the left hand side of (4.6.55) and neglecting the term $\delta M \cdot \overset{*}{e}$, we have

$$M\overset{*}{e} + (\delta M)a = \delta f \tag{4.6.56}$$

By multiplying M^{-1} on both sides of (4.6.56) , we have

$$\overset{*}{e} = M^{-1}\cdot\delta f - M^{-1}\cdot\delta M\cdot a$$

and, for $q \geq 2$

$$\begin{aligned} \|\overset{*}{e}\|_{\ell_q} &\leq \left(\|M^{-1}\|_s\|\delta f\|_{\ell_2} + \|M^{-1}\|_s\,\|\delta M\|_s\|a\|_{\ell_2}\right)\cdot h^{1/q-1/2} \\ &\leq \left(\|M^{-1}\|_s\|\delta f\|_{\ell_2} + \|M^{-1}\|_s^2\,\|\delta M\|_s\|f\|_{\ell_2}\right)\cdot h^{1/q-1/2} \end{aligned} \tag{4.6.57}$$

From (4.6.14) and (4.6.19) , we have

$$\|\delta f\|_{\ell_2} \leq 2^{1/2}\cdot K_2\cdot\|f^{[m]}\|_{L_2[0,1]}\cdot h^{m+1/2} \tag{4.6.58}$$

From (4.6.35), (4.6.43), (4.6.45) and (4.6.57), (4.6.58) becomes

$$\begin{aligned} \|\overset{*}{e}\|_{\ell q} &\leq \frac{\sqrt{2}}{p_{\min}\pi}\cdot K_2\|f^{[m]}\|_{L_2[0,1]}\cdot h^{m+1/q-1} \\ &\quad + \frac{2}{\sqrt{3}p_{\min}^2\pi^4}\cdot K_{11}\cdot\|f\|_{L_2[0,1]}\cdot h^{\min(m_{1q},m_{2q})-2+1/q} \\ &= K_{14}\cdot h^{m+1/q-1} + K_{15}\cdot h^{\min(m_{1q},m_{2q})-2+1/q} \end{aligned} \tag{4.6.59}$$

where K_{14} and K_{15} are constants independent of h for sufficiently small h .

Then from (4.6.49), we could have

$$(4.6.60)\qquad \|u_h-\overset{*}{u}_h\|_A \le (h^{-2/p} + 4h^{-2})^{1/2} \cdot \|\overset{*}{e}\|_{\ell_q}$$

$$\le (h^{-2/p} + 4h^{-2})^{1/2}(K_{14}\cdot h^{m+1/q-1} + K_{15}\cdot h^{\min(m_{1_q},m_{2_q})-2+(1/q)}$$

In particular,

(i) If $p = 1$ and $q = \infty$, we have, if $f \in H^m$, $p \in H^{m_1}$ and $q \in H^{m_2}$,

$$(4.6.61)\qquad \|u_h-\overset{*}{u}_h\|_A \le \sqrt{5}(K_{14}\cdot h^{m-2} + K_{15}\cdot h^{\min(m_1,m_2)-2})$$

$$= K_{16}\cdot h^{\min(m,m_1,m_2)-2}$$

where K_{16} is independent of h when h is sufficiently small.

(ii) If $p = 2$ and $q = 2$, we have, if $f \in H^m$, $p \in H^{m_1}$ and $q \in H^{m_2}$,

$$(4.6.62)\qquad \|u_h-\overset{*}{u}_h\|_A \le (h+4)^{1/2}\cdot(K_{14}\cdot h^{m-\frac{3}{2}} + K_{15}\cdot h^{\min(m_1,m_2)-2})$$

$$= K_{17}\cdot h^{\min(m+1/2,m_1,m_2)-2}$$

where K_{17} is independent of h when h is sufficiently small.

Consider (4.6.61), $\min(m,m_1,m_2)-2 \le \ell$ will gives a consistent choice of quadratures. Since $\ell = 1$ for $\|\cdot\|_A$, we need $\min(m,m_1,m_2) \ge 3$ i.e. $m \ge 3$, $m_1 \ge 3$, $m_2 \ge 3$, these show that the 3-exact quadrature rules in Sections 2.2a, 2.2b and 2.2c are consistent in $\|\cdot\|_A$.

CHAPTER 5

ALGORITHMS FOR SOLVING THE LARGE LINEAR SYSTEMS

5.1. LU Decomposition (Gaussian Elimination Method)

We have shown that, the solution of (1.3.3)-(1.3.4) by the linear finite element method is obtained by solving the large linear system (2.1.10)), which is $Ma = F$. The matrix M has been shown to be symmetric, positive-definite and tridiagonal. These special characteristics of M will ensure that the solution of the system by *Gaussian elimination method*, also called *LU decomposition*, without row exchanges is not only possible but also numerically stable ([56, pp.36]). By the Gaussian elimination method, the matrix M is factorized into the product $M = LU$, where L is an lower bidiagonal matrix with unit diagonal elements and U is an upper bidiagonal matrix. The linear system (2.1.10) is converted to an equivalent system $Ua = F'$, where $F' = L^{-1}F$. Thus $a = U^{-1}F' = U^{-1}L^{-1}F$. If M is symmetric, it can be easily verified that $M = LDL^T$, where D is the diagonal matrix which consists of the diagonal elements of U . For a tridiagonal matrix, the entries of L and D satisfy the recursion : $d_i = M_{i,i} - d_{i-1}\ell_{i,i-1}^2$, $d_0 = 0$, $\ell_{i+1,i} = A_{i+1,i}/d_i$. Then $F' = L^{-1}F$ satisfy $F_i' = F_i - F_{i-1}'\ell_{i,i-1}$, $F_0 = 0$ and a is obtained from back substitution by $a = F_i'/d_i - a_{i+1}\ell_{i+1,i}$, $a_{n+1} = 0$. The total operations needed are about 9n.

If the sparse system is singular or nearly singular, the LU decomposition may fail. Thus we would like to introduce the algorithm FAPIN which is able to solve singular systems.

5.2. The Algorithm FAPIN

The algorithm FAPIN developed by P.O. Frederickson first appeared in [26] was introduced in two dimensional case. We shall in this section introduce the theory of FAPIN in one dimension and in Section 6.3 demonstrate the ability of the algorithm FAPIN in solving singular linear systems.

Let $A : X \to Y$ be a large sparse linear operator. A linear operator $B : Y \to X$ is called an ρ-*approximate inverse* to A if the operator $I - AB : Y \to Y$ has a spectral radius $\rho < 1$. We shall make use of a *local* ρ-*approximate inverse* B to A to construct a ε-*approximate solution* $x \in X$ to the linear system

$$Ax = y \tag{5.2.1}$$

with the property

$$\|y-Ax\| < \varepsilon\|y\| \tag{5.2.2}$$

For a given $y \in Y$ and a given tolerance $\varepsilon < 1$.

Usually, the iterations

$$\begin{cases} r \leftarrow y - Ax \\ x \leftarrow x + Bx \end{cases} \tag{5.2.3}$$

are used to improve the approximate x^m. From (5.2.3), we have the m^{th} iterate r^m satisfies

$$r^m = (I - AB)^m r^0 \tag{5.2.4}$$

(see Chapter 3 of either Varga [58] or Young [63] for more detail).

Let $\Omega = \{i\}_{i=-n}^{n}$ be a set of $2n+1$ integer lattice points in $[0,1]$. Denote by X the space of real functions on Ω and let Y be a subspace of X. Let $A : X \to Y$, a linear operator, be a *1-local operator*, i.e.

$$(5.2.5) \qquad [(Ax)_i \neq 0] \Rightarrow [\exists\, j \in \Omega \quad |i-j| \leq 1 \text{ and } x_j \neq 0]$$

(cf. Definition 3.3.1)

The number of points in Ω is $N = 2n+1$, hence X is isomorphic to R^N, and any 1-local operator A is represented through this isomorphism by a tri-diagonal matrix having at most three nonzero elements in each row, for example, the matrix M arising from the linear finite element (RRG) approximation is a discrete 1-local operator. Corresponding to every 1-local operator $A : X \to Y$ there is an array $A_{i,j}$ such that for any point i.

$$(5.2.6) \qquad (Ax)_i = \sum_{|j|\leq 1} A_{i,j} x_{i+j}$$

Implementation of (5.2.6) allows storage of A in 3N locations and evaluation of Ax in 3N multiplications.

If an approximate inverse B to A is also 1-local, the representation (5.2.6) will be used for B as well as A. We shall discuss some techniques other than the DB_q method, developed by Benson [3],

in Section 5.3 to obtain approximate inverses of A .

Assume that we have subdivided $[0,1]$ into 2^{ℓ} number of subinterval such that $N = 2^{\ell}+1$, i.e. $|i| \le 2^{\ell-1} \quad \forall\, i \in \Omega$. We write Ω^{ℓ} for Ω and define, using the recurrence

$$\Omega^{k-1} = \{i \mid \exists\, j \ , \ |j| \le 1 \ , \ 2i+j \in \Omega^{k}\} \ , \tag{5.2.8}$$

the set Ω for $1 \le k \le \ell$, we note that $|i| \le 2^{k-1}$ if $i \in \Omega^{k}$ Denote by X^{k} the linear space of real functions of Ω^{k} and define the *collection operator* $P^{k} : X^{k} \to X^{k-1}$ by

$$r_i^{k-1} \leftarrow \sum_j t_j \, r_{2i+j}^{k} \tag{5.2.9}$$

where the coefficients t_j are binomial coefficients

$$t_j = \binom{2}{j+1} \tag{5.2.10}$$

Hence we have

$$\begin{cases} t_{-1} = t_1 = 1 \\ t_0 = 2 \\ t_j = 0 \qquad \text{otherwise} \end{cases}$$

and

$$r_i^{k-1} \leftarrow r_{2i-1}^{k} + 2r_{2i}^{k} + r_{2i+1}^{k}$$

We then use the same coefficients to define the sequence of *interpolation operators* $Q^{k} : X^{k-1} \to X^{k}$ through

$$x_i^k \leftarrow \sum_j \frac{t_{i-2j}}{2}\, x_j^{k-1} \tag{5.2.11}$$

By this definition, with (4.2.10), we have

$$\begin{cases} x_{2i}^k \leftarrow x_i^{k-1} \\ x_{2i-1}^k \leftarrow \frac{1}{2}\,[x_{i-1}^{k-1} + x_i^{k-1}] \end{cases}$$

We define the subspace Y^k of X^k by $Y^k = P^{k+1}(Y^{k+1})$, beginning with $Y^\ell = Y$ and we define the sequence of operators $A^k : X^k \to Y^k$ by

$$A_{i,j}^{k-1} \leftarrow \frac{1}{2} \sum_u \sum_v t_u\, A^k_{2i+u,2j+v}\, t_v \tag{5.2.12}$$

By this definition, we have

$$A_{i,j}^{k-1} = \begin{cases} \frac{1}{2} A^k_{2i-1,2j-1} + A^k_{2i-1,2j} + \frac{1}{2} A^k_{2i-1,2j+1} \\ + A^k_{2i,2j-1} + 2A^k_{2i,2j} + A^k_{2i,2j+1} \\ + \frac{1}{2} A^k_{2i+1,2j-1} + A^k_{2i+1,2j} + \frac{1}{2} A^k_{2i+1,2j+1} \end{cases} \qquad |j-i| \le 1$$

It can be easily verified that if A is of the form of (2.1.13), then A^k, $2 \le K \le \ell$, is also of the same form.

Implementation of P^k requires less than n multiplications and additions, and Q^k requires only n multiplications and n additions if it is well coded. Similarly, construction of all the operators A^k from the given $A^\ell = A$ requires about 3n multiplications and 12n additions.

We shall prove Theorem 1 in [26, pp.7] concerning the best approximation properties of FAPIN .

Theorem 5.2.1. (Frederickson [26], pp.7)

The operator A^{k-1} defined by equation (5.2.12) satisfies the identity

$$A^{k-1} = P^k A^k Q^k \tag{5.2.13}$$

and is the Rayleigh-Ritz-Galerkin best approximation to A^k in the subspace $U^k = Q^k(X^{k-1})$ of X^k

Proof :

The proof of equation (5.2.13) involves comparing (5.2.12) with the expression which results when the right hand side is expanded, using (5.2.11), (5.2.6) and then (5.2.9) .

To show that A^{k-1} is the RRG best approximation to A^k in the subspace $U^k = Q^k(X^{k-1})$ of X^k , we need to show the following :

(i) Given a $r^k \in X^k$, find $x^k \in X^k$ such that

$$A^k x^k = r^k \tag{5.2.14}$$

(ii) Find a $x^k \in X^k$ which minimizes the quadratic functional

$$\Phi(x^k) = \langle Ax^k - 2r^k, x^k \rangle \tag{5.2.15}$$

where $\langle x^k, y^k \rangle = \sum_{i=-n}^{n} x_i^k y_i^k$

(iii) Let $U^k = Q^k(X^{k-1})$, then find $z^k \in U^k$ which minimizes Φ .

Rewrite (5.2.11) for $Q^k : X^{k-1} \to X^k$ as

$$(5.2.16) \qquad x_j^k = (Q^k x^{k-1})_j = \sum_{i=-2^{k-2}}^{2^{k-2}} \phi_i^k(j) x_i^{k-1}$$

where

$$(5.2.17) \qquad \phi_i^k(j) = \begin{cases} 1 & \text{if } j = 2i \\ \frac{1}{2} & \text{if } |j-2i| = 1 \\ 0 & \text{otherwise} \end{cases} \qquad i = -2^{k-2}, \cdots, 0, \cdots, 2^{k-2}$$

From (4.2.15) and assume that $A^k : X^k \to X^k$ is symmetric positive definite, then we have

$$(5.2.18) \quad \Phi(x^k+\varepsilon v^k) = \Phi(x^k) + 2\varepsilon\langle A^k x^k - r^k, v^k\rangle + \varepsilon^2\langle A^k v^k, v^k\rangle$$

Thus $\Phi(x^k)$ is a minimum means that $\delta\Phi(x^k, v^k) = 0$ i.e.

$$(5.2.19) \qquad \langle Ax^k - r^k, v^k\rangle = 0 \quad \text{for any } v^k \in X^k$$

we will show (ii) iff (i), i.e.

(ii) - $x_*^k \in X^k$ which minimizes Φ in (5.2.15)

iff $\langle Ax_*^k - r^k, v^k\rangle = 0 \quad \forall\, v^k \in X^k$

iff $A^k x_*^k = r^k$ -(i)

$P^k : X^k \to X^{k-1}$ in (5.2.9) can be written as

$$(5.2.20) \qquad r_i^{k-1} = 2 \sum_{j=-2^{k-2}}^{2^{k-2}} \phi_i^k(j) r_j^k = 2\langle \phi_i^k, r^k\rangle$$

(iii) - $Az^k \in U^k = Q^k(X^{k-1})$ which minimizes Φ over U^k

iff $\quad \langle A^k z^k - r^k, \phi_i^k \rangle = 0 \quad \forall\ i = -2^{k-1}, \cdots, 0, \cdots, 2^{k-1}$

iff $\quad \langle A^k \sum_{j=-2^{k-2}}^{2^{k-2}} \phi_j^k z_j^{k-1} - r^k, \phi_i^k \rangle = 0$

iff $\quad \sum_{j=-2^{k-2}}^{2^{k-2}} z_j^{k-1} \langle A^k \phi_j^k, \phi_i^k \rangle = \langle r^k, \phi_i^k \rangle$

iff $\quad 2 \sum_{j=-2^{k-2}}^{2^{k-2}} z_j^{k-1} \langle A^k \phi_j^k, \phi_i^k \rangle = P^k r^k(i)$

iff $\quad \sum_{j=-2^{k-2}}^{2^{k-2}} A_{i,j}^{k-1} z_j^{k-1} = P^k r^k(i)$

iff $\quad A^{k-1} z^{k-1} = P^k r^k$

This completes the proof.

FAPIN should be viewed as an interative algorithm. At the beginning of each pass we have an approximate x to the solution to equation (5.2.1), which may or may not be zero during the first pass, and we have evaluated the residual vector $r \leftarrow y - Ax$. The pass really begins when we apply (5.2.9) repeatedly, creating $r^{\ell-1}, \cdots, r^{\ell}$ from $r^{\ell} = r$. Next $x^1 = B^1 r^1$ is computed, and then we work back up from $k = 2$ to $k = \ell - 1$, first interpolating and then refining this approximation :

(5.2.21)
$$\begin{cases} x^k \leftarrow Q^k x^{k-1} \\ x^k \leftarrow x^k + B^k (r^k - A^k x^k) \end{cases}$$

At the top level, $k = \ell$, these assignments are replaced by

$$\text{(5.2.22)} \qquad \begin{cases} x^{\ell} \leftarrow x^{\ell} + Q^{\ell} x^{\ell-1} \\ x^{\ell} \leftarrow x^{\ell} + B^{\ell}(y - A^{\ell} x^{\ell}) \end{cases}$$

A Subroutine FAPIN is given in Appendix 1. We make use of some other Subroutines to perform individual tasks. We shall describe a pass of FAPIN as follows with the Subroutine's name, which is used to perform the task, on the left.

$$\text{OP :} \qquad r^{\ell} \leftarrow y^{\ell} - A^{\ell}(x^{\ell})$$

$$\begin{cases} \text{P :} & \text{DO} \quad k = \ell, \ell-1, \cdots, 2 \\ & \quad r^{k-1} \leftarrow P^{k}(r^{k}) \end{cases}$$

$$\text{OP :} \qquad x^{1} \leftarrow B^{1}(r^{1})$$

$$\begin{cases} & \text{DO} \quad k = 2, 3, \cdots, \ell-1 \\ \text{Q :} & \quad x^{k} \leftarrow Q^{k}(x^{k-1}) \\ \text{OP :} & \quad r^{k} \leftarrow r^{k} - A^{k}(x^{k}) \\ \text{OP :} & \quad x^{k} \leftarrow x^{k} + B^{k}(r^{k}) \end{cases}$$

$$\begin{aligned} \text{Q :} &\qquad x^{\ell} \leftarrow x^{\ell} + Q^{\ell}(x^{\ell-1}) \\ \text{OP :} &\qquad r^{\ell} \leftarrow y^{\ell} - A^{\ell}(x^{\ell}) \\ \text{OP :} &\qquad x^{\ell} \leftarrow x^{\ell} + B^{\ell}(r^{\ell}) \end{aligned}$$

From the algorithm above, it shows clearly that FAPIN solves certain subprograms $A^k x^k = y^k$, $1 \le k \le \ell$. In order not to waste storages we store x^k and r^k in the arrays X and R, respectively. In particular, we creat two arrays of structure constants NK and MK such that MK is the dimensional of X at the level K and NK is the position of the first element of x^k in the array X. Thus we store x_i^k as X(NK+i). A subroutine STRUCT to construct these structure constants is listed in Appendix 1.

5.3. Approximate Inverses

During the numerical experiments in Chapter 6, we found that an approximate inverse B_1 of A, for the one dimensional case, obtained by the DB_q technique developed by Benson [3] is not efficient enough in the sense that the covergence of the residual r in (5.2.3) is very slow. This forces us to seek for other better approximate inverses to A . An approximate inverse B_1 of A is said to be better than an approximate inverse B_k to A if $\rho(I-B_1A) < \rho(I-B_kA)$. Following Varga [58] and Young [63], we have, for a matrix G,

$$\rho(G) = \lim_{m\to\infty} (\|G^m\|_2)^{1/m} \tag{5.3.1}$$

However, from [7, pp.269] and [58], we have

$$\frac{\|r^{m+1}\|_2}{\|r^m\|_2} \to \rho(I-B_pA) \tag{5.3.2}$$

This provides us a mean to approximate the spectral radius of $I-B_pA$. We shall also employ an algorithm of Shanks [53] to smooth the sequence $\{||r^{m+1}||_2 \;/\; ||r^m||_2\}$ and predict the limit of the sequence. This limit is taken to be an approximate spectral radius of $I-B_pA$.

For simplicity, we shall consider A to be a tridiagonal symmetric matrix with constant band elements.

Let $\hat{A}$ be the set of all tridiagonal symmetric matrix with constant band elements. Then every element $A \in \hat{A}$ can be written as (cf. Benson [3], pp.14)

$$A = (a_1 \; a_2 \; a_1) \tag{5.3.3}$$

For $A,B \in \hat{A}$, define an *convolution* operator $*$ by

$$\begin{aligned} A*B &= (a_1 \; a_2 \; a_1)*(b_1 \; b_2 \; b_1) \\ &= (a_1b_1 \;\; a_2b_1+a_1b_2 \;\; 2a_1b_1+a_2b_2 \;\; a_1b_2+a_2b_1 \;\; a_1b_1) \end{aligned}$$

(Note that $A*B \notin \hat{A}$)

The DB_q techniques [3] can be viewed as

$$B_1*A = I \tag{5.3.4}$$

i.e. we truncate B_1*A and solve the linear system :

$$\begin{cases} a_2b_{11} + a_1b_{12} = 0 \\ 2a_1b_{11} + a_2b_{12} = 1 \end{cases}$$

and obtain

$$\begin{cases} b_{11} = -\dfrac{a_1}{a_2^2-2a_1^2} \\ b_{12} = \dfrac{a_2}{a_2^2-2a_1^2} \end{cases}$$

In particular, for $A = (-1\ 2\ -1)$, we have $B_1 = (.5\ 1\ .5)$

Initiated by the DB_q technique, we have a generalization as follows : an approximate inverse B_W of A is evaluated through

(5.3.5) $$B_W * (A*W) = W$$

where $W \in A$ is a weight.

In particular, we shall consider (i) $W = A*A$ and (ii) $W = A*A*A$

Let B_2 be an approximate inverse of A evaluated through

(5.3.6) $$B_2 * (A*A) = A$$

Then we need to solve the following equations

$$\begin{cases} b_1p_1 + b_2p_2 + b_1p_3 = a_1 \\ b_1p_2 + b_2p_3 + b_1p_2 = a_2 \end{cases}$$

where

$$\begin{cases} p_1 = a_1^2 \\ p_2 = 2a_1a_2 \\ p_3 = 2a_1^2 + a_2^2 \end{cases}$$

In particular, if $A = (-1\ 2\ -1)$, then $B_2 = (0.2\ 0.6\ 0.2)$ we shall call this DB^2 techniques.

Let B_3 be an approximate inverse evaluated through

(5.3.7) $$B_3 * (A*(A*A)) = A * A$$

In this case, we need to solve a linear system

$$\begin{cases} b_1s_2 + b_2s_3 + b_1s_4 = p_2 \\ b_1s_3 + b_2s_4 + b_1s_3 = p_3 \end{cases}$$

where

$$\begin{cases} s_2 = a_2p_1 + p_2a_1 \\ s_3 = a_1p_1 + a_2p_2 + a_1p_3 \\ s_4 = a_1p_2 + a_2p_3 + a_1p_2 \end{cases}$$

$$\begin{cases} p_1 = a_1^2 \\ p_2 = 2a_1a_2 \\ p_3 = 2a_1^2 + a_2^2 \end{cases}$$

In particular, if $A = (-1\ 2\ -1)$, then $B_3 = (\frac{15}{105}\ \frac{54}{105}\ \frac{15}{105})$ we shall call this DB^3 method.

For a fixed matrix A , after we have obtained an approximate inverse, we could improve this approximate inverse by interpolation technique [42] to obtain the optimal approximate inverse B^* . We found, for $A = (-1\ 2\ -1)$, $B^* = (0.125\ 0.5\ 0.125)$ is quite close to the optimal. There are many ways of generalization. We may try, for a fixed A , varying the weight W in (5.3.5) , or we may, do a bit more calculation, consider a B_ℓ such that $\|B_\ell * (A*W) - W\|_2$ is minimum over A . However, we found that B_3 is good enough for our experiments in Chapter 6.

The following tables give a comparison of DB_q , DB^1 and DB^2 methods, in which we take

$$\text{(i)}\quad A_1 = (-1\ 2\ -1)$$

$$\text{(ii)}\quad A_2 = \frac{1}{h}(-1\ 2\ -1) - \frac{h}{6}\lambda_1(1\ 4\ 1)\ ,$$

where $\lambda_1 = 6(1 - \cos(\pi h)) / (2 + \cos(\pi h))h^2$. A_2 is a singular matrix which is arisen from the solving of an eigenvalue problem (cf. 6.3).

In the tables, L is the level of partition, i.e. we partition $[0,1]$ into $n = 2^L$ subintervals.

	DB_q	DB^1	DB^2
L	$\rho(I-B_1A_1)$	$\rho(I-B_2A_1)$	$\rho(I-B_3A_1)$
3	1.00	0.20	0.11
4	0.93	0.19	0.11
5	0.98	0.20	0.11
6	0.97	0.20	0.12
7	0.97	0.19	0.12

Table 5.3.1

	DB_q	DB^1	DB^2
L	$\rho(I-B_1A_2)$	$\rho(I-B_2A_2)$	$\rho(I-B_3A_2)$
3	1.41	0.34	0.074
4	1.11	0.34	0.092
5	0.95	0.34	0.12
6	0.98	0.33	0.12
7	0.97	0.33	0.12

Table 5.3.2

CHAPTER 6

NUMERICAL EXPERIMENTS

The numerical experiments in this chapter are carried in double precision arithmetic on the IBM/360 computer at Lakehead University. These experiments consist of (i) the verification of the rate of convergence of the RRG solution and the Global Superconvergence, (ii) the numerical evaluation of the Generalized Peano kernel function and (iii) the ability of the algorithm FAPIN in solving singular sparse systems.

Throughout the chapter L will be the level of partition i.e. we partition $[0,1]$ into $n=2^L$ number of subdivisons.

6.1. Verifications of the Rate of Convergence of the RRG Solution and the Global Superconvergence

In this experiment we solve the TPBVP : $-u''(x) = f(x)$ with $u(0) = u(1) = 0$. The RRG solution u_h is from $S_{h,0}^{1,0}$. We set $u(x) = (1-x)(1-e^x)$ and compute the H_1-norm, H_0-norm and the energy norm of the error $e_h = u-u_h$. We also construct the superconvergence approximation $s \in S_{h,0}^{3,2}$ and the various norms of the error $e_s = u-s$ are computed.

The results are listed in Table 6.1.1, Table 6.1.2 and Table 6.1.3 .

In the tables, the numbers under the name "Rate of Convergence" are computed as follows :

$$\text{Rate of convergence} = \frac{\log\left(\frac{\|\cdot\| \text{ at level } k}{\|\cdot\| \text{ at level } k+1}\right)}{\log 2}$$

Fig. 6.1.1 is a comparision of the RRG solution and the Global Superconvergence solution.

Table 6.1.1

L	$\|e_h\|_1$	$\|e_s\|_1$	Rate of Convergence	
			RRG	S
2	0.2153898(00)	0.7257666(-2)		
3	0.1080515(00)	0.7153933(-3)	0.9952310(00)	0.3342696(01)
4	0.5407050(-1)	0.6823556(-4)	0.9988053(00)	0.3390140(01)
5	0.2704085(-1)	0.6322846(-5)	0.9997007(00)	0.3431872(01)
6	0.1352113(-1)	0.5790591(-6)	0.9999250(00)	0.3448795(01)
7	0.6760649(-2)	0.5322389(-7)	0.9999814(00)	0.3443563(01)
8	0.3380336(-2)	0.4985981(-8)	0.9999945(00)	0.3416124(01)
9	0.1690169(-2)	0.4841845(-9)	0.9999986(00)	0.3364247(01)

From Table 6.1.1 , it shows that the rate of convergence of e_h and e_s in the H_1-norm is about of order $O(h^1)$ and $O(h^3)$ respectively.

Table 6.1.2

L	$\|e_h\|_A$	$\|e_s\|_A$	Rate of Convergence	
			RRG	S
2	0.2147222(00)	0.7229935(-2)		
3	0.1079672(00)	0.7140480(-3)	0.9918782(00)	0.3339890(01)
4	0.5405994(-1)	0.6817548(-4)	0.9979609(00)	0.3388695(01)
5	0.2703953(-1)	0.6320235(-5)	0.9994894(00)	0.3431202(01)
6	0.1352096(-1)	0.5789451(-6)	0.9998720(00)	0.3448479(01)
7	0.6760631(-2)	0.5321901(-7)	0.9999676(00)	0.3443412(01)
8	0.3380334(-2)	0.4985775(-8)	0.9999917(00)	0.3416051(01)
9	0.1690169(-2)	0.4841771(-9)	0.9999979(00)	0.3364210(01)

From Table 6.1.2, it shows that rate of convergence of e_h and e_s in the energy norm is about of order $O(h^1)$ and $O(h^3)$ respectively.

Table 6.1.3

L	$\|e_h\|_0$	$\|e_s\|_0$	Rate of Convergence	
			RRG	S
2	0.1694562(-1)	0.6338867(-3)		
3	0.4265890(-2)	0.4385140(-4)	0.1989993(01)	0.3853415(01)
4	0.1068335(-2)	0.2862917(-5)	0.1997482(01)	0.3937046(01)
5	0.2672004(-3)	0.1823095(-6)	0.1999368(01)	0.3973023(01)
6	0.6680739(-4)	0.1149049(-7)	0.1999842(01)	0.3987877(01)
7	0.1670231(-4)	0.7209842(-9)	0.1999958(01)	0.3994328(01)
8	0.4175605(-5)	0.4511518(-10)	0.1999989(01)	0.3998281(01)
9	0.1043903(-5)	0.2687543(-11)	0.1999995(01)	0.4069252(01)

From Table 6.1.3, it shows that the rate of convergence of e_h and e_s in the H_0-norm is about of order $O(h^2)$ and $O(h^4)$ respectively.

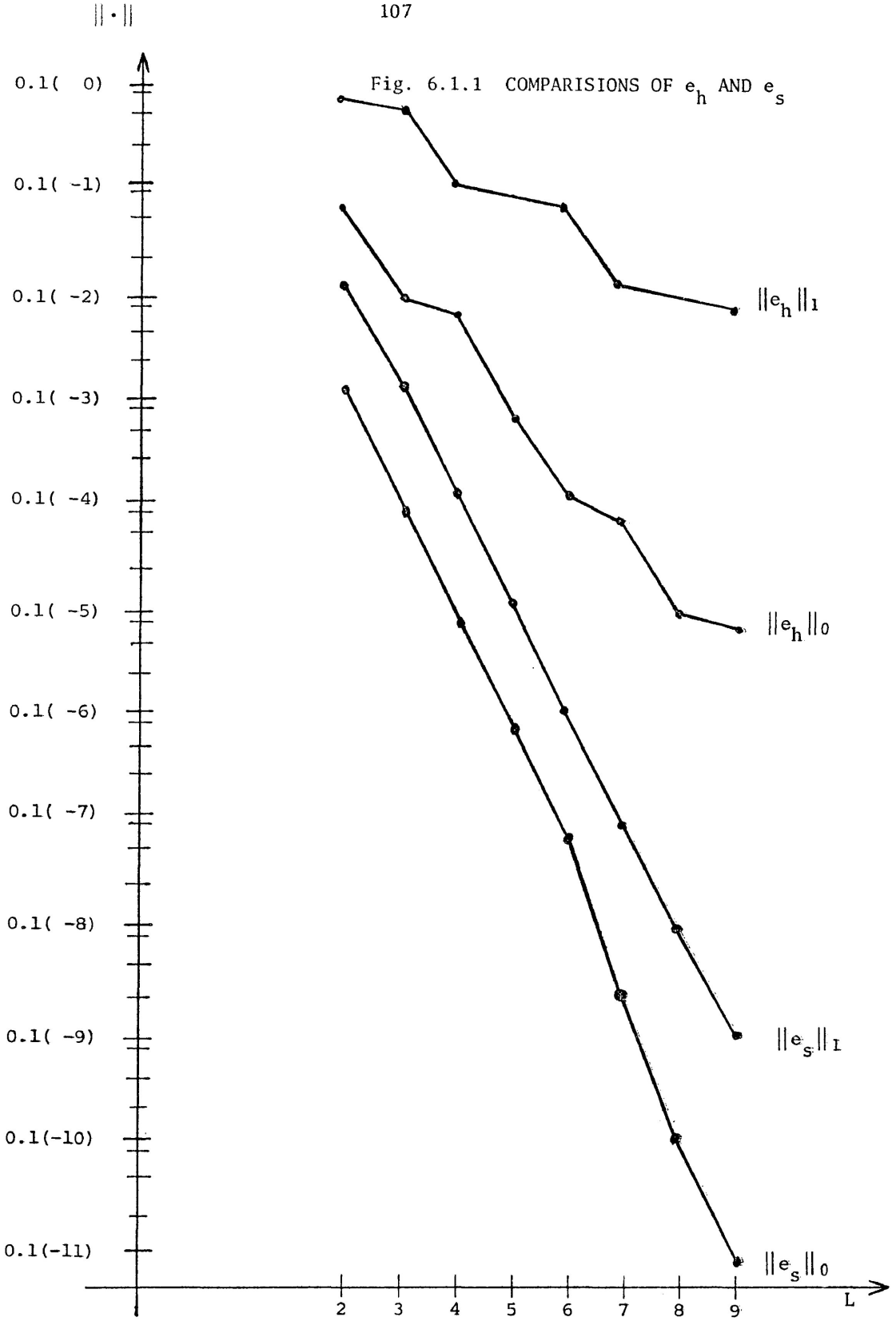

Fig. 6.1.1 COMPARISIONS OF e_h AND e_s

6.2. Numerical Evaluation of the Generalized Peano Kernel Function

In Section 3.3, we constructed an 3-exact Global Superconvergence S, i.e. $S(p) = p \quad \forall\, p \in \mathcal{P}_0^3$. In Section 4.5, we have discussed the rate of convergence of the Global Superconvergence. In this section, we shall evaluate the generalized kernel $K_2(\xi,t)$ (cf. Section 4.5) and then evaluate the quantities $k_{\infty,\infty}$ and $k_{1,\infty}$.

The results are listed in Table 6.2.1

L	$k_{\infty,\infty}$	Rate of Convergence	$k_{1,\infty}$	Rate of Convergence
2	0.13500(-1)		0.28495(-4)	
3	0.59878(-2)	0.117(01)	0.55547(-5)	0.236(01)
4	0.29089(-2)	0.104(01)	0.14471(-5)	0.194(01)
5	0.14438(-2)	0.101(01)	0.36541(-6)	0.199(01)

Table 6.2.1

From Table 6.2.1, it seems that the rate of convergence of the Global Superconvergence, for $u \in H_0^3[0,1]$, is of order $O(h^2)$.
However, from the experiment in Section 6.1, it shows that, if u is sufficiently smooth, the rate of convergence is of order $O(h^4)$ in the H_0-norm. It seems to suggest that the Global Superconvergence is of order $O(h^4)$ for a certain class of functions. Now the question remains is : "what is the class of functions for which the Global Superconvergence has an order $O(h^4)$ convergence rate?"

6.3. The Ability of the Algorithm FAPIN in Sovling Singular Systems

Consider the TPBVP $-u'' + qu = f$, $u(0) = u(1) = 0$.

If $q = -\lambda_1 = 6(1-\cos(\pi k/n))/(2+\cos(\pi k/n))h^2$, then the TPBVP becomes an eigenvalue problem. The linear sparse system arising from the linear finite element method will be singular.

Table 6.3.1 and Table 6.3.2 show the various norms of the errors when we solve the TPBVP's by LU decomposition and FAPIN with q varying around $-\lambda_1$, $u(x) = x(1-x)(1-2x)$ at levels $L = 7$ and $L = 10$

q	LU		FAPIN	
	$\|e\|_1$	$\|e\|_0$	$\|e\|_1$	$\|e\|_0$
-0.98700998592898(1)	0.183(-1)	0.503(-2)	0.781(-2)	0.249(-4)
-0.98700998592908(1)	0.184(-1)	0.506(-2)	0.781(-2)	0.249(-4)
-0.98700998592918(1)	0.287(-1)	0.837(-2)	0.781(-2)	0.249(-4)
-0.98700998592928(1)	0.278(-1)	0.807(-2)	0.781(-2)	0.249(-4)
-0.98700998592938(1)	0.338(-1)	0.999(-2)	0.781(-2)	0.249(-4)
-0.98700998592948(1)	0.398(-1)	0.119(-1)	0.781(-2)	0.249(-4)
-0.98700998592958(1)	0.568(-1)	0.171(-1)	0.781(-2)	0.249(-4)
-0.98700998592968(1)	0.280(00)	0.850(-1)	0.781(-2)	0.249(-4)
-0.98700998592978(1)	0.721(00)	0.219(00)	0.781(-2)	0.249(-4)
-0.98700998592988(1)	0.683(-1)	0.206(-1)	0.781(-2)	0.249(-4)

Table 6.3.1 , $L = 7$ (cf. Fig. 6.3.1 and Fig. 6.3.2)

$-\lambda_1 = -0.98700998592948(1)$

q	LU		FAPIN	
	$\|\|e\|\|_1$	$\|\|e\|\|_0$	$\|\|e\|\|_1$	$\|\|e\|\|_0$
-0.98696121375(1)	0.282(-2)	0.558(-3)	0.977(-3)	0.389(-6)
-0.98696121385(1)	0.259(-2)	0.727(-3)	0.977(-3)	0.389(-6)
-0.98696121395(1)	0.343(-2)	0.997(-3)	0.977(-3)	0.389(-6)
-0.98696121405(1)	0.464(-2)	0.138(-2)	0.977(-3)	0.389(-6)
-0.98696121415(1)	0.835(-2)	0.252(-2)	0.977(-3)	0.389(-6)
-0.98696121425(1)	0.480(-1)	0.146(-1)	0.977(-3)	0.389(-6)
-0.98696121435(1)	0.113(-1)	0.341(-2)	0.977(-3)	0.389(-6)
-0.98696121445(1)	0.534(-2)	0.159(-2)	0.977(-3)	0.389(-6)
-0.98696121455(1)	0.349(-2)	0.101(-2)	0.977(-3)	0.389(-6)
-0.98696121465(1)	0.265(-2)	0.745(-3)	0.977(-3)	0.389(-6)
-0.98696121475(1)	0.220(-2)	0.599(-3)	0.977(-3)	0.389(-6)

Table 6.3.2 , L = 10 (cf. Fig. 6.3.3 and Fig. 6.3.4)

$-\lambda_1 = -0.98696121425(1)$

Table 6.3.1 and Table 6.3.2 show clearly that the algorithm FAPIN is able to solve singular or nearly singular systems.

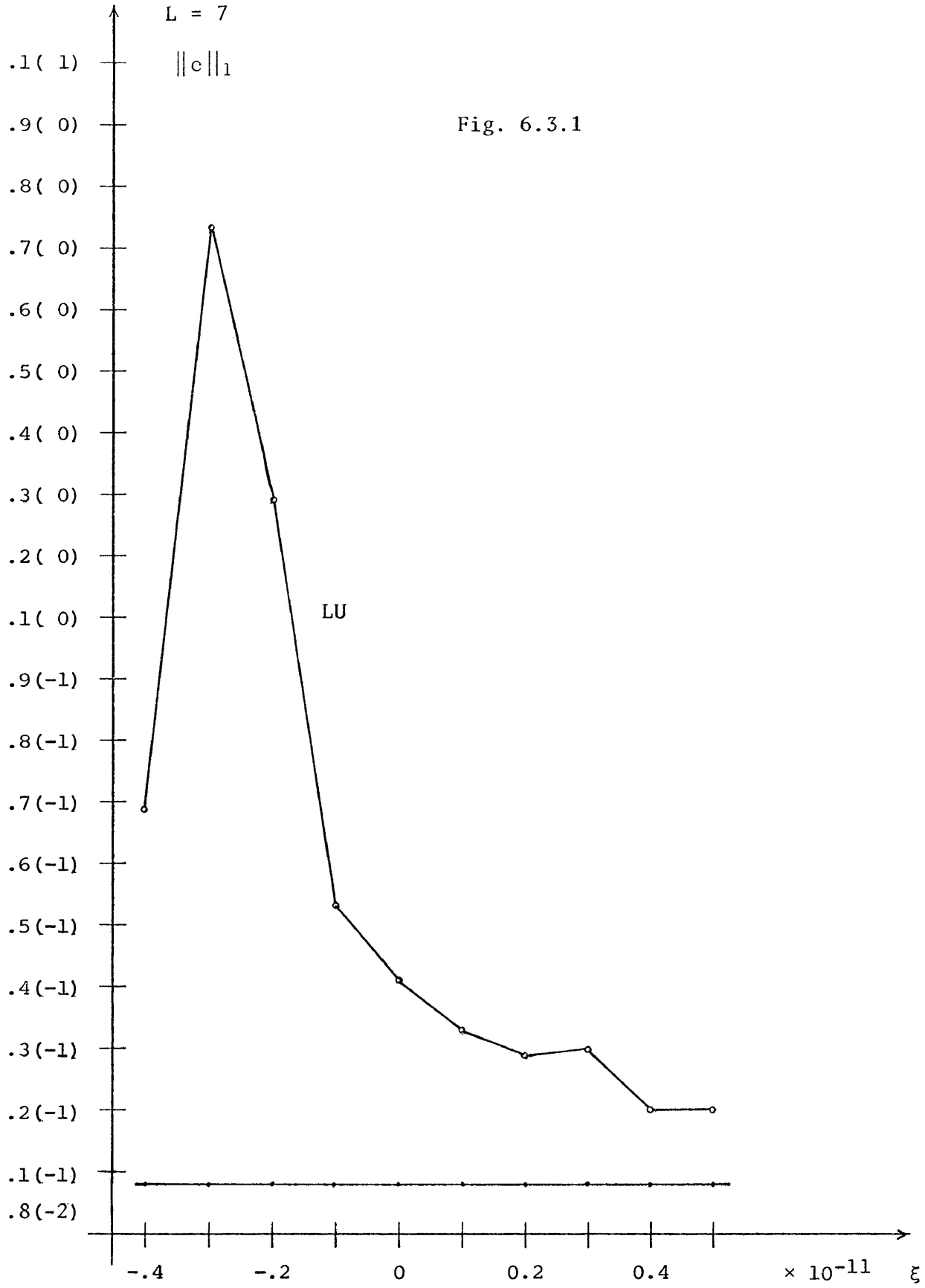

Fig. 6.3.1

$q = q_0 + \xi$; $q_0 = -0.98700998592948(1)$

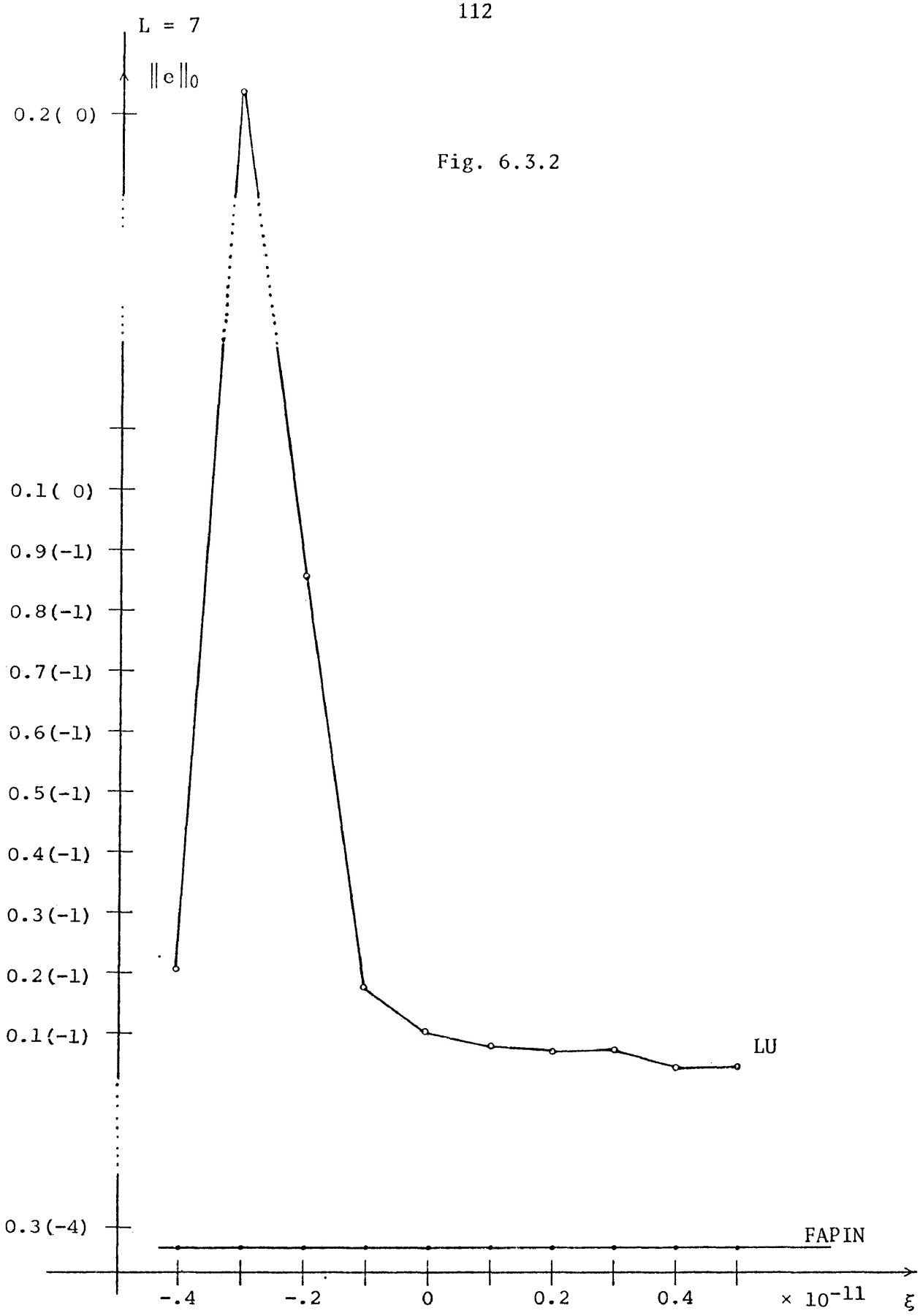

Fig. 6.3.2

$q = q_0 + \xi$; $q_0 = -0.98700998592948(1)$

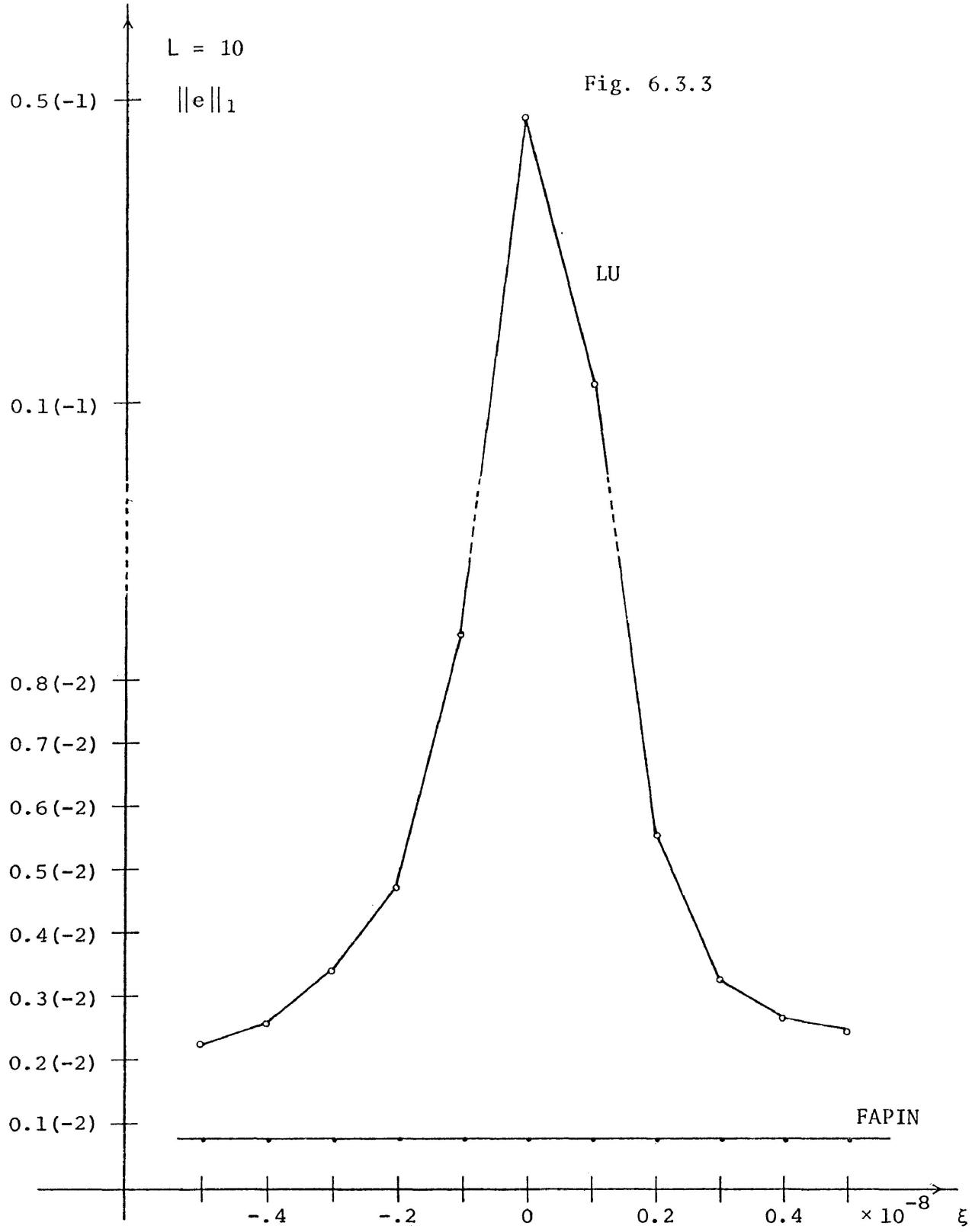

Fig. 6.3.3

q = -0.98696121425(1) + ξ

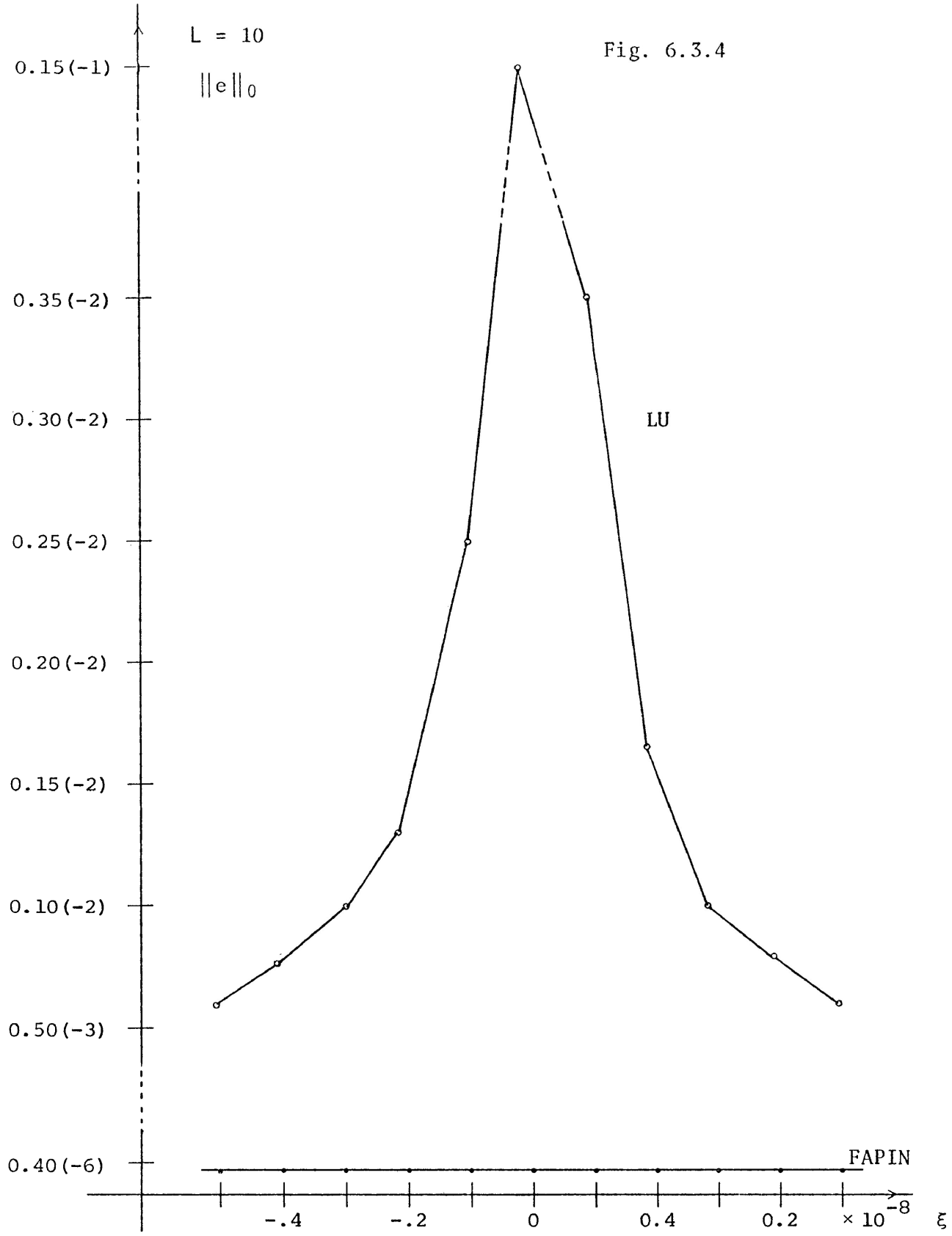

Fig. 6.3.4

q = -0.98696121425(1) + ξ

CHAPTER 7

SUMMARY AND CONCLUSIONS

The concept of Sard's theory on best quadrature formulae is extended to the integration $T = \int_0^1 fw\,dx$. The form of quadrature Q for T which we consider is $Q = \sum_{i=1}^{n'} c_i f(x_i')$. This concept of best quadrature can be applied to the quadrature H , $H = \sum_{i=1}^{n''} a_i f(x_i'')w(x_i'')$, which is considered in [34]. Furthermore, the best quadratures which we derived are consistent with the energy norm.

A quasi-inverse of the finite element method is introduced to obtain a global superconvergence solution of the TPBVP. This global superconvergence technique, as we suggested, can be applied to other methods such as the collocation method.

A generalization of the Peano kernel theorem is useful in the error analysis on the solution of TPBVP. By applying this generalized Peano kernel theorem,the rate of convergence of a 3-exact global superconvergence solution is of order $O(h^2)$. But from the experimental results, for some smooth u, the rate of convergence of the global superconvergence is of order $O(h^4)$. It still remains a question: "what is the class of functions for which the global superconvergence has an order $O(h^4)$ convergence rate?"

The algorithm FAPIN has also been used successfully in this

thesis. However, the LU decomposition is perfered if the linear system is far from singular. The approximate inverses, which used in conjunction with the algorithm FAPIN, derived in this thesis have smaller spectral radius than the DB_q approximate inverse [3]. Other techniques to obtain approximate inverses are suggested in Section 5.3 .

APPENDIX 1

In this Appendix, we list a subroutine of the algorithm FAPIN with all the relevant subroutines which have been introduced in Section 5.2.

Programs are written in WATFIV .

```
      SUBROUTINE FAPIN(F,U,R,A,C,DF,DU)
      INTEGER STR(16,2)
      COMMON STR,L
      INTEGER DF,DU
      REAL*8 F(DU),R(DU),U(DU),A(DU,3),C(DU,3)
      CALL OP(R,F,A,U,L,DF,DU,.TRUE.)
      CALL P(DF,DU,L,R)
      K=1
      CALL OP(U,U,C,R,K,DF,DU,.FALSE.)
    4 K=K+1
      IF (K.EQ.L) GO TO 5
      CALL Q(DF,DU,K,U)
      CALL OP(R,R,A,U,K,DF,DU,.TRUE.)
      CALL OP(U,U,C,R,K,DF,DU,.TRUE.)
      GO TO 4
    5 CALL Q(DF,DU,K,U)
      CALL OP(R,F,A,U,K,DF,DU,.TRUE.)
      CALL OP(U,U,C,R,K,DF,DU,.TRUE.)
      RETURN
      END

      SUBROUTINE OP(X,Y,A,Z,K,DF,DU,FLAG)
      INTEGER STR(16,2)
      COMMON STR,L
      LOGICAL FLAG
      INTEGER DF,DU
      REAL*8 A(DU,3),X(DU),Y(DU),Z(DU)
      NK=STR(K,1)
      MK2=STR(K,2)-2
      IF (FLAG) GO TO 100
      DO 50 I=1,MK2
      X(NK+I)=-A(NK+I,1)*Z(NK+I-1)-A(NK+I,2)*Z(NK+I)-A(NK+I,3)*Z(NK+I+1)
   50 CONTINUE
      RETURN
  100 DO 150 I=1,MK2
      X(NK+I)=Y(NK+I)-A(NK+I,1)*Z(NK+I-1)-A(NK+I,2)*Z(NK+I)-A(NK+I,3)*Z(
     *NK+I+1)
  150 CONTINUE
      RETURN
      END
```

```
      SUBROUTINE P(DF,DU,LL,R)
      INTEGER STR(16,2)
      COMMON STR,L
      INTEGER DF,DU
      REAL*8 R(DU)
      K=LL
    1 J=K-1
      NK=STR(K,1)
      NJ=STR(J,1)
      MJ2=STR(J,2)-2
      DO 2 I=1,MJ2
    2 R(NJ+I)=R(NK+2*I-1)+2.0*R(NK+2*I)+R(NK+2*I+1)
      K=K-1
      IF (K.GE.2) GO TO 1
      RETURN
      END



      SUBROUTINE Q(DF,DU,K,U)
      INTEGER STR(16,2)
      COMMON STR,L
      INTEGER DF,DU
      REAL*8 U(DU)
      J=K-1
      MK2=STR(K,2)-2
      NK=STR(K,1)
      NJ=STR(J,1)
      MJ2=STR(J,2)-2
      IF (K.EQ.L) GO TO 100
      U(NK+1)=0.0
      DO 50 I=1,MJ2
      U(NK+2*I-1)=U(NK+2*I-1)+.5*U(NJ+I)
      U(NK+2*I)=U(NJ+I)
   50 U(NK+2*I+1)=.5*U(NJ+I)
      RETURN
  100 DO 150 I=1,MJ2
      U(NK+2*I-1)=U(NK+2*I-1)+.5*U(NJ+I)
      U(NK+2*I)=U(NK+2*I)+U(NJ+I)
  150 U(NK+2*I+1)=U(NK+2*I+1)+.5*U(NJ+I)
      RETURN
      END



      SUBROUTINE STRUC(DF,DU)
      INTEGER STR(16,2)
      COMMON STR,L
      INTEGER DF,DU
      STR(L,1)=1
      L1=L-1
      DO 50 I=1,L1
      STR(L-I+1,2)=2**(L-I+1)+1
   50 STR(L-I,1)=STR(L-I+1,1)+STR(L-I+1,2)
      STR(1,2)=3
      DU=STR(1,1)+2
      DF=2**L+1
      RETURN
      END
```

The Subroutine AK is used to construct the matrices A^k, $1 \le k \le \ell$ (cf. Section 5.2). Note that in the Subroutine, the construction of A at level L is varied with different linear system Mu = F in 2.1.10.

```
      SUBROUTINE AK(DU,A,H,Q)
C----- TO FORM AK,THE BEST RRG APPROXIMATION TO AK+1
      INTEGER STR(16,2),DU
      COMMON STR,L
      REAL*8 A(DU,3)
      REAL*8 Q,DCOS,PI,H,DELTA
C----- CONSTRUCT A AT LEVEL L FIRST
      PI=3.1415926535897 93
      NL=STR(L,1)
      ML=STR(L,2)
      DO 11 I=1,ML
      NL1=NL+I-1
      A(NL1,1)=A(NL1,3)=-1.D0/H+Q*H/6.D0
      A(NL1,2)=2.D0/H+Q*H*4.D0/6.D0
   11 CONTINUE
C----- CONSTRUCT AK START FROM LEVEL L-1
      L1=L-1
      DO 111 KI=1,L1
      K=L-KI
      NK1=STR(K+1,1)
      NK=STR(K,1)
      MK=STR(K,2)
      MKS2=MK-2
      N1=NK+1
      K1=NK+MK-1
      K2=K1-1
      A(NK,1)=A(NK,2)=A(NK,3)=A(K1,1)=A(K1,2)=A(K1,3)=0.
      DO 112 I=1,MKS2
      NI=NK+I
      NK11=NK1+2*I-1
      NK12=NK1+2*I
      NK13=NK1+2*I+1
      A(NI,1)=A(NK11,1)+.5*A(NK11,2)+A(NK13,1)
      A(NI,2)=.5*(A(NK11,2)+A(NK13,2))
     1        +A(NK11,3)+A(NK12,1)+2.*A(NK12,2)+A(NK12,3)+A(NK13,1)
      A(NI,3)=A(NK12,3)+.5*A(NK13,2)+A(NK13,3)
  112 CONTINUE
  111 CONTINUE
      RETURN
      END
```

BIBLIOGRAPHY

[1] Alberg, J.H., Nilson, E.N. and Walsh, J.L. : " The Theory of Splines and their Applications ", Academic Press, New York, (1967).

[2] Babuska, I. and Aziz, A.K. : " Servey Lectures on the Mathematical Foundations of the Finite Element Method ", in The Mathematical Foundations of the Finite Element Method with Applications to PDE ", A.K. Aziz ed., Academic Press, (1972).

[3] Benson, M.W. : " Iterative Solution of Large Scale Linear Systems ", Thesis, Lakehead University, (1973).

[4] Birkhoff, G. : " The Numerical Solution of Elliptic Equations ", SIAM Regional Conference Series in Applied Mathematics, 1, (1971).

[5] Birkhoff, G., Schultz, M.H. and Varga, R.S. : " Piecewise Hermite Interpolation in One and Two Variables with Applications to PDE ", Numer. Math., 11, (1968), pp. 232-256.

[6] Birhkoff, G. and Rota, G.C. : " Ordinary Differential Equations ", 2nd ed., Blaisdell, Waltham, (1969).

[7] Bodewing, E. : " Matrix Calculus ", rev. ed., Interscience, New York, (1959).

[8] Ciarlet, P.G., Schultz, M.H. and Varga, R.S. : " Numerical Methods of High-order Accuracy for Non-linear Boundary Value Problems ", I. One Dimensional Problem, Numer. Math., 9, (1967), pp. 394-430.

[9] Collatz, L. : " The Numerical Treatment of Differential Equations ", 3rd ed., Springer, Berlin, (1966).

[10] Courant, R. and Hilbert, D. : " Methods of Mathematical Physics ", Interscience, New York, (1966).

[11] Curry, H.B. and Schoenberg, I.J. : " On Polya Frequency Functions IV : The Fundamental Spline Functions and their Limits ", J. d'Analyse Math., 17, (1966), pp. 71-107.

[12] Davis, P.J. : " Interpolation and Approximation ", Dover, New York, (1975).

[13] de Boor, C. : " Best Approximation Properties of Spline Functions of Odd Degree ", J. Math. Mech., 12, (1963).

[14] de Boor, C. and Fix, G.J. : " Spline Approximation by Quasi-interpolants ", J. Approx. Theory, 8, (1973), pp. 19-45.

[15] de Boor, C. : " The Quasi-interpolant as a Tool in Elementary Polynomial Spline Theory ", in " Approximation Theory ", G.G. Lorentz ed., Academic Press, New York and London, (1973).

[16] de Boor, C. and Swartz, B. : " Collocation at Gaussian Points ", SIAM J. Numer. Anal., Vol. 10, No. 4, (1973), pp. 582-606.

[17] de Boor, C. : " On Calculating with B-splines ", J. Approx. Theory, 6, (1972). pp. 50-62.

[18] de Boor, C. and Daniel, J.W. : " Splines with Nonnegative B-spline Coefficients ", Maths. of Comp., Vol. 28, No. 126, (1974), pp. 565-568.

[19] Douglas, J. and Dupont, T. : " Some Superconvergence results for Galerkin Methods for the Approximate Solution of Two-point Boundary Problems ", in " Topoics in Numerical Analysis ", Proceeding of The Royal Academy Conference on Numerical Analysis ", (1972), pp. 89-92.

[20] Douglas, J. and Dupont, T. : " Superconvergence for Galerkin Methods for the Two-point Boundary Value Problem via Local Projections ", Numer. Math., 21, (1973), pp. 270-278.

[21] Douglas, J., Dupont, T.: " Galerkin Approimations for the Two Point Boundary Problems Using Continuous, Piecewise Polynomial Spaces ", Numer. Math., 22, (1974). pp. 99-109.

[22] Douglas, J., Dupont, T. and Wahbin, L. : " Optimal L_∞ Error Estimates for Galerkin Approximations to Solutions of Two Point Boundary Value Problems ", Math. Comp., Vol. 29, No. 130, (1975), pp. 475-483.

[23] Duncan, W.J. : " The Principles of The Galerkin Method ", Rep. and Mem., No. 1848(3694), Aero. Res. Comm., (1938). 24pp..

[24] Fox, L. : " The Numerical Solution of Two-point Boundary Problems in Ordinary Differential Equations ", Oxford University, London, (1975).

[25] Frederickson. P.O. : " Quasi-interpolation, Extrapolation and Approximation on the Plane ", Proceedings of the Manitoba Conference on Numerical Mathematics, Oct 7-9, (1971), pp. 159-167.

[26] Frederickson, P.O. : " Fast Approximate Inversion of Large Sparse Linear Systems ", Math. Report #7, Lakehead University, (1975).

[27] Frederickson, P.O. : "Documentation Report For FAPIN3 ", Math. Report #8, Lakehead University, (1975).

[28] Frederickson, P.O. : " Fast Solution of Generalized Eigenvalue Problems ", Notices of The AMS, Jan. 1977, Issue 175. *742-65-23, A-172.

[29] Froberg, C.E. : " Introduction to Numerical Analysis ", 2nd ed., Addison-Wesley, Reading, (1969).

[30] Garlerkin, B.G. : " Reihenentwicklungen für einige Fälle des Gleichgewichts von Platten und Balken ", Wjestnik Ingenerow Petrograd, (1915), H.10, (Russian).

[31] Goël, J.J. : " Construction of Basic Functions for Numerical Utilisation of Ritz's Method ", Numer. Math., 12, (1968), pp. 435-447.

[32] Greville, T.N.E. : " Data Fitting by Spline Functions ", Trans. 12th Conf. Army. Mathematicians, Report 67-1, U.S.Army Research Office - Durham, Durham, N.C., (1967), pp. 65-90. Also available as MRC. Tech. Summ. Rpt. #893, Mathematics Research Center, U.S. Army, University of Wisconsin, Madison, Wisconsin, (1968).

[33] Greville, T.N.E. : " Introduction to Spline Functions ", in " Theory and Applications of Spline Functions ", T.N.E. Greville ed., Academic Press, (1969).

[34] Herbold, R.J., Schultz, M.H. and Varga, R.S. : " The Effect of Quadrature Errors in The Numerical Solution of Boundary Value Problems by Variational Techniques ", Aequ. Math., 3, (1969).

[35] Holand, I. and Bell, K. ed. : " Finite Element Methods in Stress Analysis ", Tapir, Trondheim, Norway, (1969).

[36] Holladay, J.C. : " A Smoothest Curve Approximation ", Math. Tables Aids Comp., 11, (1957), pp. 233-243.

[37] Kantorovich, L.V. and Krylov, V.I. : " Approximate Methods of Higher Analysis ", Noordhoff - Interscience, New York - Groningen, (1964).

[38] Keller, H.B. : " Numerical Methods for Two Point Boundary Value Problems ", Ginn-Blaisbell, Waltham, (1968).

[39] Keller, H.B. : " Numerical Solution of Two Point Boundary Value Problems ", SIAM Regional Conference Series in Applied Math., 24, (1976).

[40] Marchuk, G.I. : " Methods of Numerical Mathematics ", Springer-Verlag, (1975).

[41] Nitsche : " Ein Kriterium für die Quasi-Optimalitat des Ritz-schen Verfahrens ", Numer. Math., 11, (1968), pp. 346-348.

[42] Ong, H.L. : " Triangular Finite Element Solution to Boundary Value Problems ", Thesis, Lakehead University, (1977).

[43] Peano, G. : " Residue in Formulas de Quadrature ", Mathesis, 4, Vol. 34, (1914), pp. 5-10.

[44] Pian, T.H.H. and Tong, P. : " Finite Element Methods in Continuum Mechanics ", Adv. in Appl. Mech., 12, pp. 1-58.

[45] Reinsch, C.H. : " Smoothing by Spline Functions ", Numer. Math., 10, (1967), pp. 177-183.

[46] Roberts, S.M. and Shipman, J.S. : " Two-Point Boundary Value Problems: Shooting Methods ", Elsevier, New York, (1972).

[47] Ritz, W. : " Über eine neue Methode zur Lösung gewisser Variationalsprobleme der Mathematischen Physis ", Jour. f. reine und angew. Math., Vol. 135, (1908), pp. 1-61.

[48] Sard, A. : " Best Approximate Integration Formulas ; Best Appromimation Formulas ", Amer. J. Math., 71, (1949), pp. 80-91.

[49] Sard, A. : " Linear Approximation ", Math. Surveys No. 9, AMS, Providence, (1963).

[50] Schoenberg, I.J. : " Contribution to The Problem of Approxi-mation of Equidistant Data by Analytic Functions ", Quart. Appl. Math., 4, (1946), pp. 45-99, pp. 112-141.

[51] Schoenberg, I.J. : " On Interpolation by Spline Functions and Its Minimal Properties ", in " On Approximation Theory ", Proceedings of the Conf. held in The Mathematical Research Institude at Oberwolfach, Black Forest, Aug 4-10, (1963), P.L. Butzer ed., Birkhäuer Verlag, Basel, (1964), pp. 109-129.

[52] Schoenherg, I.J. : " On Spline Functions ", in " Inequalities ", Proceedings of A Symposium held at Wright-Patterson Air Force Base, Ohio, (1965), O. Shisha ed., Academic Press, New York, (1967). pp. 255-291.

[53] Shanks, D. : " Non-Linear Transforms of Divergent and Slowly Convergent Sequences ", J. Math. Phys., 34, (1958), pp.1-42.

[54] SIAM Journal on Numerical Analysis, Vol. 14, No. 1, March, (1977).

[55] Strang, G. : " Approximation in The Finite Element Method ", Numer. Math., 19, (1972), pp. 81-98.

[56] Strang, G. and Fix, G.J. : " An Analysis of The Finite Element Method ", Prentice-Hall, (1973).

[57] Stroud, A.H. : " Numerical Quadrature and Solution of Ordinary Differential Equations ", Applied Mathematical Sciences, Vol. 10, Springer-Verlag, (1974).

[58] Varga, R.S. : " Matrix Interative Analysis ", Prentice-Hall, (1962).

[59] Varga, R.S. : " Hermite Interpolation-Type Ritz Methods for Two-Point Boundary Value Problems ", in " Numerical Solution of PDE ", Proceedings of A Symposium held at University of Maryland, Collage Park, Maryland, (1965), J.H. Bramble ed., Academic Press, (1966).

[60] Wheeler, M.H. : " An Optimal L_∞ Error Estimate for Galerkin Approximations to Solutions of Two-Point Boundary Value Problems ", SIAM J. Numer. Anal., 10, (1973), pp. 914-917.

[61] Wilkinson, J.H. : " Error Analysis of Direct Methods of Matrix Inversion ", J. Assoc. Comput. Mach., 8, (1961), pp. 281-330.

[62] Yosida, K. : " Functional Analysis ", Springer-Vergla, Berlin, (1966).

[63] Young, D.M. : " Iterative Solution of Large Linear Systems ", Academic Press, (1971).