ABSTRACT

Title of dissertation: OPTIMAL APPROXIMATION SPACES FOR SOLVING PROBLEMS WITH ROUGH COEFFICIENTS

Qiaoluan Li, Doctor of Philosophy, 2009

Dissertation directed by: Professor John E. Osborn
Department of Mathematics

The finite element method has been widely used to solve partial differential equations by both engineers and mathematicians for the last several decades. This is due to its well-known effectiveness when applied to a wide variety of problems. However, it has some practical drawbacks. One of them is the need for meshing. Another is that it uses polynomials as the approximation basis functions. Commonly, polynomials are also used by other numerical methods for partial differential equations, such as the finite difference method and the spectral method. Nevertheless, polynomial approximations are not always effective, especially for problems with rough coefficients. In the dissertation, a suitable approximation space for the solution of elliptic problems with rough coefficients has been found, which is named as generalized L-spline space. Theoretically, I have developed generalized L-spline approximation spaces, where L is an operator of order m with rough coefficients, have proved the interpolation error estimate, and have also proved that the generalized L-spline space is an optimal approximation space for the problem $L^*Lu = f$

with certain operator L, by using n-widths as the criteria. Numerically, two problems have been tested and the relevant error estimate results are consistent with the shown theoretical results.

Meshless methods are newly developed numerical methods for solving partial differential equations. These methods partially eliminate the need of meshing. Meshless methods are considered to have great potential. However, the need for effective quadrature schemes is a major issue concerning meshless methods. In our recently published paper, we consider the approximation of the Neumann problem by meshless methods, and show that the approximation is inaccurate if nothing special (beyond accuracy) is assumed about the numerical integration. We then identify a condition - referred to as the zero row sum condition. This, together with accuracy, ensure the quadrature error is small. The row sum condition can be achieved by changing the diagonal elements of the stiffness matrix. Under row sum condition we derive an energy norm error estimate for the numerical solution with quadrature. In the dissertation, meshless methods are discussed and quadrature issue is explained. Two numerical experiments are presented in details. Both theoretical and numerical results indicate that the error has two components; one due to the meshless methods approximation and the other due to quadrature.

OPTIMAL APPROXIMATION SPACES FOR SOLVING PROBLEMS WITH ROUGH COEFFICIENTS

by

Qiaoluan Helen Li

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2009

Advisory Committee:
Professor John E. Osborn, Chair/Advisor
Professor Konstantina Trivisa
Professor Peter Wolfe
Professor Tobias von Petersdorff
Professor Howard Elman

Acknowledgments

I would like to thank my advisor, Professor John E. Osborn for his guidance and support over the past five years. He has always made himself available for help and advice. It has been a pleasure to work with and learn from such a great man. Thanks are due to Professor Konstantina Trivisa, Professor Peter Wolfe, Professor Tobias von Petersdorff and Professor Howard Elman for agreeing to serve on my thesis committee and for sparing their invaluable time reviewing the manuscript. I am grateful to other faculties, and fellow graduate students, for their friendship and goodwill.

I owe my deepest thanks to my parents who have always stood by me. Specially, my love and thanks go to my husband Howie, for his love, support and unwavering faith; and our daughter Vivienne, for her beautiful smiles.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Problems with Rough Coefficients

We are interested in elliptic boundary value problems with rough or highly oscillating coefficients. These type of problems arise in many applications, such as study of heterogeneous material. If the coefficients are rough, the solution will also be rough. The usual finite element methods using polynomials as basis shape function thus do not provide accurate approximations. In [13], Babuška and Osborn constructed a one-dimensional homogeneous elliptic boundary value problem, with a rough coefficient $a(x)$, which is only bounded and measurable, and with a homogeneous Dirichlet condition at one end, and a non-homogeneous Neumann condition at the other end, for which the usual finite element method converges arbitrarily slowly. The example they constructed also shows that adaptive procedures cannot improve the slow convergence. This motivates the research of developing special methods for problems with rough coefficients.

Let us first understand the mathematical background of elliptic boundary value problems with rough coefficients through a simple one-dimensional example, in which we will answer the question why usual finite elements do not approximate well for these problems. Let us consider

$$-(a(x)u')' = f(x), \quad x \in [0,1]$$
$$u(0) = u(1) = 0, \tag{1.1}$$

where $a(x)$ is measurable and $0 < \alpha \le a(x) \le \beta$. The corresponding weak formula is

$$u \in H_0^1[0,1], \quad B(u,v) = \int_0^1 au'v'\,dx = \int_0^1 fv\,dx \quad \forall v \in H_0^1[0,1].$$

B is bounded and coercive on $H_0^1[0,1]$, i.e.,

$$|B(u,v)| = |\int_0^1 au'v'\,dx| \le C\|u\|_1\|v\|_1,$$

and

$$B(u,u) = \int_0^1 au'u'\,dx \ge C\|u\|_1^2.$$

Thus, by the Lax-milgram theorem, problem (1.1) has a unique solution u. We want to approximate u. Toward this end, we let S, a finite dimensional subspace of $H_0^1[0,1]$, be the trial and test space for the Galerkin variational method. The approximate solution u_h determined by S is characterized by

$$u \in S, \quad B(u,v) = \int_0^1 au'v'\,dx = \int_0^1 fv\,dx \quad \forall v \in S.$$

As a consequence of the fact that $B(\cdot,\cdot)$ is bounded and coercive, we know that the approximation u_h is quasi-optimal, that is,

$$\|u - u_h\|_1 \le C \inf_{\chi \in S} \|u - \chi\|_1.$$

Thus the quality of the approximation, i.e., the error $\|u - u_h\|_1$, is mainly determined by the approximation properties of the trial space S. In the usual finite element method we let S be the space of continuous piecewise linear polynomials, and let χ be $I_h u$ the interpolate of u:

$$\|u - u_h\|_1 \leq Ch\|u - I_h u\|_1.$$

By using the approximation result, we have

$$\|u - I_h u\|_1 \leq C\|u\|_2,$$

providing $u \in H^2[0,1]$. If the coefficient function $a(x)$ is smooth, then by using regularity results we know that $u \in H^2[0,1]$ and

$$\|u\|_2 \leq C\|f\|_0,$$

i.e., $u \in H_0^1[0,1] \cap H^2[0,1]$. Combining these results we have

$$\|u - u_h\|_1 \leq Ch\|f\|_0.$$

This regularity property is also important in the error estimate of $u - u_h$ in $L^2[0,1]$ norm. However, when $a(x)$ is rough, solution u is also rough; to be specific, u in general is not in $H^2[0,1]$ and may not be in $H^{1+\epsilon}[0,1]$ for any $\epsilon > 0$. Thus, the usual finite element approximation is not accurate.

It is therefore natural to select the trial space S so that it maintains a good approximation to solution u. In [14], Babuška and Osborn proposed a special shape function, which reflects the local properties of the solution u, to solve a one-dimensional elliptic boundary value problem. In this dissertation, we extend the idea of this special shape function to solve problems of order $2m > 0$, mainly in one dimensional space. We have developed generalized L-spline approximation spaces, where L is an operator of order m, with rough coefficients, have proved interpolation error estimates, and have proved that the generalized L-spline space is an optimal approximation space for the problem $L^*Lu = f$ with certain operator L, by using n-widths as the criteria.

1.2 Variational Methods with Non-polynomial Approximation Spaces

Finite element methods have been widely used to solve partial differential equations by both engineers and mathematicians for the last several decades. This is due to their well-known effectiveness when applied to a wide variety of problems. Most finite element methods use polynomial shape functions for approximating functions. They are effective for many problems. However, polynomial shape functions are not always effective as shown in the previous section. In this dissertation we show the effectiveness of using certain non-polynomial shape functions.

Recently, meshless methods for solving partial differential equations have been increasingly used in the engineering community. In general, meshless methods are variational methods, which begin with a function $\phi(x)$ with compact support and use the functions $\phi_j^h(x) = \phi(\frac{x-jh}{h})$ as shape functions. These methods reduce the need of meshing and also create the freedom of choosing $\phi(x)$ to provide better approximation for certain problems. For example, for high order problems we can choose smoother $\phi(x)$. For problems whose solutions have some special features, and thus are not accurately approximated by usual shape functions, we may be able to obtain shape functions that provide accurate approximation by appropriately selecting $\phi(x)$. For general introduction to meshless methods, see [18] and [7].

Meshless methods are considered to have a large potential. However, meshless methods have to face the same question as finite element method does; in particular, the effect of numerical integration for the approximation solution. It is well known that the finite element approximation is computed by solving matrix problems whose elements involve integrals that most likely are evaluated by numerical integration, except in very simple cases. For finite element methods, the effect of numerical integration for source problems has been studied by a number of authors; we refer to [32] and [24]. Banerjee and Osborn obtained the estimation of the effect of numerical integration for the second-order selfadjoint eigenvalue problem in [15]. From those we see that the rate of convergence of the finite element approximation is preserved provided the numerical integration is sufficiently accurate. As for meshless methods, there are only a few papers treating this quadrature issue practically, such as [17] and [22]. In [8], we consider the approximation of the Neumann problem by meshless methods, and show that the approximation is inaccurate if nothing special (beyond accuracy) is assumed about the numerical integration. We then identified a condition - referred to as the zero row sum condition. This, together with accuracy, ensure the quadrature error is small. The row sum condition can be achieved by changing the diagonal elements of the stiffness matrix. Under row sum condition we derive an energy norm error estimate for the numerical solution with quadrature. See [9] for an alternative approach.

1.3 The Outline of the Dissertation

In the first three sections of Chapter 2, we recall the classical L-spline space, which is defined for operator L with smooth coefficients, and some generalization of it by various authors. When it was studied by mathematicians in the sixties, the motivation or the goal of introducing L-splines was not clearly stated. Since the problems with rough coefficients are of our interest, the generalized L-spline space is introduced in Section 2.3, which is an extension of the classical L-spline space in two ways. One is to extend the situation where the coefficients are merely measurable. Another is to use high order polynomial in constructing the L-spline space, which is useful when we are dealing with problems with smooth righthand side functions. This definition gives the advantage of generalized L-spline space over usual finite element space, and also serves as a good motivation for constructing special shape functions. The definition of generalized L-spline interpolation is given in Section 2.4. Section 2.5 shows the interpolation error estimate for generalized L-spline interpolation, which demonstrates the approximation property of the generalized L-spline space. Chapter 3 is devoted to the applications of generalized L-spline spaces to variational methods. Sections 3.1 and 3.2 discuss two possible special shape functions for a Dirichlet boundary value problem of order $2m$ with rough coefficients, and the relevant error estimates are stated and proved. In Section 3.3, two problems are computationally tested; the results are consistent with the error estimates stated in Section 3.2. Negative norm error estimate and the error estimate for eigenvalue problems with rough coefficients are established in Section 3.4 and

3.5. In Chapter 4 we briefly discuss n-width, and show that the generalized L-splines are optimal approximation spaces in the sense of n-width in certain situtions. Section 4.1 reviews n-widths theory of compact linear operator in Hilbert space. Some examples of optimal subspaces in Hilbert spaces are presented in Section 4.2. Generalized L-spline space is proved to be an optimal subspace in Section 4.3. Two dimensional optimal subspaces, consisting of special shape functions, which were introduced in [10], for a class of second order elliptic problems with rough coefficients are discussed in Sections 4.4. Chapter 5 discusses meshless methods in detail. In Section 5.1 meshless methods are introduced and the quadrature issue is explained. The construction of the meshless shape function is presented in Section 5.2. Two numerical test results are shown in Section 5.3, which indicates that the error has two components; one due to the meshless methods approximation and the other due to quadrature.

Chapter 2

Generalized L-spline Spaces

This chapter introduces generalized L-spline spaces and describes the interpolation of given functions by elements in the generalized L-spline spaces. We also discuss various properties and error bounds of these interpolant functions. We start with a brief discussion of the development of spline theory. The definition of L-spline approximation spaces and their approximation properties (cf. [49]) are recalled, which we later refer to classical L-splines. We then introduce the generalized L-spline spaces. The error estimate for the interpolation result are described in the end.

2.1 Notations

Let us begin with some notations that will be used throughout this chapter. For $-\infty < a < b < +\infty$ and for a positive integer n, let

$$\Gamma := a = x_0 < x_1 < \cdots < x_{n-1} < x_n = b$$

denote a partition or mesh of interval $[a, b]$ with knots x_j, subinterval $I_j = [x_{j-1}, x_j]$ and mesh size $h_j = x_j - x_{j-1}$. For any nonnegative integer m, let $H^m[a, b]$ be the Sobolov space, which is the collection of all real-valued functions defined on $[a, b]$ with square integrable derivatives up to order m, i.e.,

$$H^m[a, b] \equiv \{\psi \in L^2[a, b] \, : \, D^k\psi \in L^2[a, b], \; 0 \le k \le m\}.$$

Let $C^m[a,b]$ be the set of all real-valued functions which have continuous derivatives of order at least m in $[a,b]$. We then recall that $H^m[a,b]$ can be defined as the completion of the space of all real-valued functions $\psi \in C^\infty[a,b]$ with respect to the norm:

$$\|\psi\|_{H^m[a,b]} \equiv \{\sum_{j=0}^{m} \int_a^b |D^j\psi|^2\,dx\}^{\frac{1}{2}},$$

and semi-norm

$$|\psi|_{H^m[a,b]} \equiv \{\int_a^b |D^m\psi|^2\,dx\}^{\frac{1}{2}}.$$

For simplicity, we also use notations $\|\cdot\|_m$ and $|\cdot|_m$ for norm and semi-norm, respectively. Equivalently, $H^m[a,b]$ is the collection of all real-valued functions ψ defined on $[a,b]$ such that $D^{m-1}\psi$ is absolutely continuous, with $D^m\psi \in L^2[a,b]$, where $L^2[a,b]$ is the set of all square integrable functions on $[a,b]$, with norm

$$\|\psi\|_{L^2[a,b]} \equiv \{\int_a^b |\psi|^2\,dx\}^{\frac{1}{2}},$$

which can be written as $\|\cdot\|_0$. For additional notation, let $H_0^m[a,b]$ be defined as the closure of the infinitely differentiable functions compactly supported in (a,b),

$$\begin{aligned} H_0^m[a,b] &= \overline{C_0^\infty[a,b]}^{H^m} \\ &= \{u \in H^m[a,b] \,:\, D^k(a) = 0 = D^k(b) \text{ for } 0 \le k \le m-1\}, \end{aligned}$$

and let $H^{-m}[a,b]$ be the dual space of the Sobolov space $H_0^m[a,b]$ with the norm

$$\|\psi\|_{H^{-m}[a,b]} \equiv \sup_{\phi \in H^m[a,b], \|\phi\|_{H^m[a,b]} \neq 0} \frac{(\psi,\phi)}{\|\phi\|_{H^m[a,b]}}.$$

2.2 Classical L-spline Spaces

Splines are well known because of their many beautiful properties and their wide range of application to the numerical approximation of solutions of partial differential equations. They were first introduced by I.J. Schoenberg [47] in 1946, and have been in the focus of studies of many mathematicians in the 1960s and 1970s. In this section, we recall some main results regarding L-splines, which were introduced as generalization of splines, from [2], [49], [1] and [53].

For $m \geq 1$, let L be a linear differential operator of order m defined by

$$L[u(x)] = \sum_{k=0}^{m} a_k(x) D^k u(x), \quad m \geq 1, \tag{2.1}$$

where

$$a_k \in H^m[a,b] \text{ for } 0 \leq k \leq m, \tag{2.2}$$

and assume that there is a positive number α such that

$$0 < \alpha \leq a_m(x) \text{ for } x \in [a,b]. \tag{2.3}$$

The formal adjoint L^* of L is defined by

$$L^*[v(x)] = \sum_{k=0}^{m} (-1)^k D^k \{a_k(x) v(x)\}. \tag{2.4}$$

Then, given any integer satisfying $1 \leq z \leq m$, the L-spline spaces, $Sp(L, \Gamma, z)$, is the collection of all real-valued functions s defined on $[a,b]$, such that $s(x) \in H^{2m}(I_j)$, for each $1 \leq j \leq n$,

$$L^* L s = 0 \text{ on } I_j,$$

and

$$D^k(x_j-) = D^k s(x_j+) \quad \text{for } 0 \le k \le 2m-1-z,\ j = 1, 2, \cdots, n-1.$$

In other words, each $s(x)$ in L-spline space can be viewed as a piecewise smooth function, whose smoothness depends on z, locally solving $L^*Ls = 0$, i.e.,

$$Sp(L, \Gamma, z) \equiv \{s \in H^{2m-z}[a,b] : \text{For each } 1 \le j \le n,\ s|_{I_j} \in H^{2m}(I_j) \text{ and } (L^*Ls)|_{I_j} = 0\}. \tag{2.5}$$

An important special case is that when $L = D^m$, the elements in $Sp(L, \Gamma, z)$ are piecewise polynomials, and are so called polynomial splines. More specifically, when $L = D^m$ and $z = m$, $Sp(D^m, \Gamma, m)$ coincides with the Hermite spline.

Schultz and Varga defined four types of interpolation of given functions in $Sp(L, \Gamma, z)$. And they also proved the existence and uniqueness of them in [49]. Here we only discuss the one type, which we are interested in.

Definition 2.1 *Given $u(x) \in C^{m-1}[a,b]$, a function $s(x) \in Sp(L, \Gamma, z)$ is said to be a $Sp(L, \Gamma, z)$-interpolant of $u(x)$, if*

$$D^k(u-s)(x_j) = 0, \quad 0 \le k \le z-1,\ 0 < j < n,$$

$$D^k(u-s)(a) = D^k(u-s)(b) = 0, \quad 0 \le k \le m-1.$$

Simply by integration by parts, it can be shown the first integral relation in [1].

Theorem 2.1 *Let $u(x) \in H^m[a,b]$ and $s(x)$ be its interpolant in $Sp(L, \Gamma, z)$. Then the following first integral relation is valid:*

$$\int_a^b (Lu)^2\,dx = \int_a^b \{L(u-s)\}^2\,dx + \int_a^b (Ls)^2\,dx.$$

Proof. The proof can be found in [49].

By the definition of the $Sp(L,\Gamma,z)$-interpolant $s(x)$ of $u(x) \in H^m[a,b]$, s is also the unique interpolant in $Sp(L,\Gamma,z)$ for any $v \in H^m[a,b]$ for which

$$D^k(u-v)(x_j) = 0, \quad 0 \le k \le z-1,\ 0 < j < n, \tag{2.6}$$

$$D^k(u-v)(a) = D^k(u-v)(b) = 0, \quad 0 \le k \le m-1. \tag{2.7}$$

Thus, the first integral relation is valid for any v that satisfies (2.6), i.e.,

$$\int_a^b (Lv)^2\,dx = \int_a^b \{L(v-s)\}^2\,dx + \int_a^b (Ls)^2\,dx.$$

from this we have

$$\int_a^b (Lv)^2\,dx \ge \int_a^b (Ls)^2\,dx.$$

The above inequality gives a beautiful property of the $Sp(L,\Gamma,z)$-interpolant, and it can be used as an alternative definition for generalized splines.

Theorem 2.2 *Given $u \in H^m[a,b]$, let U_u be the collection of all $v \in H^m[a,b]$ which satisfy (2.6). Then*

$$\|Ls\|_{L^2[a,b]} = \inf_{v\in U_u} \|v\|_{L^2[a,b]},$$

where s is the unique $Sp(L,\Gamma,z)$-interpolant of u.

The first integral relation is important, not only because it can be interpreted as the minimization property, but also because it is the basis for the proof of the interpolant error estimate theorems.

Theorem 2.3 *Let $u \in H^m[a,b]$, let Γ be a partition of the interval $[a,b]$ with size h, and let s be the element in $Sp(L,\gamma,z)$ which interpolates u. Then there exists a* *constant C, dependent on k and m but independent of u and Γ, such that*

$$\|D^k(u-s)\|_{L^2[a,b]} \le Ch^{m-k}\|L(u-s)\|_{L^2[a,b]} \le Ch^{m-k}\|Lu\|_{L^2[a,b]},$$

for any $0 \le k \le m$.

Theorem 2.4 *Let $u \in H^{2m}[a,b]$, let Γ be a partition of the interval $[a,b]$ with size h, and let s be the element in $Sp(L,\gamma,z)$ which interpolates u. Then there exists a constant C, dependent on k and m but independent of u and Γ, such that*

$$\|D^k(u-s)\|_{L^2[a,b]} \le Ch^{2m-k}\|L^*Lu\|_{L^2[a,b]},$$

for any $0 \le k \le m$.

In [49], Schultz and Varga obtained the above two theorems, in which the upper bounds are in terms of Lu and L^*Lu instead of u. For smooth function u and the operator L with smooth coefficients, the upper bound can be changed to the corresponding norms of u, and the error estimate theorems were given as follows (cf. [53]):

Theorem 2.5 *Let $u \in H^m[a,b]$, let Γ be a partition of the interval $[a,b]$ with size h, and let s be the element in $Sp(L,\gamma,z)$ which interpolates u. Then for $2 \le q \le \infty$,*

$$\|D^k(u-s)\|_{L^q[a,b]} \le Ch^{m-k-1/2+1/q}\|u\|_{H^m[a,b]}, \quad 0 \le k \le m-1$$

For polynomial splines ($L = D^m$), $\|u\|_{H^m[a,b]}$ can be replaced by $\|D^m u\|_{L^2[a,b]}$.

Theorem 2.6 *Let $u \in H^{2m}[a,b]$, let Γ be a partition of the interval $[a,b]$ with size h, and let s be the element in $Sp(L,\gamma,z)$ which interpolates u. Then for $2 \le q \le \infty$,*

$$\|D^k(u-s)\|_{L^q[a,b]} \le Ch^{2m-k-1/2+1/q}\|u\|_{H^{2m}[a,b]}, \quad 0 \le k \le 2m-z.$$

For polynomial splines ($L = D^{2m}$), $\|u\|_{H^{2m}[a,b]}$ *can be replaced by* $\|D^{2m}u\|_{L^2[a,b]}$.

Once the term u is used in the upper bound, one can deduce the interpolation error bounds for functions u in spaces intermediate to $H^m[a,b]$ and $H^{2m}[a,b]$ from the above two theorems, by using interpolation space theory. For functions u even less smooth than C^{m-1}, a modification of the definition of interpolation in $Sp(L,\Gamma,z)$ and the corresponding error estimate results can be found in [53], where they mainly applied the notion of Lagrange polynomial interpolation (cf. [25]) and interpolation space theory.

2.3 Generalized L-spline Spaces

L-splines have been generalized in various ways based on the interests of the authors. In this section, we will review some of the generalizations and the remaining section will be devoted to the generalized L-spline spaces with rough coefficients.

We begin this section with the results of Jerome and Schumaker [34]. Let $\Lambda = \{\lambda_i\}_{i=1}^k$ be any set of linearly independent, bounded linear functionals on $H^m[a,b]$, and let $\mathbf{r} = (r_1, r_2, \cdots, r_k)^T$ be any vector of real Euclidean k-space, R^k. The so called Lg-spline interpolant is the solution of the following minimization problem

$$\|Ls\|_{L^2[a,b]} = \inf\{\|Lv\|_{L^2[a,b]} \,:\, v \in U_\Lambda(\mathbf{r})\},$$
$$\text{where } U_\Lambda(\mathbf{r}) \equiv \{v \in H^m[a,b] \,:\, \lambda_i(v) = r_i,\, 1 \le i \le k\}. \tag{2.8}$$

Thus, Lg-splines offer generalizations in the area of interpolation, but do not generalize the type of differential operator L. The next generalization we will recall is called γ-splines, which is due to Schultz [48] and Lucas [42]. Let

$$Eu \equiv \sum_{j=0}^{m} (-1)^j D^j \{p_j(x) D^j u(x)\},$$

where $p_j(x) \in H^j[a,b] \bigcap L^\infty[a,b]$, $0 \leq j \leq m$, and $p_m(x) \geq \delta > 0$ in $[a,b]$. As in previous sections, let Γ denote a partition of interval $[a,b]$ and let z again be a positive integer satisfying $1 \leq z \leq m$. Then $S(E,\Gamma,z)$, the γ-spline space, is the collection of real-valued functions w defined on $[a,b]$ such that, relative to Γ,

$$Ew(x) = 0 \quad \text{almost everywhere in each subinterval } I_j,\ 1 \leq j \leq n,$$

and

$$D^k s(x_j-) = D^k s(x_j+) \quad \text{for } 0 \leq k \leq 2m-1-z,\ j = 1,2,\cdots,n-1.$$

Thus, the generalization of γ-spline works through more general differential operators. Lucas combined these two ideas of Lg-spline and γ-spline simultaneously and obtains the error bounds theorem in [43]. One of the more interesting developments with respect to one-dimensional spline theory is due to Jerome and Pierce [33]. They were motivated by considering the numerical solution of the singular boundary value problem:

$$D^2 u(x) + \frac{\sigma}{x} Du(x) = f(x),$$

$$u(0) = \alpha,\ u(1) = \beta,$$

where $0 \leq \sigma < 1$. More detailed review on the above generalized L-splines can be found in [53].

Here, we wish to consider the numerical solution of elliptic differential equations with rough coefficients, as arise in the analysis of a laminated bar; *e.g.,* a bar consisting of a material m_1 in the interval $[a = y_0, y_1]$, a material m_2 in the interval $[y_1, y_2]$, *etc.* Then the longitudinal displacement, $u(x)$ is the solution of

$$-(a(x)u'(x))' = f(x), \text{ for } a \le x \le b,$$

where the coefficient $a(x)$ is determined on $[y_0, y_1]$ by the elastic properties of m_1, $a(x)$ on $[y_1, y_2]$ is determined by the elastic properties of m_2, *etc.* So $a(x)$ will be a step function. We are thus motivated to consider coefficients $a_k(x)$ that are merely measurable.

The generalized L-spline spaces is an extension of the classical L-spline space in two ways. One is to extend to the situation where the coefficients $a_k(x)$ have lower smoothness than assumed in [49], specifically are merely measurable. Another is to use higher order polynomials in constructing the L-spline spaces. So, instead of assuming (2.2) we assume

$$a_k(x) \text{ is measurable and bounded } (|a_k(x)| \le \beta), \quad k = 0, 1, \cdots, m. \tag{2.9}$$

In addition we continue to assume (2.3).

Throughout the dissertation we consider differential equation of the form

$$L^*Lu = f, \quad \text{on } [a, b], \tag{2.10}$$

where $f \in H^{-m}[a, b]$ = the dual space of the Sobolev space $H_0^m[a, b]$, and we seek solution $u \in H^m[a, b]$. (2.10) will be interpreted as a distribution or weak equation:

$$u \in H^m[a, b]; \quad B(u, \phi) \equiv \int Lu L\phi \, dx = f(\phi) \quad \forall \phi \in H_0^m[a, b]. \tag{2.11}$$

With $r = -1, 0, \cdots$, we define the generalized L-spline spaces to be

$$S_\Gamma^r \equiv \{\psi \in H^m[a, b] \, : L^*L\psi|_{I_j} \in P^r(I_j), \quad \text{for } j = 1, 2, \cdots, n\},$$

and

$$S_{\Gamma,0}^r \equiv \{\psi \in H_0^m[a,b] \, : L^*L\psi|_{I_j} \in P^r(I_j), \quad \text{for } j = 1, 2, \cdots, n\}.$$

If $r = -1$, we interpret $L^*L\psi|_{I_j} \in P^r(I_j)$ to mean that $L^*L\psi = 0$ on each I_j. This definition of generalized L-spline spaces is closely related to one type of generalized finite element method in [14]. We will sometimes refer to L-splines S_Γ^{-1} with smooth coefficients, *i.e.,* that satisfy the hypotheses in Section 2.1 as shown in [49], as classical L-splines. In order to investigate the properties of S_Γ^r, various equivalent norms for the space $H_0^m[a,b]$ will be introduced.

Lemma 2.1 *$|u|_m$, $\|u\|_m$ and $\|Lu\|_0$ are equivalent norms on $H_0^m[a,b]$, with equivalency constants, that depend on α and β.*

Proof. The fact that $|u|_m$ and $\|u\|_m$ are equivalent norms follows from the Poincaré inequality.

Now consider $\|u\|_m$ and $\|Lu\|_0$. Let $Lu = g$. We first show that there exists a constant C such that

$$\|u\|_m \leq C\|g\|_0, \quad \forall u \in H_0^m[a,b]. \tag{2.12}$$

In order to do so, we state a standard result for ordinary differential equations in [31]:

Let $y = y(t)$ be a solution of linear system of ordinary differential equations $y' = A(t)y + f(t)$ for $t \in [a,b]$. Then

$$|y(t)| \leq \{|y(t_0)| + \int_{t_0}^{t} |f(s)|ds\} \exp|\int_{t_0}^{t} \|A(s)\|ds|, \quad \forall\, t,\, t_0 \in [a,b],$$

where $\|A(s)\| = \sup_{|y|=1} |Ay|$ and $|y| = \max(|y^1|, \cdots, |y^d|)$.

By rewriting $Lu = g$ as a linear system and following the above result, we have

$$|U(t)| \leq \{|U(0)| + \int_0^t |\begin{pmatrix} 0 \\ \cdot \\ \cdot \\ \cdot \\ \frac{g}{a_m} \end{pmatrix}| ds\} \exp|\int_0^t \|A(s)\| ds|,$$

where $U = [u, Du, \cdots, D^{m-1}u]^T$ and $A = \begin{pmatrix} 0 & 1 & 0 & \cdot & \cdot & 0 \\ 0 & 0 & 1 & 0 & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \cdot & 0 & 1 \\ \frac{a_0}{a_m} & \frac{a_1}{a_m} & \cdot & \cdot & \cdot & \frac{a_{m-1}}{a_m} \end{pmatrix}$. Since $u \in H_0^m[a,b]$ we see that $U(a) = 0$ and from (2.2) and (2.3) we see that $\|A(s)\|$ is bounded. Thus

$$|D^k u(t)| \leq |U(t)| \leq C_1^2 \int_0^t |g(s)| ds, \quad k = 0, 1, \cdots, m-1,$$

for $0 \leq t \leq 1$, and hence

$$\begin{aligned} \int_a^b |D^k u(t)|^2 \, dt &\leq C_1^2 \int_a^b (\int_a^b |g| ds)^2 \, dt = (b-a) C_1^2 (\int_a^b |g| ds)^2 \\ &\leq (b-a) C_1^2 \int_a^b 1^2 ds \int_a^b |g|^2 ds = (C_1(b-a))^2 \|g\|_0^2, \end{aligned}$$

i.e.,

$$|u|_k \leq C_1 \|g\|_0 = C_1 \|Lu\|_0, \tag{2.13}$$

for $k = 0, 1, \cdots, m-1$ and C_1 is a generic constant. From the definition of operator L, we obtain

$$
\begin{aligned}
|D^m u(x)| &= \left| \frac{Lu(x) - \sum_{j=0}^{m-1} a_j(x) D^j u(x)}{a_m(x)} \right| \\
&\leq \frac{|Lu(x)| + \sum_{j=0}^{m-1} |a_j(x)||D^j u(x)|}{\alpha} \\
&\leq \frac{|Lu(x)|}{\alpha} + \frac{\beta}{\alpha} \sum_{j=0}^{m-1} |D^j u(x)|.
\end{aligned} \tag{2.14}
$$

(2.14), together with inequality (2.13), yields

$$
\begin{aligned}
|u|_m^2 &= \int_a^b |D^m u(x)|^2 \, dx \\
&\leq \int_a^b \frac{1}{\alpha^2} \left(|Lu| + \beta \sum_{j=0}^{m-1} |D^j u| \right)^2 dx \\
&\leq \int_a^b \frac{2}{\alpha^2} \left(|Lu|^2 + \beta^2 (\sum_{j=0}^{m-1} |D^j u|)^2 \right) dx \\
&\leq \int_a^b \frac{2}{\alpha^2} \left(|Lu|^2 + \beta^2 m \sum_{j=0}^{m-1} |D^j u|^2 \right) dx \\
&= \frac{2}{\alpha^2} \left(\|Lu\|_0^2 + \beta^2 m \sum_{j=0}^{m-1} |u|_j^2 \right) \\
&\leq \frac{2}{\alpha^2} (1 + C_1^2 \beta^2 m^2) \|Lu\|_0^2.
\end{aligned} \tag{2.15}
$$

Now from (2.13) and (2.15) we obtain

$$
\|u\|_m^2 = \sum_{j=0}^{m-1} |u|_j^2 + |u|_m^2 \leq m C_1^2 \|Lu\|_0^2 + \frac{2}{\alpha^2} (1 + C_1^2 \beta^2 m^2) \|Lu\|_0^2 = C^2 \|Lu\|_0^2,
$$

where $C^2 = m C_1^2 + \frac{2}{\alpha^2} (1 + C_1^2 \beta^2 m^2)$.

Next we see that there exists a constant $\tilde{C} = \frac{1}{(m+1)\beta^2}$ such that

$$
\tilde{C} \|Lu\|_0 \leq \|u\|_m, \tag{2.16}
$$

which is true, because

$$\begin{aligned} \|Lu\|_0^2 &= \int_a^b |\sum_{j=0}^m a_j(x)D^j u(x)|^2 dx \\ &\le \int_a^b (m+1)\sum_{j=0}^m |a_j(x)|^2 |D^j u(x)|^2 dx \\ &\le (m+1)\beta^2 \sum_{j=0}^m \int_a^b |D^j u(x)|^2 dx \\ &= (m+1)\beta^2 \|u\|_m^2. \end{aligned}$$

We see that the equivalency constants depend on α and β. □

It follows from Lemma 2.1 that the bilinear form $B(\cdot,\cdot)$, defined in (2.11), is bounded on $H^m[a,b]$ and coercive on $H_0^m[a,b]$.

Let $S_\Gamma^r(I_j) \equiv \{s \in H^m(I_j) \,:\, L^*Ls \in P^r(I_j)\}$.

Lemma 2.2 *For* $m = 1, 2, \cdots,$ *and* $r = -1, 0, 1, \cdots,$ *we have* $\dim S_\Gamma^r(I_j) = r + 2m + 1$.

Proof. Let us consider the following boundary value problem

$$\begin{cases} L^*Lu = f \text{ on } I_j, \\ D^k u(x_{j-1}) = \alpha_k,\ D^k u(x_j) = \alpha_{m+k},\ \text{for } 0 \le k \le m-1, \end{cases} \tag{2.17}$$

where $f \in H^{-m}(I_j)$, and we seek a solution in $H^m(I_j)$. Using the distribution interpretation of $L^*Lu = f$ introduced in (2.11), we see that u in (2.17) is characterized by $u \in H^m(I_j)$ such that

$$\begin{cases} B(u,\phi) = \int_{I_j} LuL\phi\, dx = f(\phi), \forall \phi \in H_0^m(I_j), \\ D^k u(x_{j-1}) = \alpha_k,\ D^k u(x_j) = \alpha_{m+k},\ \text{for } 0 \le k \le m-1, \end{cases}$$

Let $v \in H^m(I_j)$ be the $2m-1$th order polynomial such that

$$D^k v(x_{j-1}) = \alpha_k,\ D^k v(x_j) = \alpha_{m+k},$$

then $v \in H^m(I_j)$, $Lv \in L^2(I_j)$, and $L^*Lv \in H^{-m}(I_j)$. Then, if $w = u - v$ we have

$$\begin{cases} L^*Lw = L^*L(u-v) = f - L^*Lv \equiv g \text{ on } I_j, \\ D^k w(x_{j-1}) = 0 = D^k w(x_j), \text{ for } 0 \le k \le m-1. \end{cases} \tag{2.18}$$

Note that the boundary conditions in (2.18) are homogeneous. Thus w is a distribution or weak solution of (2.18) if

$$w \in H_0^m(I_j), \quad B(w,\phi) = \int_{I_j} LwL\phi\, dx = g(\phi), \quad \forall \phi \in H_0^m(I_j).$$

By Lemma 2.1, we have

$$|B(w,\phi)| = |\int_{I_j} LwL\phi\, dx| \le C\|Lw\|_0\|L\phi\|_0 \le C\|w\|_m\|\phi\|_m.$$

and

$$B(w,w) = \int_{I_j} LwLw\, dx = \|Lw\|_0^2 \ge C\|w\|_m^2,$$

i.e., the bilinear form $B(w,\phi)$ is bounded and coercive on $H_0^m(I_j)$. Thus, by Lax-Milgram theorem, problem (2.18) has a unique solution w. Since u solves (2.17) if and only if w solves (2.18), existence and uniqueness for problem (2.18) implies it for problem (2.17).

For $l = 0, 1, \cdots, 2m-1$, let w_l be the solution of

$$\begin{cases} L^*Lw_l = 0 \text{ on } I_j, \\ D^k w_l(x_{j-1}) = \alpha_k^l,\ D^k w_l(x_j) = \alpha_{m+k}^l, \text{ for } 0 \le k \le m-1, \end{cases} \tag{2.19}$$

a boundary value problem with homogeneous differential equation and non-homogeneous boundary values, where $[\alpha_0^l, \alpha_1^l, \cdots, \alpha_{2m-1}^l] = e_l$, for $l = 0, 1, \cdots, 2m-1$, and e_l is the standard basis for the vector space $\mathbb{R}^{2m}$. For $l = 0, 1, \cdots, r-2m$, let v_l be the solution of

$$\begin{cases} L^*Lv_l = p_l \text{ on } I_j, \\ D^k v_l(x_{j-1}) = 0 = D^k v_l(x_j), \text{ for } 0 \le k \le m-1, \end{cases} \tag{2.20}$$

a bounded value problem with non-homogeneous differential equation and homogeneous boundary values, where $p_l = x^l$, for $l = 0, 1, \cdots, r$. Since problems (2.19) and problems (2.20) are examples of problem (2.17) with special right hand sides f and boundary values α_k, the solutions w_l, for $l = 0, 1, \cdots, 2m-1$, and v_l, for $l = 0, 1, \cdots, r$, exist uniquely.

Then if $s \in S_\Gamma^r(I_j)$, i.e., $\exists \beta_l$ for $l = 0, \cdots, r$ and γ_l for $l = 0, \cdots, 2m-1$ such that

$$\begin{cases} L^*Ls = \sum_{l=0}^{r} \beta_l p_l \text{ on } I_j, \\ D^k s(x_{j-1}) = \sum_{l=0}^{2m-1} \gamma_l \alpha_k^l, \ D^k s(x_j) = \sum_{l=0}^{2m-1} \gamma_l \alpha_{m+k}^l, \text{ for } 0 \le k \le m-1. \end{cases} \tag{2.21}$$

Since $\sum_{l=0}^{r} \beta_l v_l + \sum_{l=0}^{2m-1} \gamma_l w_l$ also satisfies above problem, by the uniqueness of the solution, we have that $s = \sum_{l=0}^{r} \beta_l v_l + \sum_{l=0}^{2m-1} \gamma_l w_l$, i.e., s can be represented as a linear combination of $w_0, w_1, \cdots, w_{2m-1}, v_0, v_1, \cdots, v_r$. Now we only need prove that w_l and v_l are linearly independent. Assume $s = \sum_{l=0}^{r} \beta_l v_l + \sum_{l=0}^{2m-1} \gamma_l w_l = 0$. Thus $L^*Ls = \sum_{l=0}^{r} \beta_l p_l = 0$, which, because of the independence of p_l, implies $\beta_l = 0$ for $0 \le l \le r$. Next observe that

$$[s_l(x_{j-1}), s_l'(x_{j-1}), \cdots, s_l^{(m-1)}(x_{j-1}), s_l(x_j), s_l'(x_j), \cdots, s_l^{(m-1)}(x_j)] = \sum_{l=0}^{2m-1} \gamma_l e_l = 0;$$

by the independence of e_l, we have $\gamma_l = 0$ for $0 \le l \le 2m-1$. So we have $\dim S_\Gamma^r(I_j) = (r+1) + 2m = r + 2m + 1$, with $\{w_0, w_1, \cdots, w_{2m-1}, v_0, v_1, \cdots, v_r\}$ a basis. □

2.4 Generalized L-spline Interpolation

In [49], Schultz and Varga discuss the basic properties of classical L-splines, in particular establishing interpolation results including interpolation error estimates. But the desire is to obtain error bounds for the numerical solution of the problem with rough coefficients. In this section, we define the interpolation by generalized L-splines and discuss the error estimate for the corresponding interpolant in next section.

If $u(x)$ is a given function in $H^m[a,b]$, the S_Γ^r-interpolant of u is described by the following lemma.

Lemma 2.3 *Given a function $u \in H^m[a,b]$, then there is a unique $I_\Gamma^r u \in S_\Gamma^r$ such that*

$$D^k I_\Gamma^r u(x_j) = D^k u(x_j), \quad j = 0, 1, \cdots, n, \quad k = 0, 1, \cdots, m-1$$
$$\int_{I_j} (u - I_\Gamma^r u)(x - x_{j-1})^l \, dx = 0, \quad l = 0, 1, \cdots, r, \quad j = 1, 2, \cdots, n.$$

Proof. We define $I_\Gamma^r u$ locally, i.e., on each subinterval I_j. For $j = 1, 2, \cdots, n$, we seek $I_{I_j}^r u \in S_\Gamma^r(I_j)$ satisfying

$$\begin{aligned} D^k I_{I_j}^r u(x_{j-1}) &= D^k u(x_{j-1}), \quad k = 0, 1, \cdots, m-1, \\ D^k I_{I_j}^r u(x_j) &= D^k u(x_j), \quad k = 0, 1, \cdots, m-1, \\ \int_{I_j} (u - I_{I_j}^r u)(x - x_{j-1})^l \, dx &= 0, \quad l = 0, 1, \cdots, r. \end{aligned} \tag{2.22}$$

By Lemma 2.2, we have $\dim S_\Gamma^r(I_j)$ is $r + 2m + 1$. For each j, it is easy to see that the number of equations in (2.22) is the same as the number of free parameters.

Thus we have existence of $I^r_{I_j}u$ if and only if we have uniqueness. Now we prove the uniqueness. Given u, suppose $I^r_{I_j}u$ and $\widetilde{I^r_{I_j}u}$ satisfy the equations in this lemma. Let $z = I^r_{I_j}u - \widetilde{I^r_{I_j}u}$. Then

$$D^k z(x_{j-1}) = D^k z(x_j) = 0, \quad k = 0, 1, \cdots, m-1 \tag{2.23}$$

$$\int_{I_j} z(x - x_{j-1})^l \, dx = 0, \quad l = 0, 1, \cdots, r. \tag{2.24}$$

Now, $L^*Lz \in \text{span}\{1, x - x_j, \cdots, (x - x_j)^l\}$, so by (2.24) we have $\int_{I_j} zL^*Lz \, dx = 0$. Using Green's formula and (2.23), $\int_{I_j} (Lz)^2 \, dx = 0$, so $Lz = 0$. Using this and (2.23) we get $z = 0$, so $I^r_{I_j}u = \widetilde{I^r_{I_j}u}$.

Now define $I^r_\Gamma u$ on $[a, b]$

$$I^r_\Gamma u = I^r_{I_j}u, \quad x \in I^r_{I_j}.$$

Since $u \in H^m[a, b]$, we have $I^r_\Gamma u \in H^m[a, b]$. Clearly $I^r_\Gamma u$ is unique. □

Notice that if $u \in H^m_0[a, b]$ then $I^r_\Gamma u \in S^r_{\Gamma,0}$.

2.5 Interpolation Error Estimate

In this section we investigate the accuracy of the approximation $I^r_\Gamma u \simeq u$ for $u \in H^m[a, b]$.

Lemma 2.4 *For $u \in H^m_0[a, b]$, let $I^r_\Gamma u \in S^r_{\Gamma,0}$ be the interpolant of u. Then $I^r_\Gamma u$ is characterized by*

$$B(I^r_\Gamma u, v) = B(u, v) \quad \textit{for all } v \in S^r_{\Gamma,0}, \tag{2.25}$$

where $B(u, v) = \int_a^b LuLvdx$ for any u, $v \in H^m_0[a, b]$.

Proof. Let $v \in S^r_{\Gamma,0}$, i.e., $v \in H^m_0[a,b]$, $L^*Lv|_{I_j} \in P^r$, for $j = 0, 1, \cdots, n$. On each subinterval I_j, there exists a polynomial $p_j \in P^r$ such that

$$\int_{I_j} LvL\psi \, dx = \int_{I_j} p_j \psi \, dx, \quad \forall \psi \in H^m_0(I_j).$$

Let $\psi = I^r_\Gamma u - u$; then$\psi \in H^m_0(I_j)$ and by the definition of $I^r_\Gamma u$, we have

$$\int_{I_j} LvL(I^r_\Gamma u - u) \, dx = \int_{I_j} p_j(I_\Gamma u - u) \, dx = 0.$$

This is true for any $v \in S^r_{\Gamma,0}$. Here we have proved that if $I^r_\Gamma u$ satisfies the condition in Lemma 2.3, then (2.25) holds.

Now we prove that $I^r_\Gamma u$ in (2.25) is unique. Suppose $I^r_\Gamma u$, $\widetilde{I^r_\Gamma u} \in S^r_{\Gamma,0}$ and

$$B(I^r_\Gamma u - \widetilde{I^r_\Gamma u}, v) = 0 \quad \forall v \in S^r_{\Gamma,0}.$$

Let $v = I^r_\Gamma u - \widetilde{I^r_\Gamma u}$; then $v \in S^r_{\Gamma,0}$ and we have

$$\int_a^b |L(I^r_\Gamma u - \widetilde{I^r_\Gamma u})|^2 \, dx = B(I^r_\Gamma u - \widetilde{I^r_\Gamma u}, I^r_\Gamma u - \widetilde{I^r_\Gamma u}) = 0.$$

So

$$L(I^r_\Gamma u - \widetilde{I^r_\Gamma u}) = 0.$$

By the equivalence of the norms on $H^m_0[a,b]$ in Lemma 2.1, we have

$$I^r_\Gamma u - \widetilde{I^r_\Gamma u} = 0.$$

Thus we have the characterization of $I^r_\Gamma u$. □

Notice that the above lemma also holds when Γ is the whole interval $[0,1]$.

Throughout the dissertation, Green's formula, which can be found in any partial differential equation book, for example [29], is often applied in the proofs.

Theorem 2.7 *(Green's Formula) Let $u, v \in C^1[a,b]$. Then*

$$\int_a^b (Du)v\,dx = -\int_a^b u(Dv)\,dx + uv\,\Big]_a^b,$$

where $uv\,]_a^b = u(b)v(b) - u(a)v(a)$.

A general Green's formula involving operator L can be found in [49], which is

$$\int_a^b \{vL[u] - uL^*[v]\}dx = P(u(b)v(b)) - P(u(a)v(a)),$$

for any $a, b \in [0,1]$, and any $u, v \in H^m[0,1]$, where

$$P(u,v) = \sum_{j=0}^{m-1} D^{m-j-1}u(x) \sum_{k=0}^{j} (-1)^k D^k\{a_{m-j+k}(x)v(x)\}.$$

We now state two easily proved lemmas, which will be used to prove the error estimate theorem.

Lemma 2.5 *There exists a constant $0 < C(\alpha,\beta) < \infty$, such that*

$$\|u\|_m \le C\|L^*Lu\|_0,$$

for all $u \in H_0^m[a,b]$ with $L^*L\,u \in L^2[a,b]$.

Proof. By Lemma 2.1, we have

$$\|u\|_m^2 \le C_1^2\|Lu\|_0^2, \quad \forall u \in H_0^m[a,b]. \tag{2.26}$$

Since

$$\|Lu\|_0^2 = (Lu, Lu) = (L^*Lu, u) \le \|L^*Lu\|_0\|u\|_0, \tag{2.27}$$

where Green's formula is used, combining (2.26) and (2.27), we get

$$\|u\|_m^2 \le C_1^2\|Lu\|_0^2 \le C_1^2\|L^*Lu\|_0\|u\|_0.$$

Dividing both hand sides of the above equation by $\|u\|_m$ and using $\|u\|_0 \le C_2\|u\|_m$, we have the assertion that for some constant C that depends only on α and β,

$$\|u\|_m \le C\|L^*Lu\|_0.$$

□

Lemma 2.6 *$\|u\|_m$ and $\|L^*Lu\|_{-m}$ are equivalent norms on $H_0^m[a,b]$, with equivalency constants depending on α and β.*

Proof. By Lemma 2.1, we have

$$\tilde{C}\|Lu\|_0 \le \|u\|_m \le C\|Lu\|_0, \quad \forall u \in H_0^m[a,b].$$

By the definition of negative norm, Green's formula and the above inequality, we have

$$\begin{aligned} \|L^*Lu\|_{-m} &= \sup_{v \in H_0^m} \frac{|\int (L^*Lu)v dx|}{\|v\|_m} = \sup_{v \in H_0^m} \frac{|\int LuLv dx|}{\|v\|_m} \\ &\le \sup_{v \in H_0^m} \frac{\|Lu\|_0\|Lv\|_0}{\|v\|_m} \le \frac{1}{\tilde{C}}\|Lu\|_0 \le \frac{1}{\tilde{C}^2}\|u\|_m. \end{aligned}$$

On the other hand,

$$\|L^*Lu\|_{-m} \ge \frac{|\int (L^*Lu)u dx|}{\|u\|_m} = \frac{\int |Lu|^2 dx}{\|u\|_m} = \frac{\|Lu\|_0^2}{\|u\|_m} \ge \frac{1}{C}\|Lu\|_0 \ge \frac{1}{C^2}\|u\|_m.$$

□

We then have the following error bound theorem. The main idea of the proof of the theorem is to prove it locally by shifting the general problem on each subinterval $[x_{j-1}, x_j]$ to $[0, 1]$, and proving the error bound and then shifting back. Therefore, we will state the standard scaling argument first, which will be used in the proof later.

For any function $v(x),\ x \in I_j$, let $\bar{v}(\bar{x}) = \bar{v}(\frac{x-x_{j-1}}{h_j}) = v(x),\ \bar{x} = \frac{x-x_{j-1}}{h_j} \in \bar{I} = [0,\, 1]$. Then $\bar{D}\bar{v}(\bar{x}) = h_j D v(x)$ and hence $\bar{D}^k\bar{v}(\bar{x}) = h_j^k D^k v(x)$. We shift the operator L to the interval $[0,1]$ and then multiply by h_j^m to define

$$\bar{L}\bar{v} = \sum_{k=0}^{m} \bar{a}_k h_j^{m-k} \bar{D}^k \bar{v}, \quad \bar{L}^*\bar{w} = \sum_{k=0}^{m} (-1)^k h_j^{m-k} \bar{D}^k(\bar{a}_k \bar{w}), \tag{2.28}$$

where $h_j = x_j - x_{j-1}$. Then the coefficients in $\bar{L}$ and $\bar{L}^*$ are bounded by α and β:

$$|\bar{a}_k(\bar{x}) h_j^{m-k}| \le \beta, \quad \forall \bar{x},\ \forall h_j \le 1,\ k = 0, 1, \cdots, m,$$

$$|\bar{a}_m(\bar{x})| \ge \alpha, \quad \forall \bar{x}\ \forall h_j \le 1.$$

Thus Lemma 2.5 and Lemma 2.6 hold for operator $\bar{L}$ on the interval $[0,1]$. By standard scaling, we have

$$|v|_{0,I_j} = (\int_{I_j} v^2 dx)^{\frac{1}{2}} = (h_j \int_{\bar{I}} \bar{v}^2 d\bar{x})^{\frac{1}{2}} = h_j^{\frac{1}{2}} |\bar{v}|_{0,\bar{I}},$$

$$|v|_{l,I_j} = (\int_{I_j} (D^l v)^2 dx)^{\frac{1}{2}} = (h_j \int_{\bar{I}} (\bar{D}^l \bar{v} h_j^{-l})^2 d\bar{x})^{\frac{1}{2}} = h_j^{\frac{1-2l}{2}} |\bar{D}^l \bar{v}|_{0,\bar{I}} = h_j^{\frac{1}{2}-l} |\bar{v}|_{l,\bar{I}},$$

and hence

$$\|v\|_{l,I_j} \le C h_j^{\frac{1}{2}-l} \|\bar{v}\|_{l,\bar{I}}, \quad l = 0, 1, \cdots, m.$$

From this we get

$$\|u - I_\Gamma^r u\|_{l,I_j} \le C h_j^{\frac{1}{2}-l} \|\bar{u} - \overline{I_\Gamma^r u}\|_{l,\bar{I}}, \quad l = 0, 1, \cdots, m, \tag{2.29}$$

where $\overline{I_\Gamma^r u}$ is the interpolant $I_\Gamma^r u$ after shifting. Because of the uniqueness of the interpolant in $\bar{S}_\Gamma^r$, $\overline{I_\Gamma^r u}$ is the same as $\bar{I}_\Gamma^r \bar{u}$, the $\bar{S}_\Gamma^r$ interpolant of $\bar{u}$, based on the operator $\bar{L}$, as defined in (2.28), where

$$\bar{S}_\Gamma^r = \{\phi \in H^m[\bar{I}] : \bar{L}^*\bar{L}\phi = \sum_{k=0}^{m}(-1)^k h_j^{(m-k)}\bar{D}^k(\bar{a}_k \sum_{i=0}^{m} \bar{a}_i h_j^{(m-i)}\bar{D}^i\phi) \in P^r[\bar{I}]\}.$$

Since

$$\begin{aligned}\bar{L}^*\bar{L}\bar{u}(\bar{x}) &= \sum_{k=0}^{m}(-1)^k h_j^{m-k}\bar{D}^k(\bar{a}_k \sum_{i=0}^{m}\bar{a}_i h_j^{m-i}\bar{D}^i\bar{u}) \\ &= \sum_{k=0}^{m}(-1)^k h_j^{m} D^k(a_k \sum_{i=0}^{m} a_i h_j^{m} D^i u(x)) = h_j^{2m} L^*Lu(x).\end{aligned}$$

Thus

$$|\bar{L}^*\bar{L}\bar{u}|_{\mu,\bar{I}} = (\int_{\bar{I}}[(\bar{L}^*\bar{L}\bar{u})^\mu]^2 d\bar{x})^{\frac{1}{2}} = (\int_{I_j}\frac{1}{h_j}[h_j^{(\mu+2m)}(L^*Lu)^\mu]^2 dx)^{\frac{1}{2}} = h_j^{\mu+2m-\frac{1}{2}}|L^*Lu|_{\mu,I_j}, \tag{2.30}$$

for any integer $\mu \geq 0$.

Lemma 2.7 *(Bramble-Hilbert [20]) Let $u \in H^m[a,b]$, there exists a polynomial p of degree $\leq k$, such that*

$$\|u-p\|_{H^s[a,b]} \leq Ch^{m-s}\|u\|_{H^m[a,b]} \quad \text{for } 0 \leq s \leq m \leq k+1.$$

Theorem 2.8 *Let $u \in H^m[a,b]$ and $I_\Gamma^r u$ be the S_Γ^r interpolant of u. Then for $l = 0, 1, \cdots, m$, $k \geq 0$ and $r \geq -1$, we have*

$$\|u - I_\Gamma^r u\|_l \leq C(\alpha,\beta)h^{\mu+2m-l}\|L^*Lu\|_k, \tag{2.31}$$

where $\mu = \min(k, r+1)$, $h = \max h_j$ and $C(\alpha,\beta)$ is independent of u and h, but depends in general on α, β, l, k and r.

Proof. Here we will show that error bound in equation (2.31) holds locally, and then show it is true on the whole interval $[a,b]$. Now consider $u - I_\Gamma^r u$ on I_j. By standard scaling argument (2.29), we have

$$\|u - I_\Gamma^r u\|_{l,I_j} \le C h_j^{\frac{1}{2}-l} \|\bar{u} - \bar{I}_\Gamma^r \bar{u}\|_{l,\bar{I}}. \tag{2.32}$$

Since $\bar{u} - \bar{I}_\Gamma^r \bar{u} \in H_0^m(\bar{I})$, by applying Lemma 2.5 to $\bar{L}$, we get,

$$\|\bar{u} - \bar{I}_\Gamma^r \bar{u}\|_{l,\bar{I}} \le \|\bar{u} - \bar{I}_\Gamma^r \bar{u}\|_{m,\bar{I}} \le C(\alpha,\beta)\|\bar{L}^*\bar{L}(\bar{u} - \bar{I}_\Gamma^r \bar{u})\|_{0,\bar{I}},$$

for $l \le m$. Write

$$\bar{I}_\Gamma^r \bar{u} = (\bar{I}_\Gamma^r \bar{u})_1 + (\bar{I}_\Gamma^r \bar{u})_2,$$

where

$$\bar{L}^*\bar{L}(\bar{I}_\Gamma^r \bar{u})_1 = 0 \quad \text{ on } \bar{I},$$

$$D^k(\bar{I}_\Gamma^r \bar{u})_1(0) = D^k\bar{u}(0), \quad D^k(\bar{I}_\Gamma^r \bar{u})_1(1) = D^k\bar{u}(1);$$

then

$$\bar{L}^*\bar{L}[\bar{u} - (\bar{I}_\Gamma^r \bar{u})_1] = \bar{L}^*\bar{L}\bar{u}, \quad \text{ on } \bar{I},$$

$$D^k[\bar{u} - (\bar{I}_\Gamma^r \bar{u})_1](0) = 0 = D^k[\bar{u} - (\bar{I}_\Gamma^r \bar{u})_1](1).$$

Since $\bar{L}^*\bar{L}(\bar{I}_\Gamma^r \bar{u})_2 = \bar{L}^*\bar{L}(\bar{I}_\Gamma^r \bar{u}) \in P^r(\bar{I})$, using the fact that all norms on the finite-dimensional space $P^r(\bar{I})$ are equivalent and applying Lemma 2.6 to $\bar{L}$, we have

$$\|\bar{L}^*\bar{L}(\bar{I}_\Gamma^r \bar{u})\|_{0,\bar{I}} = \|\bar{L}^*\bar{L}(\bar{I}_\Gamma^r \bar{u})_2\|_{0,\bar{I}} \le C\|\bar{L}^*\bar{L}(\bar{I}_\Gamma^r \bar{u})_2\|_{-m,\bar{I}} \le C\|(\bar{I}_\Gamma^r \bar{u})_2\|_{m,\bar{I}}. \tag{2.33}$$

Applying Lemma 2.4 to $\bar{L}$, we see that $(\bar{I}_\Gamma^r \bar{u})_2$ is the Ritz projection of $\bar{u} - (\bar{I}_\Gamma^r \bar{u})_1$ onto $\bar{S}_{\Gamma,0}^r$ with respect to the form $\bar{B}(\bar{u},\bar{v}) = \int_0^1 \bar{L}\bar{u}\bar{L}\bar{v}dx$, where

$$\bar{S}_{\Gamma,0}^r = \{\phi \in \bar{S}_\Gamma^r : D^k\phi(0) = D^k\phi(1) = 0 \quad k = 0,1,\cdots,m-1\}.$$

Hence

$$\|(\bar{I}_\Gamma^r \bar{u})_2\|_{m,\bar{I}} \le C\|\bar{u} - (\bar{I}_\Gamma^r \bar{u})_1\|_{m,\bar{I}}. \tag{2.34}$$

Since $\bar{u} - (\bar{I}_\Gamma^r \bar{u})_1 \in H_0^m(\bar{I})$, we can apply Lemma 2.5 to $\bar{L}$ to get

$$\|\bar{u} - (\bar{I}_\Gamma^r \bar{u})_1\|_{m,\bar{I}} \le C\|\bar{L}^*\bar{L}\bar{u}\|_{0,\bar{I}}. \tag{2.35}$$

Now, combining (2.33), (2.34) and (2.35), we arrive at

$$\|\bar{L}^*\bar{L}(\bar{I}_\Gamma^r \bar{u})\|_{0,\bar{I}} \le C\|\bar{L}^*\bar{L}\bar{u}\|_{0,\bar{I}}. \tag{2.36}$$

From Lemma 2.5, inequalities (2.29) and (2.36), we obtain

$$\begin{aligned}\|u - I_\Gamma^r u\|_{l,I_j} &\le Ch_j^{\frac{1}{2}-l}\|\bar{u} - \bar{I}_\Gamma^r \bar{u}\|_{l,\bar{I}} \le Ch_j^{\frac{1}{2}-l}\|\bar{L}^*\bar{L}(\bar{u} - \bar{I}_\Gamma^r \bar{u})\|_{0,\bar{I}} \\ &\le Ch_l^{\frac{1}{2}-l}\|\bar{L}^*\bar{L}\bar{u}\|_{0,\bar{I}} \le Ch_j^{\frac{1}{2}-l}\|\bar{L}^*\bar{L}\bar{u}\|_{\mu,\bar{I}},\end{aligned} \tag{2.37}$$

for $0 \le l \le m$ and $\mu \ge 0$. Now

$$\|u - I_\Gamma^r u\|_{l,I_j} = \|(u-\phi) - I_\Gamma^r(u-\phi)\|_{l,I_j}, \tag{2.38}$$

for any $\phi \in S_\Gamma^r(I_j)$. Then inequality (2.37) and equation (2.38) gives

$$\|u - I_\Gamma^r u\|_{l,I_j} \le Ch^{\frac{1}{2}-l}\|\bar{L}^*\bar{L}\bar{u} - \bar{L}^*\bar{L}\bar{\phi}\|_{\mu,\bar{I}}.$$

Since $\bar{L}^*\bar{L}\bar{\phi}$ is arbitrary in $P^r(\bar{I})$ and by the Bramble-Hilbert lemma, we have

$$\|u - I_\Gamma^r u\|_{l,I_j} \le Ch_j^{\frac{1}{2}-l} \inf_{Q\in P^r} \|\bar{L}^*\bar{L}\bar{u} - Q\|_{\mu,\bar{I}} \le Ch_j^{\frac{1}{2}-l}|\bar{L}^*\bar{L}\bar{u}|_{\mu,\bar{I}},$$

for $\mu \le r+1$. Then by a further scaling argument, (2.30) yields

$$\|u - I_\Gamma^r u\|_{l,I_j} \le Ch_j^{\frac{1}{2}-l}h_j^{\mu+2m-\frac{1}{2}}|L^*Lu|_{\mu,I_j} = Ch_j^{\mu+2m-l}|L^*Lu|_{\mu,I_j} \le Ch_j^{\mu+2m-l}|L^*Lu|_{k,I_j},$$

where $k \geq \mu$. To this end we have proved the error bound on each subinterval. Then simply taking $h = \max h_j$, we have the assertion on the whole interval $[a, b]$. □

The above proof can be simplified when $r = -1$, i.e., the S_Γ^r-interpolant is the typical L-spline result, since $\bar{L}^*\bar{L}(\bar{I}_\Gamma^r \bar{u}) = 0$, and then $(\bar{I}_\Gamma^r \bar{u})_2 = 0$. The following theorem gives the error estimate for the case when $r = -1$

Theorem 2.9 *Let $u \in H^m[a, b]$ and $I_\Gamma^{-1}u$ be the S_Γ^{-1} interpolant of u, then for $l = 0, 1, \cdots, m$, we have*

$$\|u - I_\Gamma^{-1}u\|_l \leq C(\alpha, \beta)h^{2m-l}\|L^*Lu\|_0, \tag{2.39}$$

where $h = \max h_j$ and $C(\alpha, \beta)$ is independent of u and h, but depends in general on α, β, l, k and r.

Proof. Let us consider $u - I_\Gamma^{-1}u$ on I_j. After shifting to the interval $[0, 1]$, we have

$$\|u - I_\Gamma^{-1}u\|_{l, I_j} \leq Ch_j^{\frac{1}{2}-l}\|\bar{u} - \bar{I}_\Gamma^{-1}\bar{u}\|_{l,\bar{I}}, \tag{2.40}$$

Since $\bar{u} - \bar{I}_\Gamma^{-1}\bar{u} \in H_0^m(\bar{I})$ and $\bar{L}^*\bar{L}(\bar{I}_\Gamma^{-1}\bar{u}) = 0$, by applying Lemma 2.5 to $\bar{L}$, we get,

$$\|\bar{u} - \bar{I}_\Gamma^{-1}\bar{u}\|_{l,\bar{I}} \leq \|\bar{u} - \bar{I}_\Gamma^{-1}\bar{u}\|_{m,\bar{I}} \leq C(\alpha, \beta)\|\bar{L}^*\bar{L}(\bar{u} - \bar{I}_\Gamma^{-1}\bar{u})\|_{0,\bar{I}} = C(\alpha, \beta)\|\bar{L}^*\bar{L}\bar{u}\|_{0,\bar{I}},$$

for $l \leq m$. Then combining the above two inequalities, we obtain

$$\|u - I_\Gamma^{-1}u\|_{l, I_j} \leq Ch_j^{\frac{1}{2}-l}\|\bar{L}^*\bar{L}\bar{u}\|_{0,\bar{I}}.$$

By a further scaling argument (2.30), we arrive at

$$\|u - I_\Gamma^{-1}u\|_{l, I_j} \leq Ch_j^{\frac{1}{2}-l}\|\bar{L}^*\bar{L}\bar{u}\|_{0,\bar{I}} = Ch_j^{\frac{1}{2}-l}h^{2m-\frac{1}{2}}\|L^*Lu\|_{0,I_j} = Ch_j^{2m-l}\|L^*Lu\|_{0,I_j}.$$

Applying this result to the whole interval, we then have the assertion. □

Theorem 2.9 shows that we can obtain the same error estimate for interpolants in the norms $\| \cdot \|_l$, $l = 0, 1, \cdots, m$, for operators with rough coefficients as for operators with smooth coefficients. This relaxes the requirement of smooth coefficients in the usual L-spline spaces.

Chapter 3

Galerkin Methods with Generalized L-spline Spaces

In Chapter 2, generalized L-spline spaces S_Γ^r are introduced and the error estimate for the interpolant in S_Γ^r is provided. In this Chapter, we will present the applications of generalized L-spline spaces in Galerkin methods for solving one-dimensional Dirichlet boundary value problems. In Section 3.1, we begin by introducing the problem $Au = (L^*L + B)u = f$ with Dirichlet boundary condition. Error estimate results are given for the Galerkin solution with the generalized L-spline space. We state the results in two cases: $B \equiv 0$ and $B \neq 0$. In Section 3.2, for the second case we solve the problem in an approximation space V_Γ^r by using the whole operator A. The convergence of the Galerkin solution in both energy norm and L^2 norm are given. In Section 3.3, some experiments are done on a second order and two fourth order Dirichlet boundary value problems with rough coefficients. The numerical results show that a Galerkin method with a particular generalized L-spline space is very effective for problems with rough coefficients. In Section 3.4, some negative norm error estimates are given. Last, the generalized L-spline spaces are used to approximate the eigenvalue problems corresponding to the source problem we mentioned in the previous sections.

3.1 Error Estimates for the Generalized L-spline Spaces

Let us consider the following elliptic problem

$$
\begin{aligned}
&Au = L^*Lu + Bu = f, \quad a < x < b, \\
&D^k u(a) = D^k u(b) = 0, \quad k = 0, 1, \cdots, m-1,
\end{aligned}
\tag{3.1}
$$

where L is defined in the previous section and $B = \sum_{j=0}^{n} b_j D^j$, $0 \leq n \leq m$ and $|b_j(x)| \leq \gamma$, $0 \leq j \leq n$, for some positive number γ.

Let

$$
a(u, v) = \int_a^b LuLv\, dx + \int_a^b Buv\, dx. \tag{3.2}
$$

Then the weak formulation of (3.1) is

$$
\begin{cases}
\text{Seek } u \in H_0^m[a, b] \text{ such that} \\
a(u, v) = \int_a^b fv\, dx, \quad \forall v \in H_0^m[a, b].
\end{cases}
\tag{3.3}
$$

It is easily seen that

$$
|a(u, v)| \leq C\|u\|_m \|v\|_m, \quad \forall u, v \in H_0^m[a, b].
$$

When $B = 0$, we also have the coercivity

$$
a(u, v) \geq c\|u\|_m^2, \quad \forall u \in H_0^m[a, b].
$$

When $B \neq 0$, we assume that there is a constant C such that

$$
a(u, v) \geq c\|u\|_m^2, \quad \forall u \in H_0^m[a, b],
$$

The coercivity is essentially an assumption on operator B, which requires either that the signs of coefficients in B are all positive, or that the magnitude of the coefficients in B is smaller than those in L. Thus by the Lax-Milgram theorem, there exists a unique weak solution to this boundary-value problem (3.1).

The Galerkin approximation u_h to u is defined by

$$\begin{cases} u_h \in S_\Gamma^r \subset H_0^m[a,b] \\ a(u,v) = \int_a^b fv\,dx, \quad \forall v \in S_\Gamma^r, \end{cases}$$

where S_Γ^r is the generalized L-spline space,

$$S_\Gamma^r = \{\psi \in H^m[a,b] \,:\, L^*L\psi|_{I_j} \in P^r, \quad \text{for } j = 1,2,\cdots,n\}. \tag{3.4}$$

Note that we are using S_Γ^r as both trial and testing spaces.

When $B \equiv 0$, i.e., there are no extra low order terms, the Galerkin solution u_h will be the same as the interpolant $I_\Gamma^r u$ in S_Γ^r by the characterization Lemma 2.4. Directly following Theorem 2.8, we have the error estimate result.

Theorem 3.1 *For problem (3.1) with $B \equiv 0$, we have*

$$\|u - u_h\|_l \le Ch^{2m+\mu-l}\|f\|_k,$$

for $l = 0,1,\cdots,m$, $k \ge 0$ and $r \ge -1$, where $\mu = \min(k, r+1)$. When $f \in L^2[a,b]$, we have

$$\|u - u_h\|_l \le Ch^{2m-l}\|f\|_0,$$

for $l = 0,1,\cdots,m$ and $r \ge -1$.

Note that if f is assumed to just be in $L^2[a,b]$, then this error bound does not depend on r. For this case, we only use the generalized L-spline space with $r = -1$, i.e.,

$$S_\Gamma = \{\psi \in H^m[a,b] \,:\, L^*L\psi|_{I_j} = 0, \quad \text{for } j = 1,2,\cdots,n\}.$$

In general, if $B \neq 0$, the Galerkin solution u_h is different from the interpolant $I_\Gamma^r u$. The following theorem shows the error estimate for this case with $f \in L^2$.

Theorem 3.2 *For problem (3.1) with* $B = \sum_{j=0}^{n} b_j D^j$, $n \leq m$ *and* $|b_j(x)| \leq \gamma$, $0 \leq j \leq n$, *for some positive number* γ, *and* $f \in L^2$, *we have*

$$\|u - u_h\|_l \leq Ch^\mu \|f\|_0,$$

for $l = 0, 1, \cdots, m$, *where* $\mu = \min(2m - l, 2m - n)$.

Proof. Let $F = f - Bu$. Then from the coercivity of the weak formulation of the problem (3.1) and Cauchy-Schwarz inequality, for some constant C_1

$$c\|u\|_m^2 \leq a(u, u) = (Au, u) \leq C_1 \|f\|_0 \|u\|_0.$$

Dividing both hand sides by $\|u\|_m$, we have

$$\|u\|_m \leq C_2 \|f\|_0.$$

Together with the definition of the operator B and the fact that

$$\|Bu\|_0 \leq C\|u\|_m,$$

we obtain

$$\|F\|_0 = \|f - Bu\|_0 \leq \|f\|_0 + \|Bu\|_0 \leq C\|f\|_0. \tag{3.5}$$

Clearly

$$L^* Lu = F, \quad D^k u(a) = D^k u(b) = 0 \quad k = 0, 1, \cdots, m - 1.$$

From Theorem 2.8 and (3.5) we have

$$\|u - I_\Gamma^r u\|_l \leq Ch^{2m-l} \|f\|_0, \tag{3.6}$$

for $l = 0, 1, \cdots, m$. Writing

$$u_h = I_\Gamma^r u + z,$$

we immediately see that

$$a(z, z) = a(u_h - u, z) + a(u - I_\Gamma^r u, z) = a(u - I_\Gamma^r u, z)$$

$$= (L(u - I_\Gamma^r u), Lz) + (B(u - I_\Gamma^r u), z) = (B(u - I_\Gamma^r u), z),$$

for $v \in S_{\Gamma,0}^r$. Then by the coercivity we have

$$\|z\|_m^2 \leq Ca(z, z) = C(B(u - I_\Gamma^r u), z) \leq C\|B(u - I_\Gamma^r u)\|_0 \|z\|_0 \leq C\|u - I_\Gamma^r u\|_n \|z\|_0,$$

and this, together with (3.6), yields

$$\|z\|_m \leq C\|u - I_\Gamma^r u\|_n \leq Ch^{2m-n}\|f\|_0. \tag{3.7}$$

From (3.6), (3.7) and the fact that

$$\|z\|_l \leq C\|z\|_m, \quad l = 0, 1, \cdots, m,$$

we have

$$\|u - u_h\|_l \leq \|u - I_\Gamma^r u\|_l + \|z\|_l \leq C(h^{2m-l} + h^{2m-n})\|f\|_0 \leq Ch^\mu \|f\|_0,$$

where $\mu = \min(2m - l, 2m - n)$. $\square$

Theorem 3.2 gives an alternative way of using L-spline space, in which the construction of L-spline space does not involve the whole operator A. And only the rates of error in lower order norms are sacrificed for the error estimation when $n > 0$.

3.2 Error Estimates for Approximation Spaces with Operator A

In the previous section, the generalized L-spline spaces were used for the approximation spaces in the Galerkin method. There are two cases. One is that L^*L represents the problem operator, so that the numerical solution is the same as the interpolant in S_Γ^r. Another is that L^*L is only the leading part of the problem operator; then the numerical solution loses some power of h in the accuracy in the lower order norm when $B \neq 0$ with $n > 0$. In this section, we will discuss the ideal approximation space for the second case, i.e., the space V_Γ^r is constructed with the whole problem operator $A = L^*L + B$. Toward this end, let

$$V_\Gamma^r = \{\psi \in H^m[a,b] \, : A\psi|_{I_j} = (L^*L + B)\psi \in P^r(I_j), \quad \text{for } j = 1, 2, \cdots, n\}.$$

Although the numerical solution in V_Γ^r is not identical with the interpolant in V_Γ^r, the regular error estimate procedure for energy norm holds. And we can use the Nitsche argument to get the error estimate for L^2 norm. First, we still need a lemma counting the dimension of the space V_Γ^r, and a lemma to define the interpolant in V_Γ^r.

Lemma 3.1 *Let* $V_\Gamma^r(I_j) \equiv \{\psi \in H^m(I_j) \, : \, (L^*L+B)\psi \in P^r(I_j)\}$, *then* $\dim V_\Gamma^r(I_j) = r + 2m + 1$.

The proof of this lemma is similar to the proof of Lemma 2.2.

Lemma 3.2 *Given a function* $u \in H^m[a,b]$, *then there is a unique* $\tilde{I}_\Gamma^r u \in V_\Gamma^r$ *such* *that*

$$D^k \tilde{I}_\Gamma^r u(x_j) = D^k u(x_j), \quad j = 0, 1, \cdots, n, \quad k = 0, 1, \cdots, m-1$$

$$\int_{I_j} (u - \tilde{I}_\Gamma^r u)(x - x_{j-1})^l \, dx = 0, \quad l = 0, 1, \cdots, r, \quad j = 1, 2, \cdots, n.$$

Proof. We define $\tilde{I}_\Gamma^r u$ locally, i.e., on each subinterval I_j. For $j = 1, 2, \cdots, n$, we seek $\tilde{I}_{I_j}^r u \in V_\Gamma^r(I_j)$ satisfying

$$D^k \tilde{I}_{I_j}^r u(x_{j-1}) = D^k u(x_{j-1}), \quad k = 0, 1, \cdots, m-1,$$

$$D^k \tilde{I}_{I_j}^r u(x_j) = D^k u(x_j), \quad k = 0, 1, \cdots, m-1,$$

$$\int_{I_j} (u - \tilde{I}_{I_j}^r u)(x - x_{j-1})^l \, dx = 0, \quad l = 0, 1, \cdots, r. \tag{3.8}$$

According to the proof of Lemma 2.3, we have existence of $\tilde{I}_{I_j}^r u$ if and only if we have uniqueness. Now we prove the uniqueness. Given $u \in H^m(I_j)$, suppose $\tilde{I}_{I_j}^r u$ and $\widetilde{\tilde{I}_{I_j}^r u}$ satisfy the equations (3.8). Let $z = \tilde{I}_{I_j}^r u - \widetilde{\tilde{I}_{I_j}^r u}$. Then

$$D^k z(x_j^+) = D^k z(x_j^-) = 0, \quad k = 0, 1, \cdots, m-1, \tag{3.9}$$

$$\int_{I_j} z(x - x_{j-1})^l \, dx = 0, \quad l = 0, 1, \cdots, r. \tag{3.10}$$

Now, $Az \in \text{Span}\{1, x - x_j, \cdots, (x - x_j)^l\}$, so by (3.10) we have $\int_{I_j} zAz \, dx = 0$. Using the coercivity of the weak formulation of A, $0 = (Az, z) \geq c\|z\|_m^2$, so $\|z\|_m = 0$. Then we get $z = 0$, so $\tilde{I}_{I_j}^r u = \widetilde{\tilde{I}_{I_j}^r u}$. Now define $\tilde{I}_\Gamma^r u$ on $[a, b]$ by

$$\tilde{I}_\Gamma^r u = \tilde{I}_{I_j}^r u, \quad x \in I_{I_j}^r.$$

Since $u \in H^m[a, b]$, we have $\tilde{I}_\Gamma^r u \in H^m[a, b]$. Clearly $\tilde{I}_\Gamma^r u$ is unique. □

Notice that $\tilde{I}_\Gamma^r u$ does not satisfy (2.25) in Lemma 2.4, since the bilinear form for the problem (3.1) is not symmetric when $B \neq 0$. We have an interpolant error estimate, which is similar to Theorem 2.8. Before we state the next theorem, two basic lemmas involving different norms will be proved first.

Lemma 3.3 *There is a constant* $0 < C(\alpha, \beta, \gamma) < \infty$*, such that*

$$\|u\|_m \leq C\|Au\|_0 = C\|(L^*L + B)u\|_0,$$

for all $u \in H_0^m$ *with* $A\,u \in L^2[0,1]$*.*

Proof. By the coercivity of A in the Problem 3.1, we have

$$c\|u\|_m^2 \leq (Au, u) \leq C_1\|Au\|_0\|u\|_0, \quad \forall u \in H_0^m.$$

Dividing both hand sides of the above equation by $\|u\|_m$ and using the fact that $\|u\|_0 \leq C_2\|u\|_m$ for $u \in H_0^m$, we have the assertion that for some constant C that depends only on α and β, we have

$$\|u\|_m \leq C\|Au\|_0.$$

□

Lemma 3.4 $\|u\|_m$ *and* $\|Au\|_{-m}$ *are equivalent norms on* $H_0^m[a,b]$*.*

Proof. By the definition of negative norm and the fact that A is bounded, we have

$$\|Au\|_{-m} = \sup_{v \in H_0^m} \frac{|\int_a^b (Au)v dx|}{\|v\|_m} \leq \sup_{v \in H_0^m} \frac{C\|u\|_m\|v\|_m}{\|v\|_m} \leq C\|u\|_m.$$

On the other hand, by the coercivity of A,

$$\|Au\|_{-m} \geq \frac{|\int (Au)u dx|}{\|u\|_m} \geq \frac{c\|u\|_m^2}{\|u\|_m} = c\|u\|_m.$$

□

Theorem 3.3 *Let $u \in H^m[a,b]$ and $\tilde{I}_\Gamma^r u$ be the V_Γ^r interpolant of u, then for $l = 0, 1, \cdots, m$, $k \geq 0$ and $r \geq -1$, we have*

$$\|u - \tilde{I}_\Gamma^r u\|_l \leq Ch^{2m+\mu-l}\|Au\|_k, \tag{3.11}$$

where $\mu = \min(k, r+1)$, $h = \max h_j$ and C is independent of u but depends in general on α, β, γ, l, k and r.

Simply changing L^*Lu to Au in the proof of Theorem 2.8 and using the above lemmas give the proof of the above theorem.

Let $\tilde{u}_h$ be the Galerkin solution of problem (3.1) by using V_Γ^r as both trial and test spaces. The following theorem contains the energy and L^2 error estimate for $\tilde{u}_h$.

Theorem 3.4 *For problem (3.1), we have*

$$\|u - \tilde{u}_h\|_m \leq Ch^{m+\mu}\|f\|_k,$$

and

$$\|u - \tilde{u}_h\|_0 \leq Ch^{2m+\mu}\|f\|_k,$$

for $r \geq -1$, where $\mu = \min(k, r+1)$.

Proof. Because of the quasi-optimality of the Galerkin approximation $\tilde{u}_h$, we can bound the energy norm of the error by the energy norm of the interpolant error as follows

$$\|u - \tilde{u}_h\|_m \leq C \inf_{\psi \in V_\Gamma^r} \|u - \psi\|_m \leq C\|u - \tilde{I}_\Gamma^r u\|_m.$$

Together with the interpolant error estimate result for $l = m$ in Theorem 3.3, we have

$$\|u - \tilde{u}_h\|_m \leq Ch^{m+\mu}\|Au\|_k,$$

where $\mu = \min(k,\, r+1)$ and $r \geq -1$.

Since the leading term L^*L in the operator A is selfadjoint, the interpolant error estimate results in Theorem 3.3 also hold for the adjoint operator A^* of A. The proof can be obtained from the proof for the case of the operator A only with a slight change. With the help of the interpolant error estimate, we can get the convergence in the L^2 norm by using the Aubin-Nitsche argument.

We now set v to be the solution of the adjoint problem $A^*v = u - \tilde{u}_h$, i.e., $a(\psi, v) = (\psi, u - \tilde{u}_h)$ for any $\psi \in V_\Gamma^r$, then

$$\begin{aligned}
\|u - \tilde{u}_h\|_0^2 &= (u - \tilde{u}_h, u - \tilde{u}_h) \\
&= a(u - \tilde{u}_h, v) \\
&= a(u - \tilde{u}_h, v - \tilde{I}_\Gamma^r v) \\
&\leq C\|u - \tilde{u}_h\|_m \,\|v - \tilde{I}_\Gamma^r v\|_m \\
&\leq Ch^{m+\mu}\|Au\|_k \, h^m \|A^*v\|_0, \\
&= Ch^{2m+\mu}\|Au\|_k \,\|u - \tilde{u}_h\|_0.
\end{aligned}$$

Here, C is a generic positive constant, $\mu = \min(k, r+1)$, and $\tilde{I}_\Gamma^r v$ denotes the interpolant of v in space V_Γ^r. Consequently,

$$\|u - \tilde{u}_h\|_0 \leq Ch^{2m+\mu}\|f\|_k.$$

□

When problem (3.1) has extra low order terms, there are two ways of constructing the approximation space. One is to use the whole operator A to construct the approximation spaces. Alternatively, we can use only the leading term L^*L in A to construct the generalized L-spline spaces. So far, we can only give the rates of the convergence in energy and L^2 norms for the solution in space V_Γ^r. And the construction of V_Γ^r may be complicated. When $n = 0$, numerical solutions in V_Γ^r and S_Γ^r have the same rates of convergence. When $n > 0$, only the rates of error in lower order norms are sacrificed for solution in S_Γ^r. However the construction of S_Γ^r is simpler than that of V_Γ^r. When we solve a real problem, we should be able to find the balance and chose the one we want to work with.

In summary, the numerical solution in the generalized L-spline space gives the desired rates of convergence for problems with rough coefficients, in which the usual finite element method does not yield an accurate approximation. The reason for the failure of the analysis of finite element method is that the solution u to the problem with rough coefficients is also rough; to be specific, u is not in general in H^2, and may not be in $H^{1+\epsilon}$ for any $\epsilon > 0$. However, we did not use the regularity property of the numerical solution to prove the error estimate in the generalized L-spline space; instead we take advantage of the interpolant error estimation with the norm of Au as the upper bound.

3.3 Numerical Experiments

In this section we consider two numerical examples, namely a second order and a fourth order Dirichlet boundary value problems with rough coefficients. As part of the future work, we plan to test generalized L-spline spaces on more problems, specifically higher order problems with lower order terms using only the leading term in L^*L in the construction of the L-spline space.

Example 1. As an example we consider a second order elliptical equation on the unit interval $[0,1]$ with Dirichlet boundary condition:

$$-(a(x)u')' + u = 1, \quad 0 < x < 1$$

$$u(0) = u(1) = 0.$$

For definiteness, we pick a step function to be the coefficient function

$$a(x) = \begin{cases} 1 & 0 < x < \frac{1}{2}, \\ 2 & \frac{1}{2} < x < 1. \end{cases}$$

In this case we use two conforming Galerkin methods with piecewise linear finite element space and the generalized L-spline space S_Γ, only for the highest order term in the problem, which is defined by

$$S_\Gamma = \{\psi \in H^1[0,1] : (a(x)u')'|_{I_j} = 0, \quad \text{for } j = 1, 2, \cdots, n\}.$$

In all calculations test and trial spaces are the same. The mesh used in both computations is the uniform mesh Γ_n with an odd number of subintervals ($x_j = jh$, $j = 0, 1, \cdots, n$, $h = n^{-1}$, n odd), so that the discontinuity of $a(x)$ falls within a subinterval. The generalized L-spline space can also be represented as follows:

$$S_\Gamma = \{\psi \in H^1[0,1] \,:\, \text{for each } j,\, \psi|_{I_j} = \text{a linear combination of 1 and } \int^x \frac{dt}{a(t)}\}.$$

Note that the shape function of generalized L-spline coincides with the basis shape function of the usual finite element method based on piecewise linear element, when $a(x)$ is a constant. For this problem with step coefficient function, only the basis shape functions that are non-zero on the subinterval containing 0.5 are different from the hat function in the finite element method. One of these special shape functions is shown in Fig 3.1. The function shown in Fig. 3.1 might be referred to as a "broken" linear function, with a break at the jump in the step function. If the step function has two jumps, the the special shape function is a broken linear function with two breaks–at the jumps in the step function. Thus, for this given step coefficient function, only the middle two rows of stiffness matrix A by using generalized L-spline method need be modified from those of stiffness matrix A_{FEM} by using the usual finite element method. And it is similar with the calculation of the right hand side function. We use the numerical solution by over computing, i.e., $(n = 1001)$, with the finite element method as the exact solution to get the error in the energy norm. Let $\psi = \{\psi_1, \psi_2, \cdots, \psi_n\}$ be a set of basis functions of generalized L-spline space and $\phi = \{\phi_1, \phi_2, \cdots, \phi_m\}$ be a set of basis functions of finite element space. Let the exact solution be $u = \mathbf{u} \cdot \phi$ and the numerical solution be $u_h = \mathbf{u}_h \cdot \psi$, then

$$\|u - u_h\|_E^2 = \|u\|_E^2 - \|u_h\|_E^2 \approx \mathbf{u}^T A_{FEM}^m \mathbf{u} - \mathbf{u}_h^T A \mathbf{u}_h,$$

where A_{FEM}^m represents the stiffness matrix for the usual finite element method with m large.

The convergence results in relation to the element size of h show an improve-

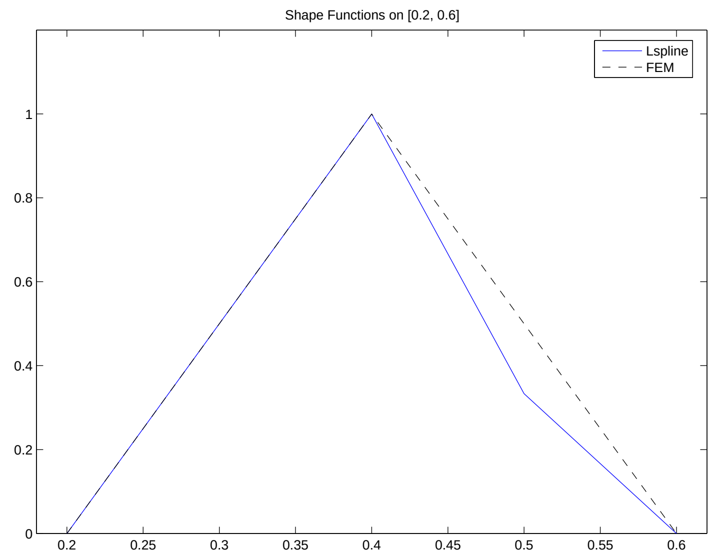

Figure 3.1: The special shape function on a subinterval

ment as one chooses the generalized L-spline space instead of piecewise linear finite element space. Since the coefficient function $a(x)$ is not smooth, we have $u \notin H^2(0,1)$. In Figure 3.2, the left end of the graph for the finite element results tilts upward, which implies the slower convergence. However, the log-log convergence curve for the generalized L-spline results has slope 1, which confirms the theoretical rate of convergence in the energy norm is $O(h)$ shown in Theorem 3.2. The detail data of the energy norm of the error are given in Table 3.1.

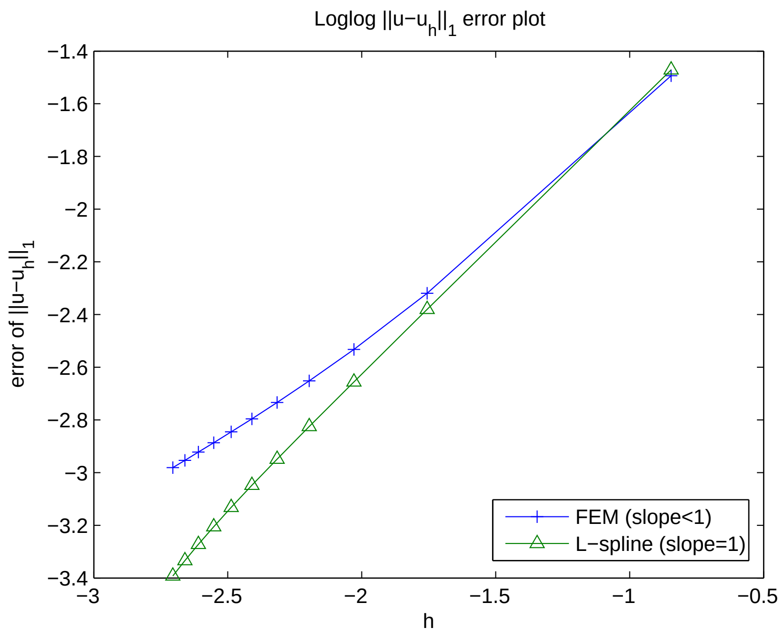

Figure 3.2: The log-log plot of $\|u - u_h\|_1$ for second order problem.

h	$\|u - u_h\|_1^{\text{FEM}}$	$\|u - u_h\|_1^{\text{Lspline}}$
1/7	3.21-2	3.37-2
1/57	4.80-3	4.16-3
1/107	2.93-3	2.21-3
1/157	2.23-3	1.50-3
1/207	1.84-3	1.12-3
1/257	1.60-3	8.94-4
1/307	1.43-3	7.38-4
1/357	1.30-3	6.23-4
1/407	1.20-3	5.34-4
1/457	1.11-3	4.64-4
1/507	1.04-3	4.05-4

Table 3.1: The energy norm error for second order problem

Example 2. As the second example we consider a fourth order elliptic equation on the unit interval $[0,1]$ with Dirichlet boundary condition:

$$\begin{aligned} &(a(x)u'')'' = 1, \quad 0 < x < 1 \\ &u(0) = u(1) = 0, \quad u'(0) = u'(1) = 0. \end{aligned} \tag{3.12}$$

We choose a smooth but sharply varying function as the coefficient function,

$$a(x) = 6 + \frac{5000(x-0.5)}{1+1000|x-0.5|},$$

which is shown in Fig 3.3. We have the analytic formula of the solution of this problem, which is

$$u(x) = \int_0^x \int_0^t \frac{s^2}{2a(s)} + \frac{cs+d}{a(s)}\,dsdt,$$

where c and d are determined by the boundary conditions and are the solution of the following linear system

$$\begin{pmatrix} \int_0^1\int_0^t \frac{s}{a(s)}dsdt & \int_0^1\int_0^t \frac{1}{a(s)}dsdt \\ \int_0^1 \frac{s}{a(s)}ds & \int_0^1 \frac{1}{a(s)}ds \end{pmatrix} \begin{pmatrix} c \\ d \end{pmatrix} = \begin{pmatrix} -\int_0^1\int_0^t \frac{s^2}{2a(s)}dsdt \\ -\int_0^1 \frac{s^2}{2a(s)}ds \end{pmatrix}$$

and

$$\begin{pmatrix} c \\ d \end{pmatrix} = \begin{pmatrix} -0.39466258071925 \\ 0.05188334464260 \end{pmatrix}.$$

In this case, we use two conforming Galerkin methods, piecewise cubic polynomials and the L-spline space S_Γ defined by

$$S_\Gamma = \{\psi \in H^1[0,1] : (a(x)u'')''|_{I_j} = 0, \quad \text{for } j = 1,2,\cdots,n\},$$

which can be written as

$$\begin{aligned} S_\Gamma \;=\; &\{\psi \in H^1[0,1] \,:\, \text{for each } j,\ \psi|_{I_j} = \text{a linear combination of} \\ &1, x, \int^x\int^s \frac{1}{a(t)}dt\,ds,\ \text{and } \int^x\int^s \frac{t}{a(t)}dt\,ds\}. \end{aligned}$$

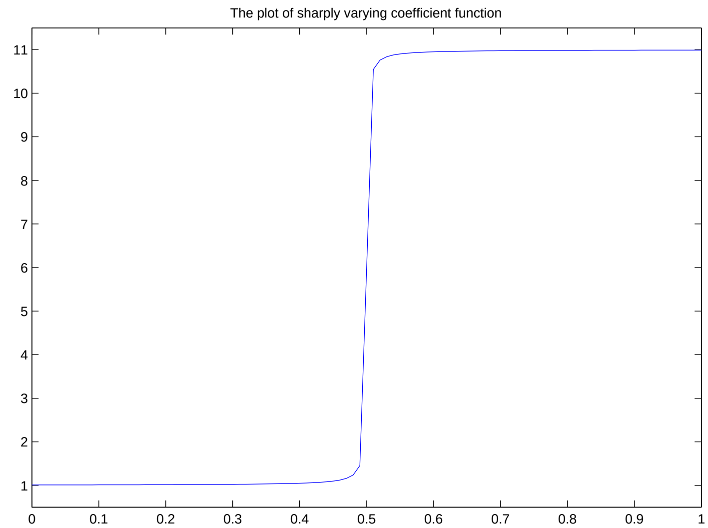

Figure 3.3: The coefficient function with a sharply varying property.

The same uniform mesh as in Example 1 is used. The slope of the convergence curve for L-spline results is 2, which confirms the theoretical result shown in Theorem 3.1. However the finite element only has convergence rate 1. In Figure 3.4 we also see a downward shift of L-spline curve compared to finite element curve. This illustrates a smaller constant C in the error estimation result, which only depends on the lower and upper bounds of the coefficient function in the generalized L-spline case, but depends on the first order derivative of the coefficient in the usual finite element case. So in this fourth order problem, generalized L-spline space provides a better approximation result not only with a faster convergence rate but also with a smaller constant C.

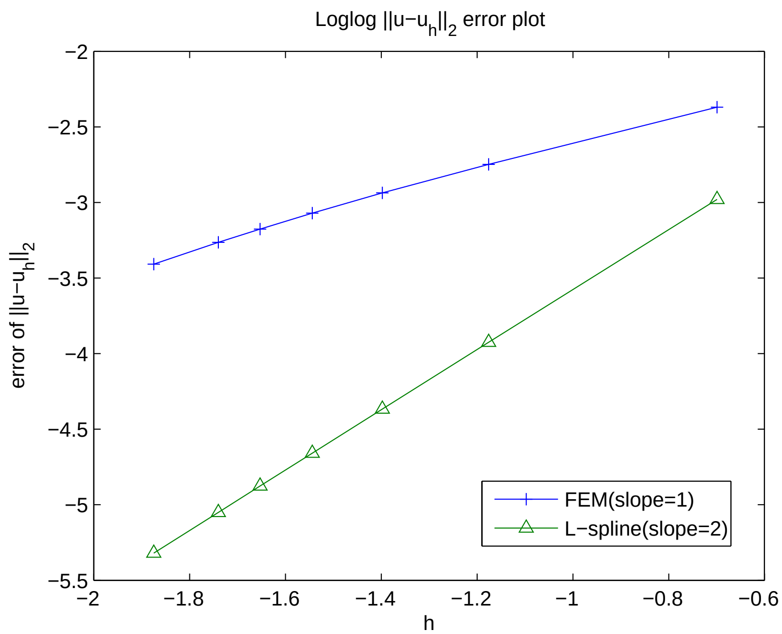

Figure 3.4: The log-log plot of $\|u - u_h\|_2$ for fourth order problem.

3.4 Negative Norm Estimates

In this section, we will provide negative norm error estimates for the problem (3.1) without lower order terms, i.e., $B \equiv 0$. In this case, the problem operator $A = L^*L$ is self-adjoint. These estimates have important applications. We are interested in using it to prove error estimates for the approximation of eigenvalues and eigenfunctions, which will be discussed in Section 3.5. Since L^*L is symmetric, we automatically have the estimate for the adjoint problem, which is used in the proof of negative norm error estimates.

For $s = 0, 1, \cdots$ the negative norm is defined as

$$\|u\|_{-s} = \sup_{v \in H_0^s} \frac{|\int_a^b uv\,dx|}{\|v\|_s}.$$

h	$\|u-u_h\|_2^{\text{FEM}}$	$\|u-u_h\|_2^{\text{Lspline}}$
1/5	4.27-3	1.05-3
1/15	1.79-3	1.19-4
1/25	1.16-3	4.30-5
1/35	8.50-4	2.19-5
1/45	6.67-4	1.33-5
1/55	5.44-4	8.90-6
1/75	3.91-4	4.79-6

Table 3.2: The energy norm error for fourth order problem

Theorem 3.5 *For problem (3.1) with* $B \equiv 0$*, let* $f \in H^k(a,b)$ *and* u_h *be the Galerkin solution in the generalized L-spline space* S_Γ^r*. Then, for* $0 \le s \le r+1$ *and* $0 \le k \le r+1$*, we have*

$$\|u-u_h\|_{-s} \le Ch^{2m+k+s}\|f\|_k.$$

Proof. By the definition of negative norm, we have

$$\|u-u_h\|_{-s} = \sup_{v \in H_0^s} \frac{|\int_a^b (u-u_h)v\,dx|}{\|v\|_s}.$$

Since

$$|(u-u_h,v)| = |(v,u-u_h)| = |a(\phi,u-u_h)| = |a(u-u_h,\phi)|,$$

where $a(\cdot,\cdot)$ is the bilinear form corresponding to the problem (3.1), and ϕ is the solution for the problem with righthand side $f=v$. By the orthogonality relation of the Galerkin solution, we obtain

$$|a(u-u_h,\phi)| = |a(u-u_h,\phi-\chi)|, \quad \forall \chi \in S_\Gamma^r.$$

Then it follows from Theorem 3.1 and Theorem 2.8 that

$$\begin{aligned}
|(u-u_h,v)| &= |a(u-u_h,\phi-\chi)| \\
&\leq C\|u-u_h\|_m \inf_{\chi\in S^r_\Gamma} \|\phi-\chi\|_m \\
&\leq Ch^{k+m}\|L^*Lu\|_k \, h^{s+m}\|L^*L\phi\|_s \\
&= Ch^{2m+k+s}\|L^*Lu\|_k \, \|v\|_s,
\end{aligned}$$

for $0 \leq s \leq r+1$. Dividing both hand sides by $\|v\|_s$, and using the definition of the negative norm, we have the assets. □

3.5 Eigenvalue Problems with Rough Coefficients

Eigenvalue problems arise in physics and engineering, in problems such as of heat conduction and of vibration of a spring or an elastic solid. They also appear in stability analysis of nonlinear problems. Generalized L-spline spaces can be also used for approximation of eigenvalue problems with rough coefficients. In this section, we will apply the generalized L-spline spaces for the eigenvalue problem corresponding to the source problem (3.1) without lower order terms, i.e., $B \equiv 0$. The error estimate results coincide with the polynomial finite element spaces for problem with smooth coefficients.

In order to carry out the analysis of error estimate, we have to review some results for eigenvalue problems. The inverse operator T of the problem differential operator $A = L^*L$ is a compact operator. A spectral approximation theory for compact operators was developed by Osborn in [44]. Further, a complete development of the spectral theory for compact operators can be found in [28]. Babuška and

Osborn gave a survey of general information on eigenvalue problems in [11].

Let $T : X \to X$ be a compact operator on a complex Banach space X. We denote by $\sigma(T)$ and $\rho(T)$ the spectrum and resolvent sets of T, respectively. For any $z \in \rho(T)$, $R_z(T) = (z - T)^{-1}$ is the resolvent operator. $\sigma(T)$ is countable and nonzero numbers in $\sigma(T)$ are eigenvalues. If zero is in $\sigma(T)$, it may or may not be an eigenvalue.

Let $\mu \in \sigma(T)$ be nonzero. The smallest integer α such that $N((\mu - T)^\alpha) = N((\mu - T)^{\alpha+1})$, where N denotes the null space, is called the ascent of $\mu - T$. $N((\mu - T)^\alpha)$ is finite dimensional and $m = \dim N((\mu - T)^\alpha)$ is called the *algebraic multiplicity of* μ. The vectors in $N((\mu - T)^\alpha)$ are called the generalized eigenvectors of T corresponding to μ. The geometric multiplicity of μ is equal to $\dim N(\mu - T)$, and is less than or equal to the *algebraic multiplicity*. The two multiplicities are equal if X is a Hilbert space and T is selfadjoint.

Throughout this section, we will consider a compact operator T and a family of compact operators $T_h : X \to X$, $0 < h \leq 1$, such that $T_h \to T$ in norm as $h \to 0$. Let μ be a nonzero eigenvalue of T with algebraic multiplicity m. Let Γ be a circle in the complex plane centered at μ which lies in $\rho(T)$ and which encloses no other points of $\sigma(T)$. The spectral projection associated with T and μ is defined by

$$E = E(\mu) = \frac{1}{2\pi i} \int_\Gamma R_z(T)\, dz.$$

E is a projection onto the space of generalized eigenvectors associated with T and the nonzero eigenvalue μ of T, i.e., $R(E) = (N(\mu - T)^\alpha)$, where R denotes the range.

For h sufficiently small, $\Gamma \subset \rho(T_h)$ and the spectral projection

$$E_h = E_h(\mu) = \frac{1}{2\pi i} \int_\Gamma R_z(T_h)\, dz$$

exists, E_h converges to E in norm, and $\dim R(E_h(\mu)) = \dim R(E(\mu)) = m$. E_h is the spectral projection associated with T_h and the eigenvalues of T_h which lie in Γ and is a projection onto the direct sum of the spaces of generalized eigenvectors corresponding to these eigenvalues. Thus, counting according to algebraic multiplicities, there are m eigenvalues of T_h in Γ; we denote these by $\mu_1(h), \cdots, \mu_m(h)$. Furthermore, if Γ' is another circle centered at μ with an arbitrarily small radius, we see that $\mu_1(h), \cdots, \mu_m(h)$ are all inside of Γ' for h sufficiently small, i.e., $\lim_{h\to 0} \mu_j(h) = \mu$ for $j = 1, \cdots, m$. $R(E)$ and $R(E_h)$ are invariant subspaces for T and T_h, respectively, and $TE = ET$ and $T_hE_h = E_hT_h$. $\{R_z\{T_h\} : z \in T, h \text{ small}\}$ is bounded.

If μ is an eigenvalue of T with algebraic multiplicity m, then μ is an eigenvalue with algebraic multiplicity m of the adjoint operator T^* on the dual space X^*. The ascent of $\mu - T^*$ will be α. E^* will be the projection operator associated with T^* and u; likewise E_h^* will be the projection operator associated with T_h^* and $\mu_1(h), \cdots, \mu_m(h)$. If $\phi \in X$ and $\phi^* \in X^*$, we will denote the value of the linear functional ϕ^* at ϕ by $[\phi, \phi^*]$.

Given two closed subspaces M and N of X, we define

$$\delta(M, N) = \sup_{x \in M, \|x\|=1} dist(x, N),$$

and

$$\hat{\delta}(M, N) = \max(\delta(M, N), \delta(N, M)).$$

$\hat{\delta}(M,N)$ is called the gap between M and N. We will use gap as a natural way to formulate results on the approximation of generalized eigenvectors.

In [44], two main convergence estimate theorems are given as the following:

Theorem 3.6 *There is a constant C_1 independent of h, such that*

$$\hat{\delta}(R(E), R(E_h)) \leq C_1 \|(T - T_h)|_{R(E)}\|$$

for small h, where $(T - T_h)|_{R(E)}$ denotes the restriction of $T - T_h$ to $R(E)$, and $\hat{\delta}$ is the gap between two spaces.

Theorem 3.7 *Let $\phi_1, \cdots, \phi_m$ be any basis for $R(E)$ and let $\phi_1^*, \cdots, \phi_m^*$ be the dual basis in $R(E^*)$. Then there is a constant C_2 such that*

$$|\mu - \hat{\mu}(h)| \leq \frac{1}{m} \sum_{j=1}^{m} |[(T - T_h)\phi_j, \phi_j^*]| + C_2 \|(T - T_h)|_{R(E)}\| \ \|(T^* - T_h^*)|_{R(E^*)}\|,$$

where $\hat{\mu}(h)$ is the average of the eigenvalues approaching μ, i.e., $\hat{\mu}(h) = \frac{1}{m}\sum_{j=1}^{m} \mu_j(h)$.

We consider the eigenvalue problem corresponding to the boundary value problem (3.1) with $B = 0$,

$$\begin{aligned} &Au = L^* L u = \lambda u, \quad a < x < b, \\ &D^k u(a) = D^k u(b) = 0, \quad k = 0, 1, \cdots, m-1, \end{aligned} \tag{3.13}$$

The variational form of this problem is

$$\begin{aligned} &\text{Seek } \lambda,\ 0 \neq u \in H_0^m(a,b) \text{ satisfying} \\ &a(u,v) = \lambda(u,v), \quad \forall v \in H_0^m(a,b). \end{aligned} \tag{3.14}$$

The form $a(\cdot,\cdot)$ as defined in (3.2) with $B = 0$ is bounded and coercive, i.e.,

$$|a(u,v)| \leq C_1 \|u\|_m \|v\|_m, \quad \forall u, v \in H_0^m(a,b)$$

and

$$Re\, a(u,u) \geq C_2 \|u\|_m^2 \quad \forall u \in H_0^m(a,b),$$

where $\alpha > 0$. We introduce the solution operator $T : L^2(a,b) \to L^2(a,b)$ for the boundary value problem (3.1), i.e., $u = Tf$ solves problem (3.1), defined by

$$Tf \in H_0^m(a,b), \quad a(Tf, v) = (f, v) \quad \forall v \in H_0^m(a,b). \tag{3.15}$$

Thus T is the inverse of the differential operator L^*L, considered on functions that satisfy the boundary conditions. It follows immediately from the Lax-Milgram theorem, that (3.15) has a unique solution Tf for each $f \in L^2(a,b)$. Let $v = Tf$ in (3.15), then by coercivity we have

$$C\|Tf\|_m^2 \leq a(Tf, Tf) = (f, Tf) \leq \|f\|_0 \|Tf\|_0 \leq \|f\|_0 \|Tf\|_m,$$

and thus for each $f \in L^2(a,b)$,

$$\|Tf\|_m \leq C\|f\|_0,$$

which means that $T : L^2(a,b) \to H_0^m(a,b)$ is bounded. Since $H_0^m(a,b)$ is compactly embedded in $L^2(a,b)$, we see that $T : L^2(a,b) \to L^2(a,b)$ is a compact operator by Rellich's theorem. It follows immediately from (3.14) and (3.15) that (λ, u) is an eigenpair of (3.14) if and only if

$$Tu = \mu u, \quad u \neq 0,$$

i.e., if and only if $(\mu = \lambda^{-1}, u)$ is an eigenpair of T. Through this correspondence, properties of the eigenvalue problem (3.14) can be derived from the spectral theory for compact operators. From the facts that T is selfadjoint on $H_0^m(a,b)$ and is positive definite, T has a countably infinite sequence of eigenvalues

$$0 \swarrow \cdots \leq \mu_2 \leq \mu_1,$$

and associated with eigenfunctions

$$u_1, u_2, \cdots,$$

which satisfy

$$a(u_i, u_j) = \mu_i^{-1}(u_i, u_j) = \delta_{ij}.$$

Furthermore, the μ_j can be characterized as various extrema of the Rayleigh quotient

$$R(u) = \frac{(Tu, u)}{(u, u)}.$$

We state *Minimum-Maximum Principle* here:

$$\mu_j = \min_{V \subset H, \dim V = j-1} \max_{u \perp V} R(u). \tag{3.16}$$

Now we consider the approximate eigenpairs defined by the Galerkin method with generalized L-spline space as trial and testing spaces, where

$$S_\Gamma^r = \{\phi \in H_0^m(a,b), L^* L\phi|_{I_j} \in P^r(I_j), \quad \text{for } j = 1, 2, \cdots, n\}.$$

We define the approximation solution operator $T_h : L^2(a,b) \to H_0^m(a,b)$ corresponding to the solution operator (3.15), by

$$T_h f \in S_\Gamma^r \subset H_0^m(a,b), \quad a(T_h f, v) = (f, v) \quad \forall v \in S_\Gamma^r. \tag{3.17}$$

Form Theorem 3.6 and Theorem 3.7, we see that we need to obtain estimates for $T - T_h$, in order to get the error estimates for the eigenvalues and eigenfunctions. The following lemma is the key to proving the error bounds for $T - T_h$.

Lemma 3.5 *Suppose $f \in H^m(a,b)$. Let T and T_h be the solution operator and approximate solution operator defined in (3.15) and (3.17) with generalized L-spline space S_Γ^r, for some r with $0 \le r+1 \le m$, respectively. Then we have*

$$|((T-T_h)f,\psi)| \le Ch^{2m+r+1+s}\|f\|_{r+1}\|\psi\|_s, \tag{3.18}$$

for $\psi \in H^s(a,b)$ with $0 \le s \le r+1$.

Proof. From the orthogonality of the Galerkin solution, we have $a((T-T_h)f,\phi) = 0$ for all $\phi \in S_\Gamma^r$. Since the operator L^*L is self-adjoint and so is the solution operator T, it follows that

$$\begin{aligned} ((T-T_h)f,\psi) &= (\psi,(T-T_h)f) \\ &= a(T\psi,(T-T_h)f) \\ &= a((T-T_h)f, T\psi-\chi) \\ &\le C\|(T-T_h)f\|_m \inf_{\chi\in S_\Gamma^r}\|T\psi-\chi\|_m. \end{aligned}$$

Applying Theorem 3.1 with $l=m$, $r+1 \le k \le m$ and $0 \le r+1 \le m$, we have

$$\|(T-T_h)f\|_m \le Ch^{m+r+1}\|f\|_{r+1},$$

and the interpolant error estimate Theorem 2.8 with $l=m$, $k=s$ and $0 \le s \le r+1$, provides that

$$\|T\psi - I_\Gamma^r(T\psi)\|_m \le Ch^{m+s}\|\psi\|_s.$$

Combining the above three arguments, we obtain that

$$|((T-T_h)f,\psi)| \le Ch^{2m+r+1+s}\|f\|_{r+1}\|\psi\|_s.$$

From the above lemma we can obtain the estimates needed to apply Theorem 3.6 and Theorem 3.7 to get the desired error estimates for the approximation of eigenpairs using generalized L-spline spaces.

Theorem 3.8 *Let λ be an eigenvalue of (3.14) of multiplicity m and $\mu = \frac{1}{\lambda}$, which is an eigenvalue of the solution operator T. Let*

$$M = M(\mu) = \{u \in H_0^m(a,b) : u \text{ is an eigenfunction of (3.14) corresponding to } \mu\}.$$

Let (μ_h, u_h) be the approximate eigenpairs determined by the Galerkin method based on generalized L-spline spaces S_Γ^r. Then $M(\mu) \subset H^m(a,b)$. We have

$$|\mu - \hat{\mu}(h)| \leq Ch^{2m+2r+2},$$

and

$$\hat{\delta}(R(E), R(E_h)) \leq Ch^{2m+r+1},$$

for $0 \leq r+1 \leq m$.

Proof. Inequality (3.18) with $r+1=0$ and $s=0$ yields

$$|((T-T_h)f, \psi)| \leq Ch^{2m}\|f\|_0\|\psi\|_0,$$

and hence

$$\|(T-T_h)f\|_0 = \sup_{\psi \in L^2(a,b), \|\psi\|_0=1} |((T-T_h)f,\psi)| \leq Ch^{2m}\|f\|_0,$$

for all $f \in L^2(a,b)$. This shows that $T_h \to T$ in $L^2(a,b)$ norm, and hence that all the results of Theorem 3.6 and Theorem 3.7 apply. Now we estimate $\|(T-T_h)\|_{R(E)}$. To this end let $f \in R(E)$. Under the assumptions the eigenvectors and generalized eigenvectors of T (or L) are in $H^m(a,b)$. Thus from (3.18) with $0 \le r+1 \le m$ and $s=0$ we get

$$|((T-T_h)f,\psi)| \le Ch^{2m+r+1}\|f\|_{r+1}\|\psi\|_0,$$

and hence

$$\begin{aligned}\|(T-T_h)_{R(E)}\| &= \sup_{f\in R(E),\psi\in L^2(a,b);\|f\|_0=\|\psi\|_0=1} |((T-T_h)f,\psi)| \\ &\le Ch^{2m+r+1} \sup_{f\in R(E),\|f\|_0=1} \|f\|_{r+1} \le Ch^{2m+r+1}. \qquad (3.19)\end{aligned}$$

Since L^*L is self-adjoint, the solution operator T is self-adjoint, and so is the approximate solution operator T_h. We have that

$$\|(T^*-T_h^*)_{R(E^*)}\|\|(T-T_h)_{R(E)}\| \le Ch^{2*(2m+r+1)}. \qquad (3.20)$$

Finally we consider $\sum_{j=1}^m |((T-T_h)\phi_i,\phi_j^*)|$. It follows immediately form (3.18) with $0 \le r+1 \le m$ and $s=r+1$ that

$$\sum_{j=1}^m |((T-T_h)\phi_i,\phi_j^*)| \le Ch^{2m+2r+2}. \qquad (3.21)$$

Thus using (3.19), (3.20) and (3.21) we see that Theorem 3.7 yields the estimate

$$|\mu-\hat{\mu}(h)| \le Ch^{2m+2r+2}.$$

Using Theorem 3.6 from (3.19) we directly have

$$\hat{\delta}(R(E),R(E_h)) \le Ch^{2m+r+1},$$

for the generalized eigenvectors. □

For the second order elliptic problems with smooth coefficients, the well known error estimate results for the approximate eigenpairs $(\tilde{\mu}_h,\tilde{u}_h)$ determined by the finite element method with piecewise polynomials are

$$|\tilde{\mu} - \tilde{\mu}_h| \leq Ch^{2k} \|u\|_{k+1}^2,$$

and

$$\|u - \tilde{u}_h\|_1 \leq Ch^k \|u\|_{k+1},$$

assuming that the generalized eigenvectors u are in H^{k+1} and the power r of the polynomial used in the finite element method is high enough, i.e., $k \leq r$. Because of the regularity of the source problem, i.e., the solution u is two orders smoother than the right hand side function f, we can assume the eigenfunction with as high smoothness as we want to have a higher convergence rate. However, for problems with rough coefficients, the regularity of the solution does not hold. Thus for $2m$th order elliptic problem, the eigenfunction is only in H_0^m. In Theorem 3.8, the power r of the polynomial, used in the generalized L-spline space, is used to control the error and the convergence rate reaches its upper limit, when $r + 1 = m$.

Chapter 4

The n-widths in Approximation Theory

A finite dimensional subspace is usually chosen in which to find the approximation solution numerically. It is natural to ask what is the best approximation subspace and how to chose it. In this chapter, first the n-width is introduced as a criterion utilized to determine the optimal n-dimensional approximation subspaces. Then some simple examples are given to demonstrate the application of n-width theorem. In the last two sections, it is used to show that the generalized L-splines are optimal subspaces for Hilbert spaces with specific norms, which suit the problems with rough coefficients, and that a special approximation space proposed in [10] is an optimal approximation space for a class of second order, two dimensional elliptic boundary value problems with rough or highly oscillating coefficients.

Let H be a normed linear space and X_n any n-dimensional subspace of H. The distance of the n-dimensional subspace X_n from $x \in H$ is defined by

$$E(x; X_n) = \inf_{y \in X_n} ||x - y||_H.$$

If A is a subset of H instead of a single element x in H, deviation is commonly used to represent how well the n-dimensional subspace X_n of H approximates the A. The definition of the *deviation* of A from X_n is

$$E(A; X_n) = \sup_{x \in A} \inf_{y \in X_n} ||x - y||_H.$$

The idea of n-width was first introduced by Kolmogorov; see [35] and [45]. It considers the possibility of allowing the n-dimensional subspace X_n to vary within H. So that it answers the question of how well one can approximate A by an n-dimensional subspaces of H.

Definition 4.1 *Let H be a normed linear space and A be a subset of H. The n-width of A in H is defined by*

$$d_n(A; H) = \inf_{X_n} \sup_{x \in A} \inf_{y \in X_n} ||x - y||_H,$$

the left most infimum being taken over all n-dimensional subspaces X_n of H.

A subspace $\overline{X}_n$ of H of dimension at most n for which the n-width takes the infimum is called an optimal subspace for $d_n(A; H)$; $\overline{X}_n$ satisfies

$$\sup_{x \in A} \inf_{y \in \overline{X}_n} ||x - y||_H \leq \sup_{x \in A} \inf_{y \in X_n} ||x - y||_H,$$

for $X_n \subset H$ with $\dim X_n = n$.

4.1 n-widths of the Image of the Unit Ball under a Compact Operator in a Hilbert Space

In general, it is difficult to obtain $d_n(A; H)$ and determine an optimal subspace X_n for $d_n(A; H)$ (if they exist) for all A and H. However, we are interested in the case that A is the image of the unit ball under a compact mapping K from H to H.

Let $K : H \to H$ be a compact operator and K^* be the adjoint operator of K, then $K^*K : H \to H$ is compact, self-adjoint and positive semi-definite. Consider the eigenpair (μ_i, ϕ_i) of K^*K, i.e., $(K^*K)\phi_i = \mu_i \phi_i$, $i = 1, 2, \ldots$, where $\sqrt{\mu_1}, \sqrt{\mu_2}, \ldots$, are the singular values of the operator K. With ϕ_i defined as above, let $\psi_i = K\phi_i$, then

$$KK^*\psi_i = KK^*K\phi_i = K\mu_i\phi_i = \mu_i\psi_i,$$

i.e., (μ_i, ψ_i) are eigenpairs of the operator KK^*. Then the n-width of the set

$$A = \{Kf \in H_2 : ||f||_H \leq 1\}$$

as a subset of H, is the $(n+1)$th singular value of K and optimal subspaces are easily constructed in terms of eigenvector subspaces. This result is stated as follows

Theorem 4.1 *Let $A = \{Kf \in H : ||f||_H \leq 1\}$. Then the n-width of A in H is given by*

$$d_n(A; H) = \mu_{n+1}^{\frac{1}{2}}.$$

Furthermore, is $X_n = span\{\psi_1, \psi_2, \ldots, \psi_n\}$ is an optimal subspace for A.

Proof. From the definition

$$d_n(A; H) = \inf_{X_n} \sup_{||h||_H \leq 1} \inf_{g \in X_n} ||Kh - g||_H,$$

we find that

$$\inf_{g \in X_n} ||Kh - g||_H = ||Kh - g_n||_H,$$

where g_n is the orthogonal projector Kh onto X_n, i.e., $Kh - g_n \in X_n^{\perp}$. Then

$$\inf_{g \in X_n} ||Kh - g||_H = ||Kh - g_n||_H = \sup_{f \perp X_n} \frac{(Kh - g_n, f)_H}{||f||_H} = \sup_{f \perp X_n} \frac{(Kh, f)_H}{||f||_H}.$$

Hence

$$\begin{aligned} d_n(A;H) &= \inf_{X_n} \sup_{||h||_H \le 1} \sup_{f \perp X_n} \frac{(Kh,f)_H}{||f||_H} = \inf_{X_n} \sup_{||h||_H \le 1} \sup_{f \perp X_n} \frac{(h,K^*f)_H}{||f||_H} \\ &= \inf_{X_n} \sup_{f \perp X_n} \sup_{||h||_H \le 1} \frac{(h,K^*f)_H}{||f||_H} = \inf_{X_n} \sup_{f \perp X_n} \frac{||K^*f||_H}{||f||_H} \\ &= \left\{ \inf_{X_n} \sup_{f \perp X_n} \frac{(KK^*f,f)_H}{||f||_H^2} \right\}^{\frac{1}{2}} = \mu_{n+1}^{\frac{1}{2}}, \end{aligned}$$

by the min-max principle for the operator KK^*. Furthermore, the infimum and the supremum in the above expression are obtained by the choice $X_n = span\{\psi_1, \psi_2, \ldots, \psi_n\}$ and $f = \psi_{n+1}$ is an optimal vector, follows from the optimality statements (3.16) for the minimum-maximum principle for eigenvalues and eigenfunctions of self-adjoint, non-negative, compact operators. □

Next suppose $V \subset\subset H$, i.e., V is compactly contained in H, and is dense in H. We then consider $d_n(A, H)$, where $A = \{u \in V : \|u\|_V \le 1\}$. The following theorem can be viewed as a variational version of Theorem 4.1.

Theorem 4.2 *Let V and H be two Hilbert spaces with V compactly and densely contained in H with inner products and norms $(u,v)_V$, $\|u\|_V$, $(u,v)_H$ and $\|u\|_H$. Then*

$$d_n(\{g \in V : ||g||_V \le 1\}; H) = \mu_{n+1}^{\frac{1}{2}}, \tag{4.1}$$

where (μ_j, u_j) are the eigenpairs of the variationally formulated eigenvalue problem

$$\begin{cases} u_j \in V,\ u_j \neq 0 \\ (u_j,\phi)_V = \mu_j^{-1}(u_j,\phi)_H \quad \forall \phi \in V; \end{cases} \tag{4.2}$$

an optimal n-dimensional subspace is given by $span\{u_1, \cdots, u_n\}$ and $u = u_{n+1}$ is an optimal vector.

Proof. We first define an operator K so that Theorem 4.1 with this K yields (4.1), and then show that the square roots of eigenvalues of (4.2) are the singular values of K.

Let $T : H \to H$ be defined by

$$\begin{cases} Tf \in V \\ (Tf, \phi)_V = (f, \phi)_H, \quad \forall \phi \in V. \end{cases}$$

For any $f \in H$, $(f, \phi)_H$ is a linear functional on V. By Riesz representation theorem, there exists a $Tf \in V$, s.t. $(Tf, \phi)_V = (f, \phi)_H$. So T is well defined. Since V is compactly embedded in H, $T : H \to V$ is bounded. The operator T can be viewed as the product of the identity operator $I : V \to H$, which we are assume to be compact, and the bounded operator $T : H \to V$, so $T : H \to H$ is compact. It is also easily seen to be symmetric and positive definite. Symmetry is seen from

$$(f, T\phi)_H = (Tf, T\phi)_V = (T\phi, Tf)_V = (\phi, Tf)_H = (Tf, \phi)_H,$$

and positive definiteness from

$$(f, Tf)_H = (Tf, Tf)_V.$$

A complete discussion of such variational problem is provided in [6].

Let $K = T^{\frac{1}{2}}$, where $T^{\frac{1}{2}} f$ is defined by

$$T^{\frac{1}{2}} f = \sum_j (f, u_j)_H \sqrt{\mu_j}\, u_j,$$

where (μ_j, u_j) are eigenpairs of operator T.

It is easily shown that $R(T^{\frac{1}{2}}) = V$ and $T^{\frac{1}{2}}$ is self-adjoint. Furthermore

$$\|T^{\frac{1}{2}} f\|_V = (T^{\frac{1}{2}} f, T^{\frac{1}{2}} f)_V = (Tf, f)_V = (f, f)_H = \|f\|_H \quad \forall f \in H,$$

so

$$\|u\|_V = \|T^{-\frac{1}{2}}u\|_H \quad \forall u \in V.$$

Hence, writing $u = Kf = T^{\frac{1}{2}}f$, we have

$$\{Kf : ||f||_H \leq 1\} = \{u \in V : ||u||_V \leq 1\},$$

and from Theorem 4.1, we have

$$d_n(\{g \in V : ||g||_V \leq 1\}; H) = \mu_{n+1}^{\frac{1}{2}},$$

where $\sqrt{\mu_{n+1}}$ is the $(n+1)$th singular value of K. By the definition of operator K, μ_{j+1} is the $(n+1)$th eigenvalue of $K^*K = (T^{\frac{1}{2}})^*T^{\frac{1}{2}} = T$, which is characterized by $Tu_i = \mu_i u_i, \quad i = 1, 2\cdots$, and can also be characterized variationally by

$$\begin{cases} u_j \in V \\ (u_j, \phi)_V = \lambda_j (u_j, \phi)_H \quad \forall \phi \in V, \end{cases} \tag{4.3}$$

where $\lambda_j = \mu_j^{-1}$. Then by Theorem 4.1, we have $span\{u_1, \cdots, u_n\}$ is an optimal n-dimensional subspace and $u = u_{n+1}$ is an optimal vector. □

Note here that the operator T is compact, symmetric and positive definite. So it has countable infinite sequence of eigenvalues

$$\mu_1 \geq \mu_2 \geq \cdots \searrow 0.$$

And on the other hand the eigenvalues λ_j of the inverse operator of T in (4.3) has countable infinite sequence of eigenvalues

$$\lambda_1 \leq \lambda_2 \leq \cdots \nearrow \infty.$$

And λ can be characterized as minimums of the Rayleigh quotient

$$\lambda_j = \min \frac{(u,u)_V}{(u,u)_H} = \min \frac{\|u\|_V^2}{\|u\|_H^2},$$

subject to certain orthogonality conditions. Specifically

$$\lambda_1 = \min_{u \in V} \frac{\|u\|_V^2}{\|u\|_H^2},$$

$$\lambda_2 = \min_{u \in V,\, (u,u_1)_H = 0} \frac{\|u\|_V^2}{\|u\|_H^2}, \cdots$$

Theorem 4.2 essentially shows that the optimal n dimensional approximation space for V is the span of the first n eigenfunctions of the eigenvalue problem corresponding to the associated two norms of the space H and the subspace V. There are two applications of Theorem 4.2: one is that we can find the optimal approximation subspace by solving an corresponding eigenvalue problem, another is that we can justify an optimal approximation subspace by finding the appropriate norms.

4.2 Examples of n-widths and Optimal Approximation Subspaces in Hilbert Spaces

In this section, five examples are given to illustrate the applications of n-width theorem. The first example shows the importance of trigonometric functions. The second and third examples indicate that the standard finite element space is an optimal approximation space for certain spaces. The space with piecewise cubic polynomial is also optimal in the sense of n-width with corresponding mesh-dependent norms. The last example presents the standard finite element space as an optimal space in two dimensional space.

Example 1. Let $H = L^2[0,1]$ and $V = H_0^1[0,1]$ with inner products and norms

$$(u,\phi)_{L_2} = \int_0^1 u\phi dx, \quad \|u\|_{L^2} = (\int_0^1 u^2 dx)^{\frac{1}{2}}$$

and

$$(u,\phi)_V = \int_0^1 u'\phi' dx, \quad \|u\|_{H_0^1} = (\int_0^1 (u')^2 dx)^{\frac{1}{2}}.$$

Then the eigenvalue problem (4.3) is to find (μ_j, u_j) such that

$$(u_j, \phi)_{H_0^1} = \mu_j^{-1}(u_j, \phi)_{L^2} \quad \forall \phi \in H_0^1,$$

i.e.

$$\int_0^1 u_j'\phi' dx = \mu_j^{-1} \int_0^1 u_j \phi dx.$$

Using integration by parts and moving the right hand side to the left, we have

$$\int_0^1 (-u_j'' - \mu_j^{-1} u_j)\phi dx = 0, \quad \forall \phi \in H_0^1,$$

which becomes a regular second order eigenvalue problem

$$-u_j'' = \mu_j^{-1} u_j, \quad u_j(0) = u_j(1) = 0.$$

By the standard ordinary differential equation result, we know that $u_j = \sqrt{2}\, sin(j\pi x)$ and $\mu_j = \frac{1}{(j\pi)^2}$ for $j = 1, 2, \cdots$ are the eigenfunctions and eigenvalues. Hence, $d_n(\{u \in H_0^1 : ||u||_{H_0^1} \leq 1\}; L^2) = \frac{1}{(n+1)\pi}$ and the optimal subspace is the span of $\{\sqrt{2}\, sin(\pi x), \sqrt{2}\, sin(2\pi x), \cdots, \sqrt{2}\, sin(n\pi x)\}$.

The finite element method has been widely used to solve partial differential equations by both engineers and mathematicians for the last several decades. It is natural to ask what makes the usual piecewise linear approximation subspace so useful. The following two examples show that the finite element approximating space is an optimal subspace for Hilbert spaces with mesh-dependent norms with and without zero boundary condition, respectively.

Example 2. Let $H = H^1[0,1]$ with inner product and norm

$$(u,\phi)_H = (u,\phi)_1 = \int_0^1 (u\phi + u'\phi')\,dx,$$

$$\|u\|_H = (\int_0^1 (u^2 + (u')^2)\,dx)^{\frac{1}{2}}.$$

For V we consider a mesh dependent space: Let $\Gamma = \{0 = x_0 < x_1 < \cdots < x_{n-1} < x_n = 1\}$ be a partition of the interval $[0,1]$, let

$$V = \{u \in H^1[0,1], u|_{I_j} \in H^2(I_j), \text{ for } j = 1, 2\cdots, n\},$$

with inner product and norm

$$(u,\phi)_V = \int_0^1 (u\phi + u'\phi')\,dx + \sum_{j=1}^{n} \int_{I_j} u''\phi''\,dx = (u,v)_H + \sum_{j=1}^{n} \int_{I_j} u''\phi''\,dx$$

and

$$\|u\|_V = (\|u\|_H^2 + \sum_{j=1}^{n} \int_{I_j} (u'')^2\,dx)^{\frac{1}{2}} = (\|u\|_H^2 + |u|_{H^2_\Gamma}^2)^{\frac{1}{2}},$$

where

$$|u|_{H^2_\Gamma}^2 = \sum_{j=1}^{n} \int_{I_j} (u'')^2\,dx.$$

We easily see that H and V are Hilbert spaces and the usual C^0, piecewise linear subspace $S_\Gamma = \{u \in H^1[0,1] : u|_{I_j} \in P^1, j = 1, \cdots, n\}$ is in V. Note that $S_\Gamma = \{u \in H^1[0,1] : |u|_{H^2_\Gamma} = 0\}$.

Theorem 4.3 *The standard finite element space* S_Γ *is the optimal subspace of* $d_n(\{u \in V : ||u||_V \leq 1\}; H)$*, where spaces* H *and* V *are defined above.*

Proof. We first prove that $V \subset\subset H$ and V is dense in H.

Since $H^2(I_j) \subset\subset H^1(I_j)$, there exists a subsequence which converges in $H^1(I_j)$ for a given bounded sequence in $H^2(I_j)$. Let $\{u_k\}$ be a bounded sequence in V. We can find a subsequence $\{u_{k_1}\}$ of $\{u_k\}$ which converges in $H^1(I_1)$, then find a subsequence $\{u_{k_2}\}$ of $\{u_{k_1}\}$ which converges in $H^1(I_2)$, ect. Finally, we can have a subsequence $\{u_{k_n}\}$ which is not only convergent in $H^1(I_n)$, but also converges in $H^1[0,1]$. This gives us the property of compact embedding.

For any function u in $H^1[0,1]$, there exists a class of sequences $\{u_k^j\}$ in $H^2(I_j)$ which converge to u in $H^1(I_j)$, since $H^2(I_j)$ is dense in $H^1(I_j)$ for $j = 1, \cdots, n$. Now we patch these sequence piece by piece from first subinterval I_1 till the last subinterval I_n to get a sequence which is in V and converges to u in $H^1[0,1]$. Specifically, let $\{u_k^1\}$ and $\{u_k^2\}$ converge to u in $H^1(I_1)$ and $H^1(I_2)$, respectively. We then let $u_j(x) = u_j^1(x)$ on I_1 and $u_j(x) = u_j^2(x) + [u_j^1(x_1^-) - u_j^2(x_1^+)]$ on I_2. It is easily seen that $\lim_{x \to x_1^+} u_j(x) = u_j^1(x_1^-) = \lim_{x \to x_1^-} u_j(x)$. So we have $u_j \in H^1(x_0, x_2)$. And $\|u_j - u\|_{H^1(x_0,x_2)} \to 0$. This is because that

$$\begin{aligned}
\|u_j - u\|^2_{H^1(x_0,x_2)} &= \|u_j - u\|^2_{H^1(I_1)} + \|u_j - u\|^2_{H^1(I_2)} \\
&= \|u_j^1 - u\|^2_{H^1(I_1)} \\
&\quad + \|u_j^2(x) + [u_j^1(x_1^-) - u_j^2(x_1^+)] - u\|^2_{H^1(I_2)} \\
&\leq \|u_j^1 - u\|^2_{H^1(I_1)} + \|u_j^2 - u\|^2_{H^1(I_2)} \\
&\quad + |u_j^1(x_1^-) - u_j^2(x_1^+)|.
\end{aligned}$$

Since every term in the right-hand side goes to 0 as $j \to \infty$, we have the convergence. By the same process, we can have $\{u_j\} \in H^1[0,1]$ so is in V and also converges to u.

Let $\lambda = \mu^{-1}$, then $(u,v)_V = \lambda(u,v)_H$. This eigenvalue problem has eigenvalues $0 < \lambda_1 \leq \lambda_2 \leq \cdots$, and corresponding eigenvectors $u_1, u_2, \cdots$, which are orthonormal in V. By Rayleigh quotient, we know that

$$\lambda_1 = \inf_{u \in V} \frac{\|u\|_V^2}{\|u\|_H^2}. \tag{4.4}$$

Since

$$\lambda_1 = \inf_{u \in V} \frac{\|u\|_V^2}{\|u\|_H^2} = \inf_{u \in V} \frac{\|u\|_H^2 + |u|_{H_\Gamma^2}^2}{\|u\|_H^2} = \inf_{u \in V} \frac{|u|_{H_\Gamma^2}^2}{\|u\|_H^2} + 1 = 1,$$

the minimum is achieved for u such that $|u|_{H_\Gamma^2} = 0$, i.e. for $u \in S_\Gamma$. So 1 is the lowest eigenvalue, and S_Γ, the associate eigenspace has dimension $n+1 =$ the number of nodes, i.e., 1 is an eigenvalue of multiplicity $n+1$, and the associate eigenfunctions are the C^0 piecewise linear functions for the mesh $\triangle$. Noting that $1 = \lambda_1 = \lambda_2 = \cdots = \lambda_{n+1}$, we apply Theorem 4.2 and obtain

$$d_n(\{u \in V : ||u||_V \leq 1\}; H) = \lambda_{n+2}^{-\frac{1}{2}}$$

and $span\{u_1, u_2, \cdots, u_{n+1}\} = S_\Gamma$ is an optimal n-dimensional subspace. □

Example 3. Next we consider $H = H_0^1[0,1]$ with norm $\|u\|_H = (\int_0^1 (u')^2 dx)^{\frac{1}{2}}$, and $V = \{u \in H_0^1[0,1], u|_{I_j} \in H^2, \text{ for } j = 1, 2 \cdots, n\}$ with norm $\|u\|_V = (\|u\|_H^2 + \sum_{j=1}^n \int_{I_j} (u'')^2\, dx)^{\frac{1}{2}}$. Proceeding as in Example 2, we find that the standard finite element space is also the optimal approximation space. For this case, it is informative to find the strong form of the eigenvalue problem as in Example 1, which was studied in [16]. I will give a brief discussion here. We start with eigenvalue problem (4.3) with V and H defined above.

$$u \in V,\ \lambda \in \mathbb{R}$$

$$(u,v)_V = \lambda (u,v)_H,\ \forall v \in V,$$

where

$$(u,v)_H = \int_0^1 u'v'\,dx,$$

and

$$(u,v)_V = \int_0^1 u'v'\,dx + \sum_{j=1}^{n} \int_{I_j} u''v''\,dx.$$

This is the same as

$$\sum_{j=1}^{n} \int_{I_j} u''v''\,dx = (\lambda - 1) \sum_{j=1}^{n} \int_{I_j} u'v'\,dx, \tag{4.5}$$

and integration by parts gives us

$$\sum_{j=1}^{n} \left\{ u''v'|_{x_{j-1}}^{x_j} - u'''v|_{x_{j-1}}^{x_j} + \int_{x_{j-1}}^{x_j} D^4uv\,dx \right\} = (\lambda - 1) \sum_{j=1}^{n} \left\{ u'v|_{x_{j-1}}^{x_j} - \int_{x_{j-1}}^{x_j} u''v\,dx \right\}. \tag{4.6}$$

The integration by parts is valid, because $v|_{I_j}$ is arbitrary in (4.5) and then $u|_{I_j} \in H^4$. Since (4.6) is true for all $v \in V$, by choosing $v \in V$ such that $v(x_{j-1}) = v(x_j) = v'(x_{j-1}) = v'(x_j) = 0$ and $v(x) = 0$ outside (x_j, x_{j-1}), we have

$$D^4u = -(\lambda - 1)u'',\ \text{on } I_j.$$

Then (4.6) becomes

$$\sum_{j=1}^{n} \left\{ u''v'|_{x_{j-1}}^{x_j} - u'''v|_{x_{j-1}}^{x_j} \right\} = (\lambda - 1) \sum_{j=1}^{n} u'v|_{x_{j-1}}^{x_j}. \tag{4.7}$$

Now we choose $v \in V$ such that $v(x_{j-1}) = v(x_j) = 0$ and $v(x) = 0$ outside (x_j, x_{j-1}), then equation (4.7) results in

$$u''v'|_{x_{j-1}}^{x_j} = 0,$$

and since $v'(x_j)$ is arbitrary we get

$$u''(x_j^-) = u''(x_{j-1}^+) = 0.$$

Using this in (4.7) we get

$$-\sum_{j=1}^{n} u'''v|_{x_{j-1}}^{x_j} = (\lambda - 1)\sum_{j=1}^{n} u'v|_{x_{j-1}}^{x_j}.$$

By choosing $v \in V$ such that $v(x_{j-1}) = v(x_{j+1}) = 0$, $v(x_j) = 1$ and $v(x) = 0$ outside (x_j, x_{j-1}), the above equation leads to

$$D^3u(x_j^+) - D^3u(x_j^-) = (\lambda - 1)[u'(x_j^-) - u'(x_j^+)].$$

Combining above results, we have the strong form of this eigenvalue problem is

$$D^4u = -(\lambda - 1)u'', \text{ on } I_j, \; j = 1, \cdots, n$$

$$u''(x_j+) = u''(x_j-) = 0, \; j = 0, \cdots, n$$

$$D^3u(x_j^+) - D^3u(x_j^-) = (\lambda - 1)[u'(x_j^-) - u'(x_j^+)], \; j = 1, \cdots, n-1$$

Then the optimal subspace is the eigenfunctions corresponding to eigenvalue $\lambda = 1$, i.e., u satisfies

$$D^4u = 0, \text{ on } I_j, \; j = 1, \cdots, n$$

$$u''(x_j) = 0, \; j = 0, \cdots, n$$

$$D^3u(x_j^+) - D^3u(x_j^-) = 0, \; j = 1, \cdots, n-1$$

It is easily seen that piecewise linear functions are solutions for the above eigenvalue problem. Thus this provides an alternative proof that the standard finite element space is the optimal subspace.

Note that we have not calculated λ_{n+1}, Theorem 2.8 provides an upper bound for

$$d_n(\{u \in V : ||u||_V \leq 1\}; H) = \lambda_{n+1}^{-\frac{1}{2}} \leq C\, h.$$

This can lead to error estimate result.

Consider the approximation of problem (1.1), we have

$$\|u - u_h\|_1 \leq C \inf_{\chi \in S} \|u - \chi\|_1,$$

where S is the approximation space. Now let $S = S_\Gamma =$ finite element space. Since S_Γ is optimal,

$$\begin{aligned}
\|u - u_h\|_1 &\leq \frac{\inf_{\chi \in S} \|u - \chi\|_1}{\|u\|_V} \|u\|_V \\
&\leq \left(\sup_{v \in V} \frac{\inf_{\chi \in S} \|u - \chi\|_1}{\|v\|_V} \right) \|u\|_V \\
&\leq \left(\sup_{v \in V,\, \|v\|_V \leq 1} \inf_{\chi \in S} \|v - \chi\|_1 \right) \|u\|_V \\
&= \left(\inf_{X_{n-1} \subset H^1,\, \dim X_{n-1} = n-1} \sup_{v \in V,\, \|v\|_V \leq 1} \inf_{\chi \in X_{n-1}} \|v - \chi\|_1 \right) \|u\|_V \\
&\leq \left(\sup_{v \in V,\, \|v\|_V \leq 1} \inf_{\chi \in X_{n-1}} \|v - \chi\|_1 \right) C \|f\|_0,
\end{aligned}$$

for all $X_{n-1} \subset H$ with $\dim X_{n-1} = n - 1$. The quantity in parenthesis is minimized for $X_{n-1} = \overline{X}_{n-1} = S_\Gamma$.

Not only standard finite element methods are optimal, higher order piecewise polynomials are also optimal for spaces with smoother properties. The fourth example demonstrates that the piecewise cubic Hermite polynomials are optimal.

Example 4. Let $H = H^2[0,1]$ with standard Sobolev norm and $V = \{u \in H^2[0,1] : u|_{I_j} \in H^4(I_j), \text{ for } j = 1, \cdots, n\}$, with norm defined by

$$\|u\|_V = (\|u\|_{H^2}^2 + \sum_{j=1}^{n} \int_{I_j} (D^4 u)^2 \, dx)^{\frac{1}{2}} = (\|u\|_{H^2}^2 + |u|_{H_\Gamma^4}^2)^{\frac{1}{2}},$$

where

$$|u|_{H_\Gamma^4}^2 = \sum_{j=1}^{n} \int_{I_j} (D^4 u)^2 \, dx.$$

Theorem 4.4 *The piecewise cubic Hermite finite element space* $S_\Gamma = \{u \in H^2[0,1] : u|_{I_j} \in P^3,\ j = 1, \cdots, n\}$ *is the optimal subspace with respect to spaces* H *and* V*, which are defined above.*

Proof. The key point of this proof is to show that $V \subset\subset H$ and V is dense in H. Once these are done, it follows from the proof of Theorem 4.3, that the optimal subspace has zero mesh-dependent semi-norm, i.e.,

$$|u|_{H_\Gamma^4}^2 = \sum_{j=1}^{n} \int_{I_j} (D^4 u)^2 \, dx = 0.$$

Thus we have that the cubic finite element space is the optimal space.

Since $H^4(I_j) \subset\subset H^2(I_j)$ for $j = 1, 2 \cdots, n$, we can prove $V \subset\subset H$ by using the technique of taking the subsequence of the subsequence as in the proof of the compact embedding property in Theorem 4.3.

Since $H^4(I_j)$ is dense in $H^2(I_j)$ for $j = 1, 2 \cdots, n$, we can find a sequence which converges to a function in $H^2[0,1]$ by choosing sequence on each subinterval I_j and patch them together as shown in Theorem 4.3. For any function u in $H^2[0,1]$, there exists a class of sequences $\{u_k^j\}$ in $H^4(I_j)$ which converges to u in $H^2(I_j)$, since $H^4(I_j)$ is dense in $H^2(I_j)$ for $j = 1, \cdots, n$. Now we patch these sequence piece by piece from first subinterval I_1 till the last subinterval I_n to get a sequence which is in V and converges to u in $H^2[0,1]$. Let $\{u_k^1\}$ and $\{u_k^2\}$ converge to u in $H^2(I_1)$ and $H^2(I_2)$, respectively. We then let $u_j(x) = u_j^1(x)$ on I_1 and $u_j(x) = u_j^2(x) + [Du_j^1(x_1^-) - Du_j^2(x_1^+)](x - x_1) + [u_j^1(x_1^-) - u_j^2(x_1^+)]$ on I_2. It is easily seen that $\lim_{x \to x_1^+} u_j(x) = u_j^1(x_1^-) = \lim_{x \to x_1^-} u_j(x)$ and $\lim_{x \to x_1^+} Du_j(x) = Du_j^1(x_1^-) = \lim_{x \to x_1^-} Du_j$. So we have $u_j \in H^2(x_0, x_2)$. And $\|u_j - u\|_{H^2(x_0,x_2)} \to 0$. This is because that

$$\begin{aligned} \|u_j - u\|^2_{H^2(x_0,x_2)} &= \|u_j - u\|^2_{H^2(I_1)} + \|u_j - u\|^2_{H^2(I_2)} \\ &= \|u_j^1 - u\|^2_{H^2(I_1)} + \|u_j^2(x) + [Du_j^1(x_1^-) - Du_j^2(x_1^+)](x - x_1) \\ &\quad + [u_j^1(x_1^-) - u_j^2(x_1^+)] - u\|^2_{H^2(I_2)} \\ &\leq \|u_j^1 - u\|^2_{H^2(I_1)} + \|u_j^2 - u\|^2_{H^2(I_2)} \\ &\quad + C|Du_j^1(x_1^-) - Du_j^2(x_1^+)| + |u_j^1(x_1^-) - u_j^2(x_1^+)|. \end{aligned}$$

Since every term in the righthand side goes to 0 as $j \to \infty$, we have the convergence. By the same process, we can have $\{u_j\} \in H^2[0,1]$ so is in V and also converges to u. $\square$

Why mesh-dependent norms? When solving a problem numerically, finite dimensional subspaces are constructed usually on a mesh of the domain. With the knowledge of the regularity of the solution and the mesh, this mesh-dependent norms are appropriate norms to use. And Theorem 4.3 shows that the usual finite element space is an optimal subspace with mesh-dependent norms in one dimensional space. The following example indicates that this is also true in two dimensional space.

Example 5. Let $H = H^1(\Omega)$, where $\Omega = [0,1] \times [0,1]$, with the inner product

$$(u,v)_H = \int_0^1 \int_0^1 uv + \nabla u \cdot \nabla v \, dxdy,$$

and corresponding norm

$$\|u\|_H = (\int_0^1 \int_0^1 \left[u^2 + (\frac{\partial u}{\partial x})^2 + (\frac{\partial u}{\partial y})^2 \right] dxdy)^{\frac{1}{2}}.$$

Let T_h be a triangulation of Ω by triangles T of diameter $\leq h$. We consider V as a mesh-dependent space

$$V = \{u \in H^1(\Omega) : u|_T \in H^2(\Omega), \text{ for } T \in T_h\},$$

with inner product and norm

$$\begin{aligned} (u,\phi)_V &= \int_0^1 \int_0^1 u\phi + \nabla u \cdot \nabla \phi \, dxdy \\ &\quad + \sum_{T \in T_h} \int_T \frac{\partial^2 u}{\partial x^2} \frac{\partial^2 \phi}{\partial x^2} + 2 \frac{\partial^2 u}{\partial x \partial y} \frac{\partial^2 \phi}{\partial x \partial y} + \frac{\partial^2 u}{\partial y^2} \frac{\partial^2 \phi}{\partial y^2} \, dxdy \\ &= (u,v)_H + \sum_{T \in T_h} \int_T \frac{\partial^2 u}{\partial x^2} \frac{\partial^2 \phi}{\partial x^2} + 2 \frac{\partial^2 u}{\partial x \partial y} \frac{\partial^2 \phi}{\partial x \partial y} + \frac{\partial^2 u}{\partial y^2} \frac{\partial^2 \phi}{\partial y^2} \, dxdy, \end{aligned}$$

$$\|u\|_V = (\|u\|_H^2 + \sum_{T \in T_h} \int_T (\frac{\partial^2 u}{\partial x^2})^2 + 2(\frac{\partial^2 u}{\partial x \partial y})^2 + (\frac{\partial^2 u}{\partial y^2})^2 \, dxdy)^{\frac{1}{2}} = (\|u\|_H^2 + |u|_{H_\Gamma^2}^2)^{\frac{1}{2}},$$

where

$$|u|_{H_\Gamma^2}^2 = \sum_{T \in T_h} \int_T (\frac{\partial^2 u}{\partial x^2})^2 + 2(\frac{\partial^2 u}{\partial x \partial y})^2 + (\frac{\partial^2 u}{\partial y^2})^2 \, dxdy.$$

It is easy to see that H and V are Hilbert space, $V \subset\subset H$, and V dense in H. Then it follows from the proof of Theorem 4.3, that the optimal subspace has zero mesh-dependent semi norm, i.e., $|u|_{H_\Gamma^2} = 0$, and the corresponding eigenvalue is $\lambda = 1$ with algebraic multiplicity the same as the dimension of the optimal subspace. By a simple calculation, we see that the piecewise linear finite element space is the optimal subspace

$$
\begin{aligned}
S_h &= \{u \in H^1(\Omega) : |u|_{H^2_\Gamma} = 0\} \\
&= \{u \in H^1(\Omega) \, : \, u_T = \text{a linear combination of} \\
&\quad 1, x, \text{ and } y, \text{ for } T \in T_h, \}.
\end{aligned}
\tag{4.8}
$$

4.3 Generalized L-spline Subspaces

Our sixth example is the generalized L-spline spaces. From the examples in the previous section, we see that the optimal subspaces depend on the choice of the norms. When the regular norms are chosen as in Example 1, the eigenfunctions to the corresponding eigenvalue problem, i.e., the basis functions of the optimal subspaces, are most likely to be defined on the whole domain. Practically, we usually work on a locally defined basis function. Thus we weaken the property of the unknown solution a little to get the optimal subspaces in the space with this mesh-dependent norm, which is mentioned in Example 2-5. In this section, a mesh-dependent norm is carefully selected based on the understanding of the unknown solution of the problems with rough coefficients, so that the special basis functions reflex the property of the unknown solution.

We discuss a special first order operator L,

$$L = a_1 u', \quad L^* L u = -(a_1^2 u')', \tag{4.9}$$

then turn to an operator L of order m. The following theorems show that the generalized L-spline space is an optimal approximation subspace with respect to the appropriate norms.

Theorem 4.5 *Let* $\Gamma = \{a = x_0 < x_1 < \cdots < x_{n-1} < x_n = b\}$ *be a partition of the interval* $[a, b]$*, and* $I_j = [x_{j-1}, x_j]$*. Let* $H = H^1[a, b]$ *and* $V = \{u \in H^1[a, b], L^*Lu|_{I_j} \in L^2(I_j)$*, for* $j = 1, 2 \cdots, n\}$*. Then the generalized L-spline space* $\{u \in H^1[a, b], L^*Lu|_{I_j} = 0$ *for* $j = 1, 2 \cdots, n\}$*, where* L *and* L^*L *are defined in (4.9), and* $a_1(x)$ *is a measurable function satisfying* $0 < \alpha \leq a_1(x) \leq \beta < \infty$*, is the optimal subspace of* $d_{(n-1)2}(\{u \in V : ||u||_V \leq 1\}; H)$*, where*

$$\|u\|_H = \left(\int_a^b u^2 + (u')^2 dx \right)^{\frac{1}{2}},$$

$$|u|_{L,\Gamma} = \left(\sum_{j=1}^{n} \int_{I_j} (L^*Lu)^2 dx \right)^{\frac{1}{2}},$$

and

$$\|u\|_V = (\|u\|_H^2 + |u|_{L,\Gamma}^2)^{\frac{1}{2}}.$$

Proof. We introduce the change of variable or mapping

$$\tilde{x}(x) = \int_a^x \frac{ds}{a_1^2(s)}, \tag{4.10}$$

and the partition of the interval $[0, \int_a^b \frac{ds}{a_1^2(s)}]$, $\tilde{\Gamma} = \{0 = \tilde{x_0} < \tilde{x_1} < \cdots < \tilde{x_{n-1}} < \tilde{x_n} = \int_a^b \frac{ds}{a_1^2(s)}\}$, which is the image of Γ under mapping (4.10). Then we see that

$$\frac{d\tilde{u}}{d\tilde{x}} = a_1^2(x)\frac{du}{dx}.$$

Since $u \in H^1[a, b]$ and $(a_1^2 u')'|_{I_j} \in L^2$, we have $\tilde{u} \in H^1[0, \int_a^b \frac{ds}{a_1^2(s)}]$ and $\tilde{u}''|_{\tilde{I_j}} \in H^2$. Thus, it is easily see that $V \subset\subset H^1[a, b]$. Then it follows from the proof of Theorem

4.3, that the optimal subspace has zero mesh-dependent semi-norm, i.e.,

$$|u|_{L,\Gamma} = \left(\sum_{j=1}^{n} \int_{I_j} (L^* L u)^2 dx \right)^{\frac{1}{2}} = 0,$$

which is the generalized L-spline space

$$S\{\phi \in H^1[a,b] \,:\, L^* L\phi|_{I_j} = 0, \quad \text{for } j = 1, 2, \cdots, n\}.$$

□

The Example 2 treated in Section 4.2 can be seen as a special case with $a = 1$ of above theorem. We now turn to an operator L of order m as defined in (2.1), with rough coefficient functions.

Theorem 4.6 *Let* $\Gamma = \{a = x_0 < x_1 < \cdots < x_{n-1} < x_n = b\}$ *be a partition of the interval* $[a, b]$*, and* $I_j = [x_{j-1}, x_j]$*. Let* $H = H^m[a, b]$ *and* $V = \{u \in H^m[a, b], L^* Lu|_{I_j} \in L^2(I_j)$*, for* $j = 1, 2 \cdots, n\}$*. Then the generalized L-spline space* $\{u \in H^m[a, b], L^* Lu|_{I_j} = 0\, for\ j = 1, 2 \cdots, n\}$*, where the coefficient functions* $a_k(x)$ *are measurable functions satisfying* $0 < \alpha \leq a_k(x) \leq \beta < \infty$ *and* $a_k(x) \in H^k(I_j)$ *for* $k = 0, 1 \cdots, m$*, is the optimal subspace of* $d_{(n-1)2m}(\{u \in V : ||u||_V \leq 1\}; H)$*, where*

$$\|u\|_H = \left(\sum_{j=0}^{m} \int_a^b (\frac{d^j u}{dx^j})^2 dx \right)^{\frac{1}{2}},$$

$$|u|_{L,\Gamma} = \left(\sum_{j=1}^{n} \int_{I_j} (L^* L u)^2 dx \right)^{\frac{1}{2}},$$

and

$$\|u\|_V = (\|u\|_H^2 + |u|_{L,\Gamma}^2)^{\frac{1}{2}}.$$

Proof. On each subinterval I_j, since $a_k(x) \in H^k(I_j)$ for $k = 0, 1 \cdots, m$, and $L^*Lu|_{I_j} \in L^2(I_j)$, we have that $u|_{I_j} \in H^{2m}(I_j) \subset\subset H^m(I_j)$. From this we have $V \subset\subset H$ and V is dense in H. Then it follows the proof of Theorem 4.3, that the optimal subspace has zero mesh-dependent semi-norm, i.e.,

$$\begin{aligned} S_\Gamma &= \{u \in H^m[a,b] : |u|_{L,\Gamma} = 0\} \\ &= \{u \in H^m[a,b] : L^*Lu|_{I_j} = 0, \text{ for } j = 1, 2, \cdots, n\}, \end{aligned}$$

which is the generalized L-spline space with dim $= (n-1)2m$. □

4.4 A Special Class of Two Dimensional Elliptic Problems with Rough Coefficients

Babuška and Osborn proposed a special approximation space for a class of second order, two-dimensional elliptic boundary value problems with rough or highly oscillating coefficients in [12]. This special shape function was discussed in detail by Babuška and Osborn in [10]. The application of this approach to one-dimensional problems can be found in [14]. In this section, n-width will be used to determine that this special approximation space is an optimal subspace.

The problems they considered are of the form

$$\begin{cases} -\frac{\partial}{\partial x}(a(x)\frac{\partial}{\partial x}u(x,y)) - \frac{\partial}{\partial y}(b(y)\frac{\partial}{\partial y}u(x,y)) = f(x,y) \quad \forall (x,y) \in \Omega, \\ u(x,y) = 0, \quad \forall (x,y) \in \partial\Omega, \end{cases} \tag{4.11}$$

where $\Omega = [0,1] \times [0,1]$, f is a function in $L^2(\Omega)$, and where the functions $a, b \in L^\infty(\Omega)$ satisfy

$$0 < \alpha \le a(x), b(y) \le \beta < \infty \quad \forall (x,y) \in \Omega,$$

where α and β are constants. The variational formulation of this problem is

$$\begin{cases} u \in H_0^1(\Omega), \\ B(u,v) = f(v), \quad \forall v(x,y) \in H_0^1(\Omega), \end{cases}$$

where

$$B(u,v) = \int_\Omega \left(a(x) \frac{\partial u}{\partial x} \frac{\partial u}{\partial y} + b(y) \frac{\partial u}{\partial y} \frac{\partial u}{\partial y} \right) dxdy,$$

and

$$f(v) = \int_\Omega fv \, dxdy.$$

A regularity result for the above problem is given in the following theorem; we refer to [10], where it was proved by applying Bernstein's theorem, which can be found in [19] and [36]. We first define the space

$$H^L(\Omega) = \{u \in H^1(\Omega) : a(x)\frac{\partial u}{\partial x}, b(y)\frac{\partial u}{\partial y} \in H^1(\Omega)\},$$

with the norm

$$\|u\|_{L,\Omega}^2 = \|u\|_{1,\Omega}^2 + |u|_{L,\Omega}^2,$$

where

$$|u|_{L,\Omega}^2 = \int_\Omega \left(\frac{a}{b} \left| \frac{\partial}{\partial x}(a \frac{\partial u}{\partial x}) \right|^2 + ab \left| \frac{\partial^2 u}{\partial x \partial y} \right|^2 + \frac{b}{a} \left| \frac{\partial}{\partial y}(b \frac{\partial u}{\partial y}) \right|^2 \right) dxdy.$$

We introduce the change of variables or mapping

$$\tilde{x}(x) = \int_0^x \frac{ds}{a(s)}, \quad \tilde{y}(y) = \int_0^y \frac{dt}{b(t)}. \tag{4.12}$$

Theorem 4.7 *(Babuška, Caloz & Osborn [10]) For each $f \in L^2(\Omega)$, problem(4.11) has a unique solution $u \in H_0^1 \cap H^L(\Omega)$. Furthermore, there is a constant $C = C(\alpha, \beta)$, depending on α and β but independent of f, such that*

$$\|u\|_{L,\Omega} \leq C(\alpha, \beta) \|f\|_{0,\Omega}.$$

For $0 < h \leq 1$, let $\mathcal{C}_h$ be a triangulation of Ω by (closed) curvilinear triangles T of diameter $\leq h$, where by a curvilinear triangle $T \subset \Omega$ we mean the preimage of an ordinary triangle $\tilde{T} \subset \tilde{\Omega} = (0, \int_0^1 \frac{ds}{a(s)}) \times (0, \int_0^1 \frac{dt}{b(t)})$ under the mapping (4.12). Corresponding to $\mathcal{C}_h$ we have a triangulation $\tilde{\mathcal{C}}_h$ of $\tilde{\Omega}$ by usual triangles. We assume that $\{\tilde{\mathcal{C}}_h\}_{0<h\leq 1}\}$ is a quasi-uniform mesh. Then we define the space

$$H_h^L = \{u \in H_0^1(\Omega) : a(x)u_x, b(y)u_y \in H^1(T), \quad \forall T \in \mathcal{C}_h\},$$

with a mesh-dependent norm

$$\|u\|_{L,h}^2 = \|u\|_{1,\Omega}^2 + |u|_{L,h}^2,$$

where

$$|u|_{L,h}^2 = \sum_{T \in \mathcal{C}_h} \int_T \left(\frac{a}{b} |\frac{\partial}{\partial x}(a \frac{\partial u}{\partial x})|^2 + ab|\frac{\partial^2 u}{\partial x \partial y}|^2 + \frac{b}{a}|\frac{\partial}{\partial y}(b\frac{\partial u}{\partial y})|^2 \right) dxdy.$$

The special approximate subspace is defined by

$$S_h = \{v \in H_0^1(\Omega) : v|_T \in \operatorname{span}\{1, \int_0^x \frac{ds}{a(s)}, \int_0^y \frac{dt}{b(t)}\}, \quad \forall T \in \mathcal{C}_h\}. \tag{4.13}$$

As a consequence of the choice for the curvilinear triangles T we see that $S_h \subset H_h^L(\Omega)$, that is, S_h is conforming. This is easily seen by noting that the functions $1, \int_0^x \frac{ds}{a(s)}, \int_0^y \frac{dt}{b(t)}$ are transformed to $1, \tilde{x}, \tilde{y}$ by (4.12). Thus, $\tilde{S}_h \equiv \{\tilde{v} : v \in S_h\}$, the image of S_h under the mapping (4.12), is the usual space of continuous piecewise linear approximation functions with respect to $\tilde{\mathcal{C}}_h$, and S_h is conforming because $\tilde{S}_h$ is. In [10], it is proved that on each curvilinear triangle T, $\|u\|_{1,T}^2 + |u|_{L,h}^2 \cong \|\tilde{u}\|_{H^2(\tilde{T})}^2$. Thus for $u \in H_h^L$, we have that $\tilde{u}|_{\tilde{T}} \in H^2(\tilde{T})$ for each $T \in \mathcal{C}_h$. Since $H^2(\tilde{T}) \subset\subset H^1(\tilde{T})$, it is true that $H_h^L \subset\subset H_0^1(\Omega)$ and H_h^L is dense in $H_0^1(\Omega)$.

Theorem 4.8 *The finite-dimensional space S_h defined in (4.13) is the optimal subspace of $d_n(\{u \in V : ||u||_V \leq 1\}; H)$, where spaces $H = H_0^1(\Omega)$ and $V = H_h^L$.*

Proof. It follows from the proof of Theorem 4.3, that the optimal subspace has zero mesh-dependent semi-norm, i.e.,

$$|u|_{L,h}^2 = \sum_{T\in\mathcal{C}_h} \int_T \left(\frac{a}{b} |\frac{\partial}{\partial x}(a\frac{\partial u}{\partial x})|^2 + ab|\frac{\partial^2 u}{\partial x \partial y}|^2 + \frac{b}{a}|\frac{\partial}{\partial y}(b\frac{\partial u}{\partial y})|^2 \right) dxdy = 0.$$

From above and $0 < \alpha \leq a(x), b(y) \leq \beta$,we have

$$\frac{\partial}{\partial x}(a\frac{\partial u}{\partial x}) = \frac{\partial}{\partial y}(b\frac{\partial u}{\partial y}) = 0,$$

and

$$ab\frac{\partial^2 u}{\partial x \partial y} = 0,$$

which can be interpreted as the following

$$\frac{\partial}{\partial y}(a(x)\frac{\partial u}{\partial x}) = \frac{\partial}{\partial x}(b(y)\frac{\partial u}{\partial y}) = 0.$$

It is easily seen that $c_1 + c_2 \int_0^x \frac{ds}{a(s)} + c_3 \int_0^y \frac{dt}{b(t)}$ represents the family of solutions to above equations. Thus, we obtain that the span of functions $1, \int_0^x \frac{ds}{a(s)}, \int_0^y \frac{dt}{b(t)}$ is the subspace with zero mesh-dependent semi norm property, i.e., S_h is the optimal subspace of $d_n(\{u \in V : ||u||_V \leq 1\}; H)$. □

Chapter 5

Quadrature Problem for Meshless Methods

Meshless methods have been increasingly used by engineers in the past few year. They can be interpreted in the context of general variational methods, in which the quality of the approximation is mainly determined by the approximation properties of the trial space. In [7] Babuška, Banerjee and Osborn provided a survey of this new field, with emphasis on mathematical analysis. For other survey of results on meshless methods we refer to Belytschko, Krongauz, Organ, Fleming and Krysl [18], Duarte [27], Li and Liu [39], Liu [40], and Scheweitzer [50].

One of the major issues concerning meshless methods is the problem of numerical quadrature. In spite of its importance, only a few papers address it, most from an implementational point of view. For instance, Beissel and Belytschko [17], Chen and Wu [22], Chen, Wu and Yoon [23], Dolbow and Belytschko [26], and Carpinteri, Ferro and Ventura [21]. In [8], we consider the approximation of the Neumann problem by meshless methods, show that the approximation is inaccurate if nothing special (beyond accuracy) is assumed about the numerical integration. We then identified a condition - referred to as the zero row sum condition. This, together with accuracy, ensure the quadrature error is small. The row sum condition can be achieved by changing the diagonal elements of the stiffness matrix. Under row sum condition we derive an energy norm error estimate for the numerical solution with quadrature. This chapter gives an introduction of meshless methods and explains why numerical integration affects the accuracy of approximation solution, i.e., quadrature issue. Then the detail of the numerical tests from [8] will be discussed.

5.1 Meshless Methods

We consider elliptic problems with purely Neumann boundary conditions,

$$\begin{aligned} -\triangle u &= f \quad \text{in } \Omega \\ \frac{\partial u}{\partial n} &= g \quad \text{on } \Gamma = \partial\Omega, \end{aligned} \tag{5.1}$$

where Ω is a bounded domain in R^d with Liptschitz boundary, $f \in L^2(\Omega)$, and $g \in L^2(\Gamma)$. The variational formulation of (5.1) is:

Find $u \in H$ satisfying

$$B(u, v) = L(v) \quad \forall v \in H,$$

where

$$B(u, v) \equiv \int_\Omega \nabla u \cdot \nabla v \, dx,$$

and

$$L(v) = \int_\Omega f v \, dx + \int_\Gamma g v \, ds.$$

For the existence of the solution to problem (5.1), we assume

$$\int_\Omega f \, dx + \int_\Gamma g \, ds = 0, \tag{5.2}$$

the solution is unique up to an additive constant. In addition, we assume that Γ, f, and g are such that $u \in H^2(\Omega)$.

We have done numerical tests in both one dimensional space and two dimensional space, i.e., $d = 1, 2$. For simplicity, we introduce the notation in two dimensional space.

We are interested in approximating u by a meshless method, which is based on uniformly distributed particles and translation invariant shape functions. Let $x_j^h = (j_1 h, j_2 h) = jh$, where $j = (j_1, j_2) \in \mathbb{Z}^2$, with $\mathbb{Z}$ the integer lattice, and $0 < h$, be a family of uniformly distributed particles. Suppose $\phi \in H^q(R^2) \cap L^\infty(R^2)$, for some $1 \leq q$; let $\zeta = supp\,\phi$, and suppose that

$$\zeta \in B_\rho \equiv \{x \in R^2 : \|x\|^2 = x_1^2 + x_2^2 < \rho^2\}.$$

ϕ is not assumed to be a piecewise polynomial. We also suppose that 0 is contained in the interior of ζ. Then we let

$$\phi_j^h(x) = \phi(x_1, x_2) \equiv \phi(\frac{x - jh}{h}) = \phi(\frac{x_1 - j_1 h}{h}, \frac{x_2 - j_2 h}{h}),$$

for $j \in \mathbb{Z}^2$ and $0 < h$. It is clear that

$$\zeta_j^h \equiv supp\,\phi_j^h = \{x : \frac{x - jh}{h} \in \zeta\} \subset B_{\rho h}^j = \{x : \|x - x_j^h\|_2 < \rho h\},$$

and $x_j^h \in \mathring{\zeta}_j^h$, where $\mathring{\zeta}_j^h$ is inside of the interior of ζ_j^h. ϕ_j^h are the associated particle shape functions. Particles and particle shape functions as defined above are translation invariant:

$$x_{j+l}^h = x_j^h + x_l^h \quad \text{and} \quad \phi_{j+1}^h = \phi_j^h(x - x_l^h).$$

We refer to $\phi(x)$ as the basic shape function. We assume that $\{\phi_j^h(x)\}_{j \in \mathbb{Z}^2}$ is reproducing of order 1, i.e.,

$$\sum_{j \in \mathbb{Z}^2} (j_1 h)^{i_1} (j_2 h)^{i_2} \phi_j^h(x_1, x_2) = x_1^{i_1} x_2^{i_2}, \quad \forall x_1,\, x_2, \tag{5.3}$$

for $0 \leq i_1, i_2$ with $i_1 + i_2 = 1$. This is the requirement that we used later in Section 5.2 to construct shape functions. Let $\omega_j^h = \mathring{\zeta}_j^h \cap \Omega$ be the interior support of ϕ_j intersect Ω. The meshless subspace is

$$V_h = span\{\phi_j^h|_\Omega : j \in N_h\},$$

where N_h is the set of indices of particles corresponding to shape functions whose supports intersect Ω, i.e., $N_h = \{j : \omega_j^h \neq \emptyset\}$; let $|N_h|$ = cardinality of N_h. We assume that $\{\phi_j^h|_\Omega : j \in N_h\}$ is linearly independent and thus a basis for V_h.

The approximation property of the meshless subspace defined above was described by Strang and Fix in [51] and Babuška in [5]. An alternative proof for sufficient part of the theorem with uniformly distributed particles in [7] was given as follows

Theorem 5.1 *Suppose $\phi \in H^q(R^n)$, with smoothness index $q \geq 0$, has compact support $\eta \in B_\rho$, and suppose $k = 0, 1, 2, \cdots$ and for $|\alpha| \leq k$,*

$$\sum_{j \in \mathbb{Z}^n} j^\alpha \phi(x - j) = \lambda x^\alpha + q^\alpha(x);$$

here $\lambda \neq 0$, and q^α is a polynomial of degree $< |\alpha|$, i.e. suppose ϕ is quasi-reproducing of order k. If $u \in H^{k'+1}(R^n)$, where $0 \leq k' \leq k$, then

$$\|u - \sum_{l \in \mathbb{Z}^n} w_l^h \phi_l^h\|_{H^s(R^n)} \leq C h^{k'+1-s} \|u\|_{H^{k'+1}(R^n)}, \quad for\ 0 \leq s \leq min\{q, k'+1\}.$$

When $k = 1$ and $q = 0$, quasi-reproducing is the same as a reproducing of order 1 (see 5.3).

The meshless approximation u_h is defined by

$$\begin{aligned} u_h &\in V_h \\ B(u_h, v) &= L(v) \quad \forall v \in V_h. \end{aligned} \tag{5.4}$$

If we express $u_h = \sum_{j\in N_h} u_j^h \phi_j^h = \mathbf{u}_h \cdot \bar{\phi}_h$, then the algebraic problem corresponding to the above problem is:

$$A\mathbf{u}_h = \mathbf{l}, \tag{5.5}$$

where $A = (a_{ij})_{i,j\in N_h}$, with

$$a_{ij} = B(\phi_i^h, \phi_j^h) = \int_\Omega \nabla\phi_i^h \cdot \nabla\phi_j^h \, dx = \int_{\omega_i^h \cap \omega_j^h} \nabla\phi_i^h \cdot \nabla\phi_j^h \, dx,$$

and $\mathbf{l} = (l_i)_{i\in N_h}$, with

$$l_i = L(\phi_i^h) = \int_\Omega f\phi_i^h \, dx + \int_\Gamma g\phi_i^h \, ds = \int_{\omega_i^h} f\phi_i^h \, dx + \int_{\Gamma\cup\varpi_i^h} g\phi_i^h \, ds = f_i^h + g_i^h.$$

Since $\{\phi_j^h\}_{j\in N_h}$ reproduce constants, i.e., $\sum_{j\in N_h} \phi_j^h = 1$, we have $\sum_{j\in N_h} a_{ij} = 0$. It is also true that $\sum_{i\in N_h} l_i = 0$, because of the assumption (5.2), which guarantees the existence of the solution to problem (5.1). This explains that the constant vector $(1, 1, \cdots, 1)$ is in the Null space of the stiffness matrix A, and the fact that solutions of problem (5.1) differ by a constant, i.e., solutions to problem (5.5) can be represented by $\mathbf{u}_h + c(1, 1, \cdots, 1)$, where $\mathbf{u}_h$ is any vector satisfying (5.5). There are many choices of $\mathbf{u}_h$; see [38] for a discussion. In order to specify a unique solution to (5.5), throughout this section we made a convenient choice of $\mathbf{u}_h$, by setting the last element of $\mathbf{u}_h$ equal to 0, i.e., $u_{|N_h|}^h = 0$.

Since the integrals $a_{ij} = B(\phi_i^h, \phi_j^h)$ and $(l_i)_{i\in N_h} = L(\phi_i^h)$ are evaluated numerically by quadrature schemes, this will affect the accuracy of the numerical solution

u_h. Our paper [8] investigates this quadrature issue and propose a row sum correction to improve the numerical solution.

Let

$$a_{ij}^* = ⨍_\Omega \nabla\phi_i^h \cdot \nabla\phi_j^h \, dx = ⨍_{\omega_i^h \cap \omega_j^h} \nabla\phi_i^h \cdot \nabla\phi_j^h \, dx,$$

and

$$l_i^* = ⨍_\Omega f\phi_i^h \, dx + ⨍_\Gamma g\phi_i^h \, ds = ⨍_{\omega_i^h} f\phi_i^h \, dx + ⨍_{\Gamma\cup\overline{\omega}_i^h} g\phi_i^h \, ds = f_i^{h*} + g_i^{h*},$$

where $⨍$ represents quadrature version of $\int$. Instead of (5.5), the linear system we really solved is

$$A^*\mathbf{u}_h^* = \mathbf{l}^*. \tag{5.6}$$

There are several possibilities: (i) system (5.6) is singular with the same structure as system (5.5), i.e., $\sum_{j\in N_h} a_{ij}^* = 0$ and $\sum_{i\in N_h} l_i^* = 0$; (ii) A^* is singular but with a different structure; and (iii) A^* is non-singular. Thus (5.6) may have infinitely many solutions, may have no solutions, or may have a unique solution. In Section 5.3 we will show the erratic behavior of the relative error $u - u_h^*$ through an example. Note that the row sum of A is zero, but the row sum of A^* is unlikely equal to zero. In our paper [8], we consider a corrected stiffness matrix A^{**}, which is defined by letting

$$a_{ij}^{**} = a_{ij}^* \quad \text{for } i \neq j,$$

and

$$a_{ii}^{**} = -\sum_{j\in N_h, j\neq i} a_{ij}^*,$$

so that it satisfies the row sum zero condition, i.e.,

$$\sum_{j \in N_h} a^*_{ij} = 0 \quad \forall i.$$

We also considered corrections to the right-hand side vectors f^{h*} and g^{h*}, but we do not consider them here. We assume that $f^{h*} = f$ and $g = 0$. Thus, the linear system we solved is

$$A^{**}\mathbf{u}^{**}_h = \mathbf{l}. \tag{5.7}$$

In this section, we focus on the computational analysis. Here, we only cite one error estimate Theorem from our paper [8]. We refer to our paper for all Axioms and Theorems and their proofs. Let

$$a^*_{ij} = a_{ij} + \eta_{ij}, \quad f^{h*}_i = f^h_i + \epsilon^h_i \quad \text{and } g^{h*}_i = g^h_i + \tau^h_i,$$

with

$$|\eta_{ij}| \leq \eta max(|a_{ij}|, \nu h^d), \quad |\epsilon^h_i| \leq max(|f^h_i|, \nu h^d \|f\|_{L^\infty(\Omega)}),$$

and

$$|\tau^h_i| \leq max(|g^h_i|, \nu h^{d-1}\|g\|_{L^\infty(\Gamma)}).$$

Theorem 5.2 *Suppose our shape functions and quadrature satisfy the Axioms in [8]. Then for small η, there is a constant C, independent of u, ϵ, τ, η, and h, such that*

$$\|u - u^{**}_h\|_E \leq C\left[h\|u\|_{2,\Omega} + \eta\|u\|_E + (\epsilon+\tau)\|f\|_{L^\infty(\Omega)} + (\epsilon+\tau)\|g\|_{L^\infty(\Gamma)}\right], \quad \forall h.$$

Note that the linear system we solved in (5.7) has $\epsilon = 0$ and $\tau = 0$. According to Theorem 5.2, we have the error estimate as following

$$\|u - u^{**}_h\|_E \leq C\left[h\|u\|_{2,\Omega} + \eta\|u\|_E\right], \quad \forall h.$$

5.2 Shape Function Construction

Let $\omega(x) \geq 0$ be a continuous function with compact support. The function $\omega(x)$ is called a weight function or window function. In one dimension, we chose the weight function as following:

$$\omega(x) = \begin{cases} e^{\frac{1}{(x+R)(x-R)}} & |x| \leq R \\ 0, & |x| \geq R, \end{cases} \tag{5.8}$$

where $R = 1.1 > 1$. In $\mathbb{R}^n$, $\omega(x)$ can be constructed from a one dimensional weight function $\omega(x)$ as $\omega(x) = \omega(\|x\|)$, where $\|x\|$ is the Euclidean length of x, or can be constructed by $\omega(x) = \prod_{j=1}^{n} \omega(x_j)$, where $x = (x_1, x_2, \cdots, x_n) \in \mathbb{R}^n$.

In general, the reproducing kernel particle (RKP) shape function $\phi_j(x)$, associated with the particle x_j, is defined by

$$\phi_j(x) = \omega_j(x) \sum_{|\alpha| \leq k} (x - x_j)^\alpha b_\alpha(x),$$

where $b_\alpha(x)$ are chosen so that

$$\sum_{j \in \mathbb{Z}} p(x_j)\phi_j(x) = p(x), \quad \text{for } x \in \mathbb{R}^n \text{ and } p \in \mathcal{P}^k(\mathbb{R}^n),$$

so that $\{\phi_j(x)\}_{j \in \mathbb{Z}}$ are reproducing of order k.

We consider the uniformly distributed particle $x_j^h = jh$, $j \in \mathbb{Z}$ as in the previous section. For each x_j^h, we define

$$\omega_j^h(x) = \omega(\frac{x - x_j^h}{h}),$$

where $\omega(x) \geq 0$ is defined in (5.8), a continuous weight function with compact support $\eta = \bar{B}_R(0)$, with $R = 1.1$. Clearly, $\eta_j^h \equiv \operatorname{supp} \omega_j^h(x) = \bar{B}_{Rh}(x_j^h)$. For simplicity, we assume $h = 1$ in construction of the basic shape function $\phi(x)$. The RKP shape function $\{\phi_j(x)\}_{j\in\mathbb{Z}}$ are also required to satisfy the linear reproducing property as in (5.3). Thus, let the basic shape function be

$$\phi(x) = \omega(x) * \sum_{\alpha\leq 1} x^\alpha b_\alpha(x) = \omega(x)(b_0(x) + b_1(x)x), \tag{5.9}$$

and

$$\phi_j(x) = \phi(x - j),$$

where $b_0(x)$ and $b_1(x)$ are chosen so that

$$\sum_{j\in\mathbb{Z}} \phi_j(x) = 1, \quad \text{and} \quad \sum_{j\in\mathbb{Z}} \phi_j(x)j = x,$$

which is the same as the following:

$$\sum_{\alpha\leq 1} m_{\alpha+\beta}(x) b_\alpha(x) = \delta_{\beta,0},$$

where $\delta_{\beta,0}$ is the Kronecker delta, and

$$m_r(x) = \sum_{j\in\mathbb{Z}} \omega_j(x)(x-j)^r = \sum_{j=-2}^{2} \omega_j(x)(x-j)^r.$$

i.e.,

$$\begin{pmatrix} m_0(x) & m_1(x) \\ m_1(x) & m_2(x) \end{pmatrix} \begin{pmatrix} b_0(x) \\ b_1(x) \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}. \tag{5.10}$$

Then the solutions $b_0(x)$ and $b_1(x)$ are

$$\begin{aligned} \begin{pmatrix} b_0(x) \\ b_1(x) \end{pmatrix} &= \tfrac{1}{m_0(x)m_2(x)-m_1^2(x)} \begin{pmatrix} m_2(x) & -m_1(x) \\ -m_1(x) & m_0(x) \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \\ &= \tfrac{1}{m_0(x)m_2(x)-m_1^2(x)} \begin{pmatrix} m_2(x) \\ -m_1(x) \end{pmatrix} \end{aligned}.$$

Note that $m_0(x)$ and $m_2(x)$ are even functions and $m_1(x)$ is an odd function. Then, from the above representations of $b_0(x)$ and $b_1(x)$, we have that $b_0(x)$ is an even function and $b_1(x)$ is an odd function. These result in the symmetry of the basic shape function $\phi(x)$, i.e.,

$$\begin{aligned}\phi(-x) &= \omega(-x)(b_0(-x) + b_1(-x)(-x)) \\ &= \omega(x)(b_0(x) + b_1(x)x) \\ &= \phi(x).\end{aligned}$$

We also construct the first order derivative of the shape function

$$\begin{aligned}\phi'(x) = &\quad \omega'(x)(b_0(x) + b_1(x)x) \\ &+ \omega(x)(b_0'(x) + b_1'(x)x + b_1(x)).\end{aligned}$$

By differentiating both hand sides of the linear system (5.10), we have $b_0'(x)$ and $b_1'(x)$ satisfying

$$(m_0 b_0)' + (m_1 b_1)' = 0$$

$$(m_1 b_0)' + (m_2 b_1)' = 0$$

which are

$$\begin{pmatrix} m_0 & m_1 \\ m_1 & m_2 \end{pmatrix} \begin{pmatrix} b_0' \\ b_1' \end{pmatrix} = - \begin{pmatrix} m_0' & m_1' \\ m_1' & m_2' \end{pmatrix} \begin{pmatrix} b_0 \\ b_1 \end{pmatrix}. \tag{5.11}$$

Solving the above linear system for b_0' and b_1', we have

$$\begin{aligned}\begin{pmatrix} b_0' \\ b_1' \end{pmatrix} &= \tfrac{-1}{m_0(x)m_2(x) - m_1^2(x)} \begin{pmatrix} m_2 & -m_1 \\ -m_1 & m_0 \end{pmatrix} \begin{pmatrix} m_0' & m_1' \\ m_1' & m_2' \end{pmatrix} \begin{pmatrix} b_0 \\ b_1 \end{pmatrix} \\ &= \tfrac{1}{m_0(x)m_2(x) - m_1^2(x)} \begin{pmatrix} (m_1' m_1 - m_0' m_2) b_0 + (m_2' m_1 - m_1' m_2) b_1 \\ (m_0' m_1 - m_1' m_0) b_0 + (m_1' m_1 - m_2' m_0) b_1 \end{pmatrix}.\end{aligned}$$

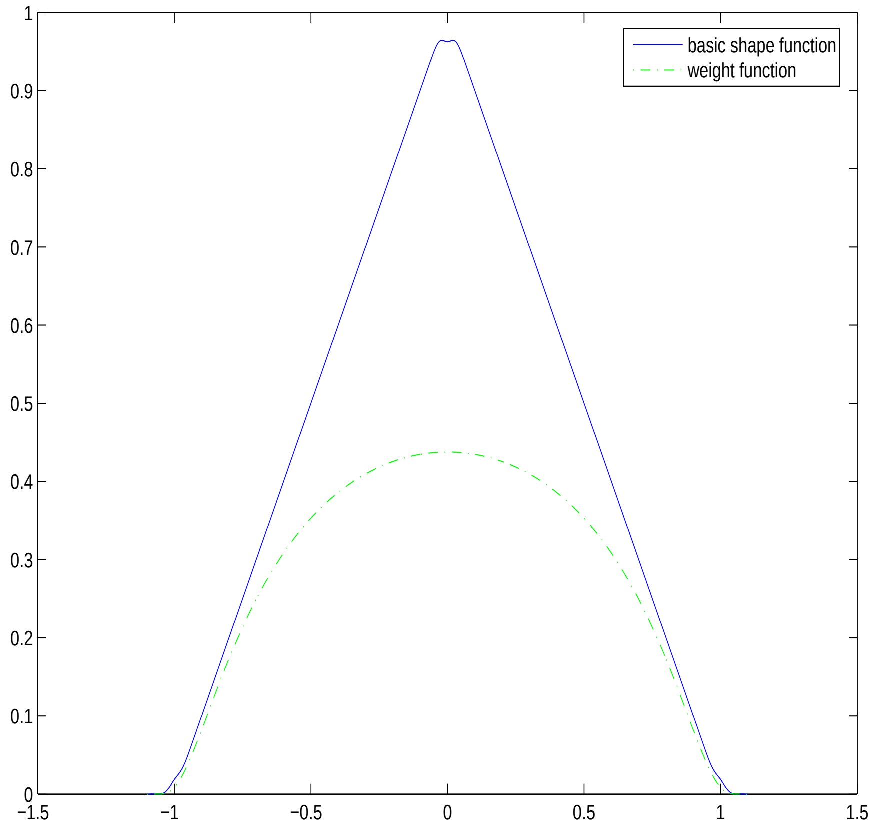

Figure 5.1: The basic shape function $\phi(x)$ and the weight function $\omega(x)$.

The basic shape function $\phi(x)$ and the weight function $\omega(x)$ are both shown on the Figure 5.1. Note that both the weight function $\omega(x)$ and the shape function $\phi(x)$ are $C^{\infty}(\mathbb{R})$ function and $\phi(x)$ is similar to the regular finite element shape function, the hat function, but with a smooth top and two smooth feet.

5.3 Numerical Test

In one dimension, seek $u(x)$ such that

$$\begin{cases} -u''(x) &= cos(x) \quad 0 \le x \le \pi \\ u'(0) &= u'(\pi) = 0 \end{cases}$$

The solution, $u(x) = cos(x)$, exists and is unique up to an additive constant.

The variational formulation for this test problem is

$$\begin{cases} u \in H^1(0,\pi) \\ B(u,v) = F(v), \quad \forall v \in H^1(0,\pi), \end{cases}$$

where $B(u,v) = \int_0^\pi u'v'\,dx$ and $F(v) = \int_0^\pi cos(x)v\,dx$. We use the basic shape function $\phi(x)$ as defined in (5.9), and a family of uniformly distributed particles $x_j^h = jh$, $j \in \mathbb{Z}$, to construct the RKP shape function

$$\phi_j^h = \phi(\frac{x}{h} - j),$$

with $h = \frac{\pi}{n}$. Clearly $\zeta_j^h = \text{supp}\phi_j^h = [x_j^h - Rh, x_j^h + Rh]$. As in Section 5.2, we consider the shape function $\{\phi_j^h(x)\}$, whose supports intersect $\Omega = (0,\pi)$, and their restriction to Ω. Let

$$\omega_j^h \equiv (x_j^h - Rh, x_j^h + Rh) \cup \Omega,$$

then $N_h = \{-1, 0, \cdots, n, n+1\}$ and $|N_h| = n+3$.

The meshless subspace is

$$V_h \equiv \text{span}\{\phi_j^h(x)\}_{j=-1}^{n+1}.$$

The solution u_h of (5.4) exists and is unique up to an additive constant. The corresponding stiffness matrix of the algebraic problem (5.5) is a $(n+3) \times (n+3)$

matrix, which has the following structure

$$\frac{1}{h}\begin{pmatrix} \cdot & \cdot & \cdot & & & & \\ \cdot & \cdot & \cdot & \cdot & & & \\ \cdot & \cdot & \cdot & \cdot & \cdot & & \\ & \alpha & \beta & \gamma & \beta & \alpha & \\ & & \cdot & \cdot & \cdot & \cdot & \cdot \\ & & & \cdot & \cdot & \cdot & \cdot \\ & & & & \cdot & \cdot & \cdot \end{pmatrix},$$

where

$$\alpha = \int_{2-R}^{R} \phi'(x)\phi'(x-2)\,dx,$$
$$\beta = \int_{1-R}^{R} \phi'(x)\phi'(x-1)\,dx,$$
$$\gamma = \int_{R}^{R} (\phi'(x))^2\,dx.$$

We then use quadrature to evaluate the integrals α, β, and γ, but for simplicity the right-hand side is evaluated exactly. The problem was tested by using three different quadrature rules to evaluate the elements in the stiffness matrix. The first one is Trapezoidal rule with m-panels

$$TP = \frac{h}{2}(f(x_0) + 2f(x_1) + \cdots + 2f(x_{m-1}) + f(x_m)),$$

where $h = \frac{b-a}{m}$ and $x_i = a + ih$. The second one is Matlab build-in function QUAD, numerically evaluating integrals by adaptive Simpson quadrature, which uses an absolute error tolerance of TOL instead of the default value 1.e-6. Larger values of TOL result in fewer function evaluations and faster computation, but less accurate results. The third one is Gaussian integration rule with p-Gaussian points. The nodes and weights of Gaussian quadrature rules can be computed by Golub-Welsch algorithm. Then, the solution u_h^* of linear system (5.6) is found by setting the last unknown to zero, eliminating the last equation and solving the resulting $n+2$ equations in $n+2$ unknowns.

In Figure 5.2 we present plots of the relative error $\frac{\|u-u_h^*\|_E}{\|u\|_E}$ with respect to h for three quadrature methods as mentioned above, i.e., the m-panel trapezoid rule (TR); the p-point Gauss rule; and MATLAB's quad (adaptive Simpson quadrature), with tolerance tol. Note that different scales for the relative error are used. As we can see that the error is erratic and that practically no reasonable accuracy was achieved. The error decreases with decreasing h, but then increases as $h \to 0$. An explanation for this behavior was given by a careful examination of an associated periodic problem in [8].

We then consider the associated approximation u_h^{**} of the linear system (5.7) with the corrected stiffness matrix as described in Section 5.1, where the stiffness matrix A^* is computed with the same quadrature methods as given above. In Figure 5.3 we present log-log plots of the relative errors $\frac{\|u_h-u_h^{**}\|_E}{\|u\|_E}$ and $\frac{\|u-u_h^{**}\|_E}{\|u\|_E}$ with respect to h.

We observe that the relative error $\frac{\|u_h-u_h^{**}\|_E}{\|u\|_E}$ becomes nearly constant as $h \to 0$; this constant reflects the accuracy of the quadrature, η. On the other hand, the relative error $\frac{\|u-u_h^{**}\|_E}{\|u\|_E}$ first decreases with decreasing h and then levels off, becoming nearly constant as $h \to 0$. These Figures illustrate the error estimate in Theorem 5.2, which indicates that the error has two components: one due to the meshless methods approximation and the other due to quadrature. If we want the relative error to converge, we have to set η, ϵ, and τ equal to $o(1)$, and if we want the relative error to be $O(h)$, we have to set η, ϵ, and τ equal to $O(h)$.

Now we consider a two dimensional problem

$$\begin{cases} -\Delta u(x) = 2cos(x)cos(y) \quad (x,y) \in \Omega = [0,\pi] \times [0,\pi] \\ \frac{\partial u}{\partial n} = 0 \quad (x,y) \in \partial\Omega. \end{cases}$$

The solution, $u(x,y) = cos(x)cos(y)$, exists and is unique up to an additive constant. The variational formulation of this problem is

$$\begin{cases} u \in H^1((0,\pi) \times (0,\pi)) \\ B(u,v) = F(v), \quad \forall v \in H^1(0,\pi). \end{cases}$$

where $B(u,v) = \int_0^\pi \int_0^\pi \bigtriangledown u \cdot \bigtriangledown v \, dxdy$ and $F(v) = \int_0^\pi \int_0^\pi cos(x)cos(y)v \, dxdy$.

The meshless basis functions we used are the tensor products of ϕ_j^h, as in one-dimensional problem, i.e.,

$$\{\phi_{i,j}^h(x,y)\}_{i,j=-1}^{n+1} \quad \phi_{i,j}^h(x,y) = \phi_i^h(x)\phi_j^h(y),$$

where $h = \frac{\pi}{n}$.

The stiffness matrix has almost the same structure as one dimensional case. The only difference is that it is a block matrix instead of an entry, and each block has the structure as in one dimensional case.

In Figure 5.4 we present log-log plots of the relative errors $\frac{\|u-u_h^*\|_E}{\|u\|_E}$ for the p-point Gauss rule without correction and $\frac{\|u-u_h^{**}\|_E}{\|u\|_E}$ and $\frac{\|u_h-u_h^{**}\|_E}{\|u\|_E}$ for p-point Gauss rule with correction. The right-hand side is computed exactly, i.e., $\epsilon = \tau = 0$.

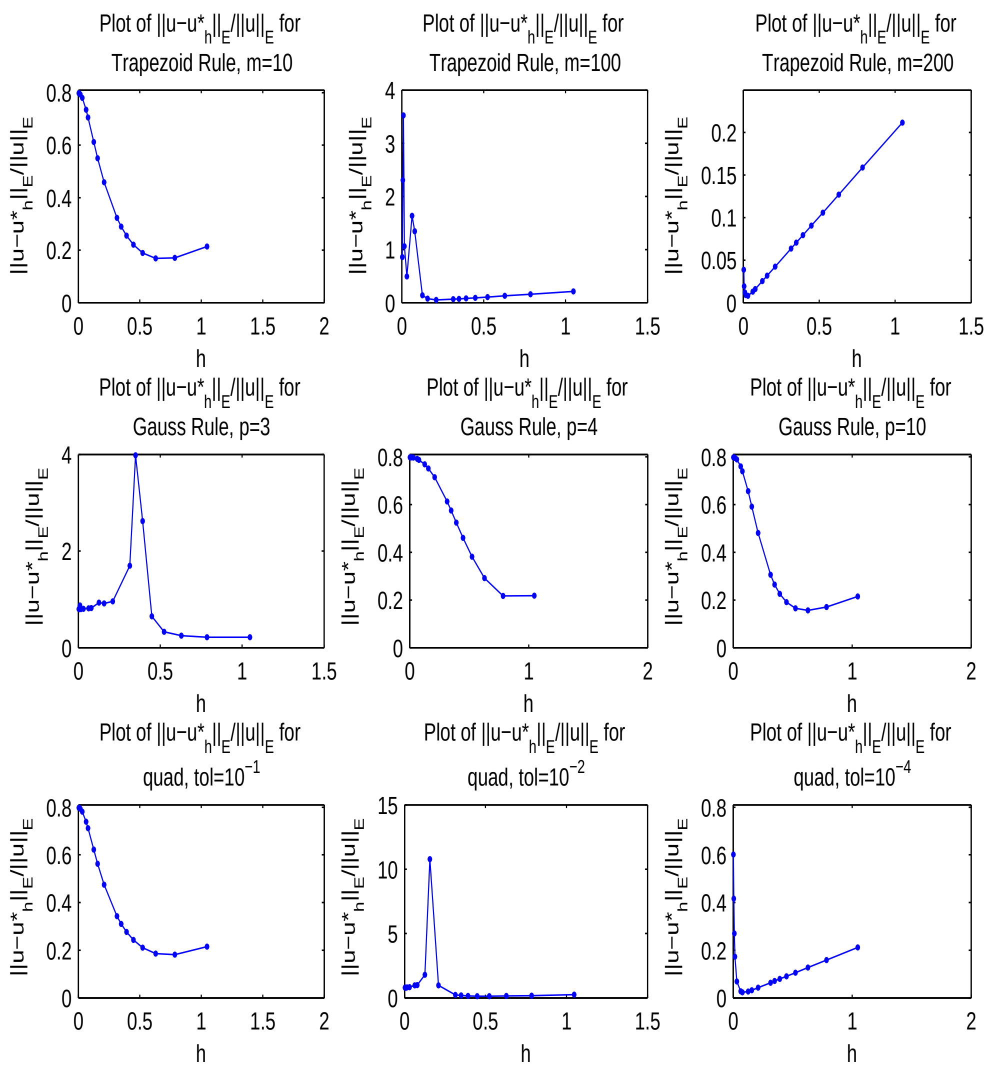

Figure 5.2: The plot of $\|u - u_h^*\|_E / \|u\|_E$ with respect to h for various quadrature schemes for one-dimensional problem.

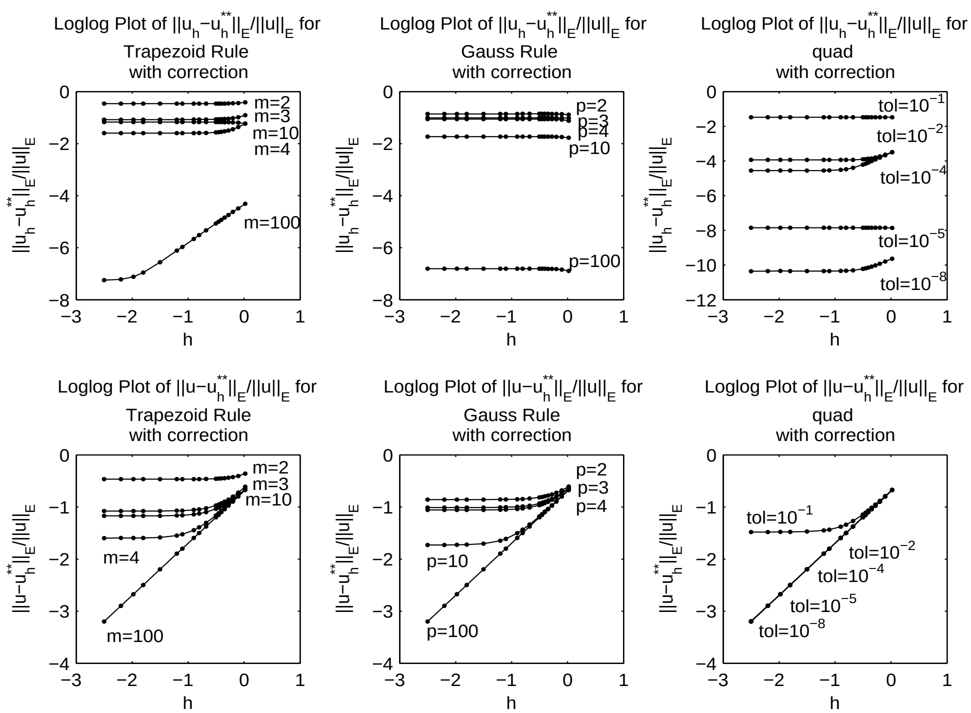

Figure 5.3: The log-log plot of $\|u_h - u_h^{**}\|_E/\|u\|_E$ and $\|u - u_h^{**}\|_E/\|u\|_E$ with respect to h with correction for various quadrature schemes for the one-dimensional problem.

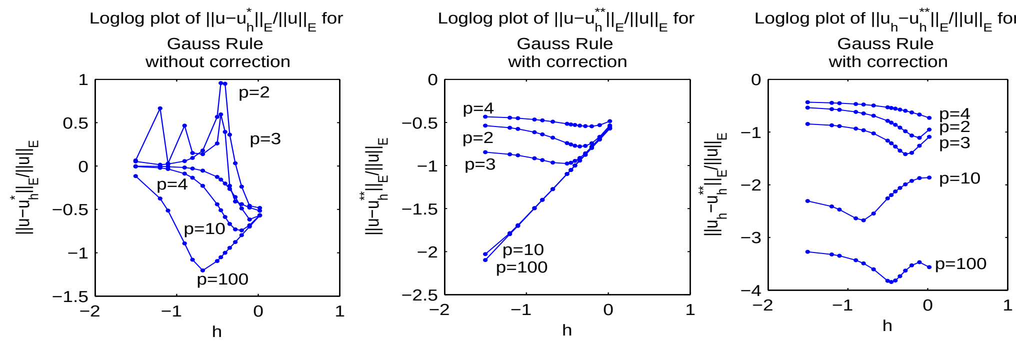

Figure 5.4: The log-log plot of $\|u - u_h^*\|_E/\|u\|_E$ for p-point Gauss rule without correction and $\|u - u_h^{**}\|_E/\|u\|_E$ and $\|u_h - u_h^{**}\|_E/\|u\|_E$ for p-point Gauss rule with correction for two-dimensional problem.

The error behavior is similar to that of the one-dimensional problem. This indicates the dimensional independence of the error estimate, as suggested by Theorem 5.2.

Bibliography

[1] J.H. Ahlberg, E.N. Nilson, "Convergence properties of generalized splines", Proc. Nat. Acad. Sci. U.S.A., Vol. 54, pp. 344-350 (1965).

[2] J.H. Ahlberg, E.N. Nilson and J.L. Walsh, *The theory of splines and their applications*, (Academic Press, New York, 1967).

[3] N.I. Akhiezer and I.M. Glazman, *Theory of Linear Operator in Hilbert Space* , (Scottish Academic Press, 1981).

[4] A.K. Aziz *The Mathematical Foundations of the Finite Element Method with Applications to Partial Differential Equations*, Acadamic Press, New York, 1972.

[5] I. Babuška, "Approximation by hill functions", Comment Math. Univ. Carolinae, 11, 1970, pp. 787-811.

[6] I. Babuška, "Error bounds for finite element methods", Numer. Math., **16**, 322-333 (1971).

[7] I. Babuška, U. Banerjee, J.E. Osborn, *Survey of Meshless and Generalized Finite Element Methods: a Unified Approach*, Acta Numerica, Cambridge University Press, 2003, pp. 1-125.

[8] I. Babuška, U. Banerjee, J.E. Osborn, and Q.H. Li, "Quadrature for Meshless Methods", International journal for numerical methods in engineering, Vol. 76, 9, 2008, pp. 1434-1470.

[9] I. Babuška, U. Banerjee, J.E. Osborn, and Q. Zhang, "Effect of Numerical Integration on Meshless Methods", Computer Methods in Applied Mechanics and Engineering, Vol 198, Issues 37-40, 2009, pp. 2886-2897.

[10] I. Babuška, G. Caloz and J.E. Osborn, "Special finite element methods for a class of second order elliptic problems with rough coefficients", SIAM J. Numer. ANAL., Vol. 31, No. 4, August (1994), pp. 945-981.

[11] I. Babuška, J.E. Osborn, "Eigenvalue Problems", in: Ciarlet, P.G., Lions, J.L. (eds.) Handbook of numerical analysis, vol. II. Finite Element Methods. Amsterdam: North Holland.

[12] I. Babuška, J.E. Osborn, “Finite Element Methods for The Solution of Problems with Rough Input Data”, Lecture Notes Mathematics, Springer–Verlag, 1985, pp. 1121-1139.

[13] I. Babuška and J.E. Osborn, “Can a finite element method perform arbitrarily badly?” Mathematics of Computation, Vol. 69, Num. 230 (1999), pp. 443-462.

[14] I. Babuška and J.E. Osborn, “Generalized finite elment methods: their perfermence and their relation to mixed methods”, SIAM J. Numer. ANAL., Vol. 20, No. 3, June (1983), pp. 322-333.

[15] U. Banerjee, J.E. Osborn,*Estimation of the Effect of Numerical Integration in Finite Element Eigenvalue Approximation*, Numer. Math. 56, 1990, 735-762.

[16] U. Banerjee and J.E. Osborn, Private Communication.

[17] S. Beissel and T. Belytschko, “Nodal Integration of the Element-free Galerkin Method”, Computer Methods in Applied Mechanics and Engineering, 139, 1996, pp. 49-74.

[18] T. Belytschko, Y. Krongauz, D. Organ, M. Fleming and P. Krysl, “Meshless Methods: An Overview and Recent Developments”, Comput. Methods Appl. Mech. Engr, 139, 1996, pp. 3-47.

[19] S.N. Berstein, “Sur la généralisation du problème de Dirichlet”, Math. Ann., 62 (1906), pp. 253-271, 69 (1910), pp. 82-136.

[20] J.H. Bramble and S.R. Hilbert, “Estimation of linear functionals on Sobolev spaces with application to Fourier transforms and spline interpolation”, SIAM J. Numer. Anal, **7**, 112-124 (1970).

[21] A. Carpinteri, G. Ferro and G. Ventura, “The Partition of Unity quadrature in meshless methods.” International Journal for Numerical Methods in Engineering, 54, 2002, pp. 987-1006.

[22] J-S. Chen and C-T Wu and S. Yoon, “A Stabilized Conformal Nodal Integration for a Galerkin Mesh-free Method.” International Journal for Numerical Methods in Engineering, 50, 2001, pp. 435-466.

[23] J-S Chen, S. Yoon and C-T Wu, “Nonlinear Version of Stabilized Conforming Nodal Integration Galerkin Mesh-free Methods.” International Journal for Numerical Methods in Engineering, 53, 2002, pp. 2587-6515.

[24] P.G. Ciarlet, *the Finite Element Method for Elliptic Problems*, North-Holland.

[25] P.J. Davis, *Interpolation and approximation*, (Blaisdell, New York, 1963).

[26] J. Dolbow and T. Belytschko, "Numerical Integration of the Galerkin Weak Form in Meshfree Methods." Computational Mechanics, 23, 1999, pp. 219-230.

[27] C.A. Duarte, "A Review of Some Meshless Methods to Solve Partial Differential Equations", Technical Report 95-06, TICAM, University of Texas at Austin.

[28] N. Dunford, J.T. Schwartz, *Linear operators, II: Spectral Theory, Selfadjoint Operators in Hilbert spaces*, (Wiley-Interscience, New York, 1963).

[29] L.C. Evans, *Partial Differential Equations*, (Graduate Studies in Mathematics, AMS, V. 19, 1998)

[30] V. Girault and P.A. Raviart, Finite element approximation of the Navier-stokes equations, Springer-Verlag, Berlin Heidelberg New York, 1979.

[31] P. Hartman, *Ordinary differential equations*, (Boston; Birkhäuser, 1982).

[32] R.J. Herbold, M.H. Schultz and R.S. Varga, "The effect of quadrature errors in the numerical solution of boundary value problems by variational techniques", Aequations Math. 3 (1969), pp. 247-270.

[33] J. Jerome and J. Pierce, "On spline functions determinded by singular self-adjoint differential operators", MRC Technical Summary Report, July 1971.

[34] J.W. Jerome and L.L. Schumaker, "On Lg-spline", J. Approx. Theory, Vol. 2, pp. 29-49 (1969).

[35] A. Kolmogorov, "über die beste Annäherung von Funktionen einer gegebenen Funktionenklasse", Ann. of Math. (2) 37 (1936), no. 1, pp. 107-110.

[36] O. Ladyzhenskya and N. Ural'tseva, *Linear and Quasilinear Elliptic Equations*, Acadamic Press, New York, 1968.

[37] P. Lancaster and K. Salkauskas, Surfaces Generated by Moving Least Squares Methods, *Mathematics of Computation*, 37, 1981, PP.141-158.

[38] R. Lehoucq and P. Bochev, "On the Finite Element Discitization of the Pure Neumann Problem", SIAM Review 47(1), 2005, pp.50-66.

[39] S. Li and W.K. Liu, "Meshfree and Particle Methods and Their Application", Applied Mechanics Review 55, 2002, pp. 1-34.

[40] G.R. Liu, *Meshless Methods*, CRC Press, Boca Raton, 2002.

[41] T. Liszka, An Interpolation Method for An Irregular Net of Nodes, *Numer. Meth.*, 20, 1984, PP. 1599-1612.

[42] T.R. Lucas, "A generalization of L-spline", Numer. Math., Vol. 15, pp. 359-370 (1970).

[43] T.R. Lucas, "A theory of generalized splines with applications to nonlinear boundary value problems", Thesis, Georgia Institute of Technology, (1970).

[44] J.E. Osborn, "Spectral Approximation for Compact Operators", Mathematics of Computation, Vol. 29, Num. 131, July 1975, pp. 712-725.

[45] A. Pinkus, *N-widths in approximation theory*, (Berlin; New York: Springer-Verlag, 1985).

[46] M. Renardy, R.C. Rogers,*An Introduction to Partial Differential Equations*, Springer-Verlag.

[47] I.J. Schoenberg, "Contributions to the problem of approximation of equidistant data by analytic functions", Parts A and B, Quart. Appl. Math., 4 (1946), pp. 45-99, 112-141.

[48] M.H. Schultz, "Elliptic spline functions and the Rayleigh-Ritz-Galerkin method", Math. Comp., Vol. 24, pp. 65-80 (1970).

[49] M.H. Schultz and R.S. Varga, "L-Splines", Numer. Math., **10**, 345-369 (1967).

[50] M.A. Schweitzer, *A Paralle Multilevel Partition of Unity Method for Elliptic Partial Differential Equations*, Lecture Notes in Computational Science and Engineering Vol. 29, Springer, 2003.

[51] G. Strang, G. Fix, *A Fourier Analysis of Finite Element Variational Method*, in *Constructive Aspects of Functional Analysis*, Edizioni Cremonese, 1973, pp. 795-840.

[52] S. Timoshenko and S. Woinowsky-Krieger, *Theory of Plates and Shells*, (New York, McGraw-Hill, 1959).

[53] R.S. Varga, *Functional analysis and approximation theory in numerical analysis*, (SIAM; Philadelphia, PA, 1971)