A Class of Immersed Finite Element Spaces and Their Application to Forward and Inverse Interface Problems

Brian D. Camp

Dissertation submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Mathematics

Tao Lin, Chair
Slimane Adjerid
Christopher Beattie
Jeff Borggaard
Robert Rogers

November 19, 2003
Blacksburg, Virginia

Keywords: elliptic interface problem, least squares, immersed finite elements, mixed equation error, inverse problems

A Class of Immersed Finite Element Spaces and Their Application to Forward and Inverse Interface Problems

Brian D. Camp

(ABSTRACT)

A class of immersed finite element (IFE) spaces is developed for solving elliptic boundary value problems that have interfaces. IFE spaces are finite element approximation spaces which are based upon meshes that can be independent of interfaces in the domain. Three different quadratic IFE spaces and their related biquadratic IFE spaces are introduced here for the purposes of solving both forward and inverse elliptic interface problems in 1D and 2D. These different spaces are constructed by (i) using a hierarchical approach, (ii) imposing extra continuity requirements or (iii) using a local refinement technique. The interpolation properties of each space are tested against appropriate testing functions in 1D and 2D. The IFE spaces are also used to approximate the solution of a forward elliptic interface problem using the Galerkin finite element method and the mixed least squares finite element method. Finally, one appropriate space is selected to solve an inverse interface problem using either an output least squares approach or the least squares with mixed equation error method.

for Katie

Acknowledgements

First, I would like to thank Jesus Christ, my Lord and Savior, for giving me the opportunity and the ability to pursue an advanced degree in Mathematics. Without His help, I would never have reached this far.

I would also like to thank my family and friends. Without all of you, I would not be the person that I am today, and I especially appreciate all of the support and encouragement that you have given to me.

I also want to thank my committee. The things that I have learned from you have certainly increased my knowledge. As one professor of mine once said, "We are expanding your capacity to do work" and you have done just that.

Dr. Lin, you are a wonderful advisor. I cannot say enough about the guidance that you have given me and the patience that has gone along with it. You have given me a much deeper appreciation for mathematics. I am very grateful to have had the opportunity to work with you and I will always be indebted to you.

Katie, without you this dissertation probably would not have happened. You have always shown me much love and encouragement and provided me with motivation when I did not think that I had any. Thank you for helping me to persevere and to achieve my dream. Most of all, thank you for teaching me more about love in the process. My love for you has only grown deeper as I have worked to finish this degree. I will love you forever.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The purpose of this thesis is to develop new finite element (FE) spaces as tools for solving both forward and inverse interface problems involving the following elliptic partial differential equation (PDE):

$$-\nabla \cdot (p\ \nabla u) = f \tag{1.1}$$

together with a certain boundary condition, say $u|_{\partial\Omega} = 0$ on a domain $\Omega \subset R^n$ (with $n = 1$ or 2) that contains a piecewise smooth interface Γ separating Ω into two sub-domains Ω_1 and Ω_2 with $\Omega = \Omega_1 \cup \Omega_2 \cup \Gamma$. The coefficient function p is piecewise defined by

$$p(\boldsymbol{x}) = \begin{cases} p_1(\boldsymbol{x}), & \text{for } \boldsymbol{x} \in \Omega_1, \\ p_2(\boldsymbol{x}), & \text{for } \boldsymbol{x} \in \Omega_2, \end{cases} \tag{1.2}$$

where $p(\boldsymbol{x}) > 0$. These types of interface problems arise in many applications such as steady-state heat flow, electrostatics, and the modelling of oil reservoirs and aquifers, to name just a few [3, 16, 19, 37, 55].

The typical forward problem associated with (1.1) is: given a coefficient p and a forcing function f together with some boundary conditions such as $u\big|_{\partial\Omega} = 0$, then solve for the unknown function u. On the other hand, the inverse problem associated with (1.1) that is considered in this thesis is the problem of determining the interface Γ based upon information about the solution u to the elliptic PDE (1.1) together with certain boundary conditions.

In this introductory chapter we give the motivation for solving problems of these types as well as their formulation. Then a brief overview of some methods for solving these problems is given. In particular, this thesis aims to develop a class of immersed finite element (IFE) spaces that can be used together with a structured mesh to solve both the forward and inverse interface problems associated with (1.1).

1.1 Applications Involving Interfaces

Elliptic equations of the form (1.1) have various applications which make solutions of these types of problems desirable. Two prominent applications arise from the fact that this equation can be considered as either the steady-state heat/diffusion equation or the electric potential equation.

With the beginning of the twenty-first century, the drive of technology is for devices to become faster, more efficient and smaller. Many of today's devices are already much smaller than their counterparts of a decade ago: computer processors, hard disk drives, cellular phones, personal digital assistants (PDA's), CD players and the like have all become increasingly compact. As devices become smaller, many questions need to be asked during their design. Do the different electronic components of a device begin to interfere with each other as they are placed in closer and closer proximity? Does the heat generated by the components cause the device to fail? What is the most efficient and cost effective way to design the device so that it performs at a high level and is safe for humans to use?

A manufacturer of these items may ask the following question:

> In designing a device, how can the cost to manufacture the device be minimized while guaranteeing that heat flux and/or EM radiation generated by the device are kept within appropriate tolerance ranges?

In other words, how can typical electronic devices or components be designed so that relevant safety and performance questions are satisfied while maintaining a moderate to low cost? More questions like these are considered in [53].

More specifically, the motivation of this thesis relates to two primary applications. The first involves electromagnetic shielding of a device, and the second involves appropriate heat dissipation of a device. A typical device or component that is being discussed here is shown in Figure 1.1. The electronic component in this device is possibly a source of heat and/or electromagnetic waves. Surrounding the component are two or more layers of shielding or dissipating materials. An important question that may arise is whether or not the device has an acceptable amount of heat flux or electromagnetic flux through the device exterior. If the flux is not within an acceptable range, then the following question arises naturally: "what shape do the shielding layers need to have such that certain (desirable) boundary conditions and manufacturing costs are satisfied"?

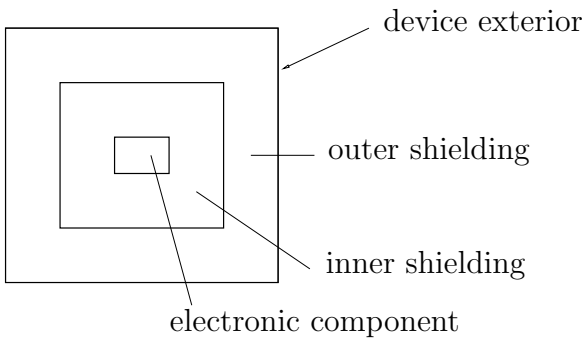


Figure 1.1: A sketch of a typical electronic device

1.1.1 Electric Flux

Consider a typical electronic device such as the one that is shown in Figure 1.1. The microchip that is shown may be a source of electromagnetic waves. Electromagnetic waves can be harmful to potential human users. They may also interact with other nearby components of the electronic device in an adverse way. For example, in the laptop computer industry, the components that make up each laptop are being used in closer and closer proximity to one another as technology advances. How close can the different microchips in a computer be before the operation of one part causes a nearby part to fail? Shielding materials can be placed around the microchip in order to control electromagnetic flux around the device. An important concern is what is the appropriate amount of shielding to be used in the device? Without loss of generality, we consider this problem with two shielding materials, see the sketch in Figure 1.1. Suppose that the outer layer of shielding is very effective but it is very expensive. On the contrary, the inner layer of shielding may be less effective but may be a very cheap material. A thick layer of the outer shielding certainly provides adequate shielding so that the device falls within the safety and design guidelines of the manufacturer. But too much outer shielding can adversely affect the cost of the product in question. On the contrary, more of the inner layer allows the cost of the device to be reasonable but the safety and operating guidelines may not be satisfied. For example, the amount of the inner shielding that is necessary by itself may make the size of the device too large.

Some devices to which the current discussion is appropriate are cell phones and PDA's. Consequently, both cellular phones and PDA's are designed with portability in mind, so that the proportion of different shielding layers is an important question. The wrong choices can adversely affect the size of these products (and thus their marketability). Manufacturers want to balance the amount of the inner and outer shielding in a device so that the cost and size of the product are competitive with other products on the market. At the same time, they also want to conform to the safety and operating goals of the device. Finding an appropriate amount of each type of shielding subject to cost constraints (and thus lowering the cost of such a device) is an important question to answer.

1.1.2 Heat Flux

Consider the same device as the one depicted in Figure 1.1. This device may be a cellular phone or possibly the central processing unit (CPU) in a laptop computer or a PDA. In both of these instances, devices can generate a significant amount of heat. Unfortunately, an excessive amount of heat can cause the device to be unsafe to use or to fail in its operation. To prevent either of these situations, an appropriate amount of heat dissipation must be included in the fabrication of the device. The same situation as with electromagnetic flux arises. The outer layer may be cheap but ineffective and the inner layer may be very effective but also very expensive. The device can easily be designed with enough inner heat dissipation material to keep the device from being dangerous to its users or from causing microchips within the device from failing. Too much inner insulation can, however, cause the product to be too expensive. Similarly, using too much of the outer insulation material can cause the device to be too cumbersome. Determining the amount of each type of insulation so that the heat dissipated satisfies the safety, operating, and marketing requirements of the device leads to the mathematical problems that are considered in this thesis.

1.2 Forward and Inverse Elliptic Interface Problems

We now give mathematical formulations for some problems that are related to the applications that are described in Section 1.1. As stated earlier, an appropriate mathematical model for these applications is the elliptic equation (1.1) with appropriate boundary conditions. This elliptic equation (1.1) can be used to represent either the electrostatic equation or the steady state heat equation in a device. For example, under certain assumptions, Maxwell's equations, which govern the electric and magnetic wave interactions inside an object, can be simplified to a differential equation for the electric potential u which has the same form as the elliptic interface problem (1.1) [23, 37]. Consider Figure 1.2 which shows a more abstract mathematical representation of the devices that are considered in Section 1.1. Then in terms of the electrostatic equation, u represents the electric potential over the domain Ω, f is a sink or a source in Ω and p is defined by

$$p(\boldsymbol{x}) = p_i, \quad \text{if } \boldsymbol{x} \in \Omega_i, \tag{1.3}$$

where the positive constants, $p_i,\ i = 1, 2, 3$, represent the conductivities of the different materials that comprise the object. The different materials in the object correspond to the subdomains that are represented by $\Omega_i,\ i = 1, 2, 3$. The assumption here is that the object or device is allowed to reach some steady operating state where it conforms to either the electric potential equation or the steady state heat equation. Here Ω_1 represents the outer shielding layer, Ω_2 represents the inner shielding layer around Ω_3 which is occupied by a microchip or some electronic component.

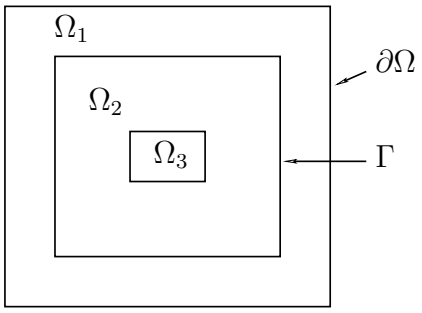


Figure 1.2: A domain $\Omega = \bar{\Omega}_1 \cup \bar{\Omega}_2 \cup \bar{\Omega}_3$ composed of three subdomains.

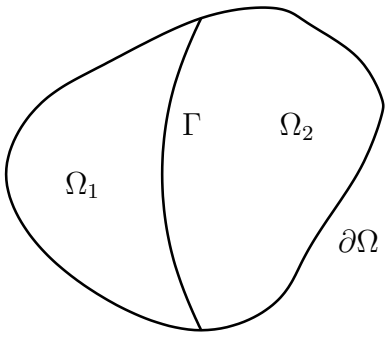


Figure 1.3: A domain $\Omega = \Omega_1 \cup \Omega_2$ with one interface Γ

We consider a device that is required to have a certain type of microchip with a fixed shape, Ω_3. Also, the overall shape of the domain $\bar{\Omega} = \bar{\Omega}_1 \cup \bar{\Omega}_2 \cup \bar{\Omega}_3$ is fixed. Then the question of optimizing the shielding layers of the object, is equivalent to determining the optimal shape of the interface Γ between Ω_1 and Ω_2 subject to an appropriate design constraint such as controlling the flux on the boundary of the object.

Since Ω_3 is assumed to be fixed, we further simplify the discussion by looking at objects that are made of only two materials such as the one that is shown in Figure 1.3. In this object, a domain Ω is composed of two subdomains Ω_1 and Ω_2 such that the function p is defined by (1.2) and hence p may be discontinuous across the interface $\Gamma = \bar{\Omega}_1 \cap \bar{\Omega}_2$. The function p represents, for example, the conductivities of the outer and inner shielding layers. Since the materials to be used in a device (which we represent by Ω_1 and Ω_2) are chosen by the manufacturer, then their physical properties are known. Hence the values of p_1 and p_2, which are the conductivities of the two materials, are known constants. In general, however, they may not be equal since the two materials may have different physical properties.

Some important problems arise from the elliptic problem (1.1) which models the physical situations in which we are interested. One such problem is to determine the shape of the interface Γ so that certain boundary criteria are satisfied along $\partial\Omega$. This can equivalently be

considered as a problem of determining the function p since Γ is a part of the definition of p. Specifically, this problem asks "if a certain flux is desired at the boundary of the domain Ω, where should the interface Γ between our two materials occur?" Another problem that may be posed is to find the function u which satisfies (1.1) subject to certain boundary conditions and a given interface Γ.

Baumeister [4] describes the difference between these two problems by giving a description of two classes of problems called forward problems and inverse problems. The difference is based upon whether one is solving for the typical inputs or outputs of a given equation or system. Inputs usually correspond to those things that are considered to be responsible for creating the outputs. Problems which use the typical inputs to determine the unknown output, he labels as forward problems. On the other hand, if information about the output is used to determine some information about some or all of the inputs, he calls this an inverse problem.

For example, in equation (1.1), p and f can both be seen traditionally as inputs to a problem since they give information about a physical situation which in turn results in a particular function u (the output) which satisfies the equation. So finding the function u (the output) is considered the forward problem. On the other hand, since u is the output, then to use information about the output such as its boundary values in order to determine information about the coefficient p is considered an inverse problem. These two problems are related to each other since the solution of one is often needed to find the solution of the other.

Now we are ready to state a forward problem and a related inverse problem that motivates the development of the immersed finite element spaces that are considered in this thesis. First we give a typical forward problem for the elliptic interface equation (1.1).

Problem (F): Find u such that

$$\begin{aligned} -\nabla \cdot (p\ \nabla u) &= f, \text{ for } \boldsymbol{x} \in \Omega, \\ u &= 0, \text{ for } \boldsymbol{x} \in \partial\Omega, \end{aligned} \tag{1.4}$$

where $f \in L^2(\Omega)$ and $p \in L^2(\Omega)$ is piecewise defined as in (1.2).

In **Problem (F)**, which we call the forward problem, p and f are known functions and the domain Ω is contained in $\mathbb{R}^n$.

Investigating **Problem (F)** in $\mathbb{R}^1$ is interesting and important for a couple of reasons. First, considering this simpler problem in one dimension may give insight into how to approach corresponding higher dimensional problems. Second, in many applications, a higher dimensional problem can be reduced to a meaningful one dimensional problem due to symmetry or other assumptions. So the one dimensional problem is worthwhile to investigate from both a

mathematical and a physical perspective. Specifically, in the typical one dimensional domain $\Omega = (0,1)$ where $\Omega_1 = (0,\alpha)$ and $\Omega_2 = (\alpha,1)$, the definition of the coefficient function p becomes

$$p(x,\alpha) = \begin{cases} p_1, & x \leq \alpha, \\ p_2, & x > \alpha, \end{cases} \tag{1.5}$$

where p_1 and p_2 are positive constants. Then **Problem (F)** reduces to

Problem (F-1): Find u such that

$$\begin{aligned} -(p\ u')' &= f, \text{ for } x \in (0,1), \\ u(0) = u(1) &= 0, \end{aligned} \tag{1.6}$$

where $f \in L^2(0,1)$ and p is piecewise defined as in (1.5).

The problems introduced in the previous section lead to the following inverse problem which is related to **Problem (F)**:

Problem (I): Find Γ such that u is the solution to (1.4), and in addition, the following flux condition is satisfied along the boundary of Ω:

$$p(\boldsymbol{x})\nabla u(\boldsymbol{x}) \cdot \boldsymbol{n} = g(\boldsymbol{x}), \quad \text{for } \boldsymbol{x} \in \partial\Omega, \tag{1.7}$$

where p is defined as in (1.2) and $\boldsymbol{n}$ is the unit outward normal on $\partial\Omega$ at $\boldsymbol{x}$.

Alternatively, the flux may only need to match the prescribed value of the function g at a discrete set of points along the boundary of Ω and so (1.7) becomes

$$p(\boldsymbol{x}_j)\nabla u(\boldsymbol{x}_j) \cdot \boldsymbol{n} = g(\boldsymbol{x}_j), \quad \text{for } \boldsymbol{x}_j \in \partial\Omega,\ j = 1,\ldots,q. \tag{1.8}$$

This latter situation may occur because of discretization or because it is not possible to measure the flux continuously on the boundary, but it is possible to measure it at a discrete set of points.

Typically in **Problem (I)** and similar problems, the interface Γ is parameterized by a curve, $\boldsymbol{l}(t)$. For example, if $\Omega \subset \mathbb{R}^2$ then $\boldsymbol{l}(t)$ may have the form

$$\boldsymbol{l}(t) = (x(t), y(t))^T. \tag{1.9}$$

There are many possible choices for $x(t)$ and $y(t)$ but one possibility is

$$\begin{aligned} x(t) &= t, \\ y(t) &= \sum_{i=1}^{q} \alpha_i \phi_i(t), \end{aligned} \tag{1.10}$$

where $\phi_i(t),\ i = 1, \ldots, q$, form a set of linearly independent basis functions. Thus in **Problem (I)**, the matter of determining Γ is reduced to the task of determining the parameter $\vec{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_q)^T$. Determining $\vec{\alpha}$ also determines the coefficient function p since the only thing that is not known about the definition of p is the location of the interface Γ. Likewise, if the definition of p is fully known, then so is Γ. Thus determining p is equivalent to determining the parameter $\vec{\alpha}$. The function u supplies some of the necessary constraints in the problem so that the coefficient p can be identified. Sometimes this problem is called the determination of the Dirichlet-to-Neumann map [34, 46]. The problem is also similar to problems that are formulated as shape optimization problems [50] or elliptic problems that are formulated using a fictitious domain approach [26].

We start the investigation of this inverse problem in one dimension. Again, this is a meaningful problem since the one dimensional problem may give insight into the solution of higher dimensional problems and because many physical domains can be reduced to one dimensional problems due to symmetry. The one dimensional **Problem (I)** on the domain $(0, 1)$ is stated as follows:

> **Problem (I-1):** Find α such that $p(0)u'(0) = g_0$ and $p(1)u'(1) = g_1$ where p is defined as in (1.5) and u satisfies **(F-1)**.

Both **Problems (I)** and **(I-1)** are parameter identification problems since the interface Γ can be determined by finding the appropriate parameter $\vec{\alpha}$.

1.3 A Brief Survey of Solution Approaches

Almost all of the solution approaches for these types of inverse problems involve minimizing a cost functional. The typical cost functional is derived from a generic least squares idea (see [47, 28, 4, 5, 37, 38, 39]). According to their specific formulations, these available solution approaches can be classified as the output least squares, least squares with equation error, and mixed least squares (see [12, 13, 45, 21, 51]). Many of these methods are implemented using various finite element schemes. Other methods are also used such as level set methods (see [11, 14, 18, 29]) and boundary element methods (see [35, 36, 48, 54]) which also use general least squares problem formulations.

1.3.1 Output Least Squares

A least squares approach (sometimes called an output least squares method) is a very simple and useful idea that can be applied to many inverse problems. This method involves minimizing a cost functional, $\mathcal{J}_o(\vec{\alpha})$, which involves the square of the difference between some

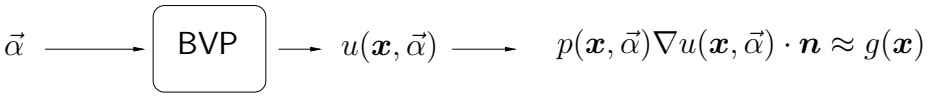


Figure 1.4: The relationship between $\vec{\alpha}$ and u in the output least squares method.

actual quantity and some desired quantity. The actual quantity depends on some unknown value or parameter, $\vec{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_q)^T$. In the process of minimizing the cost functional, the unknown is determined in a "least squares sense." "Least Squares" solutions do not always give an answer that exactly satisfies the desired requirements. If a parameter $\vec{\alpha}^*$ is found so that $\mathcal{J}_o(\vec{\alpha}^*) = 0$ then the involved requirements are then satisfied. It is often the case, however, that the minimum of $\mathcal{J}_o(\vec{\alpha})$ is not zero. When this happens, we accept the result that does the best with the given circumstances.

Describing a method as being a "least squares" method only characterizes it in the broadest sense: the square of the difference between two quantities is being minimized. How these two quantities are calculated and how the minimization is achieved are largely dependent upon the specific problem under consideration. For a background of least squares approaches to inverse problems, see [3, 47].

In a least-squares method, the actual quantity is generated from the solution of the forward problem with a value of the parameter that is to be determined. Then an iterative process (i.e. steepest descent or a Newton or quasi-Newton method) is used to change the parameter value with the goal being to decrease the square of the differences. Hopefully, this iteration procedure converges to a minimizer when appropriate starting conditions are used.

For **Problem (I)** that is considered herein, Figure 1.4 illustrates the idea of how a parameter $\vec{\alpha}$ may relate to a boundary value problem. The parameter $\vec{\alpha}$ is an input to the BVP out of which comes a solution to the BVP which depends upon $\vec{\alpha}$. Once u is obtained from the BVP, it is used in turn to calculate the flux which is required to match to the target function g in the least-squares approach.

Specifically, in **Problem (I)**, the least squares cost functional is defined by

$$\mathcal{J}_o(\vec{\alpha}) = \int_{\partial\Omega} |p(\boldsymbol{x}, \vec{\alpha})\nabla u(\boldsymbol{x}, \vec{\alpha}) \cdot \boldsymbol{n} - g(\boldsymbol{x})|^2 ds, \tag{1.11}$$

where g is a prescribed function on the boundary $\partial\Omega$ of the domain. In the **Problem (I-1)**, if g_0 and g_1 are prescribed flux values at the left and right boundaries of $\Omega = (0, 1)$ then the least squares cost functional becomes

$$\mathcal{J}_o(\alpha) = |p(0, \alpha)u'(0, \alpha) + g_0|^2 + |p(1, \alpha)u'(1, \alpha) - g_1|^2.$$

In **Problem (I-1)** we are able to represent the coefficient p as a function of the parameter α and so the unknown being solved for is α.

This method while appearing simple can be quite complex. The calculation of the first and second derivatives of $\mathcal{J}_o(\vec{\alpha})$ is by no means a trivial task. This is especially true because of the way that u depends upon the parameter $\vec{\alpha}$. Since u and $\vec{\alpha}$ are related through a boundary value problem, their relationship can be quite nonlinear (see Figure 1.4). In other words, while minimizing the functional $\mathcal{J}_o(\vec{\alpha})$ there is the constraint that $u(\boldsymbol{x}, \vec{\alpha})$ must satisfy a given boundary value problem.

1.3.2 The Method of Least Squares with Equation Error

The Method of Least Squares with Equation Error (LSEE) builds upon the output least squares idea. In the output least squares method, the forward BVP acts as a constraint placed upon the functional that is to be minimized. This constraint can be included in the cost functional via a Lagrange multiplier approach, and essentially leads to the idea of equation error. The main advantage of including the residual error of the BVP into the cost functional is that it allows the function u to be treated as a variable as opposed to a function dependent upon the parameter $\vec{\alpha}$. In other words, this idea allows us to adjust both the parameter $\vec{\alpha}$ and the solution u to the BVP simultaneously such that the requirements of both matching the target function and solving the BVP can be fulfilled in a least-squares sense.

For the **Problem (I)**, the BVP constraint leads the following residual error of the involved PDE in the cost functional:

$$\|f + \nabla \cdot (p\nabla u)\|_0^2 = \int_\Omega (f + \nabla \cdot (p\nabla u))^2 d\boldsymbol{x}, \tag{1.12}$$

and thus the cost functional becomes

$$\mathcal{J}_e(u, \vec{\alpha}) = \|f + \nabla \cdot (p\nabla u)\|_0^2 + \int_{\partial\Omega} |p\nabla u \cdot \boldsymbol{n} - g|^2 ds. \tag{1.13}$$

In **Problem (I-1)**, the equation error term is

$$\|f + (p\ u')'\|_0^2 \ = \int_0^1 (f + (pu')')^2 dx, \tag{1.14}$$

and so the cost functional becomes

$$\mathcal{J}_e(u, \alpha) = \int_0^1 (f + (pu')')^2 dx + |p(0, \alpha)u'(0) + g_0|^2 + |p(1, \alpha)u'(1) - g_1|^2. \tag{1.15}$$

One important difference between the output least squares cost functional, $\mathcal{J}_o$, and the least squares with equation error cost functional, $\mathcal{J}_e$, is that the latter involves two unknowns,

p and u, instead of just one. Both the parameter p and the function u are considered as independent variables in the method of least squares with equation error while the output least squares method treats only p as an unknown, and the function u is seen as a function of p through the BVP. An advantage of treating these two unknowns independently is that the cost functional is now a fourth order polynomial in terms of the unknowns p and u. Variable u is no longer considered to be dependent upon the parameter p. As a consequence, taking the derivative of the resulting cost functional, $\mathcal{J}_e(u, \vec{\alpha})$, with respect to both variables becomes much easier. A drawback of using this equation error method is that when a minimizer of the cost functional is found, it may produce a u which does not necessarily satisfy the original forward problem. Instead, the involved BVP is satisfied by this u in a least-squares sense. Another drawback of this method is that while the terms in the cost functionals above are easy to calculate using finite elements, they do require that C^1 finite elements be used because of the second derivative of u appearing in the formulation. This requires u to be in the space $H^2(\Omega)$.

1.3.3 The Method of Least Squares with Mixed Equation Errors

The Method of Least Squares with mixed equation errors (LSMEE) is a modification of the Method of Least Squares with Equation Error. The term "mixed" refers specifically to a procedure that reduces the PDE in the involved BVP to a system of lower order PDE's. Usually the decomposition results in two variables which both have some physical significance to the problem at hand. If this idea is used together with the idea of least squares with equation error for the inverse problems that are considered in this thesis, then the equation error terms appearing in the cost functional involve the system of lower order differential equations. By reducing the PDE into lower order equations, the need to work with higher order derivatives is removed. Therefore, finite element spaces with lower continuity requirements can then be used to minimized the cost functional numerically.

For the BVP that is considered here, by introducing a new variable for the flux, $\boldsymbol{\sigma} = p\nabla u$, we can reduce the involved PDE to the following mixed formulation:

$$\begin{aligned} p\nabla u &= \boldsymbol{\sigma}, \\ -\nabla \cdot \boldsymbol{\sigma} &= f. \end{aligned} \tag{1.16}$$

Then the equation error used in the cost functional for solving the inverse problem can be broken into two different terms:

$$\|f + \nabla \cdot \boldsymbol{\sigma}\|_0^2 + \|\boldsymbol{\sigma} - p\nabla u\|_0^2. \tag{1.17}$$

In the method of LSMEE, there are three unknowns: $\vec{\alpha}, u$ and $\boldsymbol{\sigma}$. Then the cost functional for **Problem (I)** is

$$\mathcal{J}_m(u, \boldsymbol{\sigma}, \vec{\alpha}) = \|\boldsymbol{\sigma} - p\nabla u\|_0^2 + \|f + \nabla \cdot \boldsymbol{\sigma}\|_0^2 + \int_{\partial\Omega} |p\nabla u \cdot \boldsymbol{n} - g|^2 ds. \tag{1.18}$$

For the **Problem (I-1)**, the mixed formulation of the boundary value problem is

$$\begin{array}{rcl} pu' & = & \sigma, \\ -\sigma' & = & f, \end{array} \tag{1.19}$$

and then the equation error term is

$$\|f + \sigma'\|_0^2 + \|pu' - \sigma\|_0^2. \tag{1.20}$$

For **Problem (I-1)**, the unknowns being solved for are u, σ and α. Hence, the cost functional is

$$\mathcal{J}_m(u, \sigma, \alpha) = |p(0, \alpha)u'(0) + g_0|^2 + |p(1, \alpha)u'(1) - g_1|^2 + \|f + \sigma'\|_0^2 + \|pu' - \sigma\|_0^2. \tag{1.21}$$

Similar to the method of LSEE, the cost functional in the method of LSMEE is a polynomial of order 4 in terms of p and u. Now, however, there are three independent variables $\vec{\alpha}$, u and $\boldsymbol{\sigma}$ in the cost functional that are being solved for instead of just two as in the method of LSEE. By introducing more variables, we reduce the order of the derivatives that are involved in the cost functional $\mathcal{J}_m(u, \boldsymbol{\sigma}, \vec{\alpha})$. Hence, this method allows finite elements from a lower order Sobolev space to be used in the calculation of the cost function. This could not be done with the method of LSEE unless a non-conforming finite element scheme is used. The advantage of using finite element spaces from a lower order Sobolev space is best recognized when we consider an inverse problem in a domain of $\mathbb{R}^n$ with $n > 1$. The cost functional used to solve the inverse problem determines the space in which the variable u resides. In the method of LSEE, the cost functional requires a function $u \in H^2(\Omega)$. To use a conforming finite element method to solve the inverse problem with this cost functional, we need to use a C^1 finite element space which is rather restrictive in two dimensions and higher because of the higher degree of freedom inherent to these kinds of finite element spaces. For example, let us consider the degrees of freedom of the Argyris and Bell C^1 finite element spaces when $\Omega \subset \mathbb{R}^2$. A local nodal basis function of Argyris finite element space has 21 degrees of freedom, while a local nodal basis function from the Bell finite element space has 18 degrees of freedom [8, 9]. Using these C^1 finite element spaces usually leads to complicated and time consuming implementations. On the contrary, a mixed formulation for the equation error terms such as (1.17) or (1.20) allows us to look for solutions $u \in H^1(\Omega)$. The result is that we can use the much simpler C^0 finite element spaces to solve our inverse problems.

Some references relating to the mixed formulation of a BVP and the mixed least squares formulation of the inverse problem can be found in [8, 9, 12, 13, 21, 43, 45, 51].

1.4 Mathematical Tools Needed in the Implementation of the Mixed Least Squares Method

Special finite element spaces are needed in order to solve the inverse problems **(I)** and **(I-1)** efficiently by using the method of LSMEE formulation. Note that the inverse problem considered is solved by finding the suitable location of the discontinuity in the coefficient function p. As the independent variable in the cost functional, the interface is modified in each step of a minimization method applied. On the other hand, to achieve the optimal accuracy, the traditional finite element method requires that elements are not split by the interface. Hence the traditional finite element method is not suitable for solving the inverse problems here because the partition used in the finite element has to be reformed again and again in each iteration step. Not only does this demand more computational time, but also the implementation is more complicated because of the following reasons:

- First, the numbers of unknowns from two consecutive partitions may not be the same.
- Second, the location of the node corresponding to any given unknown in two consecutive partitions may not be the same.

Therefore the number of unknowns and their locations in the discretized cost functional are undetermined when a traditional finite element methods is used.

Hence, it is desirable to utilize finite element spaces which are based upon partitions independent of the location of the interface. The recently introduced immersed finite element (IFE) spaces seem to be a natural choice [20, 40, 44]. This thesis first develops a class of quadratic IFE spaces which are extensions of those in the literature that are based on linear polynomials. As usual, these new IFE spaces can be used to solve the forward elliptic interface problem using both the Galerkin method and the least squares finite element method. Then these IFE spaces are applied to the calculation of the cost functional and its derivatives in the method of LSMEE for solving inverse elliptic interface problems.

1.4.1 Standard Finite Element Spaces

Finite element methods provide a way to approximate the solution of a BVP. Typically the BVP is rewritten as a variational or weak problem which has a solution in an appropriate Sobolev space (this is called the Galerkin method, see [8, 17]). Any problem which has a variational form can be formulated in this way. The finite element method involves approximating the solution to the variational problem in a finite dimensional subspace of the Sobolev

space where the variational solution lives. Standard finite elements involve constructing the finite dimensional subspace (or finite element space) using piecewise polynomial basis functions that have finite support. This is called a conforming finite element method if the finite element space is contained in the solution space of the variational problem. Otherwise the method is say to be nonconforming. For more on finite element methods see, for example, [8] and [9].

To form a finite element space, the domain of the BVP is first partitioned into smaller sub-domains called elements. The basis functions of the finite element space are chosen to be piecewise polynomials, each of which has a support on a small number of elements. Generally, the way this is done is to first construct a set of local basis functions for each element which span an appropriate polynomial space on the element. This may involve the introduction of a local partition on each element. Then a global finite element basis function associated with a node of the finite element partition is constructed by piecewisely defining the function on each element using the local basis functions which satisfy the appropriate nodal value specifications. Finally, the finite element space is the span of the global basis functions for each of the nodes in the partition.

As an example, consider using the first degree polynomials to form a finite element space in the typical one dimensional domain $\Omega = (0, 1)$. We begin to construct the standard linear FE space on this domain by first giving the domain a uniform partition $[x_0, x_1, \ldots, x_N]$ such that $x_i = x_0 + ih$ and $h = \dfrac{x_N - x_0}{N}$. Then the interval $e_i = [x_{i-1}, x_i]$ is the i^{th} element of the partition $\mathcal{T}_h = \bigcup_{i=1}^{N} e_i$. Thus there are N elements in the partition represented by the intervals $[x_{i-1}, x_i]$ for $i = 1, \ldots, N$. On each element, we introduce a space which is spanned by linear polynomials. This is done by introducing the local linear basis functions $\phi_{k,1}(x)$ and $\phi_{k,2}(x)$ which satisfy the following nodal value specifications:

$$\phi_{k,i}(t_{k,j}) = \begin{cases} 1, & \text{if } i = j, \\ 0, & \text{if } i \neq j, \end{cases} \tag{1.22}$$

where $t_{k,1} = x_{k-1}$ and $t_{k,2} = x_k$. Then the local basis functions on e_i are defined by

$$\phi_{k,1}(x) = \frac{x - t_{k,2}}{t_{k,1} - t_{k,2}} \text{ and } \phi_{k,2}(x) = \frac{x - t_{k,1}}{t_{k,2} - t_{k,1}}. \tag{1.23}$$

The finite element space over the domain $\Omega = (0, 1)$ is then formed by patching together the local spaces of linear polynomials on each element. For this global space, a global basis function $\Phi_i(x)$ is associated with each node x_i such that $\Phi_i(x)$ satisfies the nodal value requirement:

$$\Phi_i(x_j) = \begin{cases} 1, & \text{if } i = j, \\ 0, & \text{if } i \neq j. \end{cases} \tag{1.24}$$

The resulting basis function has support on the elements e_i and e_{i+1} and is defined by

$$\Phi_i(x) = \begin{cases} \phi_{i,2}(x), & \text{if } x \in e_i, \\ \phi_{i+1,1}(x), & \text{if } x \in e_{i+1}, \\ 0, & \text{otherwise.} \end{cases} \tag{1.25}$$

Consequently, each global basis function $\Phi_i,\ i = 0, \ldots, N$ is piecewise linear and $C^0(0,1)$. Hence the standard linear FE space is simply

$$\mathcal{S}_h = span\{\Phi_i, \text{ for } i = 0, \ldots, N\}, \tag{1.26}$$

and a typical function in $\mathcal{S}_h$ has the form

$$u_h(x) = \sum_{i=0}^{N} c_i \Phi_i(x). \tag{1.27}$$

Depending upon the BVP that is being solved, it may be desirable to use basis functions which are C^1 or stronger. One example of this is the Argyris elements mentioned in [8, 9]. The choice of the finite element space is largely dependent upon the solution space of the variational problem which arises from the BVP in question. For example, since $C^0(0,1)$ is embedded in the Sobolev space $H^1(0,1)$ then a C^0 FE space is appropriate for approximating the solution of a variational problem which occurs in $H^1(0,1)$.

The process of forming a finite element space on a domain Ω with more than one dimension is similar. First the domain is partitioned into sub-domains. A local polynomial space is defined on each element by defining appropriate local basis functions on this element. Then the finite element space is constructed over the whole domain by piecing together the polynomial spaces that are defined on each element. Global basis functions are defined by piecing together local basis functions from different elements, such that appropriate nodal value specifications analogous to (1.24) are satisfied.

1.4.2 Immersed Finite Element Spaces

The standard finite element approach is not always a good means of approximating a solution of a BVP. In particular, if the domain of the BVP contains an interface then standard finite elements do not perform as well as they should. The reason for this is connected to the fact that the solution of the BVP satisfies certain interface conditions that are derived from the BVP itself. When standard finite elements are used, the basis functions themselves do not make any allowance for these interface conditions.

Immersed finite elements are an alternative to using the standard finite elements to approximate the solution of an interface BVP. For example, an elliptic problem of form (1.1) requires that its solution be continuous and its flux be continuous across the interface [6, 16]. Immersed finite element spaces are formed in a way that is very similar to the process used for the construction of standard finite element spaces that is described above. However, immersed finite element spaces also use the jump conditions across the interface in order to define local basis functions on elements that are intersected the interface. Again, on each element, a local polynomial space is formed by defining appropriate local basis functions to span the space. However, when constructing immersed finite element spaces, the local basis functions on each element are required to satisfy appropriate interface conditions. This requirement only affects the definitions of local basis functions in elements that are touched by the interface. In other words, only basis functions whose support is over elements which contain the interface in their interior are required to satisfy the interface conditions. The meaning of this is made more precise in Chapter 2. Consequently, standard finite element basis functions are still used for the local basis functions on elements which do not contain the interface in their interior. Since this construction of a finite element space allows for the interface to be contained within or immersed within an element, we call spaces of this type immersed finite element spaces.

Here we consider the case where $\Omega = (0, 1)$. A typical elliptic interface problem is given by (1.6) where p is defined as in (1.5). This results in an interface at $x = \alpha$. Thus a function $u(x)$ which satisfies (1.4) is required to satisfy the interface conditions

$$\begin{array}{rcl} \left[u\right]_{x=\alpha} & = & 0, \\ \left[pu'\right]_{x=\alpha} & = & 0, \end{array} \tag{1.28}$$

where $\left[f\right]_{x=\alpha} = f(\alpha+) - f(\alpha-)$. We proceed in the same fashion as for standard FE spaces and construct a uniform partition $\mathcal{T}_h = \bigcup_i e_i$ where the interval $e_i = [x_{i-1}, x_i]$ is the i^{th} element of the partition. On elements which do not contain the interface, local polynomial spaces are constructed in the same manner as for standard FE spaces. However, if $x_{k-1} < \alpha < x_k$ such that the element e_k contains the interface then the local linear basis functions $\phi_{k,i}(x),\ i = 1, 2$ are defined so that they satisfy the interface conditions (1.28) as well as the following nodal value specifications:

$$\phi_{k,i}(t_{k,j}) = \begin{cases} 1, & \text{if } i = j, \\ 0, & \text{if } i \neq j. \end{cases} \tag{1.29}$$

To satisfy these conditions we require that $\phi_{k,i}(x)$ has the following general piecewise linear form

$$\phi_{k,i}(x) = \begin{cases} a_1 x + a_0, & \text{if } x < \alpha, \\ b_1 x + b_0, & \text{if } x \geq \alpha, \end{cases} \tag{1.30}$$

so that $\phi_{k,i}(x)$ has four coefficients that must be determined. It is straightforward to find values for a_1, a_0, b_1 and b_0 by using the four conditions provided by (1.28) and (1.29). Then the linear IFE space over the domain $\Omega = (0, 1)$ is constructed in the same fashion as for the standard FE space by piecing together the local linear polynomial spaces on each element. Again, a global basis function $\Phi_i(x)$ is associated with each node x_i and is required to satisfy the nodal value specification (1.24). However, in this case, on the element e_k that contains the interface, each of the global basis functions $\Phi_i(x)$ satisfies the interface conditions (1.28). Again the basis functions $\Phi_i(x),\ i = 0, \ldots, N$ are all piecewise linear and hence $C^0(0, 1)$.

One big advantage of using immersed finite element spaces is that the partition upon which the approximation space is formed is independent of the interface that occurs in the elliptic interface problem (1.4). This makes IFE spaces an ideal choice for problems which involve a moving interface since the same partition can be used throughout the process of solving the problem. Additionally, only a small number of elements are affected by the interface. Hence, on the majority of the elements, the standard FE basis functions are used. More detail is given about immersed finite element spaces and their basis functions in the next chapter.

1.5 Organization of This Thesis

The rest of this thesis is divided into five chapters. Chapter 2 discusses immersed finite element (IFE) spaces and how to construct them. In particular, quadratic and biquadratic IFE spaces are constructed in one dimension and in two dimensions with rectangular elements. Chapter 3 investigates the interpolation properties of the quadratic and biquadratic IFE spaces by performing some approximation experiments using the different IFE spaces and some appropriate test functions in one and two dimensions. Chapter 4 uses these same test functions in observing the approximation capabilities of these IFE spaces when using the Galerkin Finite Element method to solve an elliptic boundary value problems with discontinuous coefficients in both one and two dimensions. Chapter 5 uses the least squares finite element method with IFE spaces to solve elliptic boundary value problems with discontinuous coefficients in one dimension. Finally in Chapter 6, the different IFE spaces are applied to solving an inverse problem that involves an elliptic equation similar to (1.1) with a discontinuous coefficient function p. Results using both an output least squares cost functional, $\mathcal{J}_o$ and a LSMEE cost functional, $\mathcal{J}_m$, are given.

Chapter 2

A Class of Quadratic Immersed Finite Element Spaces

This chapter investigates the construction of a class of quadratic immersed finite element (IFE) spaces that are important tools in the solution of the forward and inverse interface problems introduced in Chapter 1. Finite element spaces form a cornerstone to solution methods for elliptic boundary value problems which use discretized versions of the Galerkin and Ritz methods. IFE spaces are constructed by taking an existing FE space and altering the definitions of basis functions on the elements of the partition $\mathcal{T}_h$ which intersect with the interface. On these "interface" elements, the local IFE basis functions are defined in such a way that interface conditions required by the BVP are satisfied.

Normally, for standard finite element methods to achieve optimal accuracy, the interface Γ has to be incorporated into the partition of the domain Ω. For example, consider the finite element space based on linear polynomials in $\mathbb{R}^2$. This is done by only allowing Γ to intersect the boundaries of the elements in the partition at their vertices. This restriction results in a partition in which every element is on one side of Γ or the other, and a partition with this feature is usually called a body fit partition. Several drawbacks to this approach can arise when solving an interface problem. If the body fit method is used in a boundary value problem that requires the movement of the interface in an iterative fashion (such as the least-squares type methods for an inverse problem), then with each iteration, the partition for the FE space needs to be reformulated. First, if the derivative of a cost function needs to be taken with respect to the interface (or with respect to parameters that represent the interface), it is not particularly easy, if not impossible, to calculate this derivative. In particular, if the interface is changing from one iteration to the next, this causes the partition to change. This can have the effect of both changing the total number of elements in the partition and also changing the number of elements which are nearby a given node. Therefore, the location and

the total number of the unknowns in the numerical methods based on the body fit partition for either the forward or the inverse interface problem is also changing from one iteration to the next. These difficulties give motivation for using a FE space introduced recently [22, 40, 44] that uses a fixed partition independent of the interface Γ.

Immersed finite element spaces are constructed to reduce the difficulties introduced by the use of a standard FE space as mentioned above. As a comparison, we note the following differences between a standard FE space and an IFE space for interface problems:

- The partition of a standard FE space is formed according to the interface while the partition of a IFE can be formed independent of the interface.
- The local nodal basis functions of a standard FE space are independent of the interface problem while those of a IFE space are formed according the interface jump conditions.

These features tell us that a standard FE space and an IFE space use different ways to incorporate the given interface conditions. For the problems that we consider in this thesis and many other applications, an IFE space may be a preferable alternative. For example, using an IFE space can relieve the difficulties that arise from the need to change a partition during the iterative process of minimizing a cost functional. The number of unknowns and the location of the nodes where they occur are not required to change, and taking the derivative of the basis functions with respect to the parameters that represent the interface Γ becomes a relatively straightforward process. These are desirable properties for solving inverse interface problems.

IFE spaces that use linear basis functions in one dimension, linear basis functions in two dimensions with triangular elements and bi-linear basis functions in two dimensions with rectangular elements have already been derived. Some details are given here although readers are referred to [20, 40, 41, 44] for further details on these IFE spaces. In this chapter, we extend the IFE idea to those using quadratic polynomials. IFE spaces that use quadratic basis functions in one dimension and biquadratic basis functions in two dimensions with rectangular elements are introduced. As is seen below, contrary to the case of those IFE spaces based on linear polynomial, the construction of quadratic and biquadratic IFE spaces is not unique, and so the extension is not a trivial job, and some particular strategies are presented to ensure the performance of these new IFE spaces.

As usual, the development of quadratic IFE spaces is particular to the boundary value problem that is being considered and is discussed in further detail in the following sections. Here we recall that our model elliptic interface problem is

$$\begin{aligned} -\nabla \cdot (p \nabla u) &= f, \text{ for } \boldsymbol{x} \in \Omega, \\ u &= 0, \text{ for } \boldsymbol{x} \in \partial\Omega, \end{aligned} \tag{2.1}$$

where the coefficient function p is defined by

$$p(\boldsymbol{x}) = \begin{cases} p_1, & \boldsymbol{x} \in \Omega_1, \\ p_2, & \boldsymbol{x} \in \Omega_2, \end{cases} \tag{2.2}$$

with $p_i > 0,\ i = 1, 2$ are constant. It is also required that u must satisfy the interface jump conditions

$$\begin{aligned} \big[u\big]_\Gamma &= 0, \\ \big[p\nabla u \cdot \boldsymbol{n}\big]_\Gamma &= 0, \end{aligned} \tag{2.3}$$

where $\boldsymbol{n}$ is the normal vector on Γ which points from Ω_1 to Ω_2. Here, if a function is defined piecewise by

$$f(\boldsymbol{x}) = \begin{cases} f_1(\boldsymbol{x}), & \text{if } \boldsymbol{x} \in \bar{\Omega}_1, \\ f_2(\boldsymbol{x}), & \text{if } \boldsymbol{x} \in \bar{\Omega}_2, \end{cases} \tag{2.4}$$

then $\big[f\big]_\Gamma(\boldsymbol{x})$ is defined by

$$\big[f\big]_\Gamma(\boldsymbol{x}) = f_2(\boldsymbol{x}) - f_1(\boldsymbol{x}),\ \text{for } \boldsymbol{x} \in \Gamma. \tag{2.5}$$

More generally, the domain Ω may consist of several regions so that

$$\bar{\Omega} = \bigcup_{i=1}^{n} \bar{\Omega}_i. \tag{2.6}$$

In this case, the interface Γ is the union of $\bar{\Omega}_i \cap \bar{\Omega}_j, i \neq j$. Consider a function f that is defined piecewise as

$$f(\boldsymbol{x}) = f_i(\boldsymbol{x}),\ \text{for } \boldsymbol{x} \in \bar{\Omega}_i. \tag{2.7}$$

Then, for $\boldsymbol{x} \in \Gamma$ such that $\boldsymbol{x} \in \bar{\Omega}_i \cap \bar{\Omega}_j$, we let $\big[f\big]_\Gamma(\boldsymbol{x})$ be defined by

$$\big[f\big]_\Gamma(\boldsymbol{x}) = f_j(\boldsymbol{x}) - f_i(\boldsymbol{x}),\ \text{for } \boldsymbol{x} \in \Gamma,\ j > i. \tag{2.8}$$

The IFE spaces we plan to construct are formed using basis functions that satisfy the interface jump conditions (2.3).

Since the construction of IFE spaces is also closely related to the standard polynomial FE spaces based upon Lagrangian interpolating polynomials, we introduce the following notation in order to refer to different FE spaces. Standard polynomial FE spaces are referred to as $\mathcal{S}_{h,q}(\Omega)$ where h is the size of the mesh being used and q is the degree of the local basis polynomials on each element. For example, $\mathcal{S}_{h,3}(0,1)$ is the FE space spanned with piecewise cubic Lagrangian interpolating polynomials on the domain $\Omega = (0,1)$. In two dimensions, either triangular or rectangular elements may be used. If it is not clear from the context whether triangular or rectangular elements are being used, then the standard FE spaces are referred to as either $\mathcal{S}_{h,q,\triangle}(\Omega)$ or $\mathcal{S}_{h,q,\square}(\Omega)$ respectively. With rectangular elements, q refers to the order of the polynomial pieces. For example, if $q = 1$ with rectangular elements, then the space is built upon piecewise bi-linear basis functions. Then $q = 2$ refers to biquadratic basis functions and so forth. If it is clear which domain is being used then often the FE space may simply be referred to as $\mathcal{S}_{h,q}$.

2.1 Linear Immersed Finite Element Spaces

In this section, we describe the linear IFE space $\tilde{\mathcal{S}}_{h,1}(0,1)$ along with the basic properties of the linear IFE basis functions. This IFE space was originally developed by Li [40]. For details about the linear IFE space in two dimensions with triangular elements $\mathcal{S}_{h,1,\triangle}$ and bilinear IFE space in two dimensions with rectangular elements $\mathcal{S}_{h,1,\square}$ the reader is referred to Lin, Lin, et. al. [20, 44]. The idea behind these spaces is to construct basis functions that satisfy the interface conditions (2.3) because of the discontinuous coefficient p in the elliptic interface problem (2.1).

2.1.1 Linear IFE Spaces in One Dimension

We start by considering linear IFE spaces in one dimension where we let Ω be the typical interval $(0,1)$. The interface Γ consists of just one point $x = \alpha \in (0,1)$. Then the interface problem (2.1) becomes

$$\begin{array}{rcl} -\big(p(x)u'(x)\big)' & = & f(x),\ x \in \Omega = (0,1), \\ u(0) = u(1) & = & 0, \end{array} \tag{2.9}$$

where the coefficient p is defined by

$$p(x) = \begin{cases} p_1, & x < \alpha, \\ p_2, & x \geq \alpha, \end{cases} \tag{2.10}$$

and $p_i,\ i = 1, 2$, are positive constants. If u satisfies (2.9), then it must also satisfy the interface conditions (2.3) which in one dimension are simply

$$\begin{array}{rcl} \big[u\big]_\alpha & = & u(\alpha+) - u(\alpha-) = 0, \\ \big[pu'\big]_\alpha & = & p_2u'(\alpha+) - p_1u'(\alpha-) = 0. \end{array} \tag{2.11}$$

All the discussions here can be readily extended to the case in which the interface Γ consists of multiple interface points and the coefficient function $p(x)$ is a piecewise constant function.

Now we are ready to construct the linear IFE space which is related to the elliptic interface problem (2.9). Similar to the standard FE space, building an IFE space requires several steps. First, an appropriate partition, $\mathcal{T}_h$, needs to be defined upon the domain such that the partition itself is separated into sub-domains which we call elements. Then local basis functions are defined on each element. Finally, global basis functions are defined over the whole domain by defining the global basis functions piecewise, using the local basis functions on each element. Each global basis function has local support. A common finite element space in one dimension is constructed using Lagrangian interpolating polynomials on each

element as the local basis functions. The linear IFE space that is described here is a modification of the standard FE space that uses linear Lagrangian interpolating polynomials on the each element as the local basis functions.

First, without loss of generality, we introduce a uniform mesh onto the typical domain $\Omega = (0, 1)$ with the nodes $x_i = ih$ for $i = 0, 1, \ldots, N$ so that $x_0 = 0$, $x_N = 1$ and $h = 1/N$. The elements of the domain are the intervals defined by these nodes. So the k^{th} element is $e_k = [x_{k-1}, x_k]$. Then the partition of Ω is

$$\mathcal{T}_h = \bigcup_{i=0}^{N} e_k. \tag{2.12}$$

The partition $\mathcal{T}_h$ that we form above is obviously independent of the interface α, however, to be able to construct the linear IFE space, we need to distinguish between those elements that intersect with the interface and those that do not. The following definition classifies the elements in the partition $\mathcal{T}_h$ into two types.

Definition 2.1.1. *An element $e_k = [x_{k-1}, x_k] \in \mathcal{T}_h$ is called an **interface element** if $x_{k-1} < \alpha < x_k$. Otherwise e_k is called a **non-interface element**.*

Consider the k^{th} element $e_k = [x_{k-1}, x_k]$. First we introduce the following notation for local nodes within the element e_k:

$$\begin{aligned} t_{k,1} &= x_{k-1}, \\ t_{k,2} &= x_k. \end{aligned} \tag{2.13}$$

Then if e_k is a non-interface element, the following standard local linear FE basis functions can be used in this element:

$$\phi_{k,i}(x) = \frac{x - t_{k,j}}{t_{k,i} - t_{k,j}} \quad \text{for } i, j = 1, 2 \text{ and } i \neq j. \tag{2.14}$$

These basis functions $\phi_{k,i}(x),\ i = 1, 2$ satisfy the following nodal value specifications:

$$\phi_{k,i}(t_{k,j}) = \begin{cases} 1, & \text{if } i = j, \\ 0, & \text{if } i \neq j. \end{cases} \tag{2.15}$$

On the other hand, if e_k is an interface element, then the local linear IFE basis functions are required to not only satisfy the nodal value specifications (2.15), but they also need to satisfy the interface conditions (2.11). All together, (2.15) and (2.11) give four conditions that must be satisfied by the local basis function $\phi_{k,i}(x)$. Thus, on the element e_k, the linear basis function $\phi_{k,i}$ is constructed as a piecewise linear function as follows:

$$\phi_{k,i}(x) = \begin{cases} a_1 x + a_0, & \text{if } x < \alpha, \\ b_1 x + b_0, & \text{if } x \geq \alpha, \end{cases} \tag{2.16}$$

which has four coefficients to be determined. This leads us to the following theorem about the local linear IFE basis functions $\phi_{k,i}(x)$ for $i = 1, 2$.

Theorem 2.1.2. *The local linear IFE basis functions, $\phi_{k,i}(x)$, $i = 1, 2$, are uniquely defined.*

Proof: Let $e_k \in \mathcal{T}_h$ be an interface element. Since the local linear IFE basis function $\phi_{k,1}$ is required to satisfy the conditions (2.15) and (2.11) then the coefficients of $\phi_{k,1}(x)$ satisfy the following system of equations:

$$\begin{pmatrix} t_{k,1} & 1 & 0 & 0 \\ 0 & 0 & t_{k,2} & 1 \\ \alpha & 1 & -\alpha & -1 \\ p_1 & 0 & -p_2 & 0 \end{pmatrix} \begin{pmatrix} a_1 \\ a_0 \\ b_1 \\ b_0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}. \tag{2.17}$$

The determinant of the coefficient matrix is given by

$$p_1(t_{k,2} - \alpha) + p_2(\alpha - t_{k,1}), \tag{2.18}$$

which is strictly greater than zero since $p_i > 0$ for $i = 1, 2$ and $t_{k,1} < \alpha < t_{k,2}$. Hence the coefficient matrix is nonsingular and the system of equations (2.17) has a unique solution. Similar arguments can be applied to $\phi_{k,2}$ since its coefficients also satisfy the system (2.17) but with a right hand side vector of $(0, 1, 0, 0)^T$. Therefore $\phi_{k,i}(x)$ for $i = 1, 2$ are uniquely determined. ■

The proof of this theorem gives a specific way for defining the local linear IFE basis functions on the interface element e_k. Solving the system of equations (2.17) in Theorem 2.1.2 gives the following definitions for $\phi_{k,i}(x)$ for $i = 1, 2$:

$$\phi_{k,1}(x) = \begin{cases} \dfrac{p_1(t_{k,2} - \alpha) + p_2(\alpha - x)}{k(\alpha)}, & \text{if } x < \alpha, \\ \dfrac{p_1(t_{k,2} - x)}{k(\alpha)}, & \text{if } x \geq \alpha, \end{cases} \tag{2.19}$$

and

$$\phi_{k,2}(x) = \begin{cases} \dfrac{p_2(x - t_{k,1})}{k(\alpha)}, & \text{if } x < \alpha, \\ \dfrac{p_1(x - \alpha) + p_2(\alpha - t_{k,1})}{k(\alpha)}, & \text{if } x \geq \alpha, \end{cases} \tag{2.20}$$

where $k(\alpha)$ is defined by

$$k(\alpha) = p_1(t_{k,2} - \alpha) + p_2(\alpha - t_{k,1}). \tag{2.21}$$

It should be pointed out that local basis functions are only defined in this way on interface elements. In this one dimensional case, where the location of the interface is described by α, these alternate definitions only apply on one interface element e_k where $x_{k-1} < \alpha < x_k$. The construction can be readily extended to handle the case in which the interface appears at

multiple places. For example, if an element contains more than one interface, then the local linear IFE basis function is defined by satisfying nodal value specifications (2.15) and the interface continuity conditions (2.11) at all of the interfaces within the element. Otherwise, if an element only contains one interface, then the above definitions in (2.19) and (2.20) can be used for the local immersed basis functions.

Now that the definitions for the linear immersed basis functions $\phi_{k,i}(x)$ for $i = 1, 2$ are known, we have the following theorem about their basic properties.

Theorem 2.1.3. *For the local linear IFE basis functions defined by (2.19) and (2.20), we have*

1. *The local linear IFE basis functions $\phi_{k,i}$ satisfy a partition of unity so that*

$$\phi_{k,1}(x) + \phi_{k,2}(x) \equiv 1 \tag{2.22}$$

   for all x.

2. *The local linear IFE basis function $\phi_{k,i}(x),\ i = 1, 2$ is consistent with the standard local linear FE basis function $\dfrac{x - t_{k,j}}{t_{k,i} - t_{k,j}}$ in the following ways:*

   (a) $\lim\limits_{\alpha \to t_{k,1}} \phi_{k,i}(x) = \dfrac{x - t_{k,j}}{t_{k,i} - t_{k,j}}$

   (b) $\lim\limits_{\alpha \to t_{k,2}} \phi_{k,i}(x) = \dfrac{x - t_{k,j}}{t_{k,i} - t_{k,j}}$

   (c) $\lim\limits_{p_1 \to p_2} \phi_{k,i}(x) = \dfrac{x - t_{k,j}}{t_{k,i} - t_{k,j}}$

   where $i = 1, 2$, $j = 1, 2$ and $i \neq j$.

Proof:

1. This result follows by simply adding together the two basis functions and then cancelling terms.

2. These three results can be seen by investigating the behavior of the denominator, $k(\alpha)$, which is defined in (2.21).

$$\begin{aligned} \lim_{\alpha \to t_{k,1}} k(\alpha) &= \lim_{\alpha \to t_{k,1}} p_1(t_{k,2} - \alpha) + p_2(\alpha - t_{k,1}) \\ &= p_1(t_{k,2} - t_{k,1}) \end{aligned} . \tag{2.23}$$

Similarly it can be seen that

$$\begin{aligned} \lim_{\alpha \to t_{k,2}} k(\alpha) &= p_2(t_{k,2} - t_{k,1}) \\ \lim_{p_1 \to p_2} k(\alpha) &= p_2(t_{k,2} - t_{k,1}). \end{aligned} \tag{2.24}$$

Then substituting these results into the appropriate limits for $\phi_{k,i}$ with $i = 1$ or $i = 2$ gives the result. ■

Remark 2.1.4. *Part 2 of this theorem can be restated in the following way: if the coefficients p_1 and p_2 are equal or the interface occurs at the boundary of the interface element then the local linear immersed basis functions become the standard local linear FE basis functions.*

2.1.2 Sensitivities of Linear IFE Basis Functions in One Dimension

Since the basis functions in the linear IFE space are defined in terms of the location of the interface, α, it is natural to investigate how sensitive the basis functions are to the location of the interface. This can be accomplished by studying the sensitivities of the basis functions which are simply their derivatives with respect to α. Recall that a local linear IFE basis function on the interface element $e_k = [x_{k-1}, x_k]$ has the general form

$$\phi_{k,i}(x) = \begin{cases} a_1 x + a_0, & x \leq \alpha, \\ b_1 x + b_0, & x > \alpha, \end{cases} \tag{2.25}$$

where a_1, a_0, b_1 and b_0 are coefficients that depend upon α. Then determining the sensitivity of $\phi_{k,i}(x)$ to the interface at $x = \alpha$ is equivalent to estimating the magnitudes of the derivatives of the coefficients with respect to α.

Theorem 2.1.5. *If the basis function $\phi_{k,i}(x),\ i = 1, 2$ has the general form (2.16) then the first derivative of the coefficients of $\phi_{k,i}(x)$ have the following bounds:*

$$\left|\frac{\partial a_m}{\partial \alpha}\right| = \mathcal{O}(h^{-2}) \text{ for } \ m = 0, 1 \qquad \text{and} \qquad \left|\frac{\partial b_n}{\partial \alpha}\right| = \mathcal{O}(h^{-2}) \text{ for } \ n = 0, 1, \tag{2.26}$$

for $\alpha \in e_k$ and $h = t_{k,2} - t_{k,1}$.

Proof: First we notice that it is sufficient to consider only the coefficients of the basis function $\phi_{k,1}$ since Theorem 2.1.3 asserts that $\phi_{k,1} + \phi_{k,2} \equiv 1$ and hence $\phi_{k,2} = 1 - \phi_{k,1}$. Then

$$\frac{\partial \phi_{k,1}}{\partial \alpha} = -\frac{\partial \phi_{k,2}}{\partial \alpha}, \tag{2.27}$$

and determining the sensitivities of the coefficients $\phi_{k,1}$ is equivalent to determining the sensitivities of the coefficients of $\phi_{k,2}$.

Then let's consider the basis function $\phi_{k,1}(x)$ that has the general form

$$\phi_{k,1}(x) = \begin{cases} a_1 x + a_0, & \text{if } x < \alpha, \\ b_1 x + b_0, & \text{if } x \geq \alpha. \end{cases} \tag{2.28}$$

Recall that the coefficients of $\phi_{k,1}$ satisfy the system (2.17) from Theorem 2.1.2. Then $\phi_{k,1}$ is defined as in (2.19) and so

$$\begin{array}{rcl} a_1 &=& -p_2/k(\alpha), \\ a_0 &=& (p_1(t_{k,2} - \alpha) + p_2\alpha)/k(\alpha), \\ b_1 &=& -p_1/k(\alpha), \\ b_0 &=& p_1 t_{k,2}/k(\alpha), \end{array} \tag{2.29}$$

where $k(\alpha)$ is defined in (2.21). It follows that

$$\begin{array}{rcl} \dfrac{d}{d\alpha}(a_1) &=& p_2(p_2 - p_1)k(\alpha)^{-2}, \\ \dfrac{d}{d\alpha}(a_0) &=& -p_2 t_{k,1}(p_2 - p_1)k(\alpha)^{-2}, \\ \dfrac{d}{d\alpha}(b_1) &=& p_1(p_2 - p_1)k(\alpha)^{-2}, \\ \dfrac{d}{d\alpha}(b_0) &=& -p_1 t_{k,2}(p_2 - p_1)k(\alpha)^{-2}. \end{array} \tag{2.30}$$

So each derivative has the form $C(p_2 - p_1)k(\alpha)^{-2}$ where C is a generic constant that depends upon p_1, p_2, $t_{k,1}$ and $t_{k,2}$. The term $k(\alpha)$ however is a linear function in terms of α and so it attains its minimum and maximum values at the endpoints of the interval $[t_{k,1}, t_{k,2}]$. Then evaluating $k(\alpha)$ at the endpoints of e_k we have

$$k(t_{k,1}) = p_1(t_{k,2} - t_{k,1}) = p_1 h, \tag{2.31}$$

and

$$k(t_{k,2}) = p_2(t_{k,2} - t_{k,1}) = p_2 h. \tag{2.32}$$

Therefore since $p_1 h > 0$ and $p_2 h > 0$, $k(\alpha) > 0$ for $\alpha \in e_k$. Consequently each sensitivity is

bounded away from zero. It then follows that we have

$$\begin{aligned}
\left|\frac{d}{d\alpha}(a_1)\right| &\leq p_2 P h^{-2}, \\
\left|\frac{d}{d\alpha}(a_0)\right| &\leq p_2 t_{k,1} P h^{-2}, \\
\left|\frac{d}{d\alpha}(b_1)\right| &\leq p_1 P h^{-2}, \\
\left|\frac{d}{d\alpha}(b_0)\right| &\leq p_1 t_{k,2} P h^{-2},
\end{aligned} \tag{2.33}$$

where

$$P = (p_2 - p_1) \max\left\{p_1^{-2}, p_2^{-2}\right\}, \tag{2.34}$$

and hence the theorem is proved. ■

Moreover, each of the derivatives of the coefficients is more than bounded away from zero. Due to the definition of $k(\alpha)$ we have

$$\begin{aligned}
\left|\frac{d}{d\alpha}(a_1)\right| &\geq p_2 \hat{P} h^{-2}, \\
\left|\frac{d}{d\alpha}(a_0)\right| &\geq p_2 t_{k,1} \hat{P} h^{-2}, \\
\left|\frac{d}{d\alpha}(b_1)\right| &\geq p_1 \hat{P} h^{-2}, \\
\left|\frac{d}{d\alpha}(b_0)\right| &\geq p_1 t_{k,2} \hat{P} h^{-2},
\end{aligned} \tag{2.35}$$

where

$$\hat{P} = (p_2 - p_1) \min\left\{p_1^{-2}, p_2^{-2}\right\}. \tag{2.36}$$

2.2 Quadratic Immersed Finite Element Spaces

In this section, we discuss three different types of quadratic IFE spaces whose local nodal basis functions are developed according to the following ideas: (1) hierarchical basis functions that have no extra condition explicitly imposed on the interface; (2) basis functions with extra continuity requirements; (3) basis functions which use a form of local refinement. Each of these basis functions is discussed in detail in Sections 2.2.1 and 2.2.2 below. First, the

three types of quadratic immersed basis functions are described in one dimension. Then, these definitions are extended to two dimensions to form biquadratic IFE spaces. There are infinitely many ways to define the quadratic IFE nodal basis functions that can satisfy the nodal value specification and interface jump conditions. The three IFE spaces that we present here demonstrate what happens if no extra condition is imposed and how to add extra conditions suitably so that the local immersed quadratic basis functions can be well-defined.

There are two unifying features for these three quadratic IFE spaces. One is that alternative basis functions are only used on the so-called interface elements for which the interface Γ intersects their interior. Interface elements are defined more rigorously below in Sections 2.2.1 and 2.2.2 for domains which are either one dimensional or two dimensional. On non-interface elements, these IFE spaces just use the standard quadratic nodal basis functions. Another unifying feature is that the basis functions in these IFE spaces can satisfy the interface jump conditions (2.11) because these spaces are developed for solving interface problems whose solutions are required to satisfy (2.11).

2.2.1 Quadratic IFE spaces in One Dimension

To start, we consider one dimensional quadratic IFE spaces. To facilitate the discussion, we first introduce some notations used in this and later sections. Let $\mathcal{T}_h = \bigcup_{k=1}^{n} e_k$ be a partition of $\Omega = (0, 1)$, where elements $e_k = [x_{2(k-1)}, x_{2k}]$ are intervals in Ω that are formed by node points

$$0 = x_0 < x_1 < x_2 < x_3 < x_4 < \cdots < x_{2n-1} < x_{2n} = 1,$$

such that

$$x_{2k-1} = \frac{x_{2k} - x_{2k-2}}{2}, \quad k = 1, 2, \cdots n.$$

Without loss of generality, we consider a uniform partition such that $x_{2k} = x_0 + kh$ where $x_0 = 0$ (the left endpoint) and $h = 1/n$. The local quadratic nodes on the k^{th} element e_k are denoted by $t_{k,1} = x_{2(k-1)}, t_{k,2} = x_{2k-1}$ and $t_{k,3} = x_{2k}$.

We categorize all the elements into two classes as follows.

Definition 2.2.1. *An element e_k of the partition $\mathcal{T}_h$ is called an **interface element** if the interface point α is in the interior of e_k, i.e., $x_{2(k-1)} < \alpha < x_{2k}$. Any element which is not an interface element is a **non-interface element**.*

If the k^{th} element e_k is a non-interface element, then we use the three standard Lagrange type quadratic FE local nodal functions $\phi_{k,i}(x), i = 1, 2, 3$ such that

$$\phi_{k,i}(t_{k,j}) = \begin{cases} 1, & \text{if } i = j, \\ 0, & \text{if } i \neq j, \end{cases} \tag{2.37}$$

for $j = 1, 2, 3$. The formulas for the standard local nodal basis functions $\phi_{k,i}(x),\ i = 1, 2, 3$ are easy to derive since each one is the unique quadratic Lagrange interpolant of the nodal values (2.37). The formulas for these local nodal basis functions are:

$$\phi_{k,1}(x) = \left(\frac{x - t_{k,2}}{t_{k,1} - t_{k,2}}\right)\left(\frac{x - t_{k,3}}{t_{k,1} - t_{k,3}}\right), \tag{2.38}$$

$$\phi_{k,2}(x) = \left(\frac{x - t_{k,1}}{t_{k,2} - t_{k,1}}\right)\left(\frac{x - t_{k,3}}{t_{k,2} - t_{k,3}}\right), \tag{2.39}$$

and

$$\phi_{k,3}(x) = \left(\frac{x - t_{k,1}}{t_{k,3} - t_{k,1}}\right)\left(\frac{x - t_{k,2}}{t_{k,3} - t_{k,2}}\right). \tag{2.40}$$

In the next three subsections, we give three different approaches for constructing local nodal basis functions in an interface element e_k that can satisfy the same nodal value specifications (2.37) and the interface jump conditions (2.11). Then for each node $x_i, i = 0, 1, \cdots, 2n$, we use these local basis functions to form a global Lagrange type basis function, which is used to span our IFE spaces as usual.

An 1-D Hierarchical Quadratic IFE Space

Following the same idea used for the standard quadratic local nodal basis functions, see [9, 31] for example, we demonstrate that the quadratic IFE local nodal basis functions in an interface element e_k can be constructed hierarchically by multiplying together two linear IFE interpolation functions (to be defined in the next paragraph) which are a generalization of the linear IFE local nodal basis functions, see [40] and Section 2.1. This approach can yield quadratic IFE spaces without using any extra conditions and the idea can be repeatedly used to generate IFE spaces of degree higher than two.

A quadratic local nodal basis function to be constructed in this way (referred to as $\tilde{\phi}_{k,i}$) has two roots which occur at two of the nodes in an interface element e_k, and these two nodes may or may not be on both sides of the interface point α. On the other hand, each of the two roots of this quadratic local nodal basis function needs to also be the root of one of the two linear IFE interpolation functions multiplied together. Therefore, we first need to modify the linear IFE local nodal basis functions [40] such that its nodal values can be specified at

two points that may or may not be on both sides of the interface point. Specifically, for each pair of indices (i, j) with $i = 1, 2, 3, j = 1, 2, 3$ but $i \neq j$, we form a linear IFE interpolation function that is a piecewise linear function of the following form

$$l_{i,j}(x) = \begin{cases} a_1 x + a_0, & \text{if } x < \alpha, \\ b_1 x + b_0, & \text{if } x \geq \alpha, \end{cases} \tag{2.41}$$

whose coefficients a_0, a_1, b_0, b_1 are chosen such that $l_{i,j}(x)$ can satisfy the nodal value specification and interface jump conditions as follows:

$$\begin{cases} l_{i,j}(t_{k,i}) = 1, \quad l_{i,j}(t_{k,j}) = 0, \text{ (nodal value specification),} \\ [l_{i,j}]_{x=\alpha} = 0, \quad [pl'_{i,j}]_{x=\alpha} = 0, \text{ (interface jump conditions).} \end{cases} \tag{2.42}$$

Theorem 2.2.2. *The linear immersed interpolating function $l_{i,j}(x)$ $(i \neq j)$ is uniquely determined.*

Proof: When α is not in the interval that is formed by $t_{k,i}$ and $t_{k,j}$, say $t_{k,i} < t_{k,j}$ and $\alpha < t_{k,i}$, then (2.42) leads to the following linear system about the coefficients a_0, a_1, b_0 and b_1:

$$\begin{pmatrix} 0 & 0 & t_{k,i} & 1 \\ 0 & 0 & t_{k,j} & 1 \\ \alpha & 1 & -\alpha & -1 \\ p_1 & 0 & -p_2 & 0 \end{pmatrix} \begin{pmatrix} a_1 \\ a_0 \\ b_1 \\ b_0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}. \tag{2.43}$$

Since the coefficient matrix has a determinant of $-p_1(t_{k,i} - t_{k,j})$ that is always non-zero, the system always has a unique solution. Similar arguments apply to the situations where $t_{k,i} < \alpha < t_{k,j}$ and $t_{k,i} < t_{k,j} < \alpha$. Hence the linear IFE interpolation functions are uniquely determined. ■

As the proof of Theorem 2.2.2 shows, if $t_{k,i} < t_{k,j}$ (where $i < j$) and $\alpha < t_{k,i}$ then $l_{i,j}(x)$ satisfies the system of equations (2.43). Solving this system for the coefficients a_0, a_1, b_0 and b_1 gives the following formula for the linear immersed interpolating function $l_{i,j}(x)$:

$$l_{i,j}(x) = \begin{cases} \dfrac{\alpha(p_1 - p_2) - p_1 t_{k,j} + p_2 x}{p_1(t_{k,i} - t_{k,j})}, & \text{if } x < \alpha, \\ \dfrac{t_{k,j} - x}{t_{k,j} - t_{k,i}}, & \text{if } x \geq \alpha. \end{cases} \tag{2.44}$$

The formulas of the immersed interpolating functions $l_{i,j}(x)$ for the cases when $t_{k,i} < \alpha < t_{k,j}$ and $t_{k,j} < \alpha$ can be derived in a similar fashion. Definitions for $l_{j,i}(x)$ (where $i < j$) may be handled by the following lemma.

Lemma 2.2.3. *The linear immersed interpolating functions $l_{i,j}(x)$ and $l_{j,i}(x)$ (where $i < j$) are related by the statement $l_{j,i}(x) = 1 - l_{i,j}(x)$ for all values of x.*

Proof. We need to show that $l_{i,j}(x) + l_{j,i}(x) \equiv 1$ where $l_{i,j}(x)$ has the general form

$$l_{i,j}(x) = \begin{cases} a_1 x + a_0, & \text{if } x < \alpha, \\ b_0 x + b_0, & \text{if } x \geq \alpha, \end{cases} \tag{2.45}$$

and $l_{j,i}(x)$ has the general form

$$l_{j,i}(x) = \begin{cases} \tilde{a}_1 x + \tilde{a}_0, & \text{if } x < \alpha, \\ \tilde{b}_0 x + \tilde{b}_0, & \text{if } x \geq \alpha. \end{cases} \tag{2.46}$$

Let $\alpha < t_{k,i}$. Then the function $f(x)$ which is defined by $f(x) = l_{i,j}(x) + l_{j,i}(x)$ has the general form

$$f(x) = \begin{cases} (a_1 + \tilde{a}_1)x + (a_0 + \tilde{a}_0) & \text{if } x < \alpha, \\ (b_0 + \tilde{b}_0)x + (b_0 + \tilde{b}_0) & \text{if } x \geq \alpha. \end{cases} \tag{2.47}$$

Since both $l_{i,j}(x)$ and $l_{j,i}(x)$ satisfy the interface conditions (2.11) then $f(x)$ also satisfies the interface conditions. Then $f(x)$ satisfies the following system of equations:

$$\begin{pmatrix} 0 & 0 & t_{k,i} & 1 \\ 0 & 0 & t_{k,j} & 1 \\ \alpha & 1 & -\alpha & -1 \\ p_1 & 0 & -p_2 & 0 \end{pmatrix} \begin{pmatrix} a_1 + \tilde{a}_1 \\ a_0 + \tilde{a}_0 \\ b_1 + \tilde{b}_1 \\ b_0 + \tilde{b}_0 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix}. \tag{2.48}$$

We see from the proof of Theorem 2.2.2 that this system has a unique solution since the coefficient matrix is nonsingular. Then solving this system shows that the coefficients of $f(x)$ must satisfy

$$a_1 + \tilde{a}_1 = b_1 + \tilde{b}_1 = 0, \tag{2.49}$$

and

$$a_0 + \tilde{a}_0 = b_0 + \tilde{b}_0 = 1. \tag{2.50}$$

Hence, $f(x) \equiv 1$ and so $l_{j,i}(x) = 1 - l_{i,j}(x)$. Similar arguments hold in the cases where $t_{k,i} < \alpha < t_{k,j}$ or $t_{k,j} < \alpha$. ■

Remark 2.2.4. *The linear immersed interpolation functions $l_{1,3}(x)$ and $l_{3,1}(x)$ are the usual linear IFE local nodal basis functions on the interface element e_k.*

The six different immersed interpolation functions $l_{i,j}(x)$ on the element e_k are seen in Figure 2.1 for the case where $t_{k,1} < \alpha < t_{k,2}$. Now, we can define the three quadratic IFE local nodal basis functions as follows:

$$\tilde{\phi}_{k,1}(x) = l_{1,2}(x)l_{1,3}(x), \quad \tilde{\phi}_{k,2}(x) = l_{2,1}(x)l_{2,3}(x), \quad \tilde{\phi}_{k,3}(x) = l_{3,1}(x)l_{3,2}(x). \tag{2.51}$$

First, it can be easily verified that these functions satisfy the local nodal value specifications (2.37). In addition, the following theorem shows that these functions also satisfy the interface jump conditions (2.11).

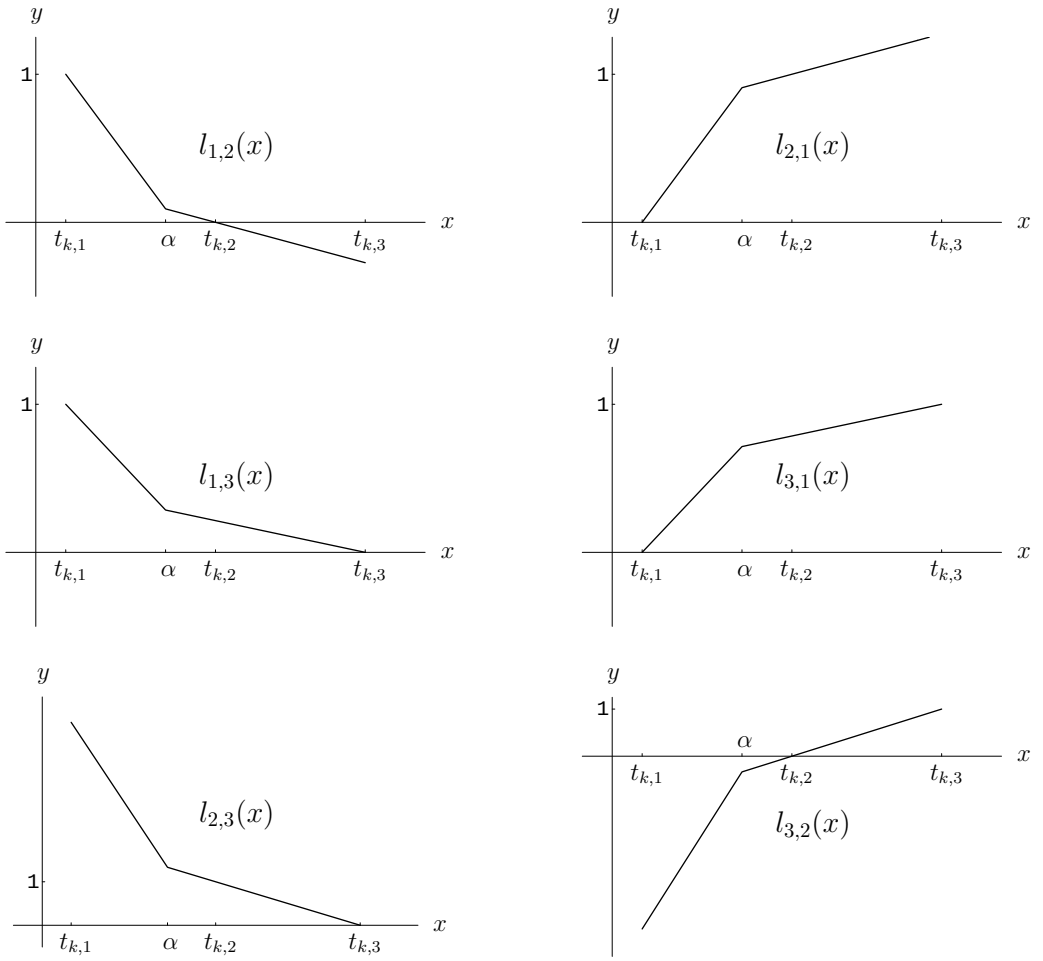


Figure 2.1: The six linear immersed interpolating functions on the element e_k.

Theorem 2.2.5. *The quadratic IFE local nodal basis functions $\tilde{\phi}_{k,i}(x), i = 1, 2, 3$ that we introduce above have the following properties:*

1. *They satisfy the interface jump conditions (2.11).*
2. $\sum_{i=1}^{3} \tilde{\phi}_{k,i}(x) \equiv 1$ *for all* $x \in e_k$.
3. *The local IFE nodal basis function $\tilde{\phi}_{k,i}(x),\ i = 1, 2, 3$ is consistent with the standard local FE nodal basis function $\phi_{k,i}(x), i = 1, 2, 3$ in the following way:*
   (a) $\lim_{\alpha \to t_{k,1}} \tilde{\phi}_{k,i}(x) = \phi_{k,i}(x)$ *for* $i = 1, 2, 3$
   (b) $\lim_{\alpha \to t_{k,3}} \tilde{\phi}_{k,i}(x) = \phi_{k,i}(x)$ *for* $i = 1, 2, 3$
   (c) *If* $p_1 = p_2 > 0$ *then* $\tilde{\phi}_{k,i}(x) = \phi_{k,i}(x)$.

Proof.

1. The continuity of $\tilde{\phi}_{k,i}(x), i = 1, 2, 3$ at $x = \alpha$ follows obviously from the continuity of $l_{i,j}(x), i, j = 1, 2, 3$. As for the continuity of the flux at $x = \alpha$, let us discuss $\tilde{\phi}_{k,1}(x)$. First, we note $\tilde{\phi}'_{k,1}(x) = l'_{1,2}(x)l_{1,3}(x) + l_{1,2}(x)l'_{1,3}(x)$. Then

$$\begin{aligned}
\left[p\tilde{\phi}'_{k,1}(\alpha)\right] &= p_1\tilde{\phi}_{k,1}(\alpha-) - p_2\tilde{\phi}_{k,1}(\alpha+) \\
&= p_1(l'_{1,2}(\alpha-)l_{1,3}(\alpha-) + l_{1,2}(\alpha-)l'_{1,3}(\alpha-)) \\
&\quad - p_2(l'_{1,2}(\alpha+)l_{1,3}(\alpha+) + l_{1,2}(\alpha+)l'_{1,3}(\alpha+)) \\
&= p_1(l'_{1,2}(\alpha-)l_{1,3}(\alpha) + l_{1,2}(\alpha)l'_{1,3}(\alpha-)) \\
&\quad - p_2(l'_{1,2}(\alpha+)l_{1,3}(\alpha) + l_{1,2}(\alpha)l'_{1,3}(\alpha-)) \\
&= \left[pl'_{1,2}(\alpha)\right] l_{1,3}(\alpha) + \left[pl'_{1,3}(\alpha)\right] l_{1,2}(\alpha) = 0,
\end{aligned}$$

   which indicates that $\tilde{\phi}_{k,1}(x)$ satisfies the flux jump condition at the interface. The same argument readily applies to both $\tilde{\phi}_{k,2}(x)$ and $\tilde{\phi}_{k,3}(x)$.

2. Let $f(x) = \tilde{\phi}_{k,1}(x) + \tilde{\phi}_{k,2}(x) + \tilde{\phi}_{k,3}(x)$. Suppose that $t_{k,1} < \alpha < t_{k,2}$ and that $x < \alpha$ then $\tilde{\phi}_{k,i}$ for $i = 1, 2, 3$ are defined by

$$\tilde{\phi}_{k,1}(x) = \frac{(R + p_2x - p_1t_{k,2})(R + p_2x - p_1t_{k,3})}{(R + p_2t_{k,1} - p_1t_{k,2})(R + p_2t_{k,1} - p_1t_{k,3})}\ , \tag{2.52}$$

$$\tilde{\phi}_{k,2}(x) = -\frac{p_2(x-t_{k,1})(R+p_2x-p_1t_{k,3})}{p_1(R+p_2t_{k,1}-p_1t_{k,2})(t_{k,2}-t_{k,3})}\ , \tag{2.53}$$

and

$$\tilde{\phi}_{k,3}(x) = \frac{p_2(x-t_{k,1})(R+p_2x-p_1t_{k,2})}{p_1(t_{k,2}-t_{k,3})(R+p_2t_{k,1}-p_1t_{k,3})}, \tag{2.54}$$

where

$$R = \alpha(p_1-p_2). \tag{2.55}$$

Putting terms over a common denominator of

$$\tilde{Q}_1 = p_1(t_{k,2}-t_{k,3})(R-p_1t_{k,2}+p_2t_{k,1})(R-p_1t_{k,3}+p_2t_{k,1})\ , \tag{2.56}$$

allows $f(x)$ to be rewritten as

$$\begin{aligned} f(x) = \big(&p_1(t_{k,2}-t_{k,3})(R-p_1t_{k,2}+p_2x)(R-p_1t_{k,3}+p_2x) \\ &-p_2(x-t_{k,1})(R-p_1t_{k,3}+p_2t_{k,1})(R-p_1t_{k,3}+p_2x) \\ &+p_2(x-t_{k,1})(R-p_1t_{k,2}+p_2t_{k,1})(R-p_1t_{k,2}+p_2x)\big)/\tilde{Q}_1. \end{aligned} \tag{2.57}$$

Since $f(x)$ for $x<\alpha$ is a sum of quadratic functions then it can also be written with the general form of

$$f(x) = \frac{1}{\tilde{Q}_1}\big(C_2x^2+C_1x+C_0\big), \quad \text{if } x<\alpha, \tag{2.58}$$

where C_2, C_1 and C_0 are constants. By expanding the terms in (2.57) it follows that

$$\begin{aligned} C_2 &= p_2^2\big(p_1(t_{k,2}-t_{k,3}) - (R-p_1t_{k,3}+p_2t_{k,1}) + (R-p_1t_{k,2}+p_2t_{k,1})\big) \\ &= 0, \end{aligned}$$

and

$$\begin{aligned} C_1 &= p_2\big(p_1(t_{k,2}-t_{k,3})(R-p_1t_{k,2}) + p_1(t_{k,2}-t_{k,3})(R-p_1t_{k,3}) \\ &\quad +(R-p_1t_{k,3}+p_2t_{k,1})(p_2t_{k,1}) - (R-p_1t_{k,3}+p_2t_{k,1})(R-p_1t_{k,3}) \\ &\quad -(R-p_1t_{k,2}+p_2t_{k,1})(p_2t_{k,1}) + (R-p_1t_{k,2}+p_2t_{k,1})(R-p_1t_{k,2})\big) \\ &= 0, \end{aligned}$$

and so $f(x) = C_0/\tilde{Q}_1$ if $x<\alpha$. However, because of the interpolation properties of $\tilde{\phi}_{k,i}(x)$ for $i=1,2,3$, it follows that $f(t_{k,1})=1$ and so $f(x)\equiv 1$ for $x<\alpha$. It remains to be shown that $f(x)\equiv 1$ for $x\geq\alpha$.

Consider, then, that the function $f(x)$ is a continuous function on the interface element e_k since the basis functions $\tilde{\phi}_{k,i}(x)$ are each continuous on e_k. In particular, this means that the function $f(x)$ also satisfies the interface conditions (2.11). This implies that $f(\alpha) = 1$. In addition, due to the nodal values of the local IFE nodal basis functions $\tilde{\phi}_{k,i}(x)$, it follows that $f(t_{k,2}) = 1$ and $f(t_{k,3}) = 1$. Then since $f(x)$ is a quadratic function on the interval $[\alpha, t_{k,3}]$ it must be the unique polynomial interpolant of degree two or less of the points $(\alpha, 1)$, $(t_{k,2}, 1)$ and $(t_{k,3}, 1)$. Thus $f(x) = 1$ for $x \geq \alpha$. It follows that $f(x) \equiv 1$ for all x on the element e_k. A similar argument can be made for the case where $t_{k,2} < \alpha < t_{k,3}$.

3. When $t_{k,1} < \alpha < t_{k,2}$, the quadratic IFE local nodal basis function $\tilde{\phi}_{k,1}(x)$ has the definition

$$\tilde{\phi}_{k,1}(x) = \begin{cases} \dfrac{(R + p_2 x - p_1 t_{k,2})(R + p_2 x - p_1 t_{k,3})}{(R + p_2 t_{k,1} - p_1 t_{k,2})(R + p_2 t_{k,1} - p_1 t_{k,3})} , & \text{if } x < \alpha, \\ \dfrac{p_1^2 (x - t_{k,2})(x - t_{k,3})}{(R - p_1 t_{k,2} + p_2 t_{k,1})(R - p_1 t_{k,3} + p_2 t_{k,2})} , & \text{if } x \geq \alpha, \end{cases} \tag{2.59}$$

where R is defined in (2.55). When $\alpha \to t_{k,1}$ then the definition of $\tilde{\phi}_{k,1}(x)$ becomes

$$\begin{aligned} \lim_{\alpha \to t_{k,1}} \tilde{\phi}_{k,1}(x) &= \lim_{\alpha \to t_{k,1}} \frac{p_1^2 (x - t_{k,2})(x - t_{k,3})}{(R - p_1 t_{k,2} + p_2 t_{k,1})(R - p_1 t_{k,3} + p_2 t_{k,2})} \\ &= \left(\frac{x - t_{k,2}}{t_{k,1} - t_{k,2}} \right) \left(\frac{x - t_{k,3}}{t_{k,1} - t_{k,3}} \right) = \phi_{k,1}(x) , \end{aligned}$$

which is shown by direct substitution and cancellation. The results for $\lim_{\alpha \to t_{k,3}} \tilde{\phi}_{k,1}(x)$ and $\lim_{p_2 \to p_1} \tilde{\phi}_{k,1}(x)$ can also be shown using direct substitution and cancellation. Similar arguments hold for $\tilde{\phi}_{k,2}(x)$ and $\tilde{\phi}_{k,3}$ and the cases when $t_{k,2} < \alpha < t_{k,3}$. ■

Plots of the local IFE basis functions $\tilde{\phi}_{k,i}(x)$, $i = 1, 2, 3$ are seen in comparison with the standard FE basis functions $\phi_{k,i}(x)$ in Figure 2.2. Plots of the local IFE basis functions $\tilde{\phi}_{k,i}(x)$, $i = 1, 2, 3$ are seen in comparison with the standard FE basis functions $\phi_{k,i}(x)$ and the other IFE basis functions in Figure 2.5 in the last part of this section (see page 50).

We now proceed to construct an IFE space using these local basis functions. First, for each element $e_k \in \mathcal{T}_h$, we let

$$\tilde{\mathcal{S}}_{h,2}(e_k) = \begin{cases} \text{span}\{\phi_{k,i}(x), i = 1, 2, 3\}, & \text{if } e_k \text{ is a non-interface element,} \\ \text{span}\{\tilde{\phi}_{k,i}(x), i = 1, 2, 3\}, & \text{if } e_k \text{ is an interface element.} \end{cases} \tag{2.60}$$

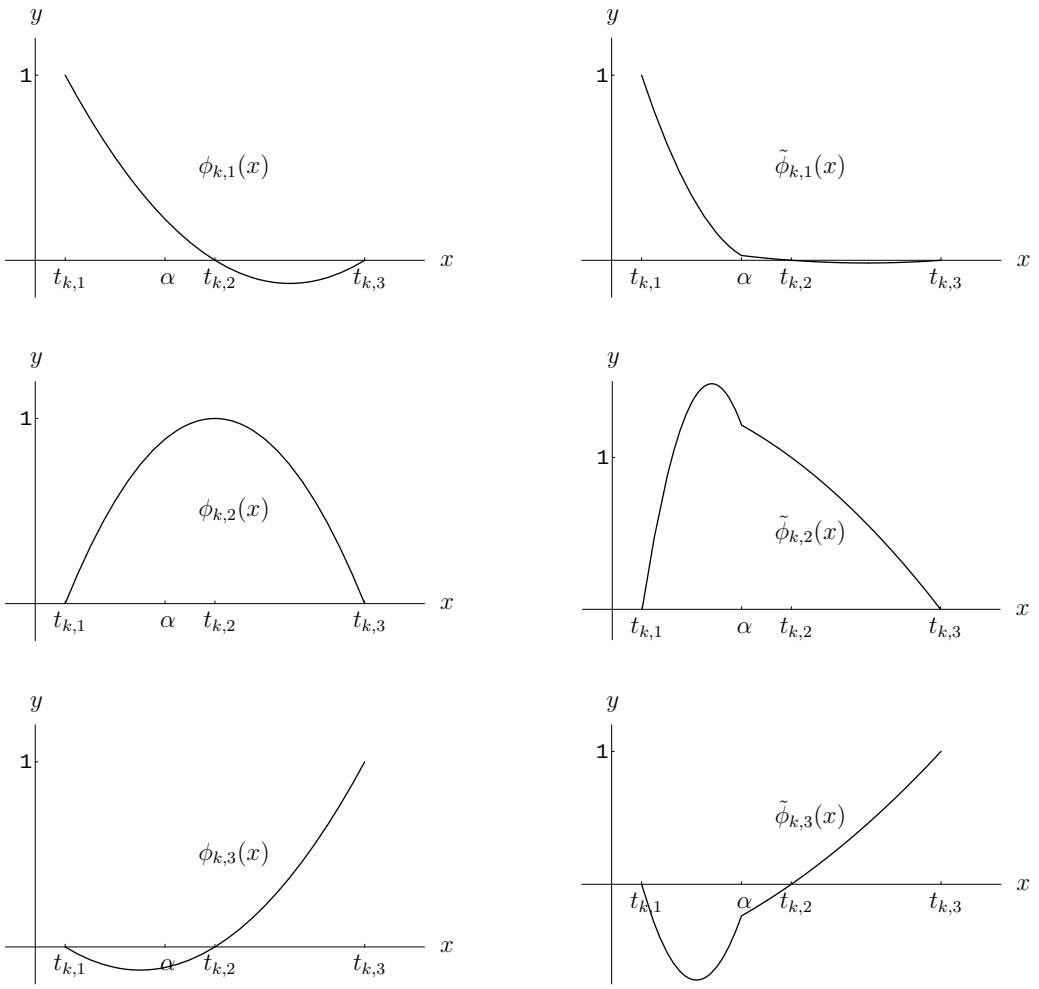


Figure 2.2: Standard quadratic basis functions $\phi_{k,i}(x)$ versus the corresponding immersed basis functions $\tilde{\phi}_{k,i}$ on the element e_k.

Then, for each node $x_i, i = 0, 2, \cdots, 2n$, we let $\tilde{\phi}_i(x)$ be a piecewise quadratic polynomial such that $\tilde{\phi}_i|_{e_k} \in \tilde{\mathcal{S}}_{h,2}(e_k)$ for any $e_k \in \mathcal{T}_h$, and

$$\tilde{\phi}_i(x_j) = \begin{cases} 1, & \text{if } i = j, \\ 0, & \text{if } i \neq j. \end{cases} \tag{2.61}$$

Finally, we can form the IFE space over the whole domain by the hierarchical local nodal basis functions as follows:

$$\tilde{\mathcal{S}}_{h,2}(\Omega) = \text{span}\{\tilde{\phi}_i, i = 0, 1, \cdots, 2n\}. \tag{2.62}$$

An 1-D Quadratic IFE Space with Extra Continuity

The approach that we present in the previous section is not the only way to determine quadratic IFE local nodal basis functions. In this section, following a similar idea as in [40, 41, 42], we form the quadratic IFE local nodal basis functions directly from piecewise quadratic polynomials using an undetermined coefficient technique.

In a typical interface element e_k, the interface point α separates e_k into two subintervals. Then we can use two quadratic polynomials that are defined on each of these subintervals to form a quadratic IFE local nodal basis function with the following form:

$$\bar{\phi}_{k,i}(x) = \begin{cases} a_2x^2 + a_1x + a_0, & \text{if } x < \alpha, \\ b_2x^2 + b_1x + b_0, & \text{if } x \geq \alpha. \end{cases} \tag{2.63}$$

Here, as usual, we need to choose the coefficients a_2, a_1, a_0, b_2, b_1 and b_0 such that the nodal value specifications (2.37) and the interface conditions (2.11) are satisfied. However, these provide only five conditions while $\bar{\phi}_{k,i}(x)$ contains six coefficients that need to be determined. So in order to determine the coefficients uniquely, one extra condition needs to be added. Although any arbitrary condition may be used to provide a sixth condition, a natural condition to impose is another continuity requirement at the interface. By definition, we know that the solution to the one dimensional interface problem (2.9), $u(x)$, and its flux, $p(x)u'(x)$, are both continuous at the interface provided that $f(x)$ has enough continuity. Moreover, it often happens that $(p(x)u'(x))'$ is continuous at the interface as well. For example, this is true when the right hand side function $f(x)$ in (2.9) is continuous over $\Omega = (0, 1)$. We therefore require that the IFE local basis functions satisfy the following extra continuity condition:

$$\left[(p(x)\bar{\phi}'_{k,i}(x))'\right]_{x=\alpha} = 0, \quad i = 1, 2, 3. \tag{2.64}$$

Quadratic IFE local basis functions which are constructed using (2.64) as an extra continuity requirement are referred to with the notation $\bar{\phi}_{k,i}$ in order to distinguish them from the other quadratic IFE local basis functions.

Theorem 2.2.6. *The quadratic IFE local basis functions,* $\bar{\phi}_{k,i}(x), \quad i = 1, 2, 3,$ *are uniquely determined on the interface element* e_k*.*

Proof. Consider the quadratic IFE local basis function $\bar{\phi}_{k,1}(x)$. If $t_{k,1} < \alpha < t_{k,2}$ then the coefficients in (2.63) satisfy the following system of equations:

$$\begin{pmatrix} t_{k,1}^2 & t_{k,1} & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & t_{k,2}^2 & t_{k,2} & 1 \\ 0 & 0 & 0 & t_{k,3}^2 & t_{k,3} & 1 \\ \alpha^2 & \alpha & 1 & -\alpha^2 & -\alpha & -1 \\ 2p_1\alpha & p_1 & 0 & -2p_2\alpha & -p_2 & 0 \\ 2p_1 & 0 & 0 & -2p_2 & 0 & 0 \end{pmatrix} \begin{pmatrix} a_2 \\ a_1 \\ a_0 \\ b_2 \\ b_1 \\ b_0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}. \tag{2.65}$$

The determinant of the coefficient matrix in this system is given by

$$\begin{aligned} \bar{k}_1(\alpha) \;=\; & -\frac{1}{4}p_1 h\big(h^2 + (\beta-1)(-2\alpha^2 \\ & \qquad +3h(\alpha - t_{k,1}) - 2t_{k,1}^2 + 4\alpha t_{k,1})\big), \end{aligned} \tag{2.66}$$

where $h = t_{k,3} - t_{k,1}$ and $\beta = p_2/p_1 > 1$. To show that this system has a unique solution, it suffices to show that this determinant is bounded away from zero on the interval $[t_{k,1}, t_{k,2}]$. The derivative $\bar{k}_1'(\alpha)$ is given by

$$\bar{k}_1'(\alpha) = -\frac{1}{4}p_1 h(\beta - 1)(-4\alpha + 3h + 4t_{k,1}), \tag{2.67}$$

which is a linear function in α that is bounded away from zero on the interval $[t_{k,1}, t_{k,2}]$. Consequently, the minimum and maximum of $\bar{k}_1(\alpha)$ must occur at the endpoints of the interval $[t_{k,1}, t_{k,2}]$. Then since

$$\bar{k}_1(t_{k,1}) = -p_1 h^3/4, \tag{2.68}$$

and

$$\bar{k}_1(t_{k,2}) = -p_1 h^3 \beta/4, \tag{2.69}$$

the determinant $\bar{k}_1(\alpha)$ is bounded away from zero on the interval $[t_{k,1}, t_{k,2}]$ and so this system has a nonsingular coefficient matrix for $t_{k,1} \leq \alpha < t_{k,2}$. Therefore the coefficients of $\bar{\phi}_{k,1}(x)$ can be uniquely determined.

When $t_{k,2} \leq \alpha \leq t_{k,3}$, the coefficients of $\bar{\phi}_{k,1}(x)$ satisfy the system

$$\begin{pmatrix} t_{k,1}^2 & t_{k,1} & 1 & 0 & 0 & 0 \\ t_{k,2}^2 & t_{k,2} & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & t_{k,3}^2 & t_{k,3} & 1 \\ \alpha^2 & \alpha & 1 & -\alpha^2 & -\alpha & -1 \\ 2p_1\alpha & p_1 & 0 & -2p_2\alpha & -p_2 & 0 \\ 2p_1 & 0 & 0 & -2p_2 & 0 & 0 \end{pmatrix} \begin{pmatrix} a_2 \\ a_1 \\ a_0 \\ b_2 \\ b_1 \\ b_0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}. \tag{2.70}$$

Here the determinant of the coefficient matrix is given by

$$\begin{aligned}\bar{k}_2(\alpha) &= -\frac{1}{4}hp_2\big(h^2+(\beta-1)(2\alpha^2 \\ &\qquad +ht_{k,1}+2t_{k,1}^2-\alpha h-4\alpha t_{k,1})\big),\end{aligned} \tag{2.71}$$

where $t_{k,3}-t_{k,1}=h$ and $\beta=p_2/p_1$. We only need to show that the coefficient matrix is nonsingular on the interval $[t_{k,2},t_{k,3}]$. The derivative $\bar{k}_2'(\alpha)$ is given by

$$\bar{k}_2'(\alpha)=-\frac{1}{4}hp_2(\beta-1)(4\alpha-h-4t_{k,1}). \tag{2.72}$$

Notice that $\bar{k}_2'(\alpha)$ is a linear function in α which is bounded away from zero meaning that the maximum and minimum of $\bar{k}_2(\alpha)$ on the interval $[t_{k,2},t_{k,3}]$ must occur at the endpoints of the interval. Consequently since

$$\bar{k}_2(t_{k,2})=-h^3\beta/4, \tag{2.73}$$

and

$$\bar{k}_2(t_{k,3})=-h^3\beta^2/4, \tag{2.74}$$

then $\bar{k}_2(\alpha)$ is bounded away from zero on the interval $[t_{k,2},t_{k,3}]$ and so the coefficient matrix is nonsingular. Then the coefficients of $\bar{\phi}_{k,1}(x)$ are uniquely determined on the interface element e_k and so the quadratic IFE local basis function $\bar{\phi}_{k,1}(x)$ is uniquely determined on the element e_k.

The coefficients for the quadratic IFE local basis functions $\bar{\phi}_{k,2}(x)$ and $\bar{\phi}_{k,3}(x)$ are determined by linear systems with the same coefficient matrix as in (2.65) and (2.70) with the right hand vectors being $(0,1,0,0,0,0)^T$ and $(0,0,1,0,0,0)^T$ respectively. Hence the coefficients of $\bar{\phi}_{k,2}(x)$ and $\bar{\phi}_{k,3}(x)$ are also uniquely determined on the element e_k. Hence we conclude that the local quadratic IFE basis functions $\bar{\phi}_{k,i}(x),\ i=1,2,3$ are uniquely determined on the element e_k. ■

The proof of Theorem 2.2.6 provides a method for constructing the quadratic IFE local basis functions $\bar{\phi}_{k,i}(x)$. For example, if $t_{k,1}<\alpha<t_{k,2}$ then solving the system of equations (2.65) for the coefficients a_2,a_1,a_0,b_2,b_1 and b_0 of $\bar{\phi}_{k,1}(x)$ from (2.63) gives

$$\bar{\phi}_{k,1}(x)=\begin{cases}\dfrac{1}{\bar{Q}_1}\big(p_1(t_{k,2}-t_{k,3})(R(\alpha-t_{k,2}-t_{k,3}) \\ \qquad +p_2x(x-t_{k,2}-t_{k,3})+p_1t_{k,2}t_{k,3})\big)\ , & \text{if } x<\alpha, \\ \dfrac{1}{\bar{Q}_1}\big(p_1^2(x-t_{k,2})(x-t_{k,3})(t_{k,2}-t_{k,3})\big)\ , & \text{if } x\geq\alpha,\end{cases} \tag{2.75}$$

where R is defined in (2.55) and

$$\bar{Q}_1 = p_1(t_{k,2} - t_{k,3})\big(R(\alpha - t_{k,2} - t_{k,3}) + p_2 t_{k,1}(t_{k,1} - t_{k,2} - t_{k,3}) + p_1 t_{k,2} t_{k,3}\big). \tag{2.76}$$

Similarly, when $t_{k,1} < \alpha < t_{k,2}$, solving for the coefficients of the quadratic IFE local basis functions $\bar{\phi}_{k,2}(x)$ and $\bar{\phi}_{k,3}(x)$ using equation (2.65) with a right hand vector of $(0,1,0,0,0,0)^T$ gives

$$\bar{\phi}_{k,2}(x) = \begin{cases} \dfrac{1}{\bar{Q}_1}\big(p_2(x - t_{k,1})(R(\alpha - x - t_{k,1}) \\ \qquad +p_2 x(x - t_{k,2} - t_{k,3}) + p_1 t_{k,2} t_{k,3})\big)\ , & \text{if } x < \alpha, \\ \dfrac{1}{\bar{Q}_1}\big(p_1(x - t_{k,3})(R(\alpha - x - t_{k,3}) \\ \qquad +p_2 t_{k,1}(-x + t_{k,1} - t_{k,3}) + p_1 x t_{k,3})\big)\ , & \text{if } x \geq \alpha. \end{cases} \tag{2.77}$$

Again using system (2.65) to solve for $\phi_{k,3}(x)$ but this time with a right hand vector of $(0,0,1,0,0,0)^T$ we find that

$$\bar{\phi}_{k,3}(x) = \begin{cases} \dfrac{1}{\bar{Q}_1}\big(p_2(x - t_{k,1})(R(x + t_{k,1} - a) \\ \qquad +p_2 x t_{k,1} + p_1 t_{k,2}(-x - t_{k,1} + t_{k,2}))\big)\ , & \text{if } x < \alpha, \\ \dfrac{1}{\bar{Q}_1}\big(-p_1(x - t_{k,2})(R(\alpha - x - t_{k,2}) \\ \qquad +p_2 t_{k,1}(-x + t_{k,1} - t_{k,2}) + p_1 x t_{k,2})\big)\ , & \text{if } x \geq \alpha. \end{cases} \tag{2.78}$$

On the other hand, if $t_{k,2} \leq \alpha < t_{k,3}$ then the coefficients of quadratic IFE local basis functions $\bar{\phi}_{k,i}(x),\ i = 1,2,3$ can be found using the system (2.70) and we have the following definitions:

$$\bar{\phi}_{k,1}(x) = \begin{cases} \dfrac{p_1}{\bar{Q}_2}(x - t_{k,2})\big(R(x + t_{k,2} - \alpha) \\ \qquad +p_2 t_{k,1} t_{k,2} + p_1 t_{k,3}(-x - t_{k,2} + t_{k,3})\big)\ , & \text{if } x < \alpha, \\ \dfrac{p_1}{p_2 \bar{Q}_2}(t_{k,3} - x)\big(R(\alpha - x - t_{k,3}) \\ \qquad +p_2 t_{k,2}(-x + t_{k,2} - t_{k,3}) + p_1 x t_{k,3}\big)\ , & \text{if } x \geq \alpha, \end{cases} \tag{2.79}$$

$$\bar{\phi}_{k,2}(x) = \begin{cases} \dfrac{1}{\bar{Q}_2}(x - t_{k,1})\big(R(\alpha - x - t_{k,1}) \\ \qquad -p_2 x t_{k,1} + p_1 t_{k,3}(x + t_{k,1} - t_{k,3})\big)\ , & \text{if } x < \alpha, \\ \dfrac{p_1}{p_2 \bar{Q}_2}(x - t_{k,3})\big(R(\alpha - x - t_{k,3}) \\ \qquad p_2 t_{k,1}(-x + t_{k,1} - t_{k,3}) + p_1 x t_{k,3}\big)\ , & \text{if } x \geq \alpha, \end{cases} \tag{2.80}$$

and

$$\bar{\phi}_{k,3}(x) = \begin{cases} \dfrac{p_2}{\bar{Q}_2}(x - t_{k,1})(x - t_{k,2})(t_{k,1} - t_{k,2}) \ , & \text{if } x < \alpha, \\ \dfrac{1}{\bar{Q}_2}(t_{k,1} - t_{k,2})\big(R(-\alpha + t_{k,1} + t_{k,2}) & \\ \qquad +p_1 x(x - t_{k,1} - t_{k,2}) + p_2 t_{k,1} t_{k,2}\big) \ , & \text{if } x \geq \alpha, \end{cases} \tag{2.81}$$

where

$$\bar{Q}_2 = (t_{k,1} - t_{k,2})(R(-\alpha + t_{k,1} + t_{k,2}) + p_2 t_{k,1} t_{k,2} + p_1 t_{k,3}(-t_{k,1} - t_{k,2} + t_{k,3})). \tag{2.82}$$

Figure 2.3 shows a comparison of the three standard quadratic FE local nodal basis functions, $\phi_{k,i}$ versus the three immersed FE local nodal basis functions, $\bar{\phi}_{k,i}$ for $i = 1, 2, 3$ when $t_{k,1} < \alpha < t_{k,2}$. Plots of the local IFE basis functions $\bar{\phi}_{k,i}(x),\ i = 1, 2, 3$ are compared with the standard FE basis functions $\phi_{k,i}(x)$ and the other IFE basis functions in Figure 2.5 at the end of this section (see page 50). The following theorem addresses some basic properties of these basis functions.

Theorem 2.2.7. *The quadratic IFE local nodal basis functions $\bar{\phi}_{k,i}(x), i = 1, 2, 3$ that we introduce above have the following properties:*

1. $\sum_{i=1}^{3} \bar{\phi}_{k,i}(x) \equiv 1$ *for all* $x \in e_k$.

2. *The local IFE nodal basis function $\bar{\phi}_{k,i}(x)$ is consistent with the standard local FE nodal basis function $\phi_{k,i}(x)$ in the following way:*

   (a) $\lim_{\alpha \to t_{k,1}} \bar{\phi}_{k,i}(x) = \phi_{k,i}(x)$ *for* $i = 1, 2, 3$

   (b) $\lim_{\alpha \to t_{k,3}} \bar{\phi}_{k,i}(x) = \phi_{k,i}(x)$ *for* $i = 1, 2, 3$

   (c) If $p_1 = p_2 > 0$ *then* $\bar{\phi}_{k,i}(x) = \phi_{k,i}(x)$.

Proof.

1. This part of the proof is very similar to the proof of Lemma 2.2.3. Consider the function $f(x) = \bar{\phi}_{k,1}(x) + \bar{\phi}_{k,2}(x) + \bar{\phi}_{k,3}(x)$. Since the quadratic IFE local basis functions $\bar{\phi}_{k,i}(x), i = 1, 2, 3$ satisfy the properties (2.42) and (2.64), it follows that $f(x)$ also satisfies these same conditions. Additionally, $f(x)$ is a piecewise quadratic function

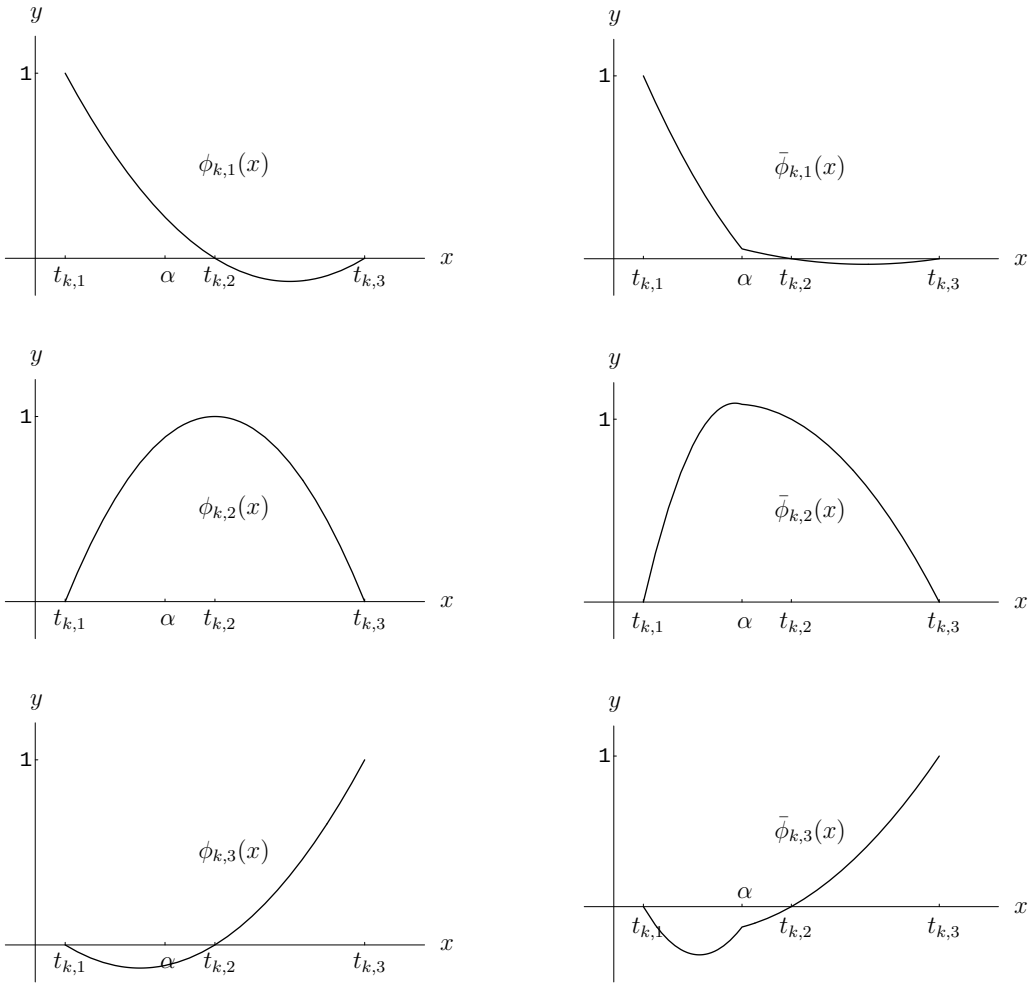


Figure 2.3: Standard quadratic basis functions $\phi_{k,i}(x)$ versus the corresponding immersed basis functions $\bar{\phi}_{k,i}$ on the element e_k

of the form (2.63) so the coefficients of $f(x)$ satisfy the following system of equations when $t_{k,1} < \alpha < t_{k,2}$:

$$\begin{pmatrix} t_{k,1}^2 & t_{k,1} & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & t_{k,2}^2 & t_{k,2} & 1 \\ 0 & 0 & 0 & t_{k,3}^2 & t_{k,3} & 1 \\ \alpha^2 & \alpha & 1 & -\alpha^2 & -\alpha & -1 \\ 2p_1\alpha & p_1 & 0 & -2p_2\alpha & -p_2 & 0 \\ 2p_1 & 0 & 0 & -2p_2 & 0 & 0 \end{pmatrix} \begin{pmatrix} a_2 \\ a_1 \\ a_0 \\ b_2 \\ b_1 \\ b_0 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}. \tag{2.83}$$

Solving this system for the coefficients a_i, b_i for $i = 1, 2, 3$, shows that $a_2 = a_1 = b_2 = b_1 = 0$ and $a_0 = b_0 = 1$. So $f(x) \equiv 1$ when $t_{k,1} \leq \alpha < t_{k,2}$. A similar argument holds when $t_{k,2} \leq \alpha \leq t_{k,3}$ and so $f(x) = \sum_{i=1}^{3} \bar{\phi}_{k,i}(x) \equiv 1$ for all $x \in e_k$.

2. When $t_{k,1} < \alpha < t_{k,2}$, the quadratic IFE local nodal basis function $\bar{\phi}_{k,1}(x)$ is defined by (2.75). First we consider the limiting value of $\bar{\phi}_{k,1}(x)$ as $\alpha \to t_{k,1}$. By direct substitution into the definitions of R, and $\bar{Q}_1$ it can be seen that

$$\lim_{\alpha \to t_{k,1}} R = t_{k,1}(p_1 - p_2), \tag{2.84}$$

$$\begin{aligned} \lim_{\alpha \to t_{k,1}} \bar{Q}_1 &= p_1(t_{k,2} - t_{k,3})(t_{k,1}(p_1 - p_2)(\alpha - t_{k,2} - t_{k,3}) \\ &\qquad + p_2 t_{k,1}(t_{k,1} - t_{k,2} - t_{k,3}) + p_1 t_{k,2} t_{k,3}) \\ &= p_1^2(t_{k,1} - t_{k,2})(t_{k,1} - t_{k,3})(t_{k,2} - t_{k,3}). \end{aligned} \tag{2.85}$$

First if $x < \alpha$ then as $\alpha \to t_{k,1}$ then also $x \to t_{k,1}$ and it can be easily seen by substituting into the definition of $\bar{\phi}_{k,1}$ that

$$\lim_{\alpha \to t_{k,1}} \bar{\phi}_{k,1}(x) = 1 = \phi_{k,1}(t_{k,1}). \tag{2.86}$$

On the other hand when $x \geq \alpha$ then we have that

$$\begin{aligned} \lim_{\alpha \to t_{k,1}} \bar{\phi}_{k,1}(x) &= \lim_{\alpha \to t_{k,1}} \frac{1}{\bar{Q}_1}\big(p_1^2(x - t_{k,2})(x - t_{k,3})(t_{k,2} - t_{k,3})\big) \\ &= \left(\frac{x - t_{k,2}}{t_{k,1} - t_{k,2}}\right)\left(\frac{x - t_{k,3}}{t_{k,1} - t_{k,3}}\right) = \phi_{k,1}(x). \end{aligned} \tag{2.87}$$

Now we consider the limiting value as $\alpha \to t_{k,3}$. First we notice that

$$\lim_{\alpha \to t_{k,3}} R = t_{k,3}(p_1 - p_2) \tag{2.88}$$

$$\lim_{\alpha \to t_{k,e}} (\bar{Q}_2) = p_2^2(t_{k,1} - t_{k,2})(t_{k,1} - t_{k,3})(t_{k,2} - t_{k,3})/p_1. \tag{2.89}$$

Then there are two cases: (i) $x \geq \alpha$ and (ii) $x < \alpha$. In the first case, as $\alpha \to t_{k,3}$ then also $x \to t_{k,3}$ and so it follows by substituting into the definition of $\phi_{k,1}$ that

$$\lim_{\alpha \to t_{k,3}} \bar{\phi}_{k,1}(x) = 1 = \phi_{k,1}(t_{k,3}). \tag{2.90}$$

On the other hand when $x < \alpha$ we have

$$\begin{aligned} \lim_{\alpha \to t_{k,3}} \bar{\phi}_{k,1}(x) &= \lim_{\alpha \to t_{k,3}} \frac{p_1}{p_2 \bar{Q}_2}(t_{k,3} - x)\big(R(\alpha - x - t_{k,3}) \\ &\qquad p_2 t_{k,2}(-x + t_{k,2} - t_{k,3}) + p_1 x t_{k,3}\big) \\ &= \left(\frac{x - t_{k,2}}{t_{k,1} - t_{k,2}}\right)\left(\frac{x - t_{k,3}}{t_{k,1} - t_{k,3}}\right) = \phi_{k,1}(x). \end{aligned} \tag{2.91}$$

Finally, we consider the case where $p_2 \to p_1$ (or equivalently $p_1 \to p_2$). First we notice that

$$\lim_{p_2 \to p_1} R = 0, \tag{2.92}$$

$$\lim_{p_2 \to p_1} \bar{Q}_1 = p_2^2(t_{k,2} - t_{k,3})(t_{k,1} - t_{k,2})(t_{k,1} - t_{k,3}), \tag{2.93}$$

$$\lim_{p_2 \to p_1} \bar{Q}_2 = p_2(t_{k,1} - t_{k,2})(t_{k,1} - t_{k,3})(t_{k,2} - t_{k,3}). \tag{2.94}$$

Then the result follows by substituting into (2.75) and (2.79) for $\bar{\phi}_{k,1}(x)$. The same arguments for all three limits can be used to establish the results for $\bar{\phi}_{k,2}(x)$ and $\bar{\phi}_{k,3}(x)$. For $\bar{\phi}_{k,2}(x)$, definitions (2.77) and (2.80) are used while for $\bar{\phi}_{k,3}(x)$, the definitions (2.78) and (2.81) are used to evaluate the limits. ■

In part 2 of this theorem, (a), (b) and (c) show that the basis functions $\bar{\phi}_{k,i}(x),\ i = 1, 2, 3$ are "close" to the standard FE basis functions in the sense that when the coefficient function is continuous or if the interface occurs at the endpoints of the interval, then the basis function becomes the standard quadratic FE basis function.

Now, we are ready to construct an IFE space by using these local basis functions. First, for each element $e_k \in \mathcal{T}_h$, we let

$$\bar{S}_h(e_k) = \begin{cases} span\{\phi_{k,i}(x), i = 1, 2, 3\}, & \text{if } e_k \text{ is a non-interface element,} \\ span\{\bar{\phi}_{k,i}(x), i = 1, 2, 3\}, & \text{if } e_k \text{ is an interface element.} \end{cases} \tag{2.95}$$

Then, for each node $x_i,\ i = 0, 1, 2, \ldots, 2N$, we let $\bar{\phi}_i(x)$ be a piecewise quadratic polynomial such that $\bar{\phi}_i|_{e_k} \in \bar{S}_h(e_k)$ for any $e_k \in \mathcal{T}_h$, and

$$\bar{\phi}_i(x_j) = \begin{cases} 1, & \text{if } i = j, \\ 0, & \text{if } i \neq j, \end{cases} \tag{2.96}$$

Finally, we can form the IFE space over the whole domain by the local nodal basis functions with extra continuity condition as follows:

$$\bar{S}_h(\Omega) = span\{\bar{\phi}_i, i = 0, 1, \ldots, 2N\}. \tag{2.97}$$

An 1-D IFE Space with Local Mesh Refinement

Another way of defining quadratic immersed basis functions is through a process that is similar to local refinement [1]. Each interface element is divided into two pieces based upon the location of the interface point $x = \alpha$. Thus the interface element e_k is broken into two sub-elements, $[t_{k,1}, \alpha]$ and $[\alpha, t_{k,3}]$. Each sub-element is given three nodes which correspond to its endpoints and the midpoints. For example, the left sub-element has the three nodes $t_{k,1}, (t_{k,1} + \alpha)/2$ and α. Then we introduce the following notation for the nodes of these new subelements excepting those that occur at the interface:

$$\begin{array}{rcl} \hat{t}_{k,1} & = & t_{k,1}, \\ \hat{t}_{k,2} & = & (t_{k,1} + \alpha)/2, \\ \hat{t}_{k,3} & = & (\alpha + t_{k,3})/2, \\ \hat{t}_{k,4} & = & t_{k,3}. \end{array} \tag{2.98}$$

The quadratic IFE local basis functions (referred to as $\hat{\phi}_{k,i}(x)$) in the interface element e_k are then defined as piecewise quadratic polynomials such as

$$\hat{\phi}_{k,i}(x) = \begin{cases} a_2x^2 + a_1x + a_0, & \text{if } t_{k,1} < x < \alpha, \\ b_2x^2 + b_1x + b_0, & \text{if } \alpha \le x < t_{k,3}. \end{cases} \tag{2.99}$$

The coefficients $a_m,\ m = 0, 1, 2$ and $b_n,\ n = 0, 1, 2$ are chosen such that $\hat{\phi}_{k,i}(x),\ i = 1, 2, 3, 4$ satisfy the following nodal value specifications

$$\hat{\phi}_{k,i}(\hat{t}_{k,j}) = \begin{cases} 1, & \text{if } i = j, \\ 0, & \text{if } i \ne j, \end{cases} \tag{2.100}$$

and the usual interface jump conditions

$$\begin{array}{rcl} [\hat{\phi}_{k,i}]_{x=\alpha} & = & 0, \\ [p\hat{\phi}'_{k,i}]_{x=\alpha} & = & 0. \end{array} \tag{2.101}$$

The following theorem establishes the existence and uniqueness of the basis functions $\hat{\phi}_{k,i}(x),\ i = 1, 2, 3, 4$ which are subject to the conditions (2.100) and (2.101).

[1] This idea is attributed to Dr. Yangpin Lin, University of Alberta, Edmonton, Alberta.

Theorem 2.2.8. *Quadratic IFE local basis functions $\hat{\phi}_{k,i}(x), i = 1, 2, 3, 4$ which are constructed as described above so that each $\hat{\phi}_{k,i}(x)$ satisfies (2.100) and (2.101) are uniquely determined.*

Proof. Let e_k be an interface element so that $t_{k,1} < \alpha < t_{k,3}$. Consider the quadratic IFE local basis function $\hat{\phi}_{k,1}(x)$ which is a piecewise quadratic function as in (2.63). Since $\hat{\phi}_{k,1}(x)$ is required to satisfy (2.100) and (2.101), then the coefficients a_2, a_1, a_0, b_2, b_1, and b_0 must satisfy the following system of equations:

$$\begin{pmatrix} (\hat{t}_{k,1})^2 & \hat{t}_{k,1} & 1 & 0 & 0 & 0 \\ (\hat{t}_{k,2})^2 & \hat{t}_{k,2} & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & (\hat{t}_{k,3})^2 & \hat{t}_{k,3} & 1 \\ 0 & 0 & 0 & (\hat{t}_{k,4})^2 & \hat{t}_{k,4} & 1 \\ \alpha^2 & \alpha & 1 & -\alpha^2 & -\alpha & -1 \\ 2p_1\alpha & p_1 & 0 & -2p_2\alpha & -p_2 & 0 \end{pmatrix} \begin{pmatrix} a_2 \\ a_1 \\ a_0 \\ b_2 \\ b_1 \\ b_0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}. \tag{2.102}$$

The determinant of the coefficient matrix is a function of the interface location given by

$$q(\alpha) = -\frac{3}{16}(\alpha - t_{k,1})^2(\alpha - t_{k,3})^2(R + p_2 t_{k,1} - p_1 t_{k,3}). \tag{2.103}$$

Since $(R + p_2 t_{k,1} - p_1 t_{k,3}) \neq 0$ for $\alpha \in e_k$ then $q(\alpha) \neq 0$ for $t_{k,1} < \alpha < t_{k,3}$. Then the determinant of the coefficient matrix is non-zero meaning that the system of equations has a unique solution for $t_{k,1} < \alpha < t_{k,3}$. Hence the coefficients are uniquely determined and $\hat{\phi}_{k,1}(x)$ is unique. The basis functions $\hat{\phi}_{k,i}(x), i = 2, 3, 4$ are also unique since their coefficients also satisfy systems that have the same nonsingular coefficient matrix as in (2.102) along with different vectors on the right hand side of the equation. So $\hat{\phi}_{k,i}, i = 1, \ldots, 4$ are uniquely determined on the interface element e_k. ■

The proof of Theorem 2.2.8 provides a format similar to that of Theorem 2.2.6 for determining the coefficients of the basis functions $\hat{\phi}_{k,i}(x)$. The following constants are useful in their descriptions:

$$\hat{Q} = 3(R + p_2 t_{k,1} - p_1 t_{k,3}), \tag{2.104}$$

$$h_l = \alpha - t_{k,1}, \tag{2.105}$$

$$h_r = t_{k,3} - \alpha. \tag{2.106}$$

Then the quadratic IFE local basis function $\hat{\phi}_{k,1}$ have the following definitions:

$$\hat{\phi}_{k,1}(x) = \begin{cases} \dfrac{1}{\hat{Q}h_l^2}(\alpha - 2x + t_{k,1})\big(p_1 t_{k,3}(2x + t_{k,1}) - 3p_2 x t_{k,1} \\ \qquad +\alpha(3R + 3p_2(x + t_{k,1}) - p_1(2x + t_{k,1} + 3t_{k,3}))\big), & \text{if } x < \alpha, \\ -\dfrac{p_1}{\hat{Q}h_r}(x - t_{k,3})(\alpha - 2x + t_{k,3}), & \text{if } x \geq \alpha. \end{cases} \tag{2.107}$$

The local IFE basis functions $\hat{\phi}_{k,i}(x),\ i=2,3,4$ have similar formulas.

$$\hat{\phi}_{k,2}(x)=\begin{cases}\dfrac{4}{\hat{Q}h_l^2}(x-t_{k,1})\big(p_1t_{k,3}(x+t_{k,1})-3p_2xt_{k,1}\\ \qquad\qquad +\alpha(2R-\alpha p_2+3p_2(x+t_{k,1})\\ \qquad\qquad -p_1(x+t_{k,1}+2t_{k,3}))\big)\ , & \text{if } x<\alpha,\\ \dfrac{4p_1}{\hat{Q}h_r}(\alpha-2x+t_{k,3})(x-t_{k,3})\ , & \text{if } x\geq\alpha,\end{cases}\tag{2.108}$$

$$\hat{\phi}_{k,3}(x)=\begin{cases}\dfrac{4p_2}{\hat{Q}h_l}(\alpha-2x+t_{k,1})(x-t_{k,1})\ , & \text{if } x<\alpha,\\ \dfrac{4}{\hat{Q}h_r^2}(x-t_{k,3})(3p_1xt_{k,3}-p_2t_{k,1}(x+t_{k,3})\\ \qquad\qquad +\alpha(2R+p_1(\alpha-3x-3t_{k,3})\\ \qquad\qquad +p_2(x+2t_{k,1}+t_{k,3}))\ , & \text{if } x\geq\alpha,\end{cases}\tag{2.109}$$

and

$$\hat{\phi}_4(x)=\begin{cases}\dfrac{-p_2}{\hat{Q}h_l}(x-t_{k,1})(\alpha-2x+t_{k,1})\ , & \text{if } x<\alpha,\\ \dfrac{1}{\hat{Q}h_r^2}(\alpha-2x+t_{k,3})(3p_1xt_{k,3}-p_2t_{k,1}(2x+t_{k,3})\\ \qquad\qquad +\alpha(3R-3p_1(x+t_{k,3})+p_2(2x+3t_{k,1}+t_{k,3}))\ , & \text{if } x\geq\alpha.\end{cases}\tag{2.110}$$

Figure 2.4 shows a comparison of the three standard FE local nodal basis functions, $\phi_{k,i}$ for $i=1,2,3$ versus the four immersed FE local nodal basis functions, $\hat{\phi}_{k,i}$ for $i=1,2,3,4$. Plots of the local IFE basis functions $\hat{\phi}_{k,i}(x),\ i=1,2,3,4$ are compared with the standard FE basis functions $\phi_{k,i}(x)$ and the other IFE basis functions in Figure 2.5 at the end of this section. Now we turn to the discussion of basic properties of the quadratic IFE local basis functions $\hat{\phi}_{k,i}(x), i=1,2,3,4$. They are summarized in the following theorem.

Theorem 2.2.9. *The quadratic IFE local nodal basis functions $\hat{\phi}_{k,i}(x), i=1,2,3,4$ that we introduce above have the following property:*

$$\sum_{i=1}^{4}\hat{\phi}_{k,i}(x)\equiv 1 \text{ for all } x\in e_k.\tag{2.111}$$

Proof. Consider the function that is defined by $f(x)=\hat{\phi}_{k,1}(x)+\hat{\phi}_{k,2}(x)+\hat{\phi}_{k,3}(x)+\hat{\phi}_{k,4}(x)$. Due to the local nodal values of the quadratic IFE local basis functions $\hat{\phi_{k,i}}(x),\ \ i=1,\ldots,4,$

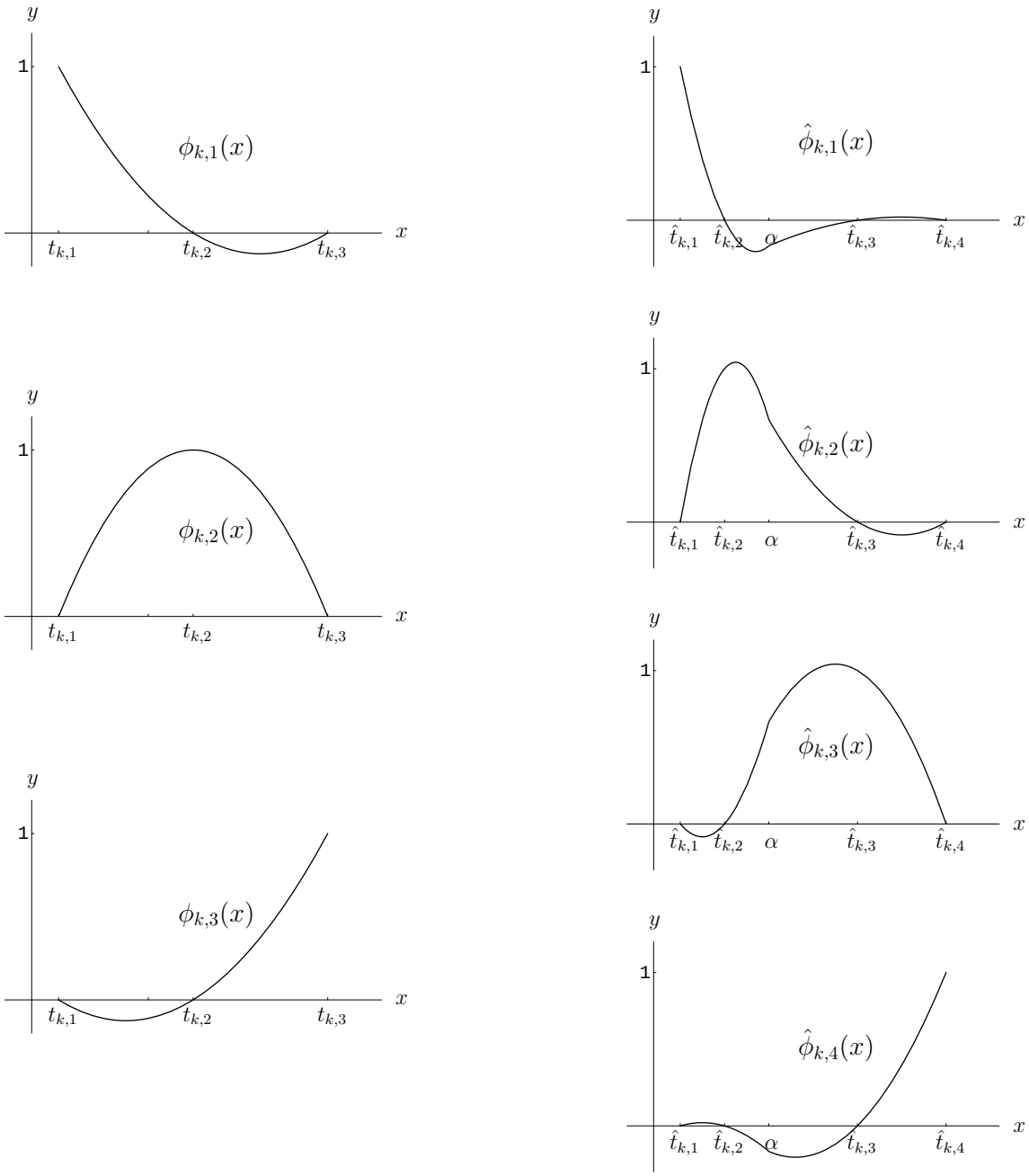


Figure 2.4: Standard quadratic basis functions $\phi_{k,i}(x)$ for $i = 1, 2, 3$ versus the corresponding immersed basis functions $\hat{\phi}_{k,i}$ for $i = 1, 2, 3, 4$ on the element e_k

it follows that the function $f(x)$ is a piecewise quadratic function of the form (2.63) and that the coefficients of $f(x)$ satisfy the following system of equations:

$$\begin{pmatrix} (\hat{t}_{k,1})^2 & \hat{t}_{k,1} & 1 & 0 & 0 & 0 \\ (\hat{t}_{k,2})^2 & \hat{t}_{k,2} & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & (\hat{t}_{k,3})^2 & \hat{t}_{k,3} & 1 \\ 0 & 0 & 0 & (\hat{t}_{j,4})^2 & \hat{t}_{j,4} & 1 \\ \alpha^2 & \alpha & 1 & -\alpha^2 & -\alpha & -1 \\ 2p_1\alpha & p_1 & 0 & -2p_2\alpha & -p_2 & 0 \end{pmatrix} \begin{pmatrix} a_2 \\ a_1 \\ a_0 \\ b_2 \\ b_1 \\ b_0 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 0 \\ 0 \end{pmatrix}. \tag{2.112}$$

We see that the coefficient matrix is nonsingular in the proof of Theorem 2.2.8. Then solving the system for the coefficients of $f(x)$ shows that

$$a_2 = a_1 = b_2 = b_1 = 0, \text{ and } a_0 = b_0 = 1, \tag{2.113}$$

and hence $f(x) \equiv 1$ for $t_{k,1} < \alpha < t_{k,3}$. ■

Remark 2.2.10. *The quadratic IFE local basis functions $\hat{\phi}_{k,i}(x)$ are not consistent with standard FE local basis functions. This means that if $p_1 = p_2$ or the interface approaches the boundary of e_k (i.e. $\alpha \to t_{k,1}$ or $\alpha \to t_{k,3}$), then the basis functions $\hat{\phi}_{k,i}(x)$ do not transform into the standard FE local basis functions $\phi_{k,i}(x)$. This is because of the way in which the interface element e_k is divided into subelements, and subsequently there are four local basis functions $\hat{\phi}_{k,i}(x), i = 1, \ldots, 4$.*

Now, we can construct an IFE space from these local basis functions. First, for each element $e_k \in \mathcal{T}_h$, we let

$$\hat{\mathcal{S}}_h(e_k) = \begin{cases} span\{\phi_{k,i}(x), i = 1, 2, 3\}, & \text{if } e_k \text{ is a non-interface element,} \\ span\{\hat{\phi}_{k,i}(x), i = 1, 2, 3, 4\}, & \text{if } e_k \text{ is an interface element.} \end{cases} \tag{2.114}$$

Then we need to redefine our global nodes to incorporate the local refinement on the interface elements. If $e_k = [x_{2(k-1)}, x_{2k}]$ is the interface element where $x_i = i(h/2),\ i = 0, 1, \ldots, 2N$ and $h = 1/N$, then the global nodes $\hat{x}_i, i = 0, 1, 2, \ldots, 2N + 1$ are

$$\hat{x}_i = \begin{cases} x_i, & \text{if } i = 0, \ldots, 2k - 2, \\ x_{2k-2} + (\alpha + x_{2k-2})/2, & \text{if } i = 2k - 1, \\ x_{2k} - (x_{2k} + \alpha)/2, & \text{if } i = 2k, \\ x_{i-1}, & \text{if } i = 2k + 1, \ldots, 2N + 1. \end{cases} \tag{2.115}$$

Then, for each node $\hat{x}_i, i = 0, 1, 2, \cdots, 2N + 1$, we let $\hat{\phi}_i(x)$ be a piecewise quadratic polynomial such that $\hat{\phi}_i|_{e_k} \in \hat{\mathcal{S}}_h(e_k)$ for any $e_k \in \mathcal{T}_h$, and

$$\hat{\phi}_i(\hat{x}_j) = \begin{cases} 1, & \text{if } i = j, \\ 0, & \text{if } i \neq j, \end{cases} \tag{2.116}$$

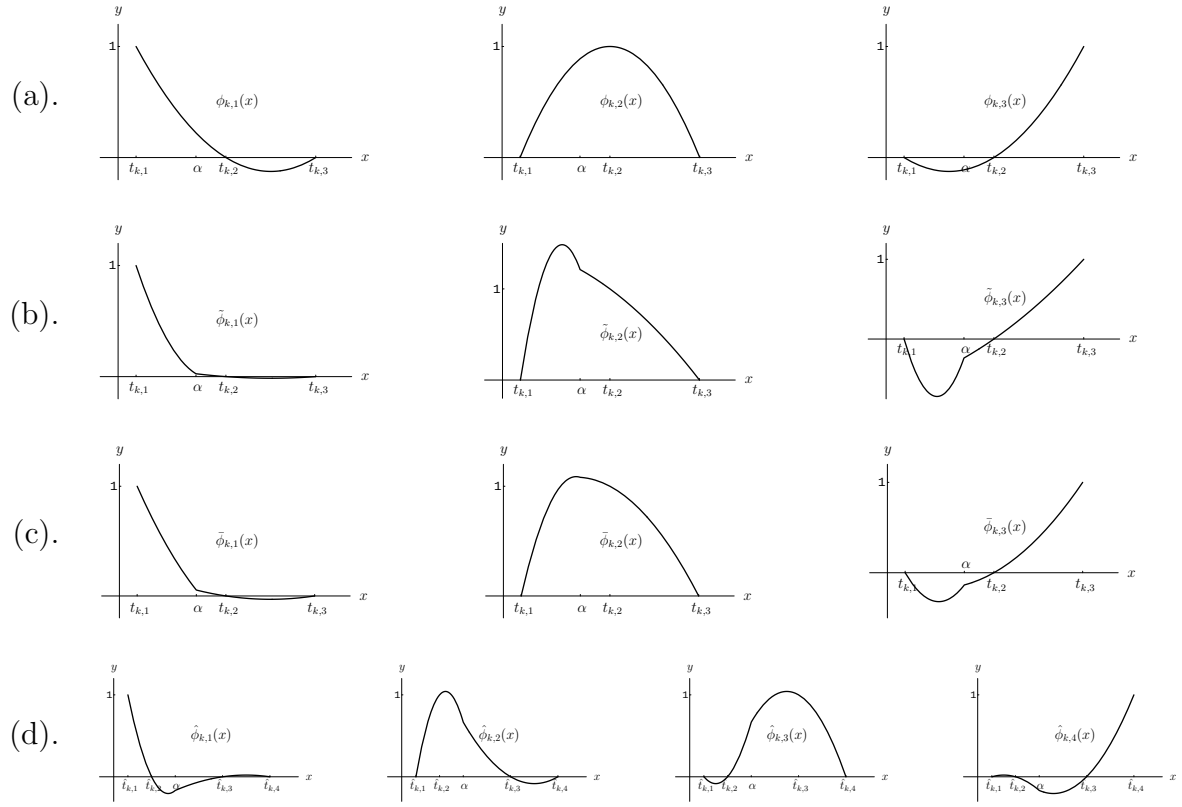


Figure 2.5: A comparison of local quadratic FE basis functions: (a) standard $\phi_{k,i}(x)$ (b) hierarchial $\tilde{\phi}_{k,i}(x)$ (c) extra continuity $\bar{\phi}_{k,i}(x)$ and (d) local refinement $\hat{\phi}_{k,i}(x)$.

Finally, we can form the IFE space over the whole domain by the local nodal basis functions with extra continuity condition as follows:

$$\hat{\mathcal{S}}_h(\Omega) = span\{\hat{\phi}_i, i = 0, 1, \cdots, 2N+1\}. \tag{2.117}$$

Brief comparison of the 1-D IFE Spaces and Their Basis Functions

Plots of the local basis functions used in the IFE spaces $\tilde{\mathcal{S}}_h(\Omega)$, $\bar{\mathcal{S}}_h(\Omega)$ and $\hat{\mathcal{S}}_h(\Omega)$ are shown in Figure 2.5. The fact that the IFE space $\hat{S}_{h,2}(\Omega)$ uses four local basis functions on an interface element while the other IFE spaces $\tilde{\mathcal{S}}_{h,2}(\Omega)$ and $\bar{\mathcal{S}}_{h,2}(\Omega)$ each only use three local IFE basis functions can be seen clearly in Figure 2.5. The fact that the IFE space $\hat{\mathcal{S}}_{h,2}(\Omega)$ uses four local basis functions instead of three is the primary reason that these basis functions are not consistent with those of the standard FE space, $\mathcal{S}_{h,2}(\Omega)$. The plots also show some of the similarities that the IFE spaces $\tilde{\mathcal{S}}_{h,2}(\Omega)$ and $\bar{\mathcal{S}}_{h,2}(\Omega)$ have with the standard FE space since

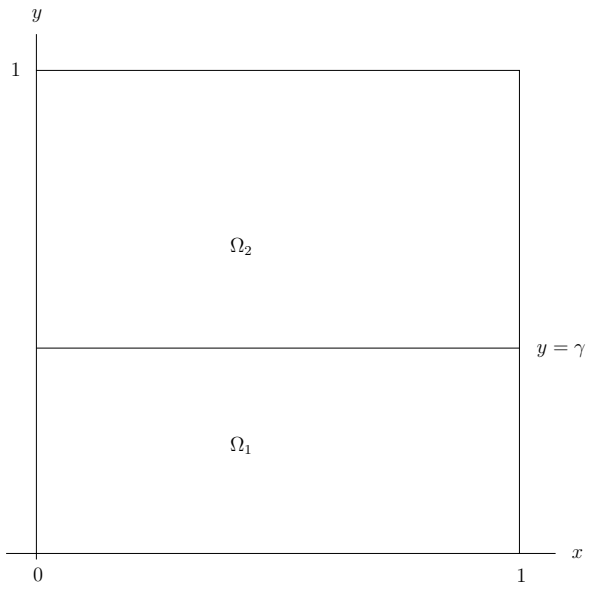


Figure 2.6: The domain of the elliptic interface problem (2.120).

the local IFE basis functions from these spaces are consistent with the standard local FE basis functions according to Theorems 2.2.5 and 2.2.7.

2.2.2 Biquadratic IFE Spaces in Two Dimensions

In this section, we extend the one dimensional quadratic IFE spaces based upon $\tilde{\phi}_{k,i}(x)$, $\bar{\phi}_{k,i}(x)$ and $\hat{\phi}_{k,i}(x)$, to two dimensional biquadratic IFE spaces based on rectangular partitions where the interface Γ is a line parallel to one of the coordinate axes. In this configuration, a local biquadratic IFE basis function in an interface element can be constructed by taking the product of a standard FE local basis function and an IFE local basis function.

Before the discussion on constructing the IFE spaces, we first restate the model BVP in two dimensions. We let the domain of the elliptic interface problem (2.1) be the typical unit square $\Omega = (0,1) \times (0,1)$. We assume that the interface Γ is a horizontal line at $y = \gamma$ (see Figure 2.6) so that Ω_1 and Ω_2 are defined as

$$\Omega_1 = \Omega \cap \{(x,y)|y < \gamma\}, \tag{2.118}$$

$$\Omega_2 = \Omega \cap \{(x,y)|y > \gamma\}. \tag{2.119}$$

It should be noted that the methods for constructing biquadratic IFE spaces discussed here can handle more general linear interfaces. For instance, the biquadratic IFE spaces in this

section are able to handle an interface that is a piecewise linear function as long as each linear piece is parallel to one of the coordinate axes and pieces in the $y-$direction occur along the gridlines of the finite element mesh. In other words, all linear pieces in the $y-$direction are not allowed to pass through elements. Choosing the $y-$direction for this restriction is arbitrary, however, and one could also restrict interface pieces in the $x-$direction to be along the gridlines and consequently allow the interface in the $y-$direction to be immersed.

Then the boundary value problem (2.1) becomes

$$\begin{aligned} -\nabla\cdot(p(\boldsymbol{x})\nabla u(\boldsymbol{x})) &= f(\boldsymbol{x}), \quad \boldsymbol{x}=(x,y)^T\in\Omega, \\ u(x,0) &= g_0(x), \\ u(x,1) &= g_1(x), \\ u(0,y) &= h_0(y), \\ u(1,y) &= h_1(y). \end{aligned} \tag{2.120}$$

As usual, $u(\boldsymbol{x})$ is also required to satisfy the interface conditions

$$\begin{aligned} \big[u\big]_\Gamma &= 0, \\ \big[\boldsymbol{n}\cdot p\nabla u\big]_\Gamma &= 0, \end{aligned} \tag{2.121}$$

where $\boldsymbol{n}=(n_1,n_2)^T$ is the unit normal at $\boldsymbol{x}\in\Gamma$ pointing from Ω_1 to Ω_2. The coefficient p is then defined by

$$p(x,y)=\begin{cases} p_1 & \text{if } y<\gamma, \\ p_2 & \text{if } y\geq\gamma. \end{cases} \tag{2.122}$$

Without loss of generality, we consider forming quadratic IFE spaces in a uniform partition consisting of axially aligned rectangular elements. First, we introduce quadratic node points of the partition in the x and y directions by letting

$$\begin{aligned} x_i &= x_0+i(h/2), \quad i=0,1,\ldots,2N, \\ y_j &= y_0+j(h/2), \quad j=0,1,\ldots,2N, \end{aligned} \tag{2.123}$$

with $h=1/N$ for an integer N. Then for each pair of integers $(i,j), i,j=1,2,\cdots,N$, we let $k=i+(j-1)N$, and let

$$e_k=[x_{2(i-1)},x_{2i}]\times[y_{2(j-1)},y_{2j}], i,j=1,2,\ldots,N, \tag{2.124}$$

as Figure 2.7 shows. Finally, we let $\mathcal{T}_h=\cup_{k=1}^{N^2}e_k$ to form a partition of Ω. We note that the k^{th} element of $\mathcal{T}_h$ locates at the i^{th} column and j^{th} row of the partition (see Figure 2.8). In particular, this means that the partition $\mathcal{T}_h$ is made up of N^2 elements. Each biquadratic element has nine local nodes, so the following notation for the local nodal points on the element e_k is adopted:

$$\begin{aligned} s_{k,1}&=x_{2(i-1)}, & t_{k,1}&=y_{2(j-1)}, \\ s_{k,2}&=x_{2i-1}, & t_{k,2}&=y_{2j-1}, \\ s_{k,3}&=x_{2i}, & t_{k,3}&=y_{2j}, \end{aligned} \tag{2.125}$$

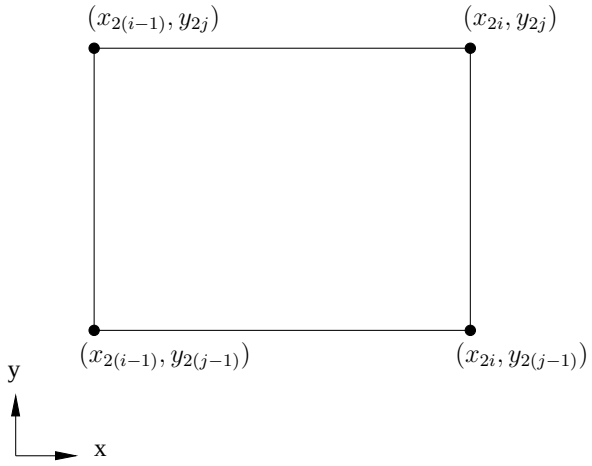


Figure 2.7: The element $e_k \in \mathcal{T}_h$ where $k = i + (j-1)N$.

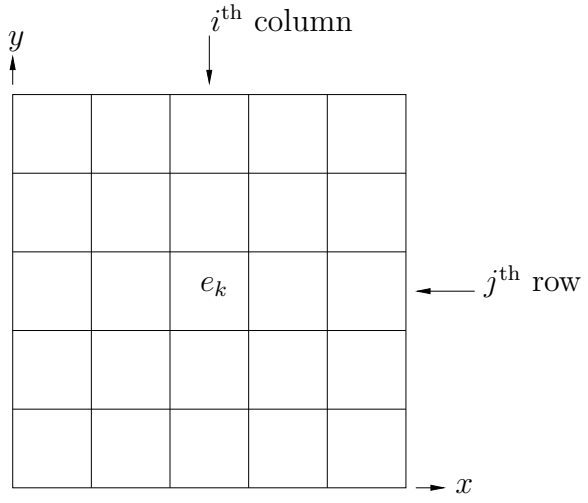
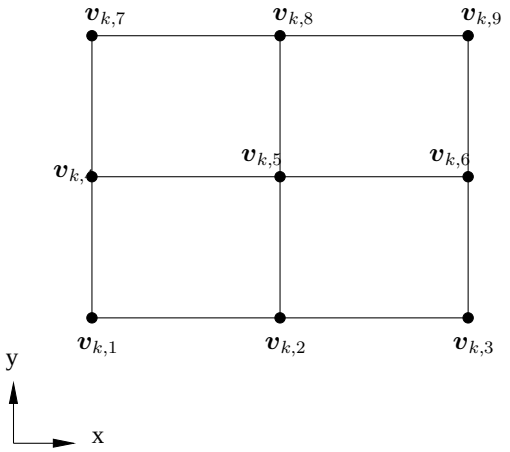


Figure 2.8: The finite element partition $\mathcal{T}_h = \bigcup_{k=1}^{N^2} e_k$ of the domain Ω.



$$
\begin{aligned}
\boldsymbol{v}_{k,1} &= (s_{k,1}, t_{k,1}),\\
\boldsymbol{v}_{k,2} &= (s_{k,2}, t_{k,1}),\\
\boldsymbol{v}_{k,3} &= (s_{k,3}, t_{k,1}),\\
\boldsymbol{v}_{k,4} &= (s_{k,1}, t_{k,2}),\\
\boldsymbol{v}_{k,5} &= (s_{k,2}, t_{k,2}),\\
\boldsymbol{v}_{k,6} &= (s_{k,3}, t_{k,2}),\\
\boldsymbol{v}_{k,7} &= (s_{k,1}, t_{k,3}),\\
\boldsymbol{v}_{k,8} &= (s_{k,2}, t_{k,3}),\\
\boldsymbol{v}_{k,9} &= (s_{k,3}, t_{k,3}).
\end{aligned}
$$

Figure 2.9: An element $e_k \in \mathcal{T}_h$ with its nine local nodes.

where $k = i + (j-1)N$. The nine local nodes on the element e_k, $\boldsymbol{v}_{k,l}$ for $l = 1, \ldots, 9$ are seen in Figure 2.9 where the node $\boldsymbol{v}_{k,l}$ is defined by $\boldsymbol{v}_{k,l} = (s_{k,m}, t_{k,n})$. As before, to form an IFE spaces, specialized local nodal basis functions are only used on interface elements which are defined as follows:

Definition 2.2.11. *Let the interface Γ occur along the horizontal line $y = \gamma$. The element $e_k = [x_{2(i-1)}, x_{2i}] \times [y_{2(j-1)}, y_{2j}]$ (where $k = i + (j-1)N$) of the partition $\mathcal{T}_h$ is called an* ***interface element*** *if $y_{2(j-1)} < \gamma < y_{2j}$. Any element which is not an interface element is a* ***non-interface element****.*

A typical interface element e_k with a horizontal interface at $y = \gamma$ is shown in Figure 2.10. Then, on a typical non-interface element, the nine biquadratic IFE local nodal basis functions are simply the standard quadratic FE local basis functions that are defined by

$$\psi_{k,l}(\boldsymbol{v}_{k,j}) = \begin{cases} 1, & \text{if } j = l, \\ 0, & \text{if } j \neq l, \end{cases} \tag{2.126}$$

for $l = m + 3(n-1)$ with $1 \leq m, n \leq 3$ and $k = i + (j-1)N$. These local basis function $\psi_{k,l}(x, y)$ can be defined in terms of the standard quadratic FE local basis functions in one dimension by

$$\psi_{k,l}(x, y) = \phi_{i,m}(x)\phi_{j,n}(y), \tag{2.127}$$

where $x_{2(i-1)} < x < x_{2i}$ and $y_{2(j-1)} < y < y_{2j}$. On an interface element, the biquadratic IFE local basis function are defined by the product of a standard FE local basis function and a IFE local basis function. More detail is given about each type of IFE local basis functions in the following subsections.

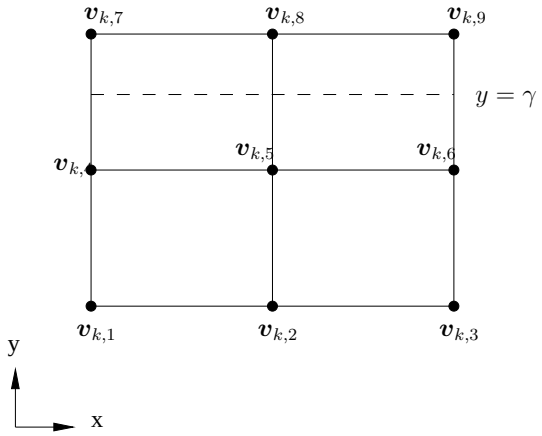


Figure 2.10: An interface element e_k when the interface Γ is a horizontal line at $y = g$.

The three types of biquadratic IFE spaces to be discussed below are all in separable function form. Additionally, each of these IFE local basis functions satisfies properties similar to those of the one dimensional IFE spaces, such as a partition of unity. Since each of the IFE local basis functions is a separable function, it is useful to introduce some lemmas regarding these types of functions here. These lemmas help establish some of the properties satisfied by these biquadratic IFE local basis functions.

The following lemma describes how a group of separable functions can satisfy a partition of unity.

Lemma 2.2.12. *Let $h_{i,j}(x, y)$ for $1 \le i \le m,\ 1 \le j \le n$, be a group of separable functions that is defined by*

$$h_{i,j}(x, y) = f_i(x)g_j(y).$$

If $f_i(x),\ i = 1, \ldots, m$ and $g_j(y),\ j = 1, \ldots, n$ are functions that satisfy

$$\sum_{i=1}^{m} f_i(x) \equiv 1 \text{ for all } x \in I_x = [x_1, x_2],$$

and

$$\sum_{j=1}^{m} g_j(y) \equiv 1 \text{ for all } y \in I_y = [y_1, y_2],$$

then

$$\sum_{i=1}^{m}\sum_{j=1}^{n} h_{i,j}(x, y) \equiv 1, \quad \text{for all } (x, y) \in \Omega_e = I_x \times I_y. \tag{2.128}$$

Proof. Notice that

$$\sum_{i=1}^{m}\sum_{j=1}^{n} h_{i,j}(x,y) \;=\; \sum_{i=1}^{m}\sum_{j=1}^{n} f_i(x)g_j(y) \;=\; \sum_{i=1}^{m}\left(f_i(x)\sum_{j=1}^{n} g_j(y)\right) \;=\; \sum_{i=1}^{m} f_i(x) = 1,$$

since $\sum_{j=1}^{n} g_j(y) \equiv 1$ for all $y \in I_y = [y_1, y_2]$ and $\sum_{i=1}^{m} f_i(x) \equiv 1$ for all $x \in I_x = [x_1, x_2]$. Hence $\sum_{i=1}^{m}\sum_{j=1}^{n} h_{i,j}(x,y) \equiv 1$ for all $(x,y) \in \big\{(x,y)|x_1 \le x \le x_2,\ y_1 \le y \le y_2\big\} = I_x \times I_y$. ■

The following lemma describes how a separable function can satisfy the two dimensional interface conditions (2.121).

Lemma 2.2.13. *Let there be an interface Γ that occurs along the line $y = \gamma$ such that*

$$p = \begin{cases} p_1, & \text{if } \ y < \gamma, \\ p_2, & \text{if } \ y \ge \gamma. \end{cases}$$

Let $h(x,y) = f(x)g(y)$ where $g(y)$ is piecewise C^1 function that is defined by

$$g(y) = \begin{cases} g_1(y), & \text{if } \ y < \gamma, \\ g_2(y), & \text{if } \ y \ge \gamma. \end{cases}$$

If $g(y)$ satisfies the conditions

$$\begin{aligned} \big[g\big]_{y=\gamma} &= 0, \\ \big[pg'(y)\big]_{y=\gamma} &= 0, \end{aligned}$$

then the function $h(x,y)$ satisfies (2.121).

Proof. Since $h(x,y)$ is separable, then it follows that

$$\begin{aligned} \big[h\big]_{\Gamma} &= \lim_{y\to\gamma^-} f(x)g(y) - \lim_{y\to\gamma^+} f(x)g(y) \\ &= f(x)\big[g(y)\big]_{y=\gamma} \\ &= 0. \end{aligned}$$

Also, since the interface is a horizontal line, then the normal vector to the interface is $\boldsymbol{n} = (0,1)^T$ and it follows that

$$\boldsymbol{n} \cdot p\nabla h = p\frac{\partial h}{\partial y}.$$

Also, since $h(x,y)$ is separable then $\dfrac{\partial h}{\partial y} = f(x)g'(y)$ and hence

$$\begin{aligned}\left[\boldsymbol{n}\cdot p\nabla h\right]_\Gamma &= \left[pf(x)g'(y)\right]_\Gamma \\ &= \lim_{y\to\gamma^-} pf(x)g'(y) - \lim_{y\to\gamma^+} pf(x)g'(y) \\ &= f(x)\left[pg'(y)\right]_{y=\gamma} \\ &= 0.\end{aligned}$$

Therefore $h(x,y)$ satisfies the conditions (2.121). ■

Now we are ready to give details about the three two dimensional IFE spaces developed from one dimensional IFE local basis functions $\tilde{\phi}_{k,i}(x)$, $\bar{\phi}_{k,i}(x)$ and $\hat{\phi}_{k,i}(x)$.

A 2-D Hierarchical Biquadratic IFE Space

The first type of biquadratic IFE space is the one that is formed by the one dimensional hierarchical quadratic IFE local basis functions $\tilde{\phi}_{q,i}(y)$ that are defined on the interval $[y_{2(q-1)}, y_{2q}]$. If the interface Γ occurs along the line $y=\gamma$ and e_k is an interface element, we define the biquadratic IFE local basis function $\tilde{\psi}_{k,l}(x,y)$ by

$$\tilde{\psi}_{k,l}(x,y) = \phi_{i,m}(x)\tilde{\phi}_{j,n}(y), \tag{2.129}$$

where $k = i+(j-1)N,\ \ 1\le\ i,j\ \le N$ and $l = m+3(n-1),\ \ 1\le\ m,n\le\ 3$. These nine local basis functions $\tilde{\psi}_{k,l}(x,y),\ l=1,\ldots,9$, are shown in Figure 2.11.

These two dimensional biquadratic IFE local basis functions $\tilde{\psi}_{k,i}(x,y)$ have properties similar to those of their one dimensional counterparts. The following theorem establishes that $\tilde{\psi}_{k,l}(x,y)$ satisfies a partition of unity and that it satisfies the interface conditions (2.121). Additionally, this theorem also points out how these IFE basis functions are related to the standard FE basis functions. There is a definite similarity between this theorem and Theorem 2.2.5 due to the fact that the basis function $\tilde{\psi}_{k,l}(x,y)$ is a separable function and one of its component functions is an one dimensional quadratic IFE local basis function $\tilde{\phi}_{q,i}(x)$, both of which are crucial to the proof of this theorem.

Theorem 2.2.14. *Let $e_k = [x_{2(i-1)}, x_{2i}]\times[y_{2(j-1)}, y_{2j}] \in \mathcal{T}_h$ (where $k = i+(j-1)N$) be an interface element. The biquadratic IFE local basis functions $\tilde{\psi}_{k,l}(x,y),\ l=1,\ldots,9$, have the following properties:*

1. *$\tilde{\psi}_{k,l}(x,y),\ l=1,\ldots,9$ satisfy the interface conditions (2.121).*

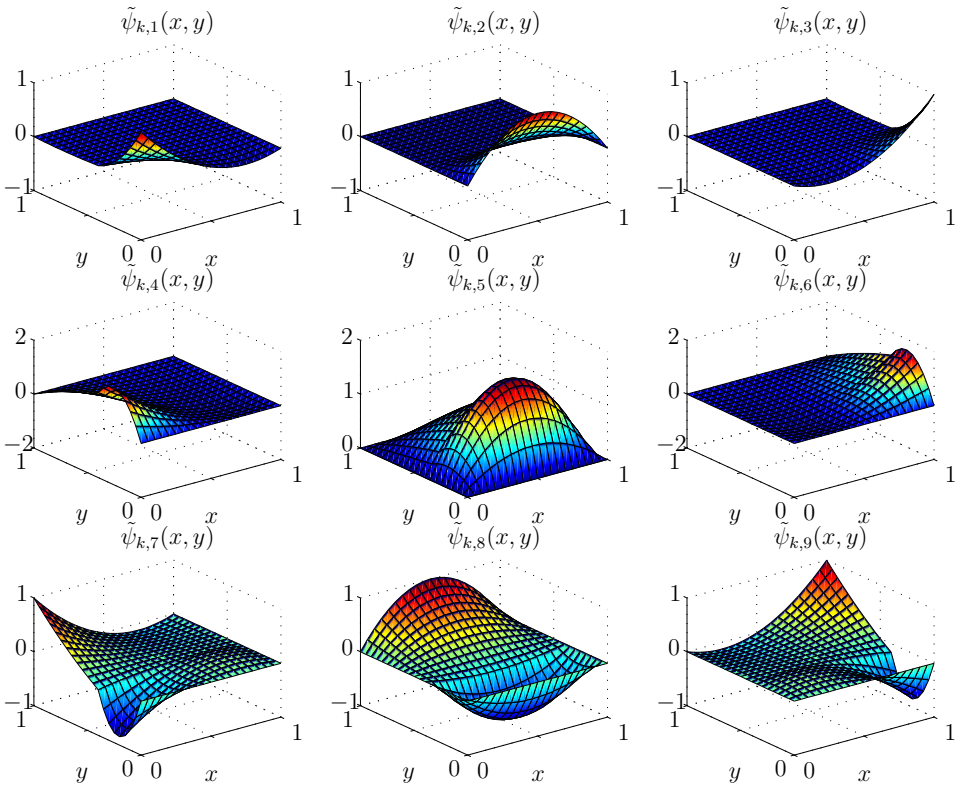


Figure 2.11: The nine local basis functions $\tilde{\psi}_{k,l},\ l = 1, \ldots, 9$, where the element e_k is the unit square and the interface occurs at $y = 0.4$.

2. $\sum_{i=l}^{9} \tilde{\psi}_{k,l}(x,y) \equiv 1$ *for all* $(x,y) \in e_k$.

3. *They are consistent with the standard biquadratic FE functions* $\psi_{k,l}(x,y),\ l = 1,\ldots,9$ *in the following ways:*

$$\lim_{\gamma \to y_{2(j-1)}} \left(\tilde{\psi}_{k,l}(x,y)\right) = \psi_{k,l}(x,y),$$
$$\lim_{\gamma \to y_{2j}} \left(\tilde{\psi}_{k,l}(x,y)\right) = \psi_{k,l}(x,y),$$
$$\lim_{p_2 \to p_1} \left(\tilde{\psi}_{k,l}(x,y)\right) = \psi_{k,l}(x,y).$$

Proof.

1. Notice that each basis function $\tilde{\psi}_{k,l}(x,y)$ is a separable function as defined in (2.129). Thus it follows from Lemma 2.2.12 that

$$\sum_{l=1}^{9} \tilde{\psi}_{k,l}(x,y) \equiv 1 \text{ for all } (x,y) \in e_k,$$

since

$$\sum_{m=1}^{3} \phi_{i,m}(x) \equiv 1 \text{ for all } x \in [x_{2(i-1)}, x_{2i}],$$

and

$$\sum_{n=1}^{3} \tilde{\phi}_{j,n}(y) \equiv 1 \text{ for all } x \in [y_{2(j-1)}, y_{2j}].$$

2. Likewise, the local basis functions $\phi_{i,m}(x),\ m = 1,2,3$ and $\tilde{\phi}_{j,n}(y),\ n = 1,2,3$ satisfy the hypotheses of Lemma 2.2.13 and so the biquadratic IFE local basis functions $\tilde{\psi}_{k,l}(x,y)$ satisfy the interface conditions (2.121).

3. Finally, we come to the relationship between the biquadratic IFE local basis function $\tilde{\psi}_{k,l}(x,y)$ and the standard biquadratic FE local basis function $\psi_{k,l}(x,y)$. Notice that

$$\begin{aligned}\lim_{\gamma \to y_{2(j-1)}} \left(\tilde{\psi}_{k,l}(x,y)\right) &= \lim_{\gamma \to y_{2(j-1)}} \phi_{i,m}(x)\tilde{\phi}_{j,n}(y) \\ &= \phi_{i,m}(x)\left(\lim_{\gamma \to y_{2(j-1)}} \left(\tilde{\phi}_{j,n}(y)\right)\right),\end{aligned}$$

since $\tilde{\psi}_{k,l}(x,y)$ is separable. However, it Theorem 2.2.5 shows that

$$\lim_{\gamma \to y_{2(j-1)}} \left(\tilde{\phi}_{j,n}(y) \right) = \phi_{j,n}(y).$$

So it follows that

$$\lim_{\gamma \to y_{2(j-1)}} = \left(\tilde{\psi}_{k,l}(x,y) \right) = \phi_{i,m}(x)\phi_{j,n}(y) = \psi_{k,l}(x,y).$$

The results for the limits $\lim_{\gamma \to y_{2j}} \left(\tilde{\psi}_{k,l}(x,y) \right)$ and $\lim_{p_2 \to p_1} \left(\tilde{\psi}_{k,l}(x,y) \right)$ follow similarly from the results of Theorem 2.2.5. ■

Following the same procedure used to construct the one dimensional IFE space $\tilde{\mathcal{S}}_{h,2}(0,1)$, we can first define the local IFE space $\tilde{\mathcal{S}}_{h,2,\square}(e_k)$ for each element $e_k \in \mathcal{T}_h$ as follows:

$$\tilde{\mathcal{S}}_{h,2,\square}(e_k) = \begin{cases} span\{\psi_{k,l}(x,y), l = 1,2,\cdots,9\}, & \text{if } e_k \text{ is a non-interface element,} \\ span\{\tilde{\psi}_{k,l}(x,y), l = 1,2,\cdots,9\}, & \text{if } e_k \text{ is an interface element.} \end{cases}$$

Then we can use $\tilde{\mathcal{S}}_{h,2,\square}(e_k)$ to form a global nodal basis function for each node in the partition. Finally, we can define the hierarchical two dimensional biquadratic IFE space $\tilde{\mathcal{S}}_{h,2,\square}((0,1) \times (0,1))$ as the linear space spanned by these global nodal basis functions.

A Biquadratic IFE space with Extra Continuity

Next, we construct a biquadratic IFE space that is derived from the one dimensional quadratic IFE basis function with extra continuity properties, $\bar{\phi}_{q,i}(y)$. Specifically, for a typical interface element $e_k \in \mathcal{T}_h$, we define nine quadratic nodal basis functions as follows:

$$\bar{\psi}_{k,l}(x,y) = \phi_{i,m}(x)\bar{\phi}_{j,n}(y), \tag{2.130}$$

where $k = i + (j-1)N$, $1 \leq i,j \leq N$ and $l = m + 3(n-1)$, $1 \leq m,n \leq 3$. These functions are shown in Figure 2.12.

The biquadratic IFE local basis function $\bar{\psi}_{k,l}(x,y)$ also has properties similar to those of the one dimensional quadratic IFE local basis function $\bar{\phi}_{q,i}(x)$. The connection between $\bar{\psi}_{k,l}(x,y)$ and $\bar{\phi}_{q,i}(x)$ is very similar to the relationship between the biquadratic local basis function $\tilde{\psi}_{k,l}(x,y)$ and the quadratic local basis function $\tilde{\phi}_{q,i}(x)$ that is shown in Theorem 2.2.14. The following theorem similarly shows that $\bar{\psi}_{k,l}(x,y)$ has the partition of unity property and satisfies the interface condition.

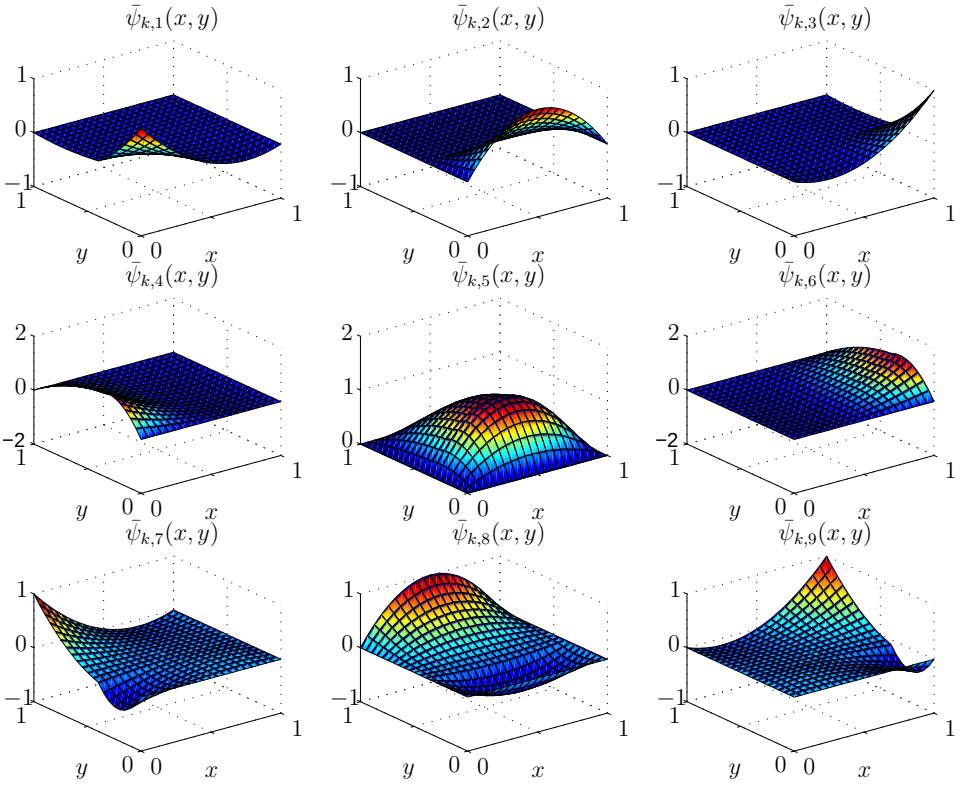


Figure 2.12: The nine local basis functions $\bar{\psi}_{k,l}(x)$, $l = 1, \ldots, 9$, where the element e_k is the unit square and the interface occurs at $y = 0.4$.

Theorem 2.2.15. *Let $e_k = [x_{2(i-1)}, x_{2i}] \times [y_{2(j-1)}, y_{2j}] \in \mathcal{T}_h$ (where $k = i + (j-1)N$) be an interface element. The biquadratic IFE local basis functions $\bar{\psi}_{k,l},\ l = 1, \ldots, 9$, have the following properties:*

1. *They satisfy the interface conditions (2.121).*
2. $\sum_{i=l}^{9} \bar{\psi}_{k,l}(x, y) \equiv 1$ *for all* $(x, y) \in e_k$.
3. *They are consistent with the standard biquadratic FE functions $\psi_{k,l}(x, y),\ l = 1, \ldots, 9$ in the following ways:*

$$\lim_{\gamma \to y_{2(j-1)}} \bar{\psi}_{k,l}(x, y) = \psi_{k,l}(x, y),$$
$$\lim_{\gamma \to y_{2j}} \bar{\psi}_{k,l}(x, y) = \psi_{k,l}(x, y),$$
$$\lim_{p_2 \to p_1} \bar{\psi}_{k,l}(x, y) = \psi_{k,l}(x, y).$$

Proof. This proof is very similar to that given for Theorem 2.2.14 and relies upon the fact that $\bar{\psi}_{k,l}(x, y)$ is defined as a separable function.

First, we notice that $\bar{\psi}_{k,l}(x, y)$ is a separable function and that its components, $\phi_{i,m}(x)$ and $\bar{\phi}_{j,n}(y)$, satisfy the hypotheses of Lemmas 2.2.12 and 2.2.13. Hence it follows in the same way as in Theorem 2.2.14 that $\bar{\psi}_{k,l}(x, y),\ l = 1, \ldots, 9$ satisfy the interface conditions (2.121) and that

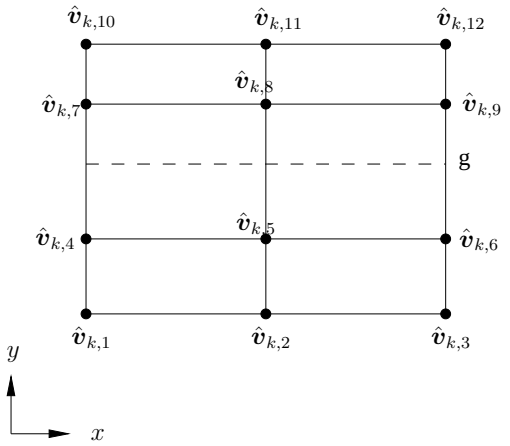
$$\sum_{l=1}^{9} \bar{\psi}_{k,l}(x, y) \equiv 1 \text{ for all } (x, y) \in e_k.$$

The rest of the proof is exactly the same as the proof for Theorem 2.2.14. ■

Now, we can define a local IFE space $\bar{\mathcal{S}}_{h,2,\square}(e_k)$ for each element $e_k \in \mathcal{T}_h$ as follows:

$$\bar{\mathcal{S}}_{h,2,\square}(e_k) = \begin{cases} span\{\psi_{k,l}(x, y), l = 1, 2, \cdots, 9\}, & \text{if } e_k \text{ is a non-interface element,} \\ span\{\bar{\psi}_{k,l}(x, y), l = 1, 2, \cdots, 9\}, & \text{if } e_k \text{ is an interface element.} \end{cases}$$

Then we can use $\bar{\mathcal{S}}_{h,2,\square}(e_k)$ to form a global nodal basis function for each node in the partition. Finally, we can define the two dimensional biquadratic IFE space $\bar{\mathcal{S}}_{h,2,\square}((0, 1) \times (0, 1))$ with extra continuity as the linear space spanned by these global nodal basis functions.



$$\begin{aligned}
\hat{\boldsymbol{v}}_{k,1} &= (s_{k,1}, \hat{t}_{k,1}), \quad & \hat{\boldsymbol{v}}_{k,7} &= (s_{k,1}, \hat{t}_{k,3}),\\
\hat{\boldsymbol{v}}_{k,2} &= (s_{k,2}, \hat{t}_{k,1}), \quad & \hat{\boldsymbol{v}}_{k,8} &= (s_{k,2}, \hat{t}_{k,3}),\\
\hat{\boldsymbol{v}}_{k,3} &= (s_{k,3}, \hat{t}_{k,1}), \quad & \hat{\boldsymbol{v}}_{k,9} &= (s_{k,3}, \hat{t}_{k,3}),\\
\hat{\boldsymbol{v}}_{k,4} &= (s_{k,1}, \hat{t}_{k,2}), \quad & \hat{\boldsymbol{v}}_{k,10} &= (s_{k,1}, \hat{t}_{k,4}),\\
\hat{\boldsymbol{v}}_{k,5} &= (s_{k,2}, \hat{t}_{k,2}), \quad & \hat{\boldsymbol{v}}_{k,11} &= (s_{k,2}, \hat{t}_{k,4}),\\
\hat{\boldsymbol{v}}_{k,6} &= (s_{k,3}, \hat{t}_{k,2}), \quad & \hat{\boldsymbol{v}}_{k,12} &= (s_{k,3}, \hat{t}_{k,4}).
\end{aligned}$$

Figure 2.13: The local quadratic nodes $\hat{\boldsymbol{v}}_{k,l}$ for $\hat{\psi}_{k,l}(x,y)$ on the interface element e_k with interface $y = \gamma$

A Biquadratic IFE Finite Element Space Using Local Refinement

The last type of biquadratic finite element space considered here is derived by using the one dimensional quadratic IFE local basis function $\hat{\phi}_{k,i}(x)$. Recall that in constructing the basis function $\hat{\phi}_{q,i}(x)$ on the element e_q, the interval $[x_{2(q-1)}, x_{2q}]$ is first subdivided into two sub-elements, $[x_{2(q-1)}, \alpha]$ and $[\alpha, x_{2q}]$, where $x = \alpha$ is the location of the interface. Then four basis functions are introduced onto the element e_q where each basis function satisfies four local nodal value specifications (two on each sub-element) and two interface conditions. In two dimensions, when there is an interface at $y = \gamma$, the interface element $e_k = [x_{2(i-1)}, x_{2i}] \times [y_{2(j-1)}, y_{2j}]$ (where $k = i + (j-1)N$) has twelve local nodes as Figure 2.13 shows.

The nodal points in the x and y directions of a typical interface element are

$$\begin{aligned}
s_{k,1} &= x_{2(i-1)}, \qquad & \hat{t}_{k,1} &= y_{2(j-1)},\\
s_{k,2} &= x_{2i-1}, \qquad & \hat{t}_{k,2} &= (y_{2(j-1)} + \gamma)/2,\\
s_{k,3} &= x_{2i}, \qquad & \hat{t}_{k,3} &= (\gamma + y_{2j})/2,\\
& & \hat{t}_{k,4} &= y_{2j}.
\end{aligned} \tag{2.131}$$

Notice that there are four nodal points in the y direction since four nodal value specifications are needed in the direction that crosses the interface so that the interface conditions may be satisfied. Then we define twelve nodal $\hat{\psi}_{k,l}(x,y),\ l = 1, \ldots, 12$ in an interface element e_k as

$$\hat{\psi}_{k,l}(x,y) = \phi_{i,m}(x)\hat{\phi}_{j,n}(y), \tag{2.132}$$

where $k = i + (j-1)N,\ 1 \le\ i, j\ \le N$ and $l = m + 4(n-1),\ 1 \le\ m \le\ 3,\ 1 \le\ n \le\ 4$. These local basis functions are shown in Figure 2.14. While the construction of the IFE local

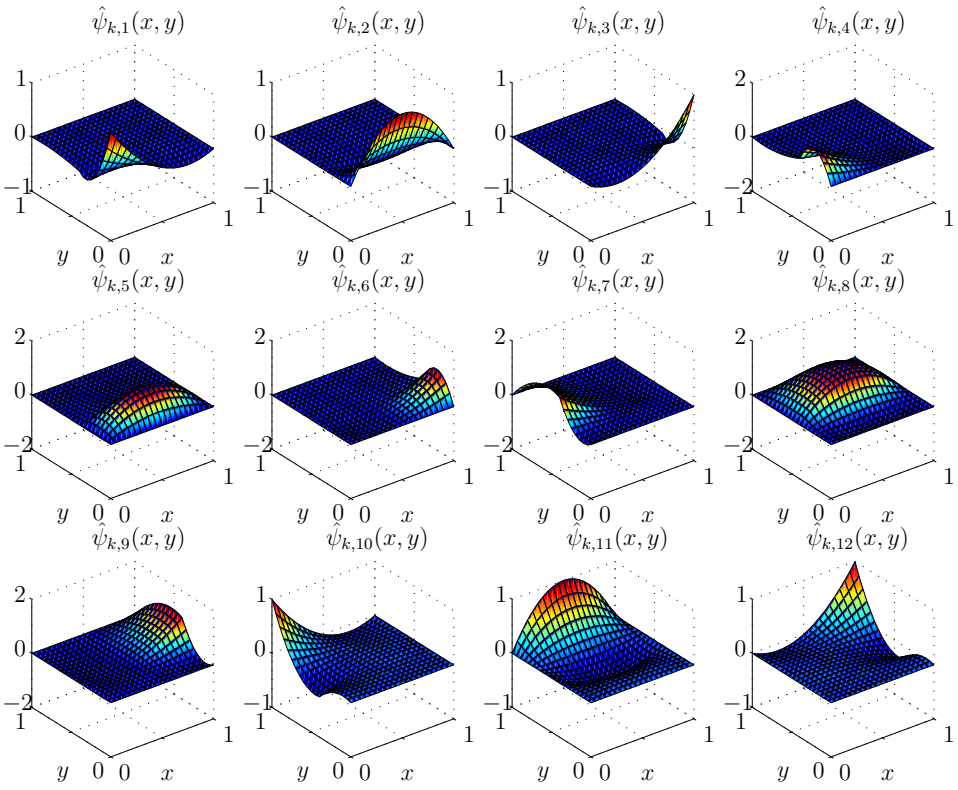


Figure 2.14: The twelve local basis functions $\hat{\psi}_{k,l}(x), \quad l = 1, \ldots, 12$, where the element e_k is the unit square and the interface occurs at $y = 0.4$.

basis function $\hat{\psi}_{k,l}(x,y)$ is similar to the construction used for both $\tilde{\psi}_{k,l}(x,y)$ and $\bar{\psi}_{k,i}(x,y)$, these local basis functions do not satisfy all of the same properties as $\tilde{\psi}_{k,l}(x,y)$ and $\bar{\psi}_{k,i}(x,y)$. For example, these local basis functions do not have a direct relationship with the standard finite element local basis functions. However, since the construction does result in a separable function, then Lemmas 2.2.12 and 2.2.13 can again be used to prove the following theorem regarding the properties of $\hat{\psi}_{k,l}(x,y)$.

Theorem 2.2.16. *Let $e_k = [x_{2(i-1)}, x_{2i}] \times [y_{2(j-1)}, y_{2j}] \in \mathcal{T}_h$ (where $k = i + (j-1)N$) be an interface element. The biquadratic IFE local basis functions $\hat{\psi}_{k,l}$, $l = 1, \ldots, 12$, have the following properties:*

1. *They satisfy the interface conditions (2.121).*
2. $\sum_{i=l}^{12} \hat{\psi}_{k,l}(x,y) \equiv 1$ *for all* $(x,y) \in e_k$.

Proof. The proof is the same as those given for Theorems 2.2.14 and 2.2.15 since $\hat{\psi}_{k,l}(x,y)$ is a separable function whose components, $\phi_{i,m}(x)$ and $\hat{\phi}_{j,n}(y)$ satisfy the hypotheses of Lemmas 2.2.12 and 2.2.13. ■

Now, we are ready to define the IFE space based on local refinement. We first define a local IFE space $\hat{\mathcal{S}}_{h,2,\square}(e_k)$ for each element $e_k \in \mathcal{T}_h$ as follows:

$$\hat{\mathcal{S}}_{h,2,\square}(e_k) = \begin{cases} span\{\psi_{k,l}(x,y), l = 1, 2, \cdots, 9\}, & \text{if } e_k \text{ is a non-interface element,} \\ span\{\hat{\psi}_{k,l}(x,y), l = 1, 2, \cdots, 12\}, & \text{if } e_k \text{ is an interface element.} \end{cases}$$

Then we can use $\hat{\mathcal{S}}_{h,2,\square}(e_k)$ to form a global nodal basis function for each node in the partition. Finally, we can define the two dimensional biquadratic IFE space $\hat{\mathcal{S}}_{h,2,\square}((0,1) \times (0,1))$ based upon local refinement as the linear space spanned by these global nodal basis functions.

Chapter 3

Interpolation Accuracy

In this chapter, we investigate the approximation capabilities of interpolants that are formed in the quadratic and biquadratic IFE spaces introduced in the previous chapter. Subsequent error results using the Galerkin Finite Element Method (Chapter 4) and the Least Squares Finite Element Method (Chapter 5) depend upon the ability of the IFE spaces to provide good interpolation errors. Since the construction of the quadratic and biquadratic basis functions is not a unique process, the different quadratic IFE spaces in one dimension and in two dimensions are tested here in order to determine their approximation capabilities numerically. Interpolation estimates for standard finite element spaces can be found in [6, 8, 9, 17].

3.1 Introduction

When standard quadratic and biquadratic basis functions are used to interpolate a function with an interface by using the body fit partition, it is well known that the errors in these interpolants typically have the following estimates [6, 9]:

$$\|u - I_h u\|_\infty \leq Ch^3, \tag{3.1}$$

$$\|u - I_h u\|_0 \leq Ch^3, \tag{3.2}$$

$$\|u - I_h u\|_1 \leq Ch^2, \tag{3.3}$$

provided that u has appropriate regularity. The behavior of interpolants from linear IFE spaces in one dimension is already established by Li in [40] using the L^∞ norm. Behavior of the interpolants from linear and bi-linear IFE spaces in two dimensions is established by Lin et. al. in [20] and [44]. The main purpose of this chapter is to find out the approximation

capability of interpolants that are formed in the quadratic IFE spaces introduced in the previous chapter.

However, from the definition directly, it is not very clear how the interpolants that are formed in these quadratic IFE spaces behave because the process for defining quadratic and biquadratic IFE spaces is not unique as evidenced by the constructions of the IFE spaces $\tilde{\mathcal{S}}_{h,2}(0,1)$, $\bar{\mathcal{S}}_{h,2}(0,1)$ and $\hat{\mathcal{S}}_{h,2}(0,1)$ in one dimension, and the IFE spaces $\tilde{\mathcal{S}}_{h,2,\square}\big((0,1)^2\big)$, $\bar{\mathcal{S}}_{h,2,\square}((0,1)^2)$ and $\hat{\mathcal{S}}_{h,2,\square}\big((0,1)^2\big)$ in two dimensions (see Chapter 2). The basis functions from each of these spaces have some similar properties but their construction shows that the basis functions themselves are fundamentally different on the interface elements. For example, because of the requirements that are imposed during the construction of the basis space $\bar{\mathcal{S}}_{h,2}(0,1)$, it is known that the fluxes of the local basis functions $\bar{\phi}_{k,i}(x)$ are continuous on each element. Hence $p\bar{\phi}''_{k,i}(x)$ is continuous. On the other hand, the local basis functions in the IFE space $\tilde{\mathcal{S}}_{h,2}(0,1)$ do not have this same feature. We can easily see this from the local nodal basis function $\tilde{\phi}_{k,1}(x) = l_{1,2}(x)l_{1,3}(x)$. It follows that $\tilde{\phi}''_{k,1}(x) = 2l'_{1,2}(x)l'_{1,3}(x)$. Then it is seen that $p\tilde{\phi}''_{k,1}(x)$ is not continuous.

In this chapter, we establish numerical results that allow us to make some conclusions about the approximation capabilities of each of the quadratic and biquadratic IFE spaces that are discussed in Chapter 2. We use $I_h u$ to denote the interpolant of a function $u(x)$ in an IFE space with a mesh size h, and use u_h to denote the finite element solution of an interface value problem in an IFE space with a mesh size h where the interface value problem has an exact solution u. Similar notations are used in the two dimensional case. To measure the errors, we use the following norms in the involved Sobolev spaces:

$$\|u\|_\infty = \operatorname*{ess\,sup}_{x\in\Omega} |u(x)|, \tag{3.4}$$

$$\|u\|_0 = \left(\int_\Omega u^2 \, d\mathbf{x}\right)^{1/2}, \tag{3.5}$$

$$\|u\|_1 = \left(\|u\|_0^2 + \|u'\|_0^2\right)^{1/2}. \tag{3.6}$$

Using these norms, we can measure the magnitudes of the errors in the interpolants with the following quantities:

$$E_{h,s}(I_h u) = \|u - I_h u\|_s, \tag{3.7}$$

with $s = 0, 1$, or $s = \infty$. Similar notation is introduced for finite element approximations in the next chapter. When it is clear which norm is used, the error is simply referred to as $E_h(I_h u)$. Similarly, if it is clear from the context that the error of the interpolant $I_h u$ is being considered then the error is simply referred to as E_h or $E_{h,s}$.

We say that an interpolant, $I_h u$, is of order r if

$$E_h \approx Ch^r, \tag{3.8}$$

for some constant C. Alternatively, one may say that E_h is $\mathcal{O}(h^r)$ which means that the norm of the residual tends to zero as fast as h^r. One can also say that the interpolant, $I_h u$, is at least of order r if

$$E_h \leq Ch^r, \tag{3.9}$$

for some constant C. Generally, equations (3.8), (3.9) and $\mathcal{O}(h^r)$ are used interchangeably. As usual, we can identify the order by carrying out a linear least squares regression on a set of values of $\log E_h$ vs. $\log h$ that is generated using various values of h.

As usual, we assess the errors numerically. The L^∞ error, $E_{h,\infty}$, is approximated by evaluating the residual of the interpolants residual, $u - I_h u$, p times on each element and then using the maximum of these values over all of the elements. The interface element e_k is treated differently by considering the two natural sub-elements: $e_{k,i} = e_k \cap \Omega_i$ for $i = 1, 2$. So the residual of the interpolant is evaluated p times on each sub-element of the interface element. The L^2-norm error, $E_{h,0}$, is approximated by using a composite Gaussian quadrature. The integral of the residual of the interpolant is calculated element-by-element using simple Gaussian quadrature with q Gaussian nodes on each element where the integer q is chosen so that the desired degree of accuracy matches the requirement determined by the order of the resulting error. For example, if the $E_{h,0} = \mathcal{O}(h^r)$ then enough Gaussian quadrature nodes should be used to ensure that the integrals are calculated with an order which is at least $\mathcal{O}(h^r)$. Again, interface elements are treated differently than non-interface elements by separating them into two pieces. Integration over the whole domain is accomplished by adding the resulting numerical integrations values over all the elements together. The H^1 error, $E_{h,1}$, is calculated in the same way as the L^2 error.

3.2 Interpolation in One Dimension using Quadratic IFE Spaces

Recall that we use $\bar{\mathcal{S}}_{h,2}(\Omega), \hat{\mathcal{S}}_{h,2}(\Omega)$ and $\tilde{\mathcal{S}}_{h,2}(\Omega)$ to denote IFE spaces that are spanned by the nodal basis functions $\bar{\phi}_i(x), \hat{\phi}_i(x)$ and $\tilde{\phi}_i(x)$, respectively. Numerical results from these IFE spaces are to be considered. First, we introduce two suitable test functions which serve to emphasize different features of the basis functions in order to provide a good testing ground for these IFE spaces. Then the errors in the interpolants of these test functions in the IFE spaces are presented.

Even though all the functions in the three IFE spaces share some fundamental properties such as being able to satisfy the jump conditions, functions in these IFE spaces are indeed different in the continuity of their flux across the interface or in the location of their nodal values. It is therefore interesting to see how each IFE spaces performs in generating approximations

to test functions whose fluxes behave differently at the interface. We consider two test functions. Since one of the goals here is to develop IFE spaces that are effective for solving the forward interface problem when used in the Galerkin FEM or the Least Squares FEM, the test functions are chosen to be the exact solution of an elliptic interface boundary value problem.

Here we restate our general elliptic boundary value problem in one dimension,

$$\begin{array}{rcl} -\big(p(x)u'(x)\big)' & = & f, \text{ if } x \in (0,1), \\ u(0) = u(1) & = & 0, \end{array} \tag{3.10}$$

where p is a piecewise constant function defined by

$$p(x) = \begin{cases} p_1, & \text{if } x < \alpha, \\ p_2, & \text{if } x \geq \alpha, \end{cases} \tag{3.11}$$

and $p_1, p_2 > 0$ are two different constants. Also, it is also required that u and pu' are continuous across the interface at $x = \alpha$ and so

$$u, \quad pu' \in C(0,1). \tag{3.12}$$

Consequently, this implies that

$$\big[u\big]_\alpha = 0 \text{ and } \big[pu'\big]_\alpha = 0. \tag{3.13}$$

The first test function $u_1(x)$ is chosen so that it is the solution of (3.10) when the right hand side function is defined by

$$f_1(x) = e^x. \tag{3.14}$$

Using simple integration along with the boundary conditions of (3.10) and the interface conditions (3.13) to evaluate the constants of integration, it follows that the function $u_1(x)$ has the following formula:

$$u_1(x) = \begin{cases} \dfrac{1}{-D}\big(e^x(\alpha p_2 + p_1) + \alpha(p_1 - p_2) - p_1 \\ \qquad +x(e^\alpha(p_1 - p_2) + p_2 - ep_1)\big)\ , & \text{if } x < \alpha, \\ \dfrac{p_1}{p_2 D}\big(e^x p_1 - \alpha(e^x - e)(p_1 - p_2) \\ \qquad +e^\alpha(p_1 - p_2)(x-1) + p_2(x-1) - ep_1 x\big)\ , & \text{if } x \geq \alpha, \end{cases} \tag{3.15}$$

where

$$D = p_1 + \alpha(p_2 - p_1). \tag{3.16}$$

Since u_1 satisfies (3.10), it follows that u_1 also satisfies (3.12). Additionally, since the right hand side function $f_1(x)$ is smooth, it follows that pu_1'' is also continuous. The next test function is defined in such a way that this is not true anymore.

The second test function, $u_2(x)$, is defined as the exact solution of the interface problem (3.10) when the right hand side function is defined by

$$f_2(x) = \begin{cases} e^x, & \text{if } x < \alpha, \\ -e^{1-x}, & \text{if } x \geq \alpha. \end{cases} \tag{3.17}$$

Clearly, this function $f_2(x)$ is discontinuous at the interface $x = \alpha$ and so it follows that although u_2 satisfies (3.12) the derivative of the flux, pu_2'' is discontinuous at the interface. The exact formula for $u_2(x)$ can be obtained by simple integration and then applying the boundary values and the interface conditions to solve for constants of integration. Hence, the function $u_2(x)$ is defined by

$$u_2(x) = \begin{cases} (-e^x + c_1 x + 1)/p_1, & \text{if } x < \alpha, \\ (e^{1-x} + c_2(x-1) - 1)/p_2, & \text{if } x \geq \alpha, \end{cases} \tag{3.18}$$

where

$$c_1 = \frac{1}{D}\left(e^{-\alpha}(e^{\alpha}(e^{\alpha}-1)(p_1+p_2) - \alpha(e^{2\alpha}-e)p_1)\right) , \tag{3.19}$$

$$c_2 = c_1 - e^{\alpha} + e^{1-\alpha} , \tag{3.20}$$

and D is defined in (3.16).

Since $u_1(x)$ and $u_2(x)$ both satisfy the elliptic interface problem, the fluxes of both functions, pu_1' and pu_2', are continuous on the domain $\Omega = (0, 1)$. However, because of the construction of u_1 and u_2, the functions differ in that pu_1'' is continuous at the interface and pu_2'' is discontinuous at the interface. This is an interesting difference to consider since the local basis functions $\tilde{\phi}_{k,i}(x)$ and $\bar{\phi}_{k,i}(x)$ also differ in this way. The IFE local basis function $\bar{\phi}_{k,i}(x)$ is constructed so that $p\bar{\phi}_{k,i}''(x)$ is continuous at the interface. On the other hand, since we know it is unique, then this implies that $p\tilde{\phi}_{k,i}''(x)$ must not be continuous at the interface.

In the numerical experiments presented below, we generate interpolants of the test functions $u_1(x)$ and $u_2(x)$ in the IFE spaces $\tilde{\mathcal{S}}_{h,2}(0,1), \bar{\mathcal{S}}_{h,2}(0,1)$ and $\hat{\mathcal{S}}_{h,2}(0,1)$. A sequence of partition sizes $h = 1/8, 1/16, 1/32, 1/64, 1/128$ and $h = 1/256$ is used to generate data for each case. Also, the interface is located at $\alpha = 1/\pi$ so that it never occurs as a node of the finite element partition. Interpolation data for the L^∞ error is generated using six equally spaced function evaluations on each element while integration on each element is done with four gaussian nodes. The interpolation of u_1 and u_2 is also explored when different coefficient values for p are specified. First, data is generated for the errors in the interpolants when the coefficient

values are $p_1 = 1$ and $p_2 = 10$. Subsequently, u_1 and u_2 are interpolated when the coefficient value p_2 is increased to $p_2 = 10000$ while keeping p_1 fixed. The variable β is used to refer to the ratio between the two coefficient values and so $\beta = p_2/p_1$. We only show data for the cases when $\beta = 10000$ if they offer significantly different results from the cases when $\beta = 10$. Orders of the error are determined by doing a linear least squares regression of the $\log h$ vs. $\log E_h$ data for the L^∞, L^2 and H^1 norms which is carried out using the `polyfit` command in MATLAB.

3.2.1 Interpolation Accuracy of the IFE Space $\tilde{\mathcal{S}}_{h,2}(0,1)$

We start by the looking at the errors of the interpolants of the test functions $u_1(x)$ and $u_2(x)$ in the hierarchical IFE space $\tilde{\mathcal{S}}_{h,2}(0,1)$. Errors in the interpolants $I_h u_1(x)$ and $I_h u_2(x)$ from $\tilde{\mathcal{S}}_{h,2}(0,1)$ are shown in Tables 3.1 and 3.2 respectively when the coefficient p has a ratio of $\beta = 10$. The data here shows that the errors do not always decrease as h decreases, although

h	$\|u_1 - I_h u_1\|_\infty$	$\|u_1 - I_h u_1\|_0$	$\|u_1 - I_h u_1\|_1$
1/8	2.398×10^{-4}	4.950×10^{-5}	2.667×10^{-3}
1/16	5.293×10^{-5}	4.353×10^{-6}	1.162×10^{-3}
1/32	2.612×10^{-5}	1.628×10^{-6}	6.682×10^{-4}
1/64	9.317×10^{-6}	5.441×10^{-7}	2.770×10^{-4}
1/128	2.177×10^{-7}	9.784×10^{-9}	8.387×10^{-6}
1/256	5.287×10^{-7}	1.763×10^{-8}	2.912×10^{-5}

Table 3.1: $E_{h,s}(I_h u_1)$ where $s = \infty, 0, 1$ and $I_h u_1 \in \tilde{\mathcal{S}}_{h,2}(0,1)$ with $\beta = 10$.

h	$\|u_2 - I_h u_2\|_\infty$	$\|u_2 - I_h u_2\|_0$	$\|u_2 - I_h u_2\|_1$
1/8	3.680×10^{-4}	7.671×10^{-5}	4.159×10^{-3}
1/16	8.547×10^{-5}	6.827×10^{-6}	1.881×10^{-3}
1/32	4.368×10^{-5}	2.714×10^{-6}	1.118×10^{-3}
1/64	1.580×10^{-5}	9.220×10^{-7}	4.695×10^{-4}
1/128	3.702×10^{-7}	1.626×10^{-8}	1.401×10^{-5}
1/256	9.018×10^{-7}	3.008×10^{-8}	4.966×10^{-5}

Table 3.2: $E_{h,s}(I_h u_2)$ where $s = \infty, 0, 1$ and $I_h u_2 \in \tilde{\mathcal{S}}_{h,2}(0,1)$ with $\beta = 10$.

the errors do tend to decrease in an oscillatory fashion which becomes more apparent when we look at the numerical results generated with a larger set of values of h. The oscillation of the errors is more visible in Figures 3.1 and 3.2 for the errors in the interpolants $I_h u_1$ and $I_h u_2$ respectively. These figures show a larger selection of the error data using a variety of h

values ranging between $h = 1/8$ and $h = 1/256$. Applying linear regression to the data in the tables, we can see that the errors in these numerical results obey the following relationships:

$$\begin{aligned} E_{h,\infty} = \|u_1 - I_h u_1\|_\infty &\approx 0.016\ h^{1.983}, \\ E_{h,0} = \|u_1 - I_h u_1\|_0 &\approx 0.006\ h^{2.436}, \\ E_{h,1} = \|u_1 - I_h u_1\|_1 &\approx 0.093\ h^{1.577}, \end{aligned} \tag{3.21}$$

when $u_1(x)$ is interpolated using the IFE space $\tilde{\mathcal{S}}_{h,2}(0,1)$ and

$$\begin{aligned} E_{h,\infty} = \|u_2 - I_h u_2\|_\infty &\approx 0.024\ h^{1.954}, \\ E_{h,0} = \|u_2 - I_h u_2\|_0 &\approx 0.009\ h^{2.408}, \\ E_{h,1} = \|u_2 - I_h u_2\|_1 &\approx 0.141\ h^{1.554}, \end{aligned} \tag{3.22}$$

when $u_2(x)$ is interpolated using the IFE space $\tilde{\mathcal{S}}_{h,2}(0,1)$. These orders are worse than expected for a FE space that is spanned by quadratic basis functions. For example, according to (3.1)-(3.3), then $E_{h,\infty}$ and $E_{h,0}$ should be $\mathcal{O}(h^3)$ while $E_{h,1}$ should be $\mathcal{O}(h^2)$. While the orders of $E_{h,0}$ and $E_{h,1}$ only appear to be a little more than one fourth of an order less than what one expects, the orders of $E_{h,\infty}$ appear to be a little more than three fourths of an order less than what is expected. Thus the IFE space $\tilde{\mathcal{S}}_{h,2}(0,1)$ tends to underperform in the interpolation of both u_1 and u_2. The data that is generated when the coefficient ratio

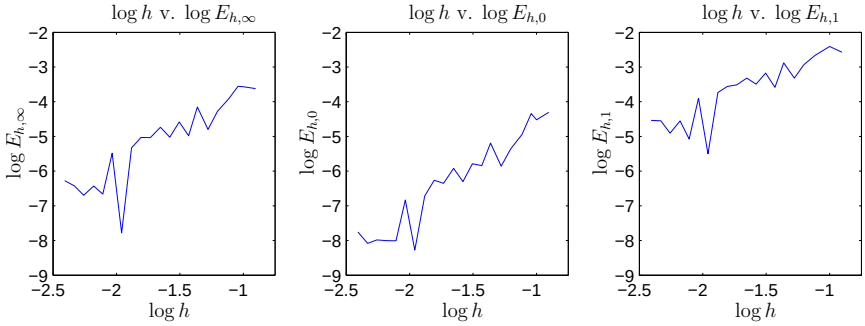


Figure 3.1: $E_{h,s}(I_h u_1)$ where $s = \infty, 0, 1$ with $I_h u_1 \in \tilde{\mathcal{S}}_{h,2}(0,1)$ and $\beta = 10$.

is changed to $\beta = 10000$ is not significantly different than when the ratio is $\beta = 10$. The resulting orders of the errors of the interpolants is comparable to (3.21) and (3.22). When $\beta = 10000$ and $I_h u_1 \in \tilde{\mathcal{S}}_{h,2}(0,1)$, the errors obey the relationships

$$\begin{aligned} E_{h,\infty} = \|u_1 - I_h u_1\|_\infty &\approx 0.006\ h^{1.936}, \\ E_{h,0} = \|u_1 - I_h u_1\|_0 &\approx 0.001\ h^{2.200}, \\ E_{h,1} = \|u_1 - I_h u_1\|_1 &\approx 0.038\ h^{1.432}. \end{aligned} \tag{3.23}$$

On the other hand, when $\beta = 10000$ and $I_h u_2 \in \tilde{\mathcal{S}}_{h,2}(\Omega)$, the errors obey the relationships

$$\begin{aligned} E_{h,\infty} = \|u_2 - I_h u_2\|_\infty &\approx 0.006\ h^{1.900}, \\ E_{h,0} = \|u_2 - I_h u_2\|_0 &\approx 0.002\ h^{2.162}, \\ E_{h,1} = \|u_2 - I_h u_2\|_1 &\approx 0.042\ h^{1.395}. \end{aligned} \tag{3.24}$$

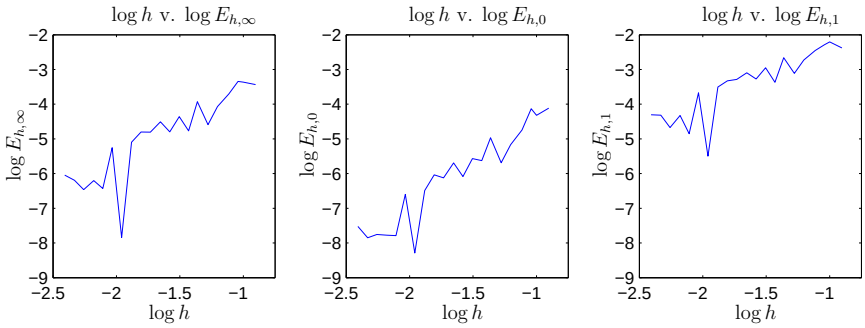


Figure 3.2: $E_{h,s}(I_h u_2)$ where $s = \infty, 0, 1$ with $I_h u_2 \in \tilde{\mathcal{S}}_{h,2}(0,1)$ and $\beta = 10$.

For the increased value in the ratio p_2/p_1, the errors in the interpolants in $\tilde{\mathcal{S}}_{h,2}(0,1)$ still tend to decrease in an oscillatory way as h decreases.

There appears to be no significant difference in the way that the IFE space $\tilde{\mathcal{S}}_{h,2}(0,1)$ is able to interpolate the function $u_1(x)$ compared to the function $u_2(x)$. The errors of the interpolants $I_h u_1$ and $I_h u_2$ have similar orders and similar oscillatory behavior. Even though the nodal basis functions $\tilde{\phi}_i(x)$ that span $\tilde{\mathcal{S}}_{h,2}(0,1)$ have a C^0 flux like the test function $u_2(x)$, this does not seem to translate into the better performance of the interpolant errors for $u_2(x)$ as we anticipate.

3.2.2 Interpolation Accuracy of the IFE Space $\bar{\mathcal{S}}_{h,2}(0,1)$

Next we consider the errors in the interpolants of the functions $u_1(x)$ and $u_2(x)$ using the quadratic IFE space $\bar{\mathcal{S}}_{h,2}(0,1)$ which is the span of the nodal IFE basis functions $\bar{\phi}_i(x)$. Note that functions $\bar{\phi}_i(x)$ spanning $\bar{\mathcal{S}}_{h,2}(0,1)$ are derived using extra continuity requirements. They are more similar to the test function u_1 than u_2 since both $p\bar{\phi}_i''(x)$ and $pu_1''(x)$ are continuous at the interface. Therefore, it is rather natural to anticipate that $I_h u_1$ has a smaller error than $I_h u_2$, and this is confirmed by the numerical experiments that we present here.

The errors of the interpolants of $u_1(x)$ and $u_2(x)$ in $\bar{\mathcal{S}}_{h,2}(0,1)$ are shown in Tables 3.3 and 3.4 respectively when the ratio of the coefficients is $\beta = 10$. According to Table 3.3, the errors in the interpolant $I_h u_1$ appear to decrease whenever h decreases while Table 3.4 shows that the errors in the interpolant $I_h u_2$ decrease only in an oscillatory fashion. Also, the errors in $I_h u_1$ appear to be a couple of orders of magnitude less than those for the interpolant $I_h u_2$. Additionally, the magnitude of the errors in $I_h u_2$ is similar to the magnitude of the errors that is obtained when using the hierarchical IFE space $\tilde{\mathcal{S}}_{h,2}(0,1)$.

h	$\|u_1 - I_h u_1\|_\infty$	$\|u_1 - I_h u_1\|_0$	$\|u_1 - I_h u_1\|_1$
1/8	1.903×10^{-5}	7.375×10^{-6}	3.844×10^{-4}
1/16	2.600×10^{-6}	9.562×10^{-7}	9.934×10^{-5}
1/32	3.293×10^{-7}	1.194×10^{-7}	2.482×10^{-5}
1/64	4.144×10^{-8}	1.499×10^{-8}	6.236×10^{-6}
1/128	5.197×10^{-9}	1.872×10^{-9}	1.554×10^{-6}
1/256	6.532×10^{-10}	2.356×10^{-10}	3.909×10^{-7}

Table 3.3: $E_{h,s}(I_h u_1)$ where $s = \infty, 0, 1$ with $I_h u_1 \in \bar{\mathcal{S}}_{h,2}(0,1)$ and $\beta = 10$.

h	$\|u_2 - I_h u_2\|_\infty$	$\|u_2 - I_h u_2\|_0$	$\|u_2 - I_h u_2\|_1$
1/8	4.360×10^{-4}	9.072×10^{-5}	4.918×10^{-3}
1/16	1.968×10^{-5}	1.790×10^{-6}	4.467×10^{-4}
1/32	1.492×10^{-5}	9.327×10^{-7}	3.829×10^{-4}
1/64	9.794×10^{-6}	5.718×10^{-7}	2.911×10^{-4}
1/128	5.703×10^{-7}	2.499×10^{-8}	2.153×10^{-5}
1/256	7.612×10^{-7}	2.539×10^{-8}	4.192×10^{-5}

Table 3.4: $E_{h,s}(I_h u_2)$ where $s = \infty, 0, 1$ with $I_h u_2 \in \bar{\mathcal{S}}_{h,2}(0,1)$ and $\beta = 10$.

Applying linear regression to the data in Table 3.3, we can see that the errors in the interpolant $I_h u_1(x)$ in the IFE space $\bar{\mathcal{S}}_{h,2}(0,1)$ when $\beta = 10$ obey the following estimates

$$\begin{aligned} E_{h,\infty} = \|u_1 - I_h u_1\|_\infty &\approx 0.010\ h^{2.973}, \\ E_{h,0} = \|u_1 - I_h u_1\|_0 &\approx 0.004\ h^{2.990}, \\ E_{h,1} = \|u_1 - I_h u_1\|_1 &\approx 0.025\ h^{1.991}. \end{aligned} \tag{3.25}$$

Similarly, the orders of the interpolant $I_h u_2(x)$ have the following estimates:

$$\begin{aligned} E_{h,\infty} = \|u_2 - I_h u_2\|_\infty &\approx 0.008\ h^{1.764}, \\ E_{h,0} = \|u_2 - I_h u_2\|_0 &\approx 0.003\ h^{2.235}, \\ E_{h,1} = \|u_2 - I_h u_2\|_1 &\approx 0.045\ h^{1.368}. \end{aligned} \tag{3.26}$$

The orders for the interpolant $I_h u_1$ agrees with our expectation for a FE space using quadratic polynomials, while the orders of $I_h u_2$ indicate that the IFE space $\tilde{\mathcal{S}}_{h,2}(0,1)$ has a weaker approximation capability than the usual quadratic FE space.

Figure 3.3 shows that there exists a very good correlation between a decrease in h and the decrease in the errors in $I_h u_1$. This figure shows error data for a larger selection of h values than the tables with h ranging between $h = 1/8$ and $h = 1/256$. The nice correlation between the errors in $I_h u_1$ and the value of h is further evidenced by the order of the errors shown in (3.25). The errors in the interpolant $I_h u_2$, on the other hand, have a decreasing oscillatory

behavior pattern (see Figure 3.4) that is very similar to that evidenced by the interpolants from the IFE space $\tilde{\mathcal{S}}_{h,2}(0,1)$. The order of $E_{h,\infty}(I_h u_2)$ is a little more than three fourths of an order less than the desired h^3.

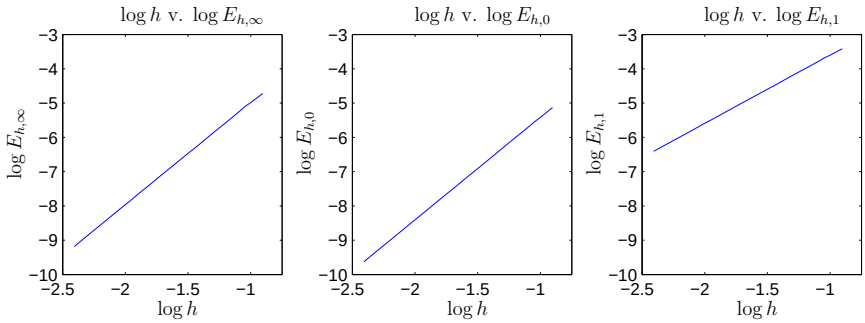


Figure 3.3: $E_{h,s}(I_h u_1)$ where $s = \infty, 0, 1$ with $I_h u_1 \in \bar{\mathcal{S}}_{h,2}(0,1)$ and $\beta = 10$.

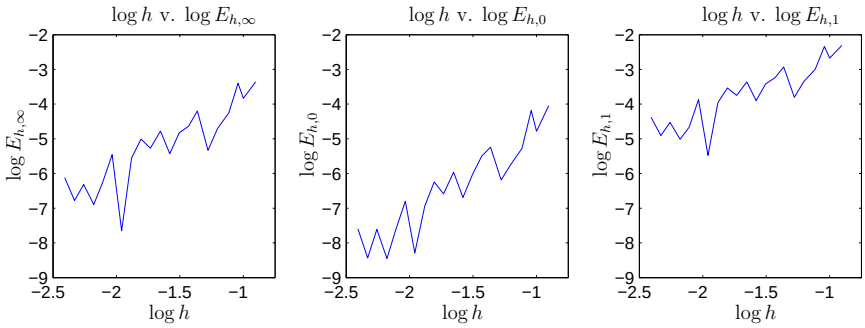


Figure 3.4: $E_{h,s}(I_h u_2)$ where $s = \infty, 0, 1$ with $I_h u_2 \in \bar{\mathcal{S}}_{h,2}(0,1)$ and $\beta = 10$.

Again, the ratio p_2/p_1 is increased to see the effects upon the errors of the interpolants for $u_1(x)$ and $u_2(x)$ with the IFE space $\bar{\mathcal{S}}_{h,2}(0,1)$. Like the behavior of the interpolants from the space $\tilde{\mathcal{S}}_{h,2}(0,1)$, the errors of the interpolants in the IFE space $\bar{\mathcal{S}}_{h,2}(0,1)$ and their orders are not significantly different from the case when $\beta = 10$.

As we expect, the IFE space $\bar{\mathcal{S}}_{h,2}(0,1)$ performs better in the interpolation of the function u_1 which has a differentiable flux similar to the basis functions of the IFE space.

3.2.3 Interpolation Accuracy of the IFE Space $\hat{\mathcal{S}}_{h,2}(0,1)$

Finally, we consider the interpolation errors of the interpolants in the quadratic IFE space $\hat{\mathcal{S}}_{h,2}(0,1)$ for the test functions $u_1(x)$ and $u_2(x)$. The errors in these interpolants are shown in Tables 3.5 and 3.6 respectively when the ratio of the coefficients is $\beta = 10$. The interpolants

h	$\|u_1 - I_h u_1\|_\infty$	$\|u_1 - I_h u_1\|_0$	$\|u_1 - I_h u_1\|_1$
1/8	1.903×10^{-5}	6.677×10^{-6}	3.499×10^{-4}
1/16	2.600×10^{-6}	9.544×10^{-7}	9.913×10^{-5}
1/32	3.293×10^{-7}	1.192×10^{-7}	2.474×10^{-5}
1/64	4.144×10^{-8}	1.490×10^{-8}	6.182×10^{-6}
1/128	5.197×10^{-9}	1.867×10^{-9}	1.551×10^{-6}
1/256	6.532×10^{-10}	2.347×10^{-10}	3.895×10^{-7}

Table 3.5: $E_{h,s}(I_h u_1)$ where $s = \infty, 0, 1$ and $I_h u_1 \in \hat{\mathcal{S}}_{h,2}(0,1)$ and $\beta = 10$.

h	$\|u_2 - I_h u_2\|_\infty$	$\|u_2 - I_h u_2\|_0$	$\|u_2 - I_h u_2\|_1$
1/8	1.903×10^{-5}	6.543×10^{-6}	3.431×10^{-4}
1/16	2.600×10^{-6}	9.417×10^{-7}	9.797×10^{-5}
1/32	3.293×10^{-7}	1.176×10^{-7}	2.440×10^{-5}
1/64	4.144×10^{-8}	1.469×10^{-8}	6.094×10^{-6}
1/128	5.197×10^{-9}	1.841×10^{-9}	1.529×10^{-6}
1/256	6.532×10^{-10}	2.314×10^{-10}	3.841×10^{-7}

Table 3.6: $E_{h,s}(I_h u_2)$ where $s = \infty, 0, 1$ with $I_h u_2 \in \hat{\mathcal{S}}_{h,2}(0,1)$ and $\beta = 10$.

perform well for both test functions, $u_1(x)$ and $u_2(x)$, as we see by the agreement between their orders with the orders that are generated in the usual quadratic FE space with a body fit partition. Specifically, using linear regression, we can see that the interpolants $I_h u_1$ satisfies the following relationships

$$\begin{aligned} E_{h,\infty} = \|u_1 - I_h u_1\|_\infty &\approx 0.010\ h^{2.973}, \\ E_{h,0} = \|u_1 - I_h u_1\|_0 &\approx 0.003\ h^{2.971}, \\ E_{h,1} = \|u_1 - I_h u_1\|_1 &\approx 0.022\ h^{1.973}, \end{aligned} \tag{3.27}$$

while the interpolant $I_h u_2$ obeys

$$\begin{aligned} E_{h,\infty} = \|u_2 - I_h u_2\|_\infty &\approx 0.010\ h^{2.973}, \\ E_{h,0} = \|u_2 - I_h u_2\|_0 &\approx 0.003\ h^{2.969}, \\ E_{h,1} = \|u_2 - I_h u_2\|_1 &\approx 0.022\ h^{1.972}. \end{aligned} \tag{3.28}$$

There is no oscillation seen in the behavior of the errors of the interpolants in this space as the value of h decreases. Similar approximation behavior of $\hat{\mathcal{S}}_{h,2}(0,1)$ is observed for the case in which the ratio of the coefficients is increased to $\beta = 10000$ for the test functions $u_1(x)$ and $u_2(x)$. Increasing the ratio p_2/p_1 of the coefficient p has little effect upon the errors of the interpolants and the resulting orders of the errors of the interpolants are virtually identical. For example, when $I_h u_2 \in \hat{\mathcal{S}}_{h,2}(0,1)$ and $\beta = 10000$ then the errors obey the

following relationships

$$\begin{aligned} E_{h,\infty} = \|u_1 - I_h u_1\|_\infty &\approx 0.010\ h^{2.973}, \\ E_{h,0} = \|u_1 - I_h u_1\|_0 &\approx 0.003\ h^{2.969}, \\ E_{h,1} = \|u_1 - I_h u_1\|_1 &\approx 0.022\ h^{1.971}. \end{aligned} \tag{3.29}$$

Similarly, when $I_h u_2 \in \hat{\mathcal{S}}_{h,2}(0,1)$ and $\beta = 10000$ then the errors obey the following relationships

$$\begin{aligned} E_{h,\infty} = \|u_2 - I_h u_2\|_\infty &\approx 0.010\ h^{2.973}, \\ E_{h,0} = \|u_2 - I_h u_2\|_0 &\approx 0.003\ h^{2.969}, \\ E_{h,1} = \|u_2 - I_h u_2\|_1 &\approx 0.022\ h^{1.971}. \end{aligned} \tag{3.30}$$

The IFE space $\hat{\mathcal{S}}_{h,2}(0,1)$ performs equally well in interpolating both test functions and displays no oscillatory behavior in the errors as h decreases.

The performance of interpolants from the IFE basis space $\hat{\mathcal{S}}_{h,2}(0,1)$ is not surprising because the IFE space is very close to a FE space based on a body fit partition. Recall that the basis functions $\hat{\phi}_i(x)$ that span the IFE space $\hat{\mathcal{S}}_{h,2}(0,1)$ are fundamentally different from the basis functions that span either $\tilde{\mathcal{S}}_{h,2}(0,1)$ or $\bar{\mathcal{S}}_{h,2}(0,1)$ because of the way that the nodal values are handled on the interface element. Because of the nodal value specifications used for the local basis functions $\hat{\phi}_{k,i}(x)$, their construction is equivalent to forming quadratic functions on each half of the element that are linear combinations of the standard FE quadratic basis functions on these two sub-elements. In the formation of these linear combinations of quadratic functions, it becomes apparent that instead of six basis functions being specified, only four are needed due to the interface jump conditions given. Finally, because the interface occurs at the "node" between the two sub-elements, this method is essentially equivalent to using a body fit scheme where the interface is located on one of the nodes of the finite element partition.

In Table 3.7, we summarize our observations about the approximation capabilities of the interpolants in the three IFE spaces:

	$\tilde{\mathcal{S}}_{h,2}(0,1)$	$\bar{\mathcal{S}}_{h,2}(0,1)$	$\hat{\mathcal{S}}_{h,2}(0,1)$
u_1	weaker than expected	as expected	as expected
u_2	weaker than expected	weaker than expected	as expected

Table 3.7: Approximation capability of interpolants in the three one dimensional IFE spaces

3.3 Interpolation in Two Dimensions using Biquadratic IFE Spaces

In this section, we report some numerical results that are generated from the two dimensional IFE spaces $\tilde{\mathcal{S}}_{h,2,\square}\big((0,1)^2\big)$, $\bar{\mathcal{S}}_{h,2,\square}\big((0,1)^2\big)$ and $\hat{\mathcal{S}}_{h,2,\square}\big((0,1)^2\big)$ which are defined in the previous chapter. As in the previous section, we first choose some test functions with different continuity features that aid in testing the approximation capabilities of these IFE spaces. After choosing the test functions, the errors in their interpolants from each of the three IFE spaces are considered using norms from the spaces $L^\infty(\big((0,1)^2\big)$, $L^2\big((0,1)^2\big)$ and $H^1\big((0,1)^2\big)$.

Since using these IFE spaces in the Galerkin FEM and Least Squares FEM to solve the forward interface problem is one of the important applications, the test functions are chosen so that they are the exact solutions of an elliptic boundary value problem. In two dimensions, the elliptic boundary value problem becomes

$$\begin{aligned} -\nabla\cdot(p\nabla u) &= f, \text{ if } \boldsymbol{x}\in\Omega=(0,1)^2, \\ u(0,y)=u(1,y) &= 0, \text{ if } 0<y<1, \\ u(x,0)=u(x,1) &= 0, \text{ if } 0<x<1, \end{aligned} \tag{3.31}$$

where the coefficient p is defined by

$$p(\boldsymbol{x})=p(x,y)=\begin{cases} p_1, & \text{if } y<\gamma, \\ p_2, & \text{if } y\geq\gamma, \end{cases} \tag{3.32}$$

so that the interface is a horizontal line occurring at $y=\gamma$ and the domain $\Omega=(0,1)\times(0,1)$. Again, u is also required to satisfy the interface jump conditions:

$$\big[u\big]_{y=\gamma}=0 \text{ and } \left[\frac{\partial u}{\partial \boldsymbol{n}}\right]_{y=\gamma}=0. \tag{3.33}$$

We take the opposite approach to forming test functions compared to the construction of test functions in the previous section. In one dimension, the test functions are constructed based upon a known right hand side function. Here in two dimensions, we first specify a test function that has the desired properties and then we calculate the formula for the right hand side function of the elliptic interface problem (3.31) by substituting the test function into (3.31).

Similar to the procedure in the previous section, we do want two test functions which have differing properties relating to the continuity of each test function's flux across the interface. The primary feature that is investigated in one dimension is the continuity of the flux pu''. In two dimensions, this corresponds to investigating the continuity of $p\frac{\partial^2}{\partial \boldsymbol{n}^2}u$ where $\boldsymbol{n}$ is the

unit normal vector on the interface Γ which points from Ω_1 to Ω_2. In two dimensions Ω_1 and Ω_2 are defined by

$$\Omega_1 = \big\{(x,y)|0 \le x \le 1 \text{ and } 0 \le y < \gamma\big\}, \tag{3.34}$$
$$\Omega_2 = \big\{(x,y)|0 \le x \le 1 \text{ and } \gamma \le y \le 1\big\}. \tag{3.35}$$

Then $\boldsymbol{n} = (n_1, n_2)^T = (0,1)^T$ and

$$p\frac{\partial^2}{\partial \boldsymbol{n}^2}u = p\frac{\partial^2 u}{\partial y^2}. \tag{3.36}$$

So we construct two test functions, $u_3(x,y)$ and $u_4(x,y)$, so that they satisfy the boundary conditions of (3.31) and the interface conditions (3.33). Additionally, we require that their fluxes are either continuous or discontinuous across the interface at $y = \gamma$.

The general form for each $u_i,\ i = 3,4$ is given by

$$u_i(x,y) = \begin{cases} \sin(\pi x)(a_5y^5 + a_4y^4 + a_3y^3 + a_2y^2 + a_1y + a_0), & y < e^{-1}, \\ \sin(\pi x)(b_5y^5 + b_4y^4 + b_3y^3 + b_2y^2 + b_1y + b_0), & y \ge e^{-1}. \end{cases} \tag{3.37}$$

The coefficients $a_i,\ i = 0,\ldots,5$ and $b_i,\ i = 0,\ldots,5$ are chosen so that the desired continuities across the interface are achieved.

Using an interface occurring along the horizontal line $y = e^{-1} \approx 0.367879$ and requiring that $p\frac{\partial^2}{\partial \boldsymbol{n}^2}u_3$ be discontinuous across the interface lead to the following coefficients of u_3:

$$\begin{aligned}
a_0 &= 0, \\
a_1 &= \tfrac{1}{2}e^{-5}\big(30e^6 + e(-30 - \pi^2(\beta-1)) - 6e^3(\beta-1) + 6\pi^2(\beta-1) \\
&\qquad -5e^4\pi^2(\beta-1) + 12e^2(2\beta-3)\big), \\
a_2 &= e^{-4}\big(18e^2 - 15e^6 + e(40 + \pi^2(\beta-1)) + 3e^3(\beta-1) \\
&\qquad -6\pi^2(\beta-1) + 5e^4\pi^2(\beta-1)\big), \\
a_3 &= \tfrac{1}{2}e^{-3}\big(2e^3 + 10e^6 + e(-60 - \pi^2(\beta-1)) + 6\pi^2(\beta-1) \\
&\qquad -5e^4\pi^2(\beta-1) - 12e^2(2\beta+1)\big), \\
a_4 &= 0, a_5 = 5, \\
b_0 &= 5, b_1 = 0, b_2 = 0, b_3 = 1, b_4 = -6, b_5 = 0.
\end{aligned} \tag{3.38}$$

where $\beta = p_2/p_1 > 1$.

Similarly, the function $u_4(x,y)$ is specified so that $p\frac{\partial^2}{\partial \boldsymbol{n}^2}u$ is continuous across the interface

at $y = \gamma = e^{-1}$. This results in the coefficients of u_4 being

$$\begin{aligned} a_0 &= 0, \\ a_1 &= e^4\big(30 + e + 9e^2\beta + 11e^3\beta + 14e^4\beta\big), \\ a_2 &= -e^{-3}\big(80 + 3e + 15e\beta + 15e^2\beta + 17e^3\beta + 21e^4\beta\big), \\ a_3 &= e^{-2}\big(60 + 3e + 15e\beta + 6e^2\beta + 6e^3\beta + 7e^4\beta\big), \\ a_4 &= -1, a_5 = -10, \\ b_0 &= e^{-3}\big(5 + 6e + 7e^2\big), \\ b_1 &= -e^{-3}\big(5 + 6e + 7e^2 + 7e^3\big), \\ b_2 &= 1, b_3 = 1, b_4 = 5, b_5 = 0. \end{aligned} \tag{3.39}$$

Interpolation errors are calculated in a similar fashion to one dimension. The primary differences are that instead of just six interpolation points used on each element, there are thirty-six interpolation points on each rectangular element (six each in the x and y directions). In the same way, sixteen gaussian nodes are used for integration instead of just four (this is four in both the x and y directions).

3.3.1 Interpolation Accuracy of the IFE Space $\tilde{\mathcal{S}}_{h,2,\square}\big((0,1)^2\big)$

We start by the looking at the numerical results generated by interpolating the functions u_3 and u_4 using the IFE space $\tilde{\mathcal{S}}_{h,2,\square}\big((0,1)^2\big)$ based on the hierarchical basis functions $\tilde{\psi}_i(x,y)$. The errors in the interpolants $I_h u_3$ and $I_h u_4$ for selected values h are shown in Tables 3.8 and 3.9 respectively when the coefficient function p has the ratio $\beta = 10$. As usual, using

h	$\|u_3 - I_h u_3\|_\infty$	$\|u_3 - I_h u_3\|_0$	$\|u_3 - I_h u_3\|_1$
1/8	4.991×10^{-2}	1.510×10^{-2}	7.972×10^{-1}
1/16	6.241×10^{-3}	1.905×10^{-3}	1.998×10^{-1}
1/32	7.801×10^{-4}	2.412×10^{-4}	5.030×10^{-2}
1/64	1.147×10^{-4}	3.091×10^{-5}	1.301×10^{-2}
1/128	1.515×10^{-5}	3.850×10^{-6}	3.925×10^{-3}
1/256	7.696×10^{-6}	4.941×10^{-7}	1.231×10^{-3}

Table 3.8: $E_{h,s}(I_h u_3)$ where $s = \infty, 0, 1$ with $I_h u_3 \in \tilde{\mathcal{S}}_{h,2}\big((0,1)^2\big)$ and $\beta = 10$.

the data from these tables, we can access the orders of the interpolants of the test functions as follows

$$\begin{aligned} E_{h,\infty} = \|u_3 - I_h u_3\|_\infty &\approx 8.789\ h^{2.633}, \\ E_{h,0} = \|u_3 - I_h u_3\|_0 &\approx 7.405\ h^{2.980}, \\ E_{h,1} = \|u_3 - I_h u_3\|_1 &\approx 35.985\ h^{1.876}, \end{aligned} \tag{3.40}$$

h	$\|u_4 - I_h u_4\|_\infty$	$\|u_4 - I_h u_4\|_0$	$\|u_4 - I_h u_4\|_1$
1/8	7.402×10^{-2}	2.224×10^{-2}	1.174×10^{0}
1/16	9.238×10^{-3}	2.807×10^{-3}	2.943×10^{-1}
1/32	1.153×10^{-3}	3.562×10^{-4}	7.421×10^{-2}
1/64	3.916×10^{-4}	4.885×10^{-5}	2.163×10^{-2}
1/128	6.653×10^{-5}	5.816×10^{-6}	1.116×10^{-2}
1/256	3.482×10^{-5}	8.864×10^{-7}	4.411×10^{-3}

Table 3.9: $E_{h,s}(I_h u_4)$ where $s = \infty, 0, 1$ with $I_h u_4 \in \tilde{\mathcal{S}}_{h,2}\big((0,1)^2\big)$ and $\beta = 10$.

$$\begin{aligned} E_{h,\infty} = \|u_4 - I_h u_4\|_\infty &\approx 4.720\ h^{2.234}, \\ E_{h,0} = \|u_4 - I_h u_4\|_0 &\approx 9.596\ h^{2.934}, \\ E_{h,1} = \|u_4 - I_h u_4\|_1 &\approx 25.053\ h^{1.606}. \end{aligned} \tag{3.41}$$

From these numerical estimates of the orders in the interpolants, we can see that from the point of view of approximating the test functions u_3 and u_4 as L^2 functions by their interpolants, the IFE space $\tilde{\mathcal{S}}_{h,2}\big((0,1)^2\big)$ behaves like the usual quadratic FE space. However, when trying to approximate u_3 and u_4 either as L^∞ or H^1 functions, this IFE space seems not to be able to generate interpolants with accuracies up to those from the usual quadratic FE space.

The errors in the interpolants of the test functions u_3 and u_4 when the ratio of the coefficient function p is increased to $\beta = 10000$ are shown in Tables 3.10 and 3.11 respectively. As

h	$\|u_3 - I_h u_3\|_\infty$	$\|u_3 - I_h u_3\|_0$	$\|u_3 - I_h u_3\|_1$
1/8	6.472×10^{1}	1.949×10^{1}	1.034×10^{3}
1/16	8.099×10^{0}	2.464×10^{0}	2.592×10^{2}
1/32	1.012×10^{0}	3.122×10^{-1}	6.517×10^{1}
1/64	1.264×10^{-1}	3.951×10^{-2}	1.643×10^{1}
1/128	6.846×10^{-2}	5.069×10^{-3}	1.486×10^{1}
1/256	2.939×10^{-2}	7.398×10^{-4}	4.456×10^{0}

Table 3.10: $E_{h,s}(I_h u_3)$ where $s = \infty, 0, 1$ with $I_h u_3 \in \tilde{\mathcal{S}}_{h,2}\big((0,1)^2\big)$ and $\beta = 1 \times 10^4$.

we see in these Tables, increasing the ratio of the values of the coefficient function p to $\beta = 10000$ causes the errors in the interpolants to become greater in magnitude for all of the functions regardless of which norm is used. The most dramatic effect is seen in the H^1 errors however. The orders of the errors in the interpolants that are obtained using the data in Tables 3.10 through 3.11 for each of the interpolants are shown in equations (3.42) through

h	$\|u_4 - I_h u_4\|_\infty$	$\|u_4 - I_h u_4\|_0$	$\|u_4 - I_h u_4\|_1$
1/8	7.354×10^{1}	2.199×10^{1}	1.166×10^{3}
1/16	9.160×10^{0}	2.778×10^{0}	2.923×10^{2}
1/32	1.143×10^{0}	3.521×10^{-1}	7.349×10^{1}
1/64	1.428×10^{-1}	4.455×10^{-2}	1.852×10^{1}
1/128	1.114×10^{-1}	5.819×10^{-3}	2.368×10^{1}
1/256	4.797×10^{-2}	9.580×10^{-4}	7.169×10^{0}

Table 3.11: $E_{h,s}(I_h u_4)$ where $s = \infty, 0, 1$ with $I_h u_4 \in \tilde{\mathcal{S}}_{h,2}\big((0,1)^2\big)$ and $\beta = 1 \times 10^4$.

(3.43) below for functions u_3 and u_4 respectively.

$$\begin{aligned} E_{h,\infty} = \|u_3 - I_h u_3\|_\infty &\approx 3987.962\ h^{2.262}, \\ E_{h,0} = \|u_3 - I_h u_3\|_0 &\approx 8687.576\ h^{2.948}, \\ E_{h,1} = \|u_3 - I_h u_3\|_1 &\approx 17835.198\ h^{1.533}, \end{aligned} \tag{3.42}$$

$$\begin{aligned} E_{h,\infty} = \|u_4 - I_h u_4\|_\infty &\approx 3230.317\ h^{2.143}, \\ E_{h,0} = \|u_4 - I_h u_4\|_0 &\approx 8945.813\ h^{2.917}, \\ E_{h,1} = \|u_4 - I_h u_4\|_1 &\approx 14527.339\ h^{1.417}. \end{aligned} \tag{3.43}$$

As we mention above, it can be seen by the orders of the errors $E_{h,1}$ for each interpolant that the order has decreased slightly upon increasing the ratio of the values of the coefficient function p. Although the errors are worse in magnitude when the coefficient ratio is increased, the IFE space $\tilde{\mathcal{S}}_{h,2}\big((0,1)^2\big)$ still performs well in the L^2 with functions u_3 and u_4 as the order of the error in the interpolants in these cases remains close to three.

3.3.2 Interpolation Accuracy of the IFE Space $\bar{\mathcal{S}}_{h,2,\square}\big((0,1)^2\big)$

Next we consider the errors in the interpolants of the test functions u_3 and u_4 using the IFE $\bar{\mathcal{S}}_{h,2,\square}\big((0,1)^2\big)$ space whose basis functions $\bar{\psi}_i(x)$ have extra continuity requirements. The errors in these interpolants when the value of the ratio of the coefficient function p is $\beta = 10$ are shown in Tables 3.12 and 3.13. Using the data in these tables, we generate estimates of the orders of the interpolants $I_h u_3$ and $I_h u_4$ in the following equations:

$$\begin{aligned} E_{h,\infty} = \|u_3 - I_h u_3\|_\infty &\approx 3.112\ h^{2.172}, \\ E_{h,0} = \|u_3 - I_h u_3\|_0 &\approx 6.247\ h^{2.898}, \\ E_{h,1} = \|u_3 - I_h u_3\|_1 &\approx 16.171\ h^{1.563}, \end{aligned} \tag{3.44}$$

$$\begin{aligned} E_{h,\infty} = \|u_4 - I_h u_4\|_\infty &\approx 37.933\ h^{3.001}, \\ E_{h,0} = \|u_4 - I_h u_4\|_0 &\approx 11.131\ h^{2.988}, \\ E_{h,1} = \|u_4 - I_h u_4\|_1 &\approx 73.998\ h^{1.993}. \end{aligned} \tag{3.45}$$

h	$\|u_3 - I_h u_3\|_\infty$	$\|u_3 - I_h u_3\|_0$	$\|u_3 - I_h u_3\|_1$
1/8	4.991×10^{-2}	1.508×10^{-2}	7.971×10^{-1}
1/16	6.241×10^{-3}	1.895×10^{-3}	1.996×10^{-1}
1/32	8.275×10^{-4}	2.426×10^{-4}	5.041×10^{-2}
1/64	9.014×10^{-4}	5.514×10^{-5}	2.915×10^{-2}
1/128	3.677×10^{-5}	3.911×10^{-6}	6.452×10^{-3}
1/256	2.833×10^{-5}	6.480×10^{-7}	3.552×10^{-3}

Table 3.12: $E_{h,s}(I_h u_3)$ where $s = \infty, 0, 1$ with $I_h u_3 \in \bar{\mathcal{S}}_{h,2}\big((0,1)^2\big)$ and $\beta = 10$.

h	$\|u_4 - I_h u_4\|_\infty$	$\|u_4 - I_h u_4\|_0$	$\|u_4 - I_h u_4\|_1$
1/8	7.402×10^{-2}	2.224×10^{-2}	1.174×10^{0}
1/16	9.238×10^{-3}	2.807×10^{-3}	2.943×10^{-1}
1/32	1.153×10^{-3}	3.552×10^{-4}	7.407×10^{-2}
1/64	1.441×10^{-4}	4.494×10^{-5}	1.872×10^{-2}
1/128	1.801×10^{-5}	5.649×10^{-6}	4.686×10^{-3}
1/256	2.251×10^{-6}	7.061×10^{-7}	1.172×10^{-3}

Table 3.13: $E_{h,s}(I_h u_4)$ where $s = \infty, 0, 1$ with $I_h u_4 \in \bar{\mathcal{S}}_{h,2}\big((0,1)^2\big)$ and $\beta = 10$.

These orders indicate that the interpolant $I_h u_3 \in \bar{\mathcal{S}}_{h,2}\big((0,1)^2\big)$ can not approximate u_3 as well as the interpolant generated from the standard quadratic FE space even rough the error of $I_h u_3$ in the L^2 norm seems to be up to our expectation. On the other hand, for the test function u_4, the IFE space $\bar{\mathcal{S}}_{h,2}\big((0,1)^2\big)$ seems to have an approximation capability comparable to that of the standard quadratic FE space.

When the coefficient ratio is increased to $\beta = 10000$, the errors of the interpolants of u_3 and u_4 are shown in Tables 3.14 and 3.15. And the orders of these interpolants are given by the

h	$\|u_3 - I_h u_3\|_\infty$	$\|u_3 - I_h u_3\|_0$	$\|u_3 - I_h u_3\|_1$
1/8	6.472×10^{1}	1.949×10^{1}	1.034×10^{3}
1/16	8.099×10^{0}	2.463×10^{0}	2.592×10^{2}
1/32	1.012×10^{0}	3.122×10^{-1}	6.517×10^{1}
1/64	1.264×10^{-1}	3.950×10^{-2}	1.642×10^{1}
1/128	2.721×10^{-2}	4.996×10^{-3}	7.018×10^{0}
1/256	2.517×10^{-2}	7.105×10^{-4}	3.852×10^{0}

Table 3.14: $E_{h,s}(I_h u_3)$ where $s = \infty, 0, 1$ with $I_h u_3 \in \bar{\mathcal{S}}_{h,2}\big((0,1)^2\big)$ and $\beta = 1 \times 10^4$.

h	$\|u_4 - I_h u_4\|_\infty$	$\|u_4 - I_h u_4\|_0$	$\|u_4 - I_h u_4\|_1$
1/8	7.354×10^{1}	2.199×10^{1}	1.166×10^{3}
1/16	9.160×10^{0}	2.778×10^{0}	2.923×10^{2}
1/32	1.143×10^{0}	3.521×10^{-1}	7.349×10^{1}
1/64	1.428×10^{-1}	4.455×10^{-2}	1.852×10^{1}
1/128	1.785×10^{-2}	5.619×10^{-3}	4.661×10^{0}
1/256	2.231×10^{-3}	7.023×10^{-4}	1.165×10^{0}

Table 3.15: $E_{h,s}(I_h u_4)$ where $s = \infty, 0, 1$ with $I_h u_4 \in \bar{\mathcal{S}}_{h,2}\big((0,1)^2\big)$ and $\beta = 1 \times 10^4$.

following estimates:

$$\begin{aligned} E_{h,\infty} = \|u_3 - I_h u_3\|_\infty &\approx 5815.693\ h^{2.408}, \\ E_{h,0} = \|u_3 - I_h u_3\|_0 &\approx 8946.894\ h^{2.958}, \\ E_{h,1} = \|u_3 - I_h u_3\|_1 &\approx 24527.924\ h^{1.656}, \end{aligned} \tag{3.46}$$

$$\begin{aligned} E_{h,\infty} = \|u_4 - I_h u_4\|_\infty &\approx 37701.975\ h^{3.001}, \\ E_{h,0} = \|u_4 - I_h u_4\|_0 &\approx 10958.006\ h^{2.986}, \\ E_{h,1} = \|u_4 - I_h u_4\|_1 &\approx 73407.840\ h^{1.992}. \end{aligned} \tag{3.47}$$

Again the interpolants $I_h u_3$ tend to underperform in both the L^∞ and H^1 norms although errors in the L^2 norms appear optimal. On the other hand, increasing the value of the ratio β to 10000 does not affect the orders of the interpolants $I_h u_4 \in \bar{\mathcal{S}}_{h,2}\big((0,1)^2\big)$. These orders remain optimal for the errors in all three norms that are considered.

3.3.3 Interpolation Accuracy of the IFE Space $\hat{\mathcal{S}}_{h,2,\square}\big((0,1)^2\big)$

Finally, we consider the errors in the interpolants of the test functions u_3 and u_4 generated from the IFE space $\hat{\mathcal{S}}_{h,2,\square}\big((0,1)^2\big)$ whose basis functions $\hat{\psi}_i(x)$ are based on the idea of local refinement on the interface elements. The errors in the interpolants of these two test functions when the coefficient ratio is $\beta = 10$ are shown in Tables 3.16 and 3.17.

We can use these data tables to generate estimates of the orders of these interpolants in the following equations:

$$\begin{aligned} E_{h,\infty} = \|u_3 - I_h u_3\|_\infty &\approx 25.555\ h^{3.000}, \\ E_{h,0} = \|u_3 - I_h u_3\|_0 &\approx 7.458\ h^{2.985}, \\ E_{h,1} = \|u_3 - I_h u_3\|_1 &\approx 50.038\ h^{1.992}, \end{aligned} \tag{3.48}$$

$$\begin{aligned} E_{h,\infty} = \|u_4 - I_h u_4\|_\infty &\approx 37.933\ h^{3.001}, \\ E_{h,0} = \|u_4 - I_h u_4\|_0 &\approx 10.980\ h^{2.985}, \\ E_{h,1} = \|u_4 - I_h u_4\|_1 &\approx 73.688\ h^{1.992}. \end{aligned} \tag{3.49}$$

h	$\|u_3 - I_h u_3\|_\infty$	$\|u_3 - I_h u_3\|_0$	$\|u_3 - I_h u_3\|_1$
1/8	4.991×10^{-2}	1.501×10^{-2}	7.962×10^{-1}
1/16	6.241×10^{-3}	1.892×10^{-3}	1.994×10^{-1}
1/32	7.801×10^{-4}	2.396×10^{-4}	5.009×10^{-2}
1/64	9.752×10^{-5}	3.039×10^{-5}	1.262×10^{-2}
1/128	1.219×10^{-5}	3.837×10^{-6}	3.183×10^{-3}
1/256	1.524×10^{-6}	4.796×10^{-7}	7.958×10^{-4}

Table 3.16: $E_{h,s}(I_h u_3)$ where $s = \infty, 0, 1$ with $I_h u_3 \in \hat{\mathcal{S}}_{h,2}\big((0,1)^2\big)$ and $\beta = 10$.

h	$\|u_4 - I_h u_4\|_\infty$	$\|u_4 - I_h u_4\|_0$	$\|u_4 - I_h u_4\|_1$
1/8	7.402×10^{-2}	2.209×10^{-2}	1.172×10^{0}
1/16	9.238×10^{-3}	2.786×10^{-3}	2.935×10^{-1}
1/32	1.153×10^{-3}	3.528×10^{-4}	7.375×10^{-2}
1/64	1.441×10^{-4}	4.474×10^{-5}	1.858×10^{-2}
1/128	1.801×10^{-5}	5.649×10^{-6}	4.686×10^{-3}
1/256	2.251×10^{-6}	7.061×10^{-7}	1.171×10^{-3}

Table 3.17: $E_{h,s}(I_h u_4)$ where $s = \infty, 0, 1$ with $I_h u_4 \in \hat{\mathcal{S}}_{h,2}\big((0,1)^2\big)$ and $\beta = 10$.

The performance of the interpolants generated from the IFE space $\hat{\mathcal{S}}_{h,2}\big((0,1)^2\big)$ for both test functions is obviously comparable to those from the standard FE space. There is no significant change in the errors of the interpolants from $\hat{\mathcal{S}}_{h,2}\big((0,1)^2\big)$ when $\beta = 10000$ other than the magnitude of the errors increasing as is the case for the other biquadratic IFE spaces. Overall, from these numerical experiments, we notice that the biquadratic IFE spaces behave very much like their corresponding one dimensional IFE spaces.

Chapter 4

The Galerkin IFE Method for the Forward Problem

In this chapter, we investigate the performance of the IFE spaces introduced in Chapter 2 when they are used to approximate the solutions of an elliptic interface problem such as

$$\begin{aligned} -\nabla \cdot (p(\boldsymbol{x})\nabla u(\boldsymbol{x})) &= f(\boldsymbol{x}), \text{ if } \boldsymbol{x} \in \Omega, \\ u(\boldsymbol{x}) &= 0, \text{ if } \boldsymbol{x} \in \partial\Omega, \end{aligned} \tag{4.1}$$

by using the Galerkin Finite Element Method. Instead of solving (4.1) directly, the Galerkin FE method attempts to approximate a solution of a variational form of (4.1). By using a variational form of the elliptic interface problem, we look for a solution among a larger class of functions than when the elliptic interface problem is solved directly. Using a larger space of functions also allows us to use IFE spaces as the space for our approximations even though they do not satisfy the requirements of the space that contains the classical solution. For more information about the Galerkin method, there are many good resources. Both Braess [8] and Brenner and Scott [9] offer very good overviews of the Galerkin method and its application to elliptic boundary value problems.

4.1 Background of the Galerkin FE Method

To apply the Galerkin method to solve the elliptic boundary value problem (4.1) we first state a weak form of the equation. This can be done by multiplying (4.1) with a test function in a suitable space and then applying integration by parts to the resulting equation using Green's formulas.

To determine what space of functions may be suitable for test functions, it is helpful to consider the space of functions that a classical solution of (4.1) comes from. If p is differentiable on the domain Ω then a classical solution of the BVP comes from the space $C^2(\Omega) \cap C^0(\bar{\Omega})$ [8]. The coefficient p in our BVP is discontinuous so the classical solution comes from the space

$$\left\{u \;\middle|\; u \in C^0, (pu') \in C^1(\Omega)\right\} \cap C^0(\bar{\Omega}). \tag{4.2}$$

Consequently, if u is in this space then $u \in C^2(\Omega_1)$ and $u \in C^2(\Omega_2)$ although it is generally not differentiable at the interface. To weaken the requirements on our solution, it is natural to try to find a space where the solution u is not required to have as many derivatives. This leads us to Sobolev spaces which are a good choice since they are also Hilbert Spaces. The Sobolev space $H^k(\Omega)$ with $\Omega \subset \mathbb{R}^n$ is the space of L^2 functions that have derivatives up to order k which are also L^2 functions.

$$H^k(\Omega) = \left\{u | D^a u \in L^2(\Omega) \;\text{ for } |a| \leq k\right\}, \tag{4.3}$$

where

$$\begin{aligned} a &= (a_1, a_2, \ldots, a_n) \geq 0, \qquad a_i \in \mathbb{N}, && (4.4)\\ |a| &= a_1 + a_2 + \ldots a_n, && (4.5) \end{aligned}$$

and so

$$D^\alpha = \prod_{i=1}^{n} \frac{\partial^{a_i}}{\partial x_i^{a_i}}. \tag{4.6}$$

All Sobolev spaces are Hilbert spaces meaning that they are complete inner product spaces. More about Sobolev spaces can be found in Adams [1], Kesavan [33] and in many texts on partial differential equations such as Rogers and Renardy [52].

Again, if u is a classical solution of the BVP when p is differentiable then u is at least twice differentiable. Then if p is differentiable, an appropriate Sobolev space to consider is the space of functions $H_0^1(\Omega)$ which requires less differentiability for u. However, since p is discontinuous and only piecewise differentiable, then we choose instead the following function space:

$$\begin{aligned} \mathcal{V} = \big\{ & u \in L^2(\Omega) \big| \;\; u|_{\Omega_1} \in H^1(\Omega_1), u|_{\Omega_2} \in H^1(\Omega_2), \\ & u|_{\partial\Omega} = 0, \left[u\right]_\Gamma = 0, \left[pu'\right]_\Gamma = 0 \big\}. \end{aligned} \tag{4.7}$$

Now if we multiply the BVP (4.1) by $v \in V$ and then integrate over the domain Ω we have

$$-\int_\Omega v \nabla \cdot (p \nabla u) d\boldsymbol{x} = \int_\Omega v f d\boldsymbol{x} \text{ for all } v \in \mathcal{V}. \tag{4.8}$$

According to Green's theorem we can integrate the left hand side of (4.8) by parts to obtain

$$-\int_{\partial\Omega} v(p\nabla u \cdot \boldsymbol{n})ds + \int_{\Omega} (p\nabla u)\cdot\nabla v d\boldsymbol{x} = \int_{\Omega} fv d\boldsymbol{x}, \text{ for all } v \in \mathcal{V}. \tag{4.9}$$

However, since $v\big|_{\partial\Omega} = 0$, we arrive at the weak form of (4.1):

$$a(u,v) = (f,v), \text{ for all } v \in \mathcal{V}, \tag{4.10}$$

where the symmetric, bilinear form, $a : \mathcal{V}\times\mathcal{V} \to \mathbb{R}$ is defined by

$$a(u,v) = \int_{\Omega} p(\nabla u \cdot \nabla v)d\boldsymbol{x}. \tag{4.11}$$

and $(\cdot,\cdot)$ denotes the L^2 inner product of u and v. Then we have the following definition of a weak solution of the BVP:

Definition 4.1.1. *If $u \in \mathcal{V}$ satisfies*

$$a(u,v) = (f,v), \text{ for all } v \in \mathcal{V}, \tag{4.12}$$

then u is a weak solution of the BVP (4.1).

In order to establish that a unique weak solution exists we need to describe some of the properties of the bilinear form, a.

Definition 4.1.2. *Let $\mathcal{H}$ be a Hilbert space and $a : \mathcal{H}\times\mathcal{H} \to \mathbb{R}$. If there exists a constant $M > 0$ such that*

$$\|a(u,v)\| \le M\|u\|\|v\|, \text{ for all } u,v \in \mathcal{H}, \tag{4.13}$$

then we say that a is continuous. Also, if there exists a constant $b > 0$ such that

$$a(v,v) \ge b\|v\|^2, \text{ for every } v \in \mathcal{H}, \tag{4.14}$$

then we say that a is $\mathcal{H}$-elliptic.

Then the following theorem establishes the existence and uniqueness of a weak solution of (4.1).

Theorem 4.1.3 (Lax-Milgram). *Let $\mathcal{H}$ be a Hilbert space and let $a : \mathcal{H}\times\mathcal{H} \to \mathbb{R}$ be continuous and $\mathcal{H}$-elliptic. Then given $f \in H$, there exists a unique $u \in \mathcal{V}'$ which satisfies (4.12). In other words, there exists a unique weak solution of (4.1).*

Proof. See Brenner and Scott, [9, p.60ff]. ■.

There are two important things to notice here. First, a solution of (4.12) quires one less derivative than the solution to (4.1). Second, the boundary conditions from (4.1) are now incorporated into the solution space of the elliptic interface problem, $\mathcal{V}$. So if $u \in \mathcal{V}$ is a weak solution of (4.12) then it satisfies the interface conditions

$$\big[u\big]_\Gamma = 0 \text{ and } \left[p\frac{\partial u}{\partial \boldsymbol{n}}\right]_\Gamma = 0. \tag{4.15}$$

Remark 4.1.4. *If the bilinear form $a(u,v)$ is symmetric and $u^* \in \mathcal{V}$ is the weak solution of (4.1) then it is shown in [8] that u^* is also the unique minimizer of the functional $\mathcal{J} : \mathcal{V} \to I\!R$ where*

$$\mathcal{J}(u) = \frac{1}{2}a(u,u) - (f,u). \tag{4.16}$$

In other words, $\mathcal{J}(u^*) < \mathcal{J}(v)$ for all $v \in \mathcal{V}$.

In our case, $a(u,v)$ is symmetric since

$$a(u,v) = \int_\Omega p(\nabla u \cdot \nabla v)d\boldsymbol{x} = a(v,u), \tag{4.17}$$

and hence the remark simply says that another way to find the weak form of the elliptic interface problem (4.1) is to minimize the functional $\mathcal{J}(u)$. A necessary condition for $\mathcal{J}(u)$ to achieve its minimum is that its first variation is equal to zero. Here, the first variational of $\mathcal{J}(u)$ is given by

$$\partial\mathcal{J}(u;v) = \lim_{t\to 0}\frac{\mathcal{J}(u+tv) - \mathcal{J}(u)}{t}, \tag{4.18}$$

where $u, u+tv \in \mathcal{V}$ for all $t \in I\!R$. The function v is called an admissible variation. Alternatively, the first variation of $\mathcal{J}$ may be written as $\delta\mathcal{J}(u)\delta u$ where $\delta u = v$ is the admissible variation. In particular, v must be chosen so that $u + tv$ still satisfies the boundary conditions of the space $\mathcal{V}$. Hence, it is appropriate to ensure that $v\big|_{\partial\Omega} = 0$. Additionally, $u + tv$ must satisfy the interface conditions (4.15) and so it follows that $v \in \mathcal{V}$. Additionally, the requirement that the first variation equals to zero just leads to the weak form (4.10). This is not generally true if $a(u,v)$ is not symmetric. Then the admissible variation v is really just the test function that we originally introduced to find the weak form. This further confirms to the choice of function spaces above for v in the weak form.

The Galerkin method then is to seek an approximation of the weak solution. Let $\mathcal{V}_h$ be a finite dimensional subset of $\mathcal{V}$. The Galerkin approximation problem is to find $u_h \in \mathcal{V}_h$ such that

$$a(u_h, v) = (f,v), \text{ for all } v \in \mathcal{V}_h. \tag{4.19}$$

It follows from the Lax-Milgram Theorem that this finite dimensional problem also has a unique solution since $\mathcal{V}_h$ is a closed subset of $\mathcal{V}$ and hence it is also a Hilbert space to which the Lax-Milgram Theorem applies. The following Lemma by Céa helps describe how close the approximate weak solution u_h is to the weak solution u.

Lemma 4.1.5 (Céa). *Let $\mathcal{H}$ be a Hilbert space and $a : \mathcal{H} \times \mathcal{H} \to \mathbb{R}$ be a continuous, $\mathcal{H}-$elliptic bilinear form. Let u satisfy (4.12) and u_h satisfy (4.19) then*

$$\|u - u_h\|_{\mathcal{V}} \leq \frac{M}{b} \min_{v \in \mathcal{V}_h} \|u - v\|_{\mathcal{V}}, \tag{4.20}$$

where M and b are the continuity and $\mathcal{H}-$elliptic constants for a in (4.13) and (4.14) respectively.

Proof. See Braess, [8], Brenner and Scott, [9] or Kesavan [33], for example. ■.

The main implication of Céa's Lemma is that studying the error between u and u_h now becomes the investigation of interpolation properties of the space $\mathcal{V}_h$ [33]. This naturally leads us to consider using finite element spaces which are based upon piecewise polynomial basis function for our choice of $\mathcal{V}_h$ since the interpolation properties of such spaces are well known ([8], [9], [17], etc.).

Specifically, let $\mathcal{V}_h$ be an IFE space such that $\mathcal{V}_h \subset \mathcal{V}$ If the basis of $\mathcal{V}_h$ is $\{\phi_i\}, i = 1, \ldots, N$, then we are looking for a function $u_h \in \mathcal{V}_h$ that satisfies (4.19) for any $v \in \mathcal{V}_h$, and the function u_h has the basic formula

$$u_h = \sum_{j=1}^{N} c_j \phi_j(x). \tag{4.21}$$

In particular, we want (4.19) to be satisfied for $v = \phi_i$ for $i = 1, \ldots, N$. This leads to a system of N equations and N unknowns as follows

$$K\boldsymbol{c} = \boldsymbol{F}, \tag{4.22}$$

from which we can solve for the coefficients $\boldsymbol{c} = (c_1, \ldots, c_N)^T$ for the approximate weak solution u_h. In the system (4.22), the matrix K corresponds to the bilinear form $a(u, v)$ and is given by

$$K = \big(a(\phi_i, \phi_j)\big)_{i,j=1}^{N} = \left(\int_{\Omega} p(\nabla \phi_i \cdot \nabla \phi_j) d\boldsymbol{x} \right). \tag{4.23}$$

Similarly, the right hand side vector $\boldsymbol{F}$ corresponds to the inner product (f, v) and so is given by

$$\boldsymbol{F} = \big(\phi_i, f\big)_{i=1}^{N} = \left(\int_{\Omega} f \phi_i d\boldsymbol{x} \right). \tag{4.24}$$

The matrix K has some nice properties because of the choice of $\mathcal{V}_h$ and the properties of the bilinear form $a(u, v)$. These are summarized in the following remark.

Remark 4.1.6. *If the bilinear form $a(u,v)$ is symmetric then so is K. If the bilinear form $a(u,v)$ is $\mathcal{V}-$elliptic then K is positive definite. Since each basis function ϕ_i has compact support, the matrix K is sparse. That is to say, for any two basis functions ϕ_i and ϕ_j of $\mathcal{V}_h$, if $supp\phi_i \cap supp\phi_j = \emptyset$ then $a(\phi_i,\phi_j)=0$. Consequently, the system (4.22) has a unique solution and it can be solved efficiently using techniques such as the preconditioned conjugate gradient method since K is a symmetric and positive definite matrix.*

In the remainder of this chapter, the IFE spaces developed in Chapter 2 are applied to solving interface problems whose exact solutions are the testing functions used in Chapter 3. Having established some results for the interpolation properties of IFE spaces in Chapter 3, we aim here to explore the approximation capability of these spaces when they are used in the Galerkin FE method to solve interface problems.

4.1.1 Notation

Errors are calculated numerically in the same way as in Chapter 3 where we investigate the interpolation errors of the IFE spaces. Also, most notations used here are also the same as those of Chapter 3. There are, however, some exceptions. The error between the exact solution u_i and its FE approximation $u_{i,h}$ is denoted by $E_{h,s}(u_{i,h})$ for $s=\infty,0,1$. It is usually clear from the context which function is being approximated, and in these cases, the approximation is referred to simply as u_h. Similarly, unless it is not clear from the context which types of errors are being discussed (either interpolation or FE approximations) then the FE error is simply referred to as $E_{h,s}$. Again, in calculating $E_{h,\infty}$, we evaluate the residual, $u-u_h$, at six equally spaced points in the x and y directions on each element. When $\Omega=(0,1)$ then this means that there are six points on each element where the residual is calculated. On the other hand, when $\Omega=(0,1)^2$ then there are 36 points on each element at which the residual is calculated. Similarly, for $E_{h,0}$, and $E_{h,1}$, we evaluate the residual at four gaussian nodes on each element when $\Omega=(0,1)$. When $\Omega=(0,1)^2$ then four gaussian nodes are used in both the x and y directions resulting in 16 gaussian nodes on each element. As before, interface elements are split into two pieces on both sides of the interface.

4.2 The Galerkin IFE Method in One Dimension

Now, we look at the errors in the Galerkin finite element solutions that are generated by the quadratic IFE spaces for interface problems that have the test functions $u_1(x)$ and $u_2(x)$

introduced in the previous chapter as are their exact solutions. Recall that in one dimension, our elliptic interface problem is given by,

$$\begin{array}{rcl} -(p(x)u'(x))' & = & f(x), \text{ if } x \in (0,1), \\ u(0) = u(1) & = & 0, \end{array} \tag{4.25}$$

where p is a discontinuous constant function as defined in (3.11). Again, we use partitions ranging in size from $h = 1/8$ to $h = 1/256$. Data is included so that the orders of the IFE solutions can be determined using linear regression on the $\log h$ vs. $\log E_h$ plots.

4.2.1 IFE Solution Accuracy Using the Space $\tilde{\mathcal{S}}_{h,2}(0,1)$

First we consider approximations that are generated from the IFE space $\tilde{\mathcal{S}}_{h,2}(0,1)$ for both $u_1(x)$ and $u_2(x)$ using the Galerkin finite element method. The errors in u_h are given in Tables 4.1 and 4.2 respectively when the coefficient function p has a ratio of $\beta = p_2/p_1 = 10$.

The errors in u_h approximating u_1 and u_2 are similar to the errors in the interpolants

h	$\|u_1 - u_h\|_\infty$	$\|u_1 - u_h\|_0$	$\|u_1 - u_h\|_1$
1/8	1.516×10^{-4}	3.268×10^{-5}	1.521×10^{-3}
1/16	2.553×10^{-5}	4.933×10^{-6}	8.411×10^{-4}
1/32	1.217×10^{-5}	1.604×10^{-6}	3.895×10^{-4}
1/64	3.396×10^{-6}	2.856×10^{-7}	1.090×10^{-4}
1/128	2.173×10^{-7}	9.802×10^{-9}	8.431×10^{-6}
1/256	1.704×10^{-7}	6.591×10^{-9}	9.739×10^{-6}

Table 4.1: IFEM approximation error, $u_1 - u_h$, where $u_h \in \tilde{\mathcal{S}}_{h,2}(0,1)$ with $\beta = 10$.

h	$\|u_2 - u_h\|_\infty$	$\|u_2 - u_h\|_0$	$\|u_2 - u_h\|_1$
1/8	2.429×10^{-4}	5.109×10^{-5}	2.389×10^{-3}
1/16	4.116×10^{-5}	7.848×10^{-6}	1.366×10^{-3}
1/32	2.034×10^{-5}	2.676×10^{-6}	6.518×10^{-4}
1/64	5.756×10^{-6}	4.833×10^{-7}	1.845×10^{-4}
1/128	3.689×10^{-7}	1.632×10^{-8}	1.416×10^{-5}
1/256	2.907×10^{-7}	1.124×10^{-8}	1.660×10^{-5}

Table 4.2: IFEM approximation error, $u_2 - u_h$, where $u_h \in \tilde{\mathcal{S}}_{h,2}(0,1)$ with $\beta = 10$.

$I_h u_1$ and $I_h u_2$ when using the IFE space $\tilde{\mathcal{S}}_{h,2}(0,1)$. As before, errors in u_h decrease as h decreases in an oscillatory way. The data in Table 4.1 can be used to find the orders of u_h

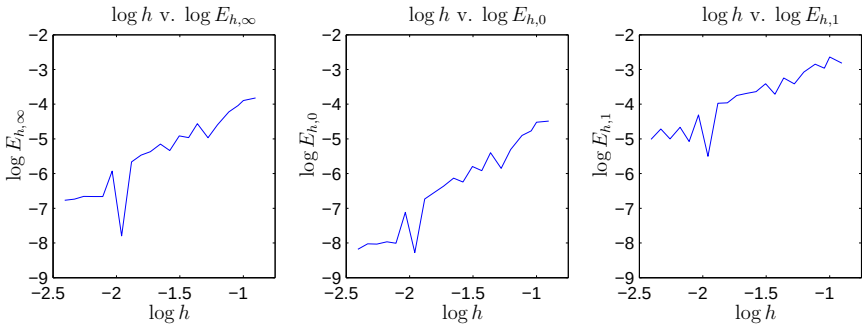


Figure 4.1: IFEM approximation error, $u_1 - u_h$, where $u_h \in \tilde{\mathcal{S}}_{h,2}(0,1)$ with $\beta = 10$.

as approximation to the test function $u_1(x)$ with the IFE space $\tilde{\mathcal{S}}_{h,2}(0,1)$ because these data obey

$$\begin{aligned} \tilde{E}_{h,\infty} = \|u_1 - u_h\|_\infty &\approx 0.010\ h^{2.042}, \\ \tilde{E}_{h,0} = \|u_1 - u_h\|_0 &\approx 0.008\ h^{2.594}, \\ \tilde{E}_{h,1} = \|u_1 - u_h\|_1 &\approx 0.073\ h^{1.663}. \end{aligned} \tag{4.26}$$

Similarly, we can see the order of u_h as approximation of the test function $u_2(x)$ with the IFE space $\tilde{\mathcal{S}}_{h,2}(0,1)$ from the following relationships:

$$\begin{aligned} \tilde{E}_{h,\infty} = \|u_2 - u_h\|_\infty &\approx 0.016\ h^{2.022}, \\ \tilde{E}_{h,0} = \|u_2 - u_h\|_0 &\approx 0.012\ h^{2.570}, \\ \tilde{E}_{h,1} = \|u_2 - u_h\|_1 &\approx 0.111\ h^{1.641}. \end{aligned} \tag{4.27}$$

The orders for approximating u_1 and u_2 are both consistent with the interpolation errors exhibited by the IFE space $\tilde{\mathcal{S}}_{h,2}(0,1)$, and these orders are lower than the orders that are generated by the usual quadratic FE method based on body fit partitions. Figures 4.1 and 4.2 show plots of $\log h$ versus $\log E_h$ for various norms and again the errors are seen to oscillate as they decrease in a pattern which is similar to the ones shown in Figures 3.1 and 3.2 in Chapter 3.

When the coefficient p has a ratio of $\beta = p_1/p_2 = 10000$, the errors in u_h approximating u_1 and u_2 are shown in Tables 4.3 and 4.4 respectively. Interestingly, the errors appear to decrease in a regular fashion as h decreases. These trends are seen in Figure 4.3 which shows a nice linear relationship between $\log h$ and $\log E_{h,s}$ for $s = \infty, 0, 1$. Figure 4.4 shows similar behavior for the approximation of u_2 using the IFE method when $\beta = 10000$. The order of

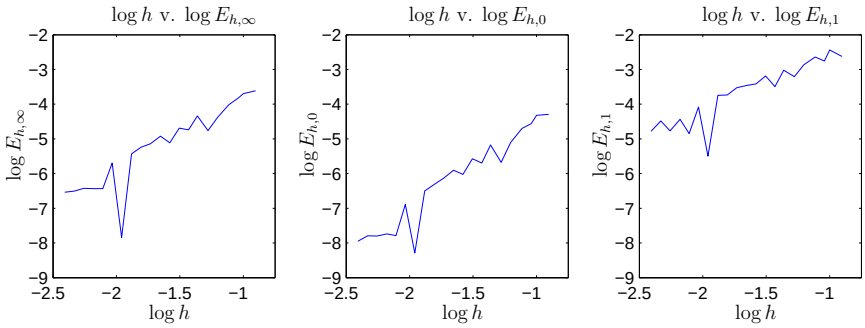


Figure 4.2: IFEM approximation error, $u_2 - u_h$, where $u_h \in \tilde{\mathcal{S}}_{h,2}(0,1)$ with $\beta = 10$.

h	$\|u_1 - u_h\|_\infty$	$\|u_1 - u_h\|_0$	$\|u_1 - u_h\|_1$
1/8	1.893×10^{-5}	6.423×10^{-6}	3.368×10^{-4}
1/16	2.593×10^{-6}	9.256×10^{-7}	9.609×10^{-5}
1/32	3.289×10^{-7}	1.157×10^{-7}	2.401×10^{-5}
1/64	4.141×10^{-8}	1.446×10^{-8}	5.999×10^{-6}
1/128	5.195×10^{-9}	1.811×10^{-9}	1.505×10^{-6}
1/256	6.531×10^{-10}	2.280×10^{-10}	3.782×10^{-7}

Table 4.3: IFEM approximation error, $u_1 - u_h$, where $u_h \in \tilde{\mathcal{S}}_{h,2}(0,1)$ with $\beta = 1 \times 10^4$.

the IFE approximations of u_1 using the IFE space $\tilde{\mathcal{S}}_{h,2}(0,1)$ have the numerical estimates

$$\begin{aligned} \tilde{E}_{h,\infty} = \|u_1 - u_h\|_\infty &\approx 0.010\ h^{2.971}, \\ \tilde{E}_{h,0} = \|u_1 - u_h\|_0 &\approx 0.003\ h^{2.969}, \\ \tilde{E}_{h,1} = \|u_1 - u_h\|_1 &\approx 0.022\ h^{1.971}, \end{aligned} \tag{4.28}$$

while the order of IFE approximations of u_2 from the space $\tilde{\mathcal{S}}_{h,2}(0,1)$ have the numerical

h	$\|u_2 - u_h\|_\infty$	$\|u_2 - u_h\|_0$	$\|u_2 - u_h\|_1$
1/8	1.893×10^{-5}	6.420×10^{-6}	3.369×10^{-4}
1/16	2.593×10^{-6}	9.264×10^{-7}	9.626×10^{-5}
1/32	3.289×10^{-7}	1.159×10^{-7}	2.404×10^{-5}
1/64	4.141×10^{-8}	1.447×10^{-8}	6.002×10^{-6}
1/128	5.195×10^{-9}	1.810×10^{-9}	1.505×10^{-6}
1/256	6.531×10^{-10}	2.282×10^{-10}	3.785×10^{-7}

Table 4.4: IFEM approximation error, $u_2 - u_h$, where $u_h \in \tilde{\mathcal{S}}_{h,2}(0,1)$ with $\beta = 1 \times 10^4$.

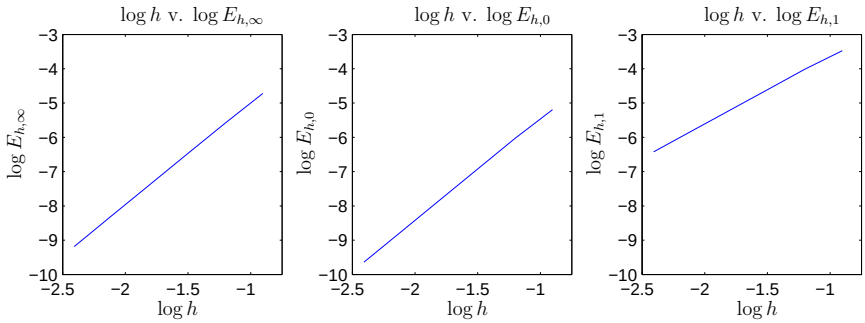


Figure 4.3: IFEM approximation error, $u_1 - u_h$, where $u_h \in \tilde{\mathcal{S}}_{h,2}(0,1)$ with $\beta = 1 \times 10^4$.

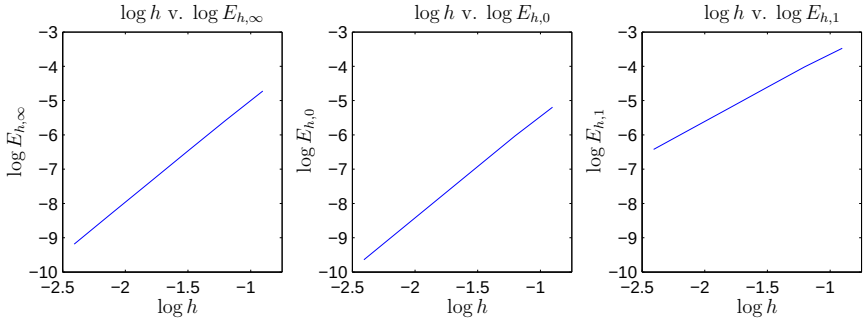


Figure 4.4: IFEM approximation error, $u_2 - u_h$, where $u_h \in \tilde{\mathcal{S}}_{h,2}(0,1)$ with $\beta = 1 \times 10^4$.

estimates

$$\begin{aligned} \tilde{E}_{h,\infty} = \|u_2 - u_h\|_\infty &\approx 0.010\ h^{2.971}, \\ \tilde{E}_{h,0} = \|u_2 - u_h\|_0 &\approx 0.003\ h^{2.969}, \\ \tilde{E}_{h,1} = \|u_2 - u_h\|_1 &\approx 0.022\ h^{1.971}. \end{aligned} \tag{4.29}$$

For this case in which the coefficient p has a large ratio, the IFE solutions that are generated from this space have orders of convergence comparable to the standard quadratic FE solutions. For the coefficient p with ratio $\beta = 1 \times 10^4$, the IFE solutions to the model interface problem seem to have a better approximation capability than interpolations from this IFE space. Apparently increasing the ratio β has the effect of enhancing the performance of the Galerkin FEM using the IFE space $\tilde{\mathcal{S}}_{h,2}(0,1)$.

4.2.2 IFE Solution Accuracy Using the Space $\bar{\mathcal{S}}_{h,2}(0,1)$

Next we look at the IFE solutions u_h that are generated in $\bar{\mathcal{S}}_{h,2}(0,1)$ for approximating u_1 and u_2 which are exact solutions to the elliptic interface problem (4.25) with the right hand side suitably chosen. Tables 4.5 and 4.6 list the errors in u_h approximating u_1 and u_2 for the

h	$\|u_1 - u_h\|_\infty$	$\|u_1 - u_h\|_0$	$\|u_1 - u_h\|_1$
1/8	1.893×10^{-5}	7.019×10^{-6}	3.678×10^{-4}
1/16	2.593×10^{-6}	9.557×10^{-7}	9.926×10^{-5}
1/32	3.289×10^{-7}	1.194×10^{-7}	2.481×10^{-5}
1/64	4.141×10^{-8}	1.496×10^{-8}	6.219×10^{-6}
1/128	5.195×10^{-9}	1.870×10^{-9}	1.552×10^{-6}
1/256	6.532×10^{-10}	2.350×10^{-10}	3.901×10^{-7}

Table 4.5: IFEM approximation error, $u_1 - u_h$, where $u_h \in \bar{\mathcal{S}}_{h,2}(0,1)$ with $\beta = 10$.

h	$\|u_2 - u_h\|_\infty$	$\|u_2 - u_h\|_0$	$\|u_2 - u_h\|_1$
1/8	3.048×10^{-4}	7.453×10^{-5}	4.114×10^{-3}
1/16	1.738×10^{-5}	1.621×10^{-6}	4.293×10^{-4}
1/32	1.238×10^{-5}	8.270×10^{-7}	3.628×10^{-4}
1/64	6.330×10^{-6}	4.248×10^{-7}	2.337×10^{-4}
1/128	5.404×10^{-7}	2.766×10^{-8}	2.445×10^{-5}
1/256	3.670×10^{-7}	1.457×10^{-8}	2.711×10^{-5}

Table 4.6: IFEM approximation error, $u_2 - u_h$, where $u_h \in \bar{\mathcal{S}}_{h,2}(0,1)$with $\beta = 10$.

coefficient p with $\beta = 10$. The behavior of the errors in the IFE solutions of u_1 generated from $\bar{\mathcal{S}}_{h,2}(0,1)$ is comparable to that of the standard FE solutions. On the other hand, similar to their interpolation counter part, the errors in the IFE solutions of u_2 generated from $\bar{\mathcal{S}}_{h,2}(0,1)$ decrease only in oscillatory way as h decreases. Using the data in these two tables, we can see the orders in the IFE solutions of u_1 using the IFE space $\bar{\mathcal{S}}_{h,2}(0,1)$ from

$$\begin{aligned} \bar{E}_{h,\infty} = \|u_1 - u_h\|_\infty &\approx 0.010\ h^{2.971}, \\ \bar{E}_{h,0} = \|u_1 - u_h\|_0 &\approx 0.004\ h^{2.981}, \\ \bar{E}_{h,1} = \|u_1 - u_h\|_1 &\approx 0.024\ h^{1.983}, \end{aligned} \tag{4.30}$$

and the orders of the IFE solutions of u_2 using the IFE space $\bar{\mathcal{S}}_{h,2}(0,1)$ are

$$\begin{aligned} \bar{E}_{h,\infty} = \|u_2 - u_h\|_\infty &\approx 0.007\ h^{1.842}, \\ \bar{E}_{h,0} = \|u_2 - u_h\|_0 &\approx 0.003\ h^{2.291}, \\ \bar{E}_{h,1} = \|u_2 - u_h\|_1 &\approx 0.046\ h^{1.408}. \end{aligned} \tag{4.31}$$

Again, the orders of these IFE solutions are consistent with those in the interpolants. The IFE solution of u_1 using the space $\bar{\mathcal{S}}_{h,2}(0,1)$ behaves as expected from a FE space using quadratic polynomials, while the orders of IFE solutions of u_2 are lower than those of a

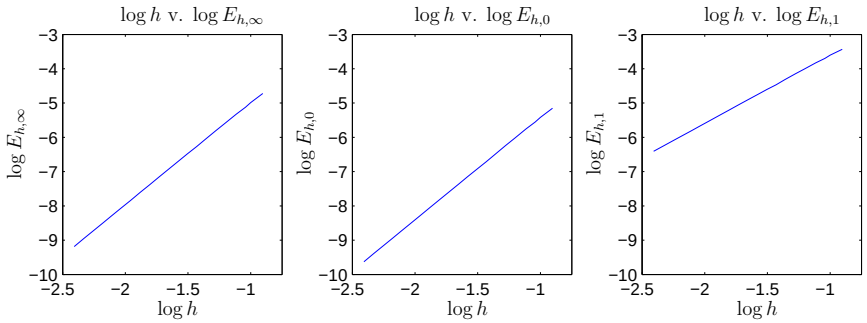


Figure 4.5: IFEM approximation error, $u_1 - u_h$, where $u_h \in \bar{\mathcal{S}}_{h,2}(0,1)$with $\beta = 10$.

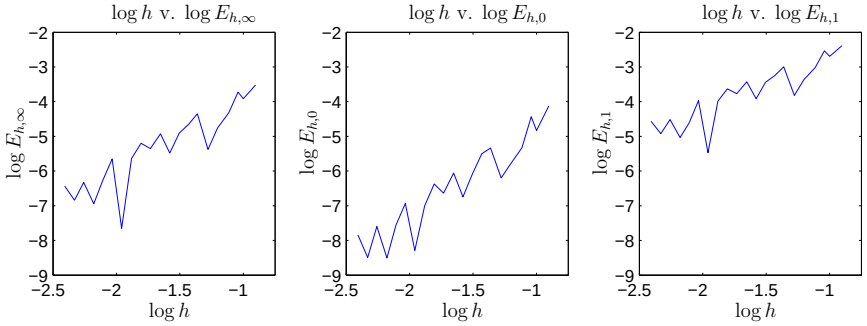


Figure 4.6: IFEM approximation error, $u_2 - u_h$, where $u_h \in \bar{\mathcal{S}}_{h,2}(0,1)$ with $\beta = 10$.

typical quadratic FE space. Figures 4.5 and 4.6 show plots of $\log h$ versus $\log E_h$. Again it is seen here that the errors in the IFE solutions of $u_1(x)$ generated from $\bar{\mathcal{S}}_{h,2}(0,1)$ behave very nicely while the errors in the IFE solutions to the test function $u_2(x)$ from the same IFE space tend to decrease albeit in an oscillatory fashion.

When the ratio of the coefficient p is increased to $\beta = 10000$ then the errors in the IFE approximations of the test functions $u_1(x)$ and $u_2(x)$ by the space $\bar{\mathcal{S}}_{h,2}(0,1)$ are shown in Tables 4.7 and 4.8. While the FE approximation errors for u_1 show no significant change,

h	$\|u_1 - u_h\|_\infty$	$\|u_1 - u_h\|_0$	$\|u_1 - u_h\|_1$
1/8	1.893×10^{-5}	6.422×10^{-6}	3.368×10^{-4}
1/16	2.593×10^{-6}	9.248×10^{-7}	9.591×10^{-5}
1/32	3.289×10^{-7}	1.156×10^{-7}	2.398×10^{-5}
1/64	4.141×10^{-8}	1.445×10^{-8}	5.996×10^{-6}
1/128	5.195×10^{-9}	1.811×10^{-9}	1.504×10^{-6}
1/256	6.531×10^{-10}	2.277×10^{-10}	3.779×10^{-7}

Table 4.7: IFEM approximation error, $u_1 - u_h$, where $u_h \in \bar{\mathcal{S}}_{h,2}(0,1)$ with $\beta = 1 \times 10^4$.

the errors in approximating u_2 with the IFE space $\bar{\mathcal{S}}_{h,2}(0,1)$ seem to behave better with less

h	$\|u_2 - u_h\|_\infty$	$\|u_2 - u_h\|_0$	$\|u_2 - u_h\|_1$
1/8	1.893×10^{-5}	6.420×10^{-6}	3.369×10^{-4}
1/16	3.594×10^{-6}	9.493×10^{-7}	1.522×10^{-4}
1/32	4.646×10^{-7}	1.191×10^{-7}	2.886×10^{-5}
1/64	4.141×10^{-8}	1.459×10^{-8}	6.119×10^{-6}
1/128	5.195×10^{-9}	1.810×10^{-9}	1.505×10^{-6}
1/256	8.535×10^{-10}	2.303×10^{-10}	3.840×10^{-7}

Table 4.8: IFEM approximation error, $u_2 - u_h$, where $u_h \in \bar{\mathcal{S}}_{h,2}(0,1)$ with $\beta = 1 \times 10^4$.

oscillation. The order of the errors in the FE approximations of u_1 using the IFE space $\bar{\mathcal{S}}_{h,2}(0,1)$ are

$$\begin{aligned} \bar{E}_{h,\infty} = \|u_1 - u_h\|_\infty &\approx 0.010\ h^{2.971}, \\ \bar{E}_{h,0} = \|u_1 - u_h\|_0 &\approx 0.003\ h^{2.969}, \\ \bar{E}_{h,1} = \|u_1 - u_h\|_1 &\approx 0.022\ h^{1.971}, \end{aligned} \tag{4.32}$$

while the order of the errors in the FE approximations of u_2 using the IFE space $\bar{\mathcal{S}}_{h,2}(0,1)$ are

$$\begin{aligned} \bar{E}_{h,\infty} = \|u_2 - u_h\|_\infty &\approx 0.011\ h^{2.971}, \\ \bar{E}_{h,0} = \|u_2 - u_h\|_0 &\approx 0.003\ h^{2.970}, \\ \bar{E}_{h,1} = \|u_2 - u_h\|_1 &\approx 0.030\ h^{2.032}. \end{aligned} \tag{4.33}$$

4.2.3 IFE Solution Accuracy Using the Space $\hat{\mathcal{S}}_{h,2}(0,1)$

Finally we look at errors of IFE solutions for u_1 and u_2 that are generated from $\hat{\mathcal{S}}_{h,2}(0,1)$ in Tables 4.9 and 4.10 respectively, where u_1 and u_2 are the exact solutions of the elliptic interface problem (4.25) when the coefficient p has a ratio of $\beta = 10$. The data from these

h	$\|u_1 - u_h\|_\infty$	$\|u_1 - u_h\|_0$	$\|u_1 - u_h\|_1$
1/8	1.893×10^{-5}	6.676×10^{-6}	3.499×10^{-4}
1/16	2.593×10^{-6}	9.540×10^{-7}	9.912×10^{-5}
1/32	3.289×10^{-7}	1.192×10^{-7}	2.474×10^{-5}
1/64	4.141×10^{-8}	1.490×10^{-8}	6.182×10^{-6}
1/128	5.195×10^{-9}	1.867×10^{-9}	1.551×10^{-6}
1/256	6.537×10^{-10}	2.347×10^{-10}	3.895×10^{-7}

Table 4.9: IFEM approximation error, $u_1 - u_h$, where $u_h \in \hat{\mathcal{S}}_{h,2}(0,1)$ with $\beta = 10$.

h	$\|u_2 - u_h\|_\infty$	$\|u_2 - u_h\|_0$	$\|u_2 - u_h\|_1$
1/8	1.893×10^{-5}	6.543×10^{-6}	3.431×10^{-4}
1/16	2.593×10^{-6}	9.409×10^{-7}	9.795×10^{-5}
1/32	3.289×10^{-7}	1.175×10^{-7}	2.440×10^{-5}
1/64	4.141×10^{-8}	1.469×10^{-8}	6.094×10^{-6}
1/128	5.195×10^{-9}	1.841×10^{-9}	1.529×10^{-6}
1/256	6.531×10^{-10}	2.314×10^{-10}	3.841×10^{-7}

Table 4.10: IFEM approximation error, $u_2 - u_h$, where $u_h \in \hat{\mathcal{S}}_{h,2}(0,1)$ with $\beta = 10$.

table indicate that the order of the IFE approximation of u_1 using the space $\hat{\mathcal{S}}_{h,2}(0,1)$ are

$$\begin{aligned} \hat{E}_{h,\infty} = \|u_1 - u_h\|_\infty &\approx 0.010\ h^{2.971}, \\ \hat{E}_{h,0} = \|u_1 - u_h\|_0 &\approx 0.003\ h^{2.971}, \\ \hat{E}_{h,1} = \|u_1 - u_h\|_1 &\approx 0.022\ h^{1.973}, \end{aligned} \tag{4.34}$$

and the order of the IFE approximations of u_2 are seen from

$$\begin{aligned} \hat{E}_{h,\infty} = \|u_2 - u_h\|_\infty &\approx 0.010\ h^{2.971}, \\ \hat{E}_{h,0} = \|u_2 - u_h\|_0 &\approx 0.003\ h^{2.969}, \\ \hat{E}_{h,1} = \|u_2 - u_h\|_1 &\approx 0.022\ h^{1.972}. \end{aligned} \tag{4.35}$$

In both cases, the IFE solutions generated from the space $\hat{\mathcal{S}}_{h,2}(0,1)$ for the test interface problems perform as well as is expected for the usual body fit quadratic FE spaces, and their errors and orders here are consistent with the results shown for the errors in the interpolants $I_h u_1$ and $I_h u_2$.

When the ratio of the coefficient p is increased to $\beta = 10000$, the errors in the FE approximations u_h of the functions u_1 and u_2 are shown in Tables 4.11 and 4.12 respectively. Based upon the data in these Tables, the order of errors in the FE approximations of u_1 are

$$\begin{aligned} \hat{E}_{h,\infty} = \|u_1 - u_h\|_\infty &\approx 0.010\ h^{2.971}, \\ \hat{E}_{h,0} = \|u_1 - u_h\|_0 &\approx 0.003\ h^{2.969}, \\ \hat{E}_{h,1} = \|u_1 - u_h\|_1 &\approx 0.022\ h^{1.971}, \end{aligned} \tag{4.36}$$

and the order of the errors in the FE approximations of u_2 are

$$\begin{aligned} \hat{E}_{h,\infty} = \|u_2 - u_h\|_\infty &\approx 0.010\ h^{2.971}, \\ \hat{E}_{h,0} = \|u_2 - u_h\|_0 &\approx 0.003\ h^{2.969}, \\ \hat{E}_{h,1} = \|u_2 - u_h\|_1 &\approx 0.022\ h^{1.971}. \end{aligned} \tag{4.37}$$

h	$\|u_1 - u_h\|_\infty$	$\|u_1 - u_h\|_0$	$\|u_1 - u_h\|_1$
1/8	1.893×10^{-5}	6.422×10^{-6}	3.368×10^{-4}
1/16	2.593×10^{-6}	9.248×10^{-7}	9.590×10^{-5}
1/32	3.289×10^{-7}	1.156×10^{-7}	2.398×10^{-5}
1/64	4.141×10^{-8}	1.445×10^{-8}	5.996×10^{-6}
1/128	5.195×10^{-9}	1.811×10^{-9}	1.504×10^{-6}
1/256	6.531×10^{-10}	2.277×10^{-10}	3.779×10^{-7}

Table 4.11: IFEM approximation error, $u_1 - u_h$, where $u_h \in \hat{\mathcal{S}}_{h,2}(0,1)$ with $\beta = 1 \times 10^4$.

h	$\|u_2 - u_h\|_\infty$	$\|u_2 - u_h\|_0$	$\|u_2 - u_h\|_1$
1/8	1.893×10^{-5}	6.422×10^{-6}	3.368×10^{-4}
1/16	2.593×10^{-6}	9.248×10^{-7}	9.590×10^{-5}
1/32	3.289×10^{-7}	1.156×10^{-7}	2.398×10^{-5}
1/64	4.141×10^{-8}	1.445×10^{-8}	5.996×10^{-6}
1/128	5.195×10^{-9}	1.811×10^{-9}	1.504×10^{-6}
1/256	6.531×10^{-10}	2.277×10^{-10}	3.779×10^{-7}

Table 4.12: IFEM approximation error, $u_2 - u_h$, where $u_h \in \hat{\mathcal{S}}_{h,2}(0,1)$ with $\beta = 1 \times 10^4$.

We finally summarize the approximation capability of the IFE solutions of the one dimensional test interface problems in Table 4.13:

	$\tilde{\mathcal{S}}_{h,2}(0,1)$	$\bar{\mathcal{S}}_{h,2}(0,1)$	$\hat{\mathcal{S}}_{h,2}(0,1)$
u_1	weaker than expected	as expected	as expected
u_2	weaker than expected	weaker than expected	as expected

Table 4.13: Approximation capability of the IFE solutions of the one dimensional test interface problems.

4.3 The Galerkin IFE Method in Two Dimensions

In this section, we investigate the errors in the Galerkin IFE solutions to the interface problem defined in the typical rectangular domain

$$\Omega = (0,1) \times (0,1) = (0,1)^2 \subset \mathbb{R}^2. \tag{4.38}$$

In this domain, the elliptic interface problem then becomes

$$\begin{array}{rcl} -\nabla \cdot (p(\boldsymbol{x})\nabla u(\boldsymbol{x})) &=& f(\boldsymbol{x}), \quad \boldsymbol{x} = (x,y)^T \in \Omega, \\ u(0,y) = u(1,y) &=& 0, \text{ if } 0 < y < 1, \\ u(x,0) = u(x,1) &=& 0, \text{ if } 0 < x < 1, \end{array} \tag{4.39}$$

where $p(\boldsymbol{x})$ is a piecewise constant function as defined in (3.32). In particular, the test functions $u_3(\boldsymbol{x})$ and $u_4(\boldsymbol{x})$ that are defined in (3.37) with the coefficients (3.38) and (3.39) respectively are used as the exact solution of (4.39). Hence the interface that is determined by the coefficient function p occurs along the horizontal line $y = e^{-1}$. Also, the right hand side function $f_i(\boldsymbol{x})$, $i = 3, 4$ in (4.39) is defined by

$$f_i(x,y) = \begin{cases} \pi P_a(y)\cos(\pi x) + P_a'(y)\sin(\pi x), & \text{if } y < e^{-1}, \\ \pi P_b(y)\cos(\pi x) + P_b'(y)\sin(\pi x), & \text{if } y \geq e^{-1}, \end{cases} \tag{4.40}$$

where $P_a(y)$ and $P_b(y)$ are polynomials defined by

$$P_a(y) = \sum_{j=0}^{5} a_j y^j \text{ and } P_b(y) = \sum_{j=0}^{5} b_j y^j. \tag{4.41}$$

For $f_3(x,y)$, the coefficients $a_j, b_j,\ j = 0, \ldots, 5$ are defined in (3.38) while for $f_4(x,y)$, the coefficients are defined in (3.39).

The three IFE spaces $\tilde{\mathcal{S}}_{h,2}\big((0,1)^2\big)$, $\bar{\mathcal{S}}_{h,2}\big((0,1)^2\big)$ and $\hat{\mathcal{S}}_{h,2}\big((0,1)^2\big)$ are used as the finite dimensional approximating space, $\mathcal{V}_h$, for the Galerkin finite element method. Again, we use different mesh sizes ranging from $h = 1/8$ to $h = 1/256$ in the same way as when these three biquadratic IFE spaces are tested for their interpolation capabilities in Chapter 3. In the same way, we evaluate the residual $r_h(x,y) = u(x,y) - u_h(x,y)$ at 36 points in a 6 by 6 mesh in each element for numerically determining the L^∞ error of the IFE solutions. For the L^2 and H^1 errors, a 4 by 4 mesh of gaussian nodes is used on each element. Of course, the interface elements are subdivided into two pieces that lie on both sides of the interface Γ and each sub-element is treated as described above. For calculating the orders of the errors, the command `polyfit` in MATLAB is again used to find the linear least squares fit for the $\log h$ vs. $\log E_h$ data. All orders are calculated using the data in the presented tables. The errors in the tables are presented with the primary purpose of showing the behavior of the errors as the size of the elements, h is reduced by factors of 2. On the other hand, some figures show more data than is presented in the tables so that a more complete trend of the errors can be seen. Again, the Galerkin IFE method is applied to the interface problems whose coefficient p have both a moderate ratio $\beta = p_2/p_1 = 10$ and a larger ratio $\beta = 10000$. If there is no substantial difference between the two cases, then only data from the case in which $\beta = 10$ is presented.

4.3.1 IFE Solution Accuracy Using the Space $\tilde{\mathcal{S}}_{h,2,\square}\big((0,1)^2\big)$

First we consider the Galerkin IFE approximations u_h when the finite dimensional subspace $\mathcal{V}_h$ is the hierarchical IFE space $\tilde{\mathcal{S}}_{h,2,\square}\big((0,1)^2\big)$. When $\beta = 10$, the data for the approximations u_h of the function u_3 are shown in Table 4.14. Likewise, the data for when $u_h \in \tilde{\mathcal{S}}_{h,2}\big((0,1)^2\big)$ approximating the function u_4 is shown in Table 4.15. Linear regression

h	$\|u_3 - u_h\|_\infty$	$\|u_3 - u_h\|_0$	$\|u_3 - u_h\|_1$
1/8	4.990×10^{-2}	1.505×10^{-2}	7.963×10^{-1}
1/16	6.241×10^{-3}	1.897×10^{-3}	1.994×10^{-1}
1/32	7.802×10^{-4}	2.402×10^{-4}	5.012×10^{-2}
1/64	9.761×10^{-5}	3.055×10^{-5}	1.269×10^{-2}
1/128	1.220×10^{-5}	3.854×10^{-6}	3.570×10^{-3}
1/256	3.507×10^{-6}	4.943×10^{-7}	9.597×10^{-4}

Table 4.14: $E_{h,s} = \|u_3 - u_h\|_s, \quad s = \infty, 0, 1$ where $u_h \in \tilde{\mathcal{S}}_{h,2}\big((0,1)^2\big)$ and $\beta = 10$.

h	$\|u_4 - u_h\|_\infty$	$\|u_4 - u_h\|_0$	$\|u_4 - u_h\|_1$
1/8	7.401×10^{-2}	2.216×10^{-2}	1.172×10^{0}
1/16	9.233×10^{-3}	2.796×10^{-3}	2.936×10^{-1}
1/32	1.154×10^{-3}	3.547×10^{-4}	7.388×10^{-2}
1/64	2.200×10^{-4}	4.630×10^{-5}	1.926×10^{-2}
1/128	3.541×10^{-5}	5.860×10^{-6}	8.569×10^{-3}
1/256	1.583×10^{-5}	8.887×10^{-7}	2.699×10^{-3}

Table 4.15: $E_{h,s} = \|u_4 - u_h\|_s, \quad s = \infty, 0, 1$ where $u_h \in \tilde{\mathcal{S}}_{h,2}\big((0,1)^2\big)$ and $\beta = 10$.

applied to the data shows that the orders of the error in approximating u_3 are

$$\begin{aligned} E_{h,\infty} = \|u_3 - u_h\|_\infty &\approx 15.247\ h^{2.828}, \\ E_{h,0} = \|u_3 - u_h\|_0 &\approx 7.351\ h^{2.979}, \\ E_{h,1} = \|u_3 - u_h\|_1 &\approx 43.049\ h^{1.939}, \end{aligned} \tag{4.42}$$

while the orders of approximating u_4 obey the relationships

$$\begin{aligned} E_{h,\infty} = \|u_4 - u_h\|_\infty &\approx 9.265\ h^{2.498}, \\ E_{h,0} = \|u_4 - u_h\|_0 &\approx 9.487\ h^{2.933}, \\ E_{h,1} = \|u_4 - u_h\|_1 &\approx 36.617\ h^{1.744}. \end{aligned} \tag{4.43}$$

We note that, in the L^2 norm, the performance of these IFE solutions match that generated from a standard quadratic FE space. For test functions u_4, in both L^∞ and H^1 norms,

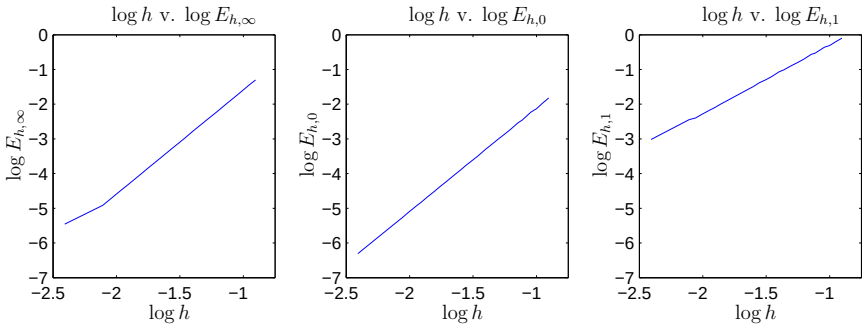


Figure 4.7: $E_{h,s} = \|u_3 - u_h\|_s, \quad s = \infty, 0, 1$ where $u_h \in \tilde{\mathcal{S}}_{h,2}\big((0,1)^2\big)$ and $\beta = 10$.

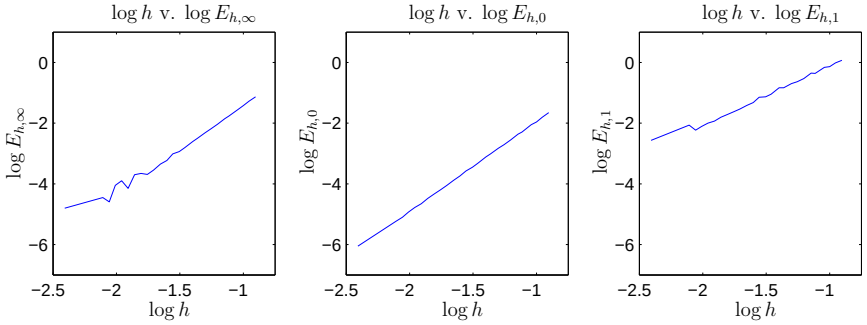


Figure 4.8: $E_{h,s} = \|u_4 - u_h\|_s, \quad s = \infty, 0, 1$ where $u_h \in \tilde{\mathcal{S}}_{h,2}\big((0,1)^2\big)$ and $\beta = 10$.

the orders of the IFE solutions seem to be lower than those from a standard quadratic FE space. It is interesting to see that the IFE solution to u_4 has higher orders than the IFE interpolant of u_4. It is seen that the $\log h$ vs. $\log E_h$ data for the FE approximation of the function u_3 has a nice linear trend as indicated in (4.42). This is supported by Figure 4.8 which shows the *logh* vs. $\log E_h$ trends of the L^∞, L^2 and H^1 norm errors of the Galerkin IFEM approximations of the function u_3.

When the coefficient ratio β is increased to 10000, the errors in the IFE solutions of u_3 and u_4, the solutions of the BVP (4.39), are shown in Tables 4.16 and 4.17 respectively for various element sizes h. The most notable difference in the data for the case when $\beta = 10000$ compared to the case when $\beta = 10$ is that the errors seem to have increased by about a factor of 10^3. Although the errors are greater in magnitude, it is seen that the errors do decrease as $h \to 0$. In fact, when linear regression is applied to the $\log h$ vs. $\log E_h$ data, the order of the IFE solutions approximating u_3 are found to be

$$\begin{aligned} E_{h,\infty} = \|u_3 - u_h\|_\infty &\approx 33192.606\ h^{3.000}, \\ E_{h,0} = \|u_3 - u_h\|_0 &\approx 9642.636\ h^{2.984}, \\ E_{h,1} = \|u_3 - u_h\|_1 &\approx 64924.097\ h^{1.992}. \end{aligned} \tag{4.44}$$

h	$\|u_3 - u_h\|_\infty$	$\|u_3 - u_h\|_0$	$\|u_3 - u_h\|_1$
1/8	6.478×10^{1}	1.943×10^{1}	1.033×10^{3}
1/16	8.100×10^{0}	2.452×10^{0}	2.588×10^{2}
1/32	1.012×10^{0}	3.108×10^{-1}	6.502×10^{1}
1/64	1.264×10^{-1}	3.945×10^{-2}	1.639×10^{1}
1/128	1.580×10^{-2}	4.983×10^{-3}	4.133×10^{0}
1/256	1.976×10^{-3}	6.228×10^{-4}	1.033×10^{0}

Table 4.16: $E_{h,s} = \|u_3 - u_h\|_s, \quad s = \infty, 0, 1$ where $u_h \in \tilde{\mathcal{S}}_{h,2}\big((0,1)^2\big)$ and $\beta = 1 \times 10^4$.

h	$\|u_4 - u_h\|_\infty$	$\|u_4 - u_h\|_0$	$\|u_4 - u_h\|_1$
1/8	7.343×10^{1}	2.191×10^{1}	1.165×10^{3}
1/16	9.150×10^{0}	2.766×10^{0}	2.918×10^{2}
1/32	1.142×10^{0}	3.505×10^{-1}	7.332×10^{1}
1/64	1.428×10^{-1}	4.448×10^{-2}	1.848×10^{1}
1/128	1.785×10^{-2}	5.619×10^{-3}	4.661×10^{0}
1/256	2.231×10^{-3}	7.023×10^{-4}	1.165×10^{0}

Table 4.17: $E_{h,s} = \|u_4 - u_h\|_s, \quad s = \infty, 0, 1$ where $u_h \in \tilde{\mathcal{S}}_{h,2}\big((0,1)^2\big)$ and $\beta = 1 \times 10^4$.

Similarly, the orders of IFE solutions approximating u_4 are found to be

$$\begin{aligned} E_{h,\infty} = \|u_4 - u_h\|_\infty &\approx 37618.216\ h^{3.001}, \\ E_{h,0} = \|u_4 - u_h\|_0 &\approx 10876.001\ h^{2.984}, \\ E_{h,1} = \|u_4 - u_h\|_1 &\approx 73224.583\ h^{1.992}. \end{aligned} \tag{4.45}$$

These orders indicate that increasing the value of β to 10000 has the effect of removing the deficiencies in the orders of the approximations that are present when $\beta = 10$. At this higher coefficient ratio, the Galerkin IFE method now appears to have the typical orders of $\mathcal{O}(h^3)$ for the L^∞ and L^2 errors and $\mathcal{O}(h^2)$ for the H^1 errors as expected from biquadratic finite element spaces. A larger set of data for the $\log h$ vs. $\log E_h$ data are shown in Figures 4.9 and 4.10 for the IFE solutions of u_3 and u_4 respectively when $u_h \in \tilde{\mathcal{S}}_{h,2}\big((0,1)^2\big)$. Here it is seen that there are virtually no oscillations present in the $\log h$ vs. $\log E_h$ plots. On the other hand, minor oscillations are detected when the coefficient p has a ratio of $\beta = 10$.

When the coefficient function p has a ratio of $\beta = 10$ or $\beta = 10000$, it is interesting to note that the IFE solutions to the interface problems perform better than the corresponding interpolation errors using the same IFE space $\tilde{\mathcal{S}}_{h,2}\big((0,1)^2\big)$. This feature is especially noticeable however, when coefficient ratio β is increased from 10 to 10000. In the IFE approximations, the involved BVP seems to smooth the errors out and remove most of the oscillation in the $\log h$ vs. $\log E_h$ plots. On the other hand, in the interpolation errors, increasing the

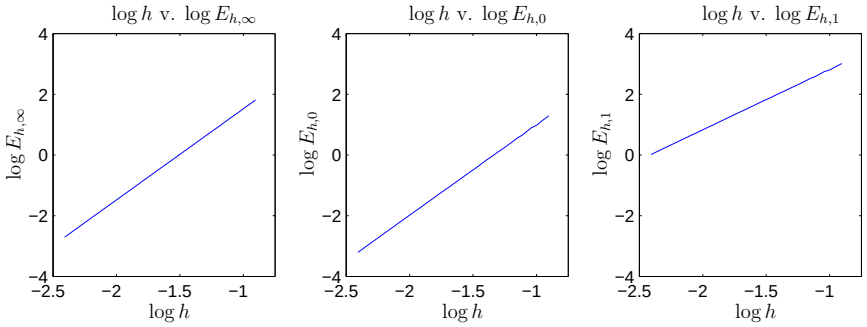


Figure 4.9: $E_{h,s} = \|u_3 - u_h\|_s, \quad s = \infty, 0, 1$ where $u_h \in \tilde{\mathcal{S}}_{h,2}\big((0,1)^2\big)$ and $\beta = 1 \times 10^4$.

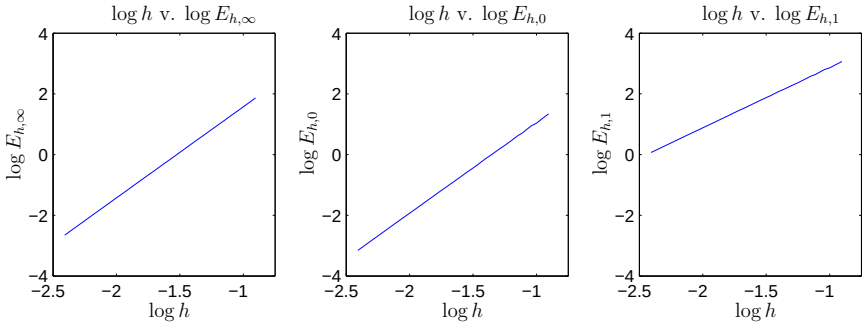


Figure 4.10: $E_{h,s} = \|u_4 - u_h\|_s, \quad s = \infty, 0, 1$ where $u_h \in \tilde{\mathcal{S}}_{h,2}\big((0,1)^2\big)$ and $\beta = 1 \times 10^4$.

ratios does not have any noticeable effect upon the errors or their orders when using the hierarchical IFE space.

4.3.2 IFE Solution Accuracy Using the Space $\bar{\mathcal{S}}_{h,2,\square}\big((0,1)^2\big)$

Next we consider the errors of the IFE solutions to (4.19) when either u_3 or u_4 is the exact solution of the BVP (4.39) and the weak solution is in the biquadratic IFE space $\bar{\mathcal{S}}_{h,2,\square}\big((0,1)^2\big)$ that uses an extra continuity requirement. The errors in these IFE solutions are shown in Tables 4.18 and 4.19 for the exact solutions u_3 and u_4 respectively with a coefficient ratio of $\beta = 10$ used for the coefficient function p. Linear regression on the $\log h$

h	$\|u_3 - u_h\|_\infty$	$\|u_3 - u_h\|_0$	$\|u_3 - u_h\|_1$
1/8	4.990×10^{-2}	1.503×10^{-2}	7.962×10^{-1}
1/16	6.241×10^{-3}	1.887×10^{-3}	1.994×10^{-1}
1/32	7.813×10^{-4}	2.430×10^{-4}	5.115×10^{-2}
1/64	6.143×10^{-4}	4.837×10^{-5}	2.463×10^{-2}
1/128	3.307×10^{-5}	3.893×10^{-6}	6.221×10^{-3}
1/256	2.430×10^{-5}	6.157×10^{-7}	3.370×10^{-3}

Table 4.18: $E_{h,s} = \|u_3 - u_h\|_s, \quad s = \infty, 0, 1$ where $u_h \in \bar{\mathcal{S}}_{h,2}\big((0,1)^2\big)$ and $\beta = 10$.

h	$\|u_4 - u_h\|_\infty$	$\|u_4 - u_h\|_0$	$\|u_4 - u_h\|_1$
1/8	7.400×10^{-2}	2.217×10^{-2}	1.172×10^{0}
1/16	9.233×10^{-3}	2.796×10^{-3}	2.936×10^{-1}
1/32	1.153×10^{-3}	3.538×10^{-4}	7.383×10^{-2}
1/64	1.442×10^{-4}	4.481×10^{-5}	1.864×10^{-2}
1/128	1.801×10^{-5}	5.649×10^{-6}	4.686×10^{-3}
1/256	2.251×10^{-6}	7.061×10^{-7}	1.172×10^{-3}

Table 4.19: $E_{h,s} = \|u_4 - u_h\|_s, \quad s = \infty, 0, 1$ where $u_h \in \bar{\mathcal{S}}_{h,2}\big((0,1)^2\big)$ and $\beta = 10$.

vs $\log E_h$ data in these tables generates the following estimates on the orders of these IFE solutions when approximating the solution u_3

$$\begin{aligned} E_{h,\infty} = \|u_3 - u_h\|_\infty &\approx 3.456\ h^{2.230}, \\ E_{h,0} = \|u_3 - u_h\|_0 &\approx 6.417\ h^{2.913}, \\ E_{h,1} = \|u_3 - u_h\|_1 &\approx 16.916\ h^{1.585}, \end{aligned} \tag{4.46}$$

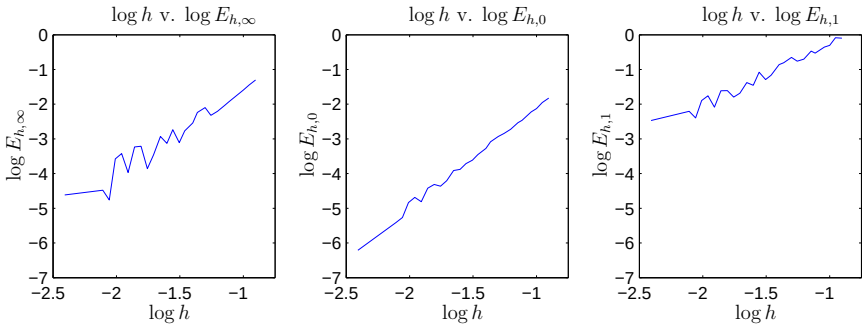


Figure 4.11: $E_{h,s} = \|u_3 - u_h\|_s, \quad s = \infty, 0, 1$ where $u_h \in \bar{\mathcal{S}}_{h,2}\big((0,1)^2\big)$ and $\beta = 10$.

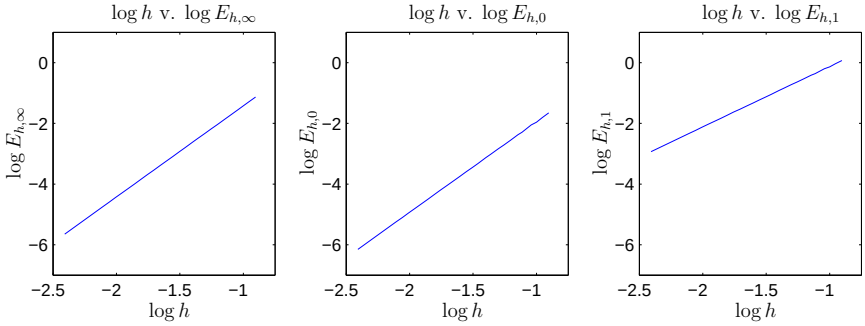


Figure 4.12: $E_{h,s} = \|u_4 - u_h\|_s, \quad s = \infty, 0, 1$ where $u_h \in \bar{\mathcal{S}}_{h,2}\big((0,1)^2\big)$ and $\beta = 10$.

and when approximating the solution u_4

$$\begin{aligned} E_{h,\infty} = \|u_4 - u_h\|_\infty &\approx 37.916\ h^{3.001}, \\ E_{h,0} = \|u_4 - u_h\|_0 &\approx 11.056\ h^{2.987}, \\ E_{h,1} = \|u_4 - u_h\|_1 &\approx 73.738\ h^{1.992}. \end{aligned} \tag{4.47}$$

In both L^∞ and H^1 norms, the order of the IFE solutions from $\bar{\mathcal{S}}_{h,2}\big((0,1)^2\big)$ for u_3 are lower than those generated with the standard biquadratic FE space using a body fit partition while L^2 norm errors of the IFE approximation of u_3 seem to have the expected orders. On the other hand, the orders of the IFE solutions from $\bar{\mathcal{S}}_{h,2}\big((0,1)^2\big)$ for u_4 match the expected orders generated by approximations from a standard biquadratic FE space using a body fit partition.

When the coefficient function p has a ratio of $\beta = 10000$, the errors in the IFE solutions $u_h \in \bar{\mathcal{S}}_{h,2}\big((0,1)^2\big)$ are shown in Tables 4.20 and 4.21 when u_3 and u_4 are the exact solutions of (4.39) respectively. As in the case when the hierarchical IFE space $\tilde{\mathcal{S}}_{h,2}\big((0,1)^2\big)$ is used, the errors seem to have increased in magnitude by a factor of 10^3. Again, estimates of the approximation orders of u_h are obtained by performing linear regression on the $\log h$ vs. $\log E_h$ data from the tables. The orders for the approximations of u_3 when $u_h \in \bar{\mathcal{S}}_{h,2}\big((0,1)^2\big)$

h	$\|u_3 - u_h\|_\infty$	$\|u_3 - u_h\|_0$	$\|u_3 - u_h\|_1$
1/8	6.478×10^{1}	1.943×10^{1}	1.033×10^{3}
1/16	8.100×10^{0}	2.452×10^{0}	2.588×10^{2}
1/32	1.012×10^{0}	3.108×10^{-1}	6.502×10^{1}
1/64	1.264×10^{-1}	3.945×10^{-2}	1.639×10^{1}
1/128	1.581×10^{-2}	4.984×10^{-3}	4.542×10^{0}
1/256	1.976×10^{-3}	6.231×10^{-4}	1.051×10^{0}

Table 4.20: $E_{h,s} = \|u_3 - u_h\|_s, \quad s = \infty, 0, 1$ where $u_h \in \bar{\mathcal{S}}_{h,2}\big((0,1)^2\big)$ and $\beta = 1 \times 10^4$.

h	$\|u_4 - u_h\|_\infty$	$\|u_4 - u_h\|_0$	$\|u_4 - u_h\|_1$
1/8	7.343×10^{1}	2.191×10^{1}	1.165×10^{3}
1/16	9.150×10^{0}	2.766×10^{0}	2.918×10^{2}
1/32	1.142×10^{0}	3.505×10^{-1}	7.332×10^{1}
1/64	1.428×10^{-1}	4.448×10^{-2}	1.848×10^{1}
1/128	1.785×10^{-2}	5.619×10^{-3}	4.661×10^{0}
1/256	2.231×10^{-3}	7.023×10^{-4}	1.165×10^{0}

Table 4.21: $E_{h,s} = \|u_4 - u_h\|_s, \quad s = \infty, 0, 1$ where $u_h \in \bar{\mathcal{S}}_{h,2}\big((0,1)^2\big)$ and $\beta = 1 \times 10^4$.

obey the relationships

$$\begin{aligned} E_{h,\infty} = \|u_3 - u_h\|_\infty &\approx 33188.998\ h^{3.000}, \\ E_{h,0} = \|u_3 - u_h\|_0 &\approx 9639.443\ h^{2.984}, \\ E_{h,1} = \|u_3 - u_h\|_1 &\approx 62439.150\ h^{1.977}, \end{aligned} \tag{4.48}$$

while the orders of the approximations of u_4 are estimated to be

$$\begin{aligned} E_{h,\infty} = \|u_4 - u_h\|_\infty &\approx 37620.639\ h^{3.001}, \\ E_{h,0} = \|u_4 - u_h\|_0 &\approx 10876.027\ h^{2.984}, \\ E_{h,1} = \|u_4 - u_h\|_1 &\approx 73227.272\ h^{1.992}. \end{aligned} \tag{4.49}$$

There is no great change in the orders of the IFE solutions when the exact solution is u_4 instead of u_3. The IFE space $\bar{\mathcal{S}}_{h,2}\big((0,1)^2\big)$ continues to perform as well as a biquadratic FE space. On the other hand, while the IFE solutions that approximate u_3 have deficiencies in their orders, increasing the value of the ratio β seems to correct this problem. This is confirmed by looking at Figure 4.13 which shows the $\log h$ vs. $\log E_h$ plots when the FE method is used to approximate the exact solution u_3 for more element sizes h than those presented in Table 4.20. Any oscillations that are present when $\beta = 10$ are either reduced or removed from the trend. The effect of the errors being of greater magnitude is seen by the shifting of the trend-lines in the vertical direction on the plots in comparison with the data that are shown in Figure 4.11.

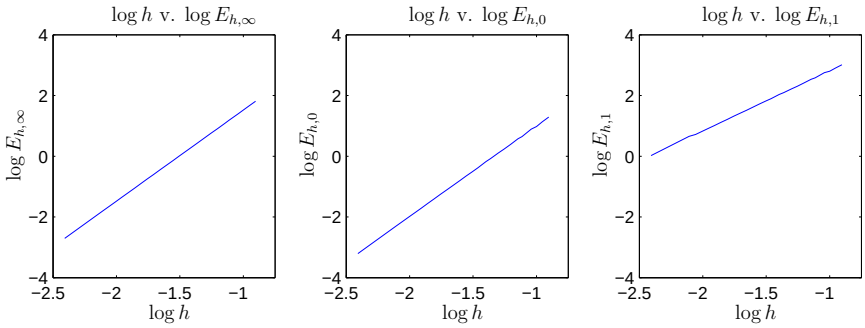


Figure 4.13: $E_{h,s} = \|u_3 - u_h\|_s, \quad s = \infty, 0, 1$ where $u_h \in \bar{\mathcal{S}}_{h,2}\big((0,1)^2\big)$ and $\beta = 1 \times 10^4$.

Again, the IFE solutions seem to show orders which are better than the corresponding interpolation orders. This time, the IFE space $\bar{\mathcal{S}}_{h,2}\big((0,1)^2\big)$ performs better in solving the interface problem with a solution of u_3 using the Galerkin finite element method than compared with the interpolation errors that we see in the previous chapter. On the other hand, the approximations of the function u_4 continue to show errors that display a nice correlation between $\log h$ and $\log E_h$ for both the coefficient ratios of $\beta = 10$ and $\beta = 10000$. Additionally, the orders of the errors in approximating u_4 when $u_h \in \bar{\mathcal{S}}_{h,2}\big((0,1)^2\big)$ which are shown in (4.47) and (4.49) are consistent with the interpolation errors that are exhibited by this space.

4.3.3 IFE Solution Accuracy Using the Space $\hat{\mathcal{S}}_{h,2,\square}\big((0,1)^2\big)$

Finally, we look at the IFE solutions that are generated from the space $\hat{\mathcal{S}}_{h,2,\square}\big((0,1)^2\big)$ for test functions u_3 and u_4. The errors of the IFE solutions to (4.19) are shown in Tables 4.22 and 4.23 when u_3 and u_4 are the exact solutions of the model interface problem (4.39) respectively with a coefficient ratio $\beta = p_2/p_1 = 10$. Linear regression of the $\log h$ vs. $\log E_h$ data that

h	$\|u_3 - u_h\|_\infty$	$\|u_3 - u_h\|_0$	$\|u_3 - u_h\|_1$
1/8	4.990×10^{-2}	1.499×10^{-2}	7.962×10^{-1}
1/16	6.241×10^{-3}	1.891×10^{-3}	1.994×10^{-1}
1/32	7.802×10^{-4}	2.396×10^{-4}	5.009×10^{-2}
1/64	9.752×10^{-5}	3.039×10^{-5}	1.262×10^{-2}
1/128	1.219×10^{-5}	3.837×10^{-6}	3.183×10^{-3}
1/256	1.524×10^{-6}	4.796×10^{-7}	7.958×10^{-4}

Table 4.22: $E_{h,s} = \|u_3 - u_h\|_s, \quad s = \infty, 0, 1$ where $u_h \in \hat{\mathcal{S}}_{h,2}\big((0,1)^2\big)$ and $\beta = 10$.

can be generated from these tables provides estimates of the orders of these IFE solutions.

h	$\|u_4 - u_h\|_\infty$	$\|u_4 - u_h\|_0$	$\|u_4 - u_h\|_1$
1/8	7.400×10^{-2}	2.207×10^{-2}	1.172×10^{0}
1/16	9.231×10^{-3}	2.785×10^{-3}	2.935×10^{-1}
1/32	1.152×10^{-3}	3.527×10^{-4}	7.375×10^{-2}
1/64	1.441×10^{-4}	4.474×10^{-5}	1.858×10^{-2}
1/128	1.801×10^{-5}	5.649×10^{-6}	4.686×10^{-3}
1/256	2.251×10^{-6}	7.061×10^{-7}	1.171×10^{-3}

Table 4.23: $E_{h,s} = \|u_4 - u_h\|_s, \quad s = \infty, 0, 1$ where $u_h \in \hat{\mathcal{S}}_{h,2}\big((0,1)^2\big)$ and $\beta = 10$.

When the exact solution is the testing function u_3, then the FE solutions $u_h \in \hat{\mathcal{S}}_{h,2}\big((0,1)^2\big)$ have orders of

$$\begin{aligned} E_{h,\infty} = \|u_3 - u_h\|_\infty &\approx 25.551\ h^{3.000}, \\ E_{h,0} = \|u_3 - u_h\|_0 &\approx 7.448\ h^{2.985}, \\ E_{h,1} = \|u_3 - u_h\|_1 &\approx 50.038\ h^{1.992}, \end{aligned} \tag{4.50}$$

while the FE solutions have the orders

$$\begin{aligned} E_{h,\infty} = \|u_4 - u_h\|_\infty &\approx 37.907\ h^{3.001}, \\ E_{h,0} = \|u_4 - u_h\|_0 &\approx 10.966\ h^{2.985}, \\ E_{h,1} = \|u_4 - u_h\|_1 &\approx 73.687\ h^{1.992}, \end{aligned} \tag{4.51}$$

when the exact solution is given by the testing function u_4. There is nothing remarkable to say here other than that again the performance of the IFE space $\hat{\mathcal{S}}_{h,2}\big((0,1)^2\big)$ is comparable to the standard biquadratic FE space when used to solve the model interface problems. When the ratio β is increased to $\beta = 10000$, the orders do not change. However, as is the case with the interpolation errors shown in Chapter 3, the magnitudes of the errors in the IFE solutions do increase by about a factor of 10^3 although this increase is also seen in the errors generated using the IFE spaces $\tilde{\mathcal{S}}_{h,2}\big((0,1)^2\big)$ and $\bar{\mathcal{S}}_{h,2}\big((0,1)^2\big)$. More importantly, the orders of the errors here are consistent with the interpolation errors in Chapter 3 that arise when using the space $\hat{\mathcal{S}}_{h,2}\big((0,1)^2\big)$.

Overall, we conclude that, when used to solve the model interface problems, the approximation capability of the two dimensional quadratic IFE spaces is similar to that of the one dimensional IFE spaces. When the coefficient ratio is $\beta = 10$, then the approximation capability of each IFE space tends to agree with the interpolation errors that are generated in Chapter 3. Accordingly, the spaces $\tilde{\mathcal{S}}_{h,2}$ and $\bar{\mathcal{S}}_{h,2}$ do not perform equally as well in the approximation of both functions u_3 and u_4. When the coefficient ratio is increased to $\beta = 10000$, both of these spaces $\tilde{\mathcal{S}}_{h,2}$ and $\bar{\mathcal{S}}_{h,2}$ return good approximation errors for both functions u_3 and u_4. This is not expected as the interpolation errors that are shown by both of these functions does not predict this improved behavior. On the other hand, the IFE space $\hat{\mathcal{S}}_{h,2}$ performs exactly as well as its interpolation errors shown in Chapter 3 regardless of which value for the coefficient ratio β is used.

Chapter 5

The Mixed Least Squares Immersed Finite Element Method

This chapter details how the least squares FE method (LSFEM) can be applied to an elliptic boundary value problem such as (1.1) when the coefficient p is discontinuous and hence introduces an interface Γ into the domain Ω. The extensive survey article by Bochev and Gunzburger [7] and the references therein provide a broad introduction to the subject of LSFE methods. Additionally, Jiang [30] provides a thorough treatment of the topic of the LSFEM and contrasts the method with the mixed Galerkin method. The basic idea of this method is to rewrite the elliptic BVP as an equivalent system of first order partial differential equations and then to minimize an appropriate quadratic cost functional that is related to the residuals of these first order PDEs. A necessary condition for the functional to have a minimum is that its first variation is equal to zero at the minimum. Taking the first variation of a quadratic cost functional is straightforward and this leads to a variational problem which can then be discretized. The usual finite element techniques can be applied to approximate the solution of the variational problem.

Since an interface separates the domain into two distinct regions, the IFE spaces introduced previously (see Chapter 2) can be used in the finite element approximation while using a fixed grid for the finite element partition. Using the IFE spaces is important for two reasons. First, the Least Squares Finite Element Method has not been implemented yet using immersed finite element spaces. Second, using a fixed partition is important for moving interface problems such as **Problem (I)** which we consider in this thesis. As we see in the next chapter, the Least Squares Finite Element Method is an important component of the mixed least squares cost functional that can be used to solve **Problem (I)** in one and two dimensions.

5.1 Background

First, recall our model interface problem which is to find u such that

$$\begin{aligned} -\nabla \cdot (p\nabla u) &= f, \\ u\big|_{\partial\Omega} &= 0, \end{aligned} \tag{5.1}$$

where the coefficient function p is a piecewise constant function of the form (1.2).

In Chapter 4, the Galerkin method is used as a method for approximating the solution u of the interface problem. In essence, the Galerkin method first reformulates the BVP (5.1) as a weaker problem so that u can be found in a larger space than if it is a classical solution. Then a finite dimensional subspace of the proposed solution space is introduced and the "best approximation" u_h is found for the solution u in this subspace. One drawback of the Galerkin method, however, occurs when information about ∇u is needed. For example, accurate information about $p\nabla u$ is needed in the inverse **Problem (I)**, since the flux of the boundary is one of the constraints imposed by **Problem (I)**. In other applications, ∇u is important since it represents a physical quantity such as electromagnetic flux, which is usually more important than the electromagnetic potential. In general, approximations of ∇u, through the differentiation of the Galerkin finite element solution u_h, lose one order of accuracy as evidenced by the fact that typically $E_{h,0} = \mathcal{O}(h^3)$ while $E_{h,1} = \mathcal{O}(h^2)$ when quadratic or biquadratic finite element functions are used, for example. Additionally, with the Galerkin method, ∇u_h is then obtained by taking the derivative of u_h. However, u_h is only piecewise smooth if it is generated from C^0 finite element space, and so ∇u_h turns out to only be piecewise continuous.

If better information is needed about ∇u, then one way to obtain a continuous approximation of ∇u is to decompose the boundary value problem into an equivalent system of first order differential equations [10, 8]. The so-called mixed formulation of the elliptic interface problem (5.1) accomplishes this by introducing a new variable $\boldsymbol{\sigma} = \nabla u$ which is precisely what we need better information for. Then the boundary value problem (5.1) can be recast as

$$\begin{aligned} \nabla u(\boldsymbol{x}) - \boldsymbol{\sigma}(\boldsymbol{x}) &= 0, \text{ if } \boldsymbol{x} \in \Omega, \\ -\nabla \cdot (p(\boldsymbol{x})\boldsymbol{\sigma}(\boldsymbol{x})) &= f(\boldsymbol{x}), \text{ if } \boldsymbol{x} \in \Omega, \\ u(\boldsymbol{x}) &= 0, \text{ if } \boldsymbol{x} \in \partial\Omega. \end{aligned} \tag{5.2}$$

A variational problem is then obtained in the usual fashion by multiplying both equations in (5.2) with appropriate test functions, integrating the resulting equations over the domain Ω, and then using integration by parts where applicable which leads us to a mixed variational problem (sometimes called a dual variational problem) [8, 9, 10]. The appropriate spaces for the test functions and the weak solutions are

$$\begin{aligned} \mathcal{V} = \big\{ u \in L^2(\Omega) \big| \;\; & u|_{\Omega_1} \in H^1(\Omega_1), u|_{\Omega_2} \in H^1(\Omega_2), \\ & u|_{\partial\Omega} = 0, \big[u\big]_\Gamma = 0, \big[p\tfrac{\partial u}{\partial \boldsymbol{n}}\big]_\Gamma = 0 \big\}, \end{aligned} \tag{5.3}$$

and

$$\mathcal{W} = \big\{(\sigma_1, \sigma_2)\big|\ \sigma_i \in L^2(\Omega),\ i = 1, 2\big\}, \tag{5.4}$$

where $\Omega \subset \mathbb{R}^n$. Then the mixed variational problem is to find $(u, \boldsymbol{\sigma}) \in \mathcal{V} \times \mathcal{W}$ such that

$$\begin{array}{rcll} a(\boldsymbol{\sigma}, \boldsymbol{\xi}) & + & b(\boldsymbol{\xi}, u) = & 0 \quad \text{for all } \boldsymbol{\xi} \in \mathcal{W}, \\ b(\boldsymbol{\sigma}, v) & & = -(v, f) & \text{for all } v \in \mathcal{V}, \end{array} \tag{5.5}$$

where $(\cdot, \cdot)$ is the L^2 inner product on Ω and the bilinear forms a and b are given by

$$a(\boldsymbol{\sigma}, \boldsymbol{\xi}) = \int_\Omega \boldsymbol{\sigma}^T \boldsymbol{\xi} d\boldsymbol{x}, \tag{5.6}$$

$$b(\boldsymbol{\sigma}, v) = -\int_\Omega \boldsymbol{\sigma}^T (p \nabla v) d\boldsymbol{x}. \tag{5.7}$$

This results in a saddle point problem. So, in order for (5.5) to have a unique solution, the bilinear form a must be $\mathcal{W}$-elliptic and b must satisfy the "inf $-$ sup" condition [8, 9, 10, 33].

Definition 5.1.1. *Let $b : \mathcal{W} \times \mathcal{V} \to \mathbb{R}$ be a bilinear continuous form. Then b satisfies the "*inf $-$ sup*" condition if there exists a $\beta > 0$ such that*

$$\inf_{v \in \mathcal{V}} \sup_{\boldsymbol{\xi} \in \mathcal{W}} \frac{b(\boldsymbol{\xi}, v)}{\|v\| \|\boldsymbol{\xi}\|} \geq \beta. \tag{5.8}$$

This is often called the Ladyshenskaja-Babuška-Brezzi condition or the LBB condition for short.

This statement about the bilinear form b is dependent upon the choice of the spaces $\mathcal{V}$ and $\mathcal{W}$. Accordingly, this gives a kind of ellipticity condition for the bilinear form b which is seen more clearly by

$$\sup_{\boldsymbol{\xi} \in \mathcal{W}} \frac{b(\boldsymbol{\xi}, v)}{\|\boldsymbol{\xi}\|_{\mathcal{W}}} \geq \beta \|v\|_{\mathcal{V}} \text{ for all } v \in \mathcal{V}, \tag{5.9}$$

which is equivalent to (5.8). The importance of this condition is first observed by Aziz and Bubuška [2] since it helps to establish a theorem (i.e. see Braess [8] p.120) for the existence and uniqueness of abstract variational problems. As it turns out, the Lax-Milgram Theorem is a specific case of this existence and uniqueness theorem. The mixed variational problem (5.5) is then discretized and an approximate solution can be sought using finite element spaces.

There are several reasons why this may not be a desirable approach to solve interface problems. First, because of the requirement that the LBB condition be satisfied, the finite element spaces that are chosen to approximate u and $\boldsymbol{\sigma}$ have to be compatible, and usually they cannot be of equal order. One typical choice for u_h is to use piecewise polynomials of

degree r and then to use piecewise polynomials of degree $r-1$ for $\boldsymbol{\sigma_h}$ (see Braess [8]). In other words, increasing the order of the approximation of ∇u may not be realized. Second, the linear system that results is not guaranteed to be positive definite anymore (Jiang [30]) which makes it difficult to use the typical efficient iterative techniques for the large linear system involved.

The Least Squares Finite Element Method addresses both of these issues. Typically, the least squares method can be applied directly to a BVP. So one possible formulation is to consider minimizing the functional

$$\mathcal{J}(u) = \|\nabla \cdot (p\nabla u) + f\|_0^2. \tag{5.10}$$

However, by taking the first variation of this cost functional, it is seen that u is required to have a weak second derivative. Hence the natural solution space for u in this formulation is $u \in H_0^2(\Omega)$. Unfortunately, this further requires C^1 finite elements to be used in finding u_h if the conforming finite element approach is employed [7].

On the other hand, the least squares finite element method can also be applied to a mixed formulation of the BVP similar to (5.11). Much of the following derivation follows Jiang [30] and Bochev and Gunzburger [7]. By introducing the variable $\boldsymbol{\sigma} = p\nabla u$ we can rewrite the model elliptic interface problem as

$$\begin{aligned} p\nabla u - \boldsymbol{\sigma} &= 0, \\ -\nabla \cdot \boldsymbol{\sigma} &= f. \end{aligned} \tag{5.11}$$

Alternatively, this can be rewritten using operator notation as

$$\mathcal{A}\boldsymbol{\omega} = \boldsymbol{f}, \tag{5.12}$$

where the operator

$$\mathcal{A}\boldsymbol{\omega} = \begin{pmatrix} p\nabla u - \boldsymbol{\sigma} \\ -\nabla \cdot \boldsymbol{\sigma} \end{pmatrix}, \tag{5.13}$$

and the vectors are $\boldsymbol{f}$ and $\boldsymbol{\omega}$ are given by

$$\boldsymbol{f} = \begin{pmatrix} \mathbf{0} \\ f \end{pmatrix} \text{ and } \boldsymbol{\omega} = \begin{pmatrix} u \\ \boldsymbol{\sigma} \end{pmatrix}. \tag{5.14}$$

Then the cost functional that we are interested in minimizing is

$$\mathcal{J}(\boldsymbol{\omega}) = \|p\nabla u - \boldsymbol{\sigma}\|_0^2 + \|-\nabla \cdot \boldsymbol{\sigma} - f\|_0^2 = \|A\boldsymbol{\omega} - \boldsymbol{f}\|_0^2. \tag{5.15}$$

A necessary condition for this quadratic cost functional to attain its minimum is that its first variation is equal to zero [30]. This leads us to the following variational problem: find $\boldsymbol{\omega} \in \mathcal{V} \times \mathcal{W}$ such that

$$(\mathcal{A}\boldsymbol{\omega}, \mathcal{A}\boldsymbol{\nu}) = (\boldsymbol{f}, \mathcal{A}\boldsymbol{\nu}), \text{ for all } \boldsymbol{\nu} \in \mathcal{V} \times \mathcal{W}, \tag{5.16}$$

where the spaces $\mathcal{V}$ and $\mathcal{W}$ are defined by

$$\begin{aligned} \mathcal{V} &= \tilde{H}_0^1(\Omega) = \big\{ u \in L^2(\Omega) \;\big|\; u_{\Omega_s} \in H^1(\Omega_s) \text{ for } s = 1, 2, \\ &\qquad u_{\partial\Omega} = 0, [u]_\Gamma = 0, [p\nabla u \cdot \boldsymbol{n}]_\Gamma = 0 \big\}, \end{aligned} \tag{5.17}$$

$$\mathcal{W} = H^1(\Omega)^n. \tag{5.18}$$

The discretized version of problem (5.16) involves looking for a solution $(u_h, \boldsymbol{\sigma}_h) \in \mathcal{V}_h \times \mathcal{W}_h$ where $\mathcal{V}_h \subset \mathcal{V}$ and $\mathcal{W}_h \subset \mathcal{W}$. In this case, we choose an immersed finite element space to represent $\mathcal{V}_h$ a and a standard finite element space to represent $\mathcal{W}_h$. One of the advantages of the mixed least squares formulation is realized here since $\mathcal{V}_h$ and $\mathcal{W}_h$ do not need to be chosen so that the LBB condition is satisfied.

This leads to a linear system that has a symmetric and positive-definite system of equations [7, 30]. Also, there is no restriction upon the types of finite element spaces that are chosen, so equal order finite elements are permissable for example. If $\Omega \subset \mathbb{R}$, then, when equal order elements are used for both u and σ, the orders for the errors $E_{h,1}(u_h)$ and $E_{h,1}(\boldsymbol{\sigma})$ are optimal. Here we are using optimality in the sense that if the errors given by the interpolation in the finite element spaces $\mathcal{V}_h$ and $\mathcal{W}_h$ are $\|u - I_h u\|_1 \leq Ch^r$ and $\|\sigma - I_h \sigma\|_1 \leq Ch^r$, respectively, then the finite element solutions have the same orders for their errors. For more details about these error estimates the reader is referred to Jiang, [30].

In two dimensions and three dimensions, however, the least squares formulation given above does not lead to optimal solutions. This is due to the fact that the variational problem that arises from the cost functional (5.15) does not lead to a bilinear form that has the appropriate ellipticity [7, 30]. This problem can be solved by considering the system

$$\begin{aligned} p\nabla u - \boldsymbol{\sigma} &= 0 \text{ in } \Omega, \\ -\nabla \cdot \boldsymbol{\sigma} &= f \text{ in } \Omega, \\ \nabla \times \boldsymbol{\sigma} &= 0 \text{ in } \Omega, \\ u &= 0 \text{ on } \partial\Omega, \\ \boldsymbol{n} \times \boldsymbol{\sigma} &= 0 \text{ on } \partial\Omega. \end{aligned} \tag{5.19}$$

This system can be shown to be equivalent to (5.11) by an appropriate use of dummy variables [30]. Then the optimal least squares cost functional is

$$\mathcal{J}(u, \boldsymbol{\sigma}) = \|p\nabla u - \boldsymbol{\sigma}\|_0^2 + \|-\nabla \cdot \boldsymbol{\sigma} - f\|_0^2 + \|\nabla \times \boldsymbol{\sigma}\|_0^2. \tag{5.20}$$

This leads to a variational problem to which the finite element method can be applied by seeking the minimum of $\mathcal{J}$. Again, it is necessary for the first variation of $\mathcal{J}$ to be equal to zero at the point $(u, \boldsymbol{\sigma})$ where it attains its minimum. This formulation restores the optimality of the orders of the errors when equal order elements are used. Again as in the one dimensional case, immersed finite elements can be used for the approximation of u and standard finite elements can be used for the approximation space for $\boldsymbol{\sigma}$ when an interface problem is to be solved.

5.2 The Mixed Least Squares IFE Method in One Dimension

In one dimension the mixed form of the interface problem (5.1) is given by

$$\begin{aligned} p(x)u'(x) - \sigma(x) &= 0, \text{ if } x \in (0,1), \\ -\sigma'(x) &= f(x), \text{ if } x \in (0,1), \\ u(0) = u(1) &= 0, \end{aligned} \tag{5.21}$$

where the coefficient p is defined as

$$p(x) = \begin{cases} p_1, & \text{if } x < \alpha, \\ p_2, & \text{if } x \geq \alpha, \end{cases} \tag{5.22}$$

and the domain is the typical $\Omega = (0,1) = \Omega_1 \cup \Omega_2$. Due to the definition of p, the interface occurs at $x = \alpha$ so $\Omega_1 = (0,\alpha)$ and $\Omega_2 = (\alpha,1)$. Then we let the space $\mathcal{V}$ be defined by

$$\begin{aligned} \mathcal{V} = \tilde{H}_0^1(0,1) = \big\{ u \in H^0(\Omega) \;\big|\; u|_{\Omega_1} \in H^1(0,\alpha), u|_{\Omega_2} \in H^1(\alpha,1), \\ u(0) = u(1) = 0, [u]_\alpha = 0, [pu']_\alpha = 0 \big\}, \end{aligned} \tag{5.23}$$

and the space $\mathcal{W}$ be defined by

$$\mathcal{W} = H^1(0,1). \tag{5.24}$$

The mixed least squares cost functional $\mathcal{J} : \mathcal{V} \times \mathcal{W} \to \mathbb{R}$, is defined by

$$\mathcal{J}(u,\sigma) = \|pu' - \sigma\|_0^2 + \|-\sigma' - f\|_0^2. \tag{5.25}$$

Since the first variation of the cost functional needs to be zero in order for it to attain its minimum we obtain the following variational problem: find $(u,\sigma) \in \mathcal{V} \times \mathcal{W}$ such that

$$\begin{aligned} &\int_0^1 (pu' - \sigma)p\nu' \, dx = 0 \text{ for all } \nu \in \mathcal{V}, \\ &\int_0^1 \big\{ (pu' - \sigma)(-\xi) + (-\sigma' - f)(-\xi') \big\} dx = 0 \text{ for all } \xi \in \mathcal{W}. \end{aligned} \tag{5.26}$$

Then we choose our finite element space $\mathcal{V}_h \subset \mathcal{V}$ to be an IFE space that satisfies the interface conditions required by the solution space $\mathcal{V}$ in (5.23). On the other hand, the finite dimensional subspace $\mathcal{W}_h \subset \mathcal{W}$ can be chosen as a standard finite element space. Here we use the notation $\bar{\phi}_k$ to represent the nodal basis functions of $\mathcal{V}_h$ and ψ_k to represent the basis

functions of the space $\mathcal{W}_h$. Then we let $u \approx u_h \in \mathcal{V}_h$ and $\sigma \approx \sigma_h \in \mathcal{W}_h$ where u_h and σ_h are given by

$$u_h(x) = \sum_{j=1}^{N} U_j \bar{\phi}_j(x), \tag{5.27}$$

$$\sigma_h(x) = \sum_{k=1}^{M} \sigma_k \psi_k(x). \tag{5.28}$$

Using the spaces $\mathcal{V}_h$ and $\mathcal{W}_h$ to discretize (5.26), we arrive at the following finite element approximation problem: find $\vec{U} = (U_1, \ldots, U_N)^T$ and $\vec{S} = (S_1, \ldots, S_M)^T$ so that

$$\begin{aligned} &\int_0^1 (pu_h' - \sigma_h) p\bar{\phi}_i' \, dx = 0 \text{ for } i = 1, \ldots, N, \\ &\int_0^1 \{(pu_h' - \sigma_h)(-\psi_i) + (-\sigma_h' - f)(-\psi_i')\} \, dx = 0 \text{ for } i = 1, \ldots, M. \end{aligned} \tag{5.29}$$

Alternatively this system of linear equations can be written in matrix form as the following system

$$\left(\begin{array}{c|c} K_{1,1} & K_{1,2} \\ \hline K_{2,1} & K_{2,2} \end{array} \right) \left(\begin{array}{c} \vec{U} \\ \hline \vec{S} \end{array} \right) = \left(\begin{array}{c} \mathbf{0} \\ \hline \boldsymbol{F} \end{array} \right), \tag{5.30}$$

where the $K_{i,j}$ blocks are defined by

$$K_{1,1} = \left(p\bar{\phi}_i', p\bar{\phi}_j' \right)_{\substack{i=1,\ldots,N \\ j=1,\ldots,N}}, \tag{5.31}$$

$$K_{1,2} = \left(p\bar{\phi}_i, -\psi_k \right)_{\substack{i=1,\ldots,N \\ k=1,\ldots,M}}, \tag{5.32}$$

$$K_{2,1} = K_{1,2}^T, \tag{5.33}$$

$$K_{2,2} = \left(\psi_i, \psi_k \right)_{\substack{i=1,\ldots,M \\ k=1,\ldots,M}} + \left(\psi_i', \psi_k' \right)_{\substack{i=1,\ldots,M \\ k=1,\ldots,M}}, \tag{5.34}$$

and the vector $\boldsymbol{F}$ is defined by

$$\boldsymbol{F} = \left(-\psi_i', f \right)_{i=1,\ldots,M}. \tag{5.35}$$

All of these vectors and matrices can be assembled using standard finite element techniques. As expected, the coefficient matrix K is symmetric, positive-definite and sparse. This method is then called the Mixed Least Squares Immersed Finite Element Method (or MLSIFEM)

since it involves a mixed formulation of the BVP, a least squares cost functional and finally IFE spaces are used in the finite element approximations.

In the rest of this section we present the numerical results when we solve the system (5.30) for $\vec{U}$ and $\vec{S}$ so that we obtain the solutions u_h and σ_h to the discretized mixed variational problem (5.29). The primary reason for the numerical investigation of the MLSIFEM that we present here is to explore the accuracy of the IFE spaces that can potentially be used to solve **Problem (I-1)** (see Chapter 6). Since solving **Problem (I-1)** involves a moving interface, it does not make sense to consider the use of the IFE space $\hat{\mathcal{S}}_{h,2}(0,1)$ that requires a slight modification of the partition with each iteration. Additionally, the right hand side function that is used in Chapter 6 is $f(x) = e^x$. Recall from Chapter 3 that this right hand side function corresponded to the exact solution u_1. Since the IFE space $\bar{\mathcal{S}}_{h,2}(0,1)$ has better results in approximating the function u_1 than the IFE space $\tilde{\mathcal{S}}_{h,2}(0,1)$, then this is the quadratic IFE space that we use in the numerical experiments here. The linear IFE space $\tilde{\mathcal{S}}_{h,1}(0,1)$ is also used since it is the unique linear IFE space. So either the linear IFE space $\tilde{\mathcal{S}}_{h,1}(0,1)$ or the quadratic IFE space $\bar{\mathcal{S}}_{h,2}(0,1)$ is chosen as the finite dimensional space $\mathcal{V}_h$. Standard finite element spaces based upon C^0 piecewise polynomials are used as the approximation space $\mathcal{W}_h$.

As we mention above, we restrict our attention to the case when right hand side function, f, is defined by $f(x) = e^x$ as this is also the function used as the right hand side function in the one dimensional inverse problems that we present in the next chapter. Since the right hand side is chosen in this fashion, the exact solution of the BVP is again the testing function $u_1(x)$ which is defined in (3.15).

Much of the same notation and methodology are used in the numerical results given here as those used in Chapter 4 when the Galerkin finite element method is used to approximate the solution of the boundary value problem. Again, 4 gaussian nodes are used on each element for the purposes of numerical integration and 6 interpolation points are chosen on each element for the purpose of approximating the L^∞ error. Errors are again calculated in the L^∞, L^2 and H^1 norms. MLSIFEM errors are referred to as $E_{h,s}(u_h)$ and $E_{h,s}(\sigma_h)$ (for $s = \infty, 0, 1$) or more simply as E_h and $E_{h,s}$ if it is clear from the context which function the error refers to. Numerical experiments are carried out with a variety of mesh sizes ranging from $h = 1/8$ to $h = 1/512$. Only a selection of the data is presented in the tables to reduce the page space. Specifically, data is included for the partition sizes of $h = 2^{-n}$ for $n = 3, \ldots, 9$ so that the behavior of the errors when the partition size is halved can be easily observed. The full data is also presented in figures where appropriate and the orders of the errors are calculated using the full data in the figures.

5.2.1 MLSIFEM: $u_h \in \tilde{\mathcal{S}}_{h,1}$ and $\sigma_h \in \mathcal{S}_{h,1}$

In this section, numerical results of the MLSIFEM are presented when the linear IFE space $\tilde{\mathcal{S}}_{h,1}(0,1)$ is chosen as the approximation space $\mathcal{V}_h$ and the standard linear lagrangian finite element space $\mathcal{S}_{h,1}(0,1)$ is chosen as $\mathcal{W}_h$. Data in Tables 5.1 and 5.2 show the MLSIFEM errors for approximating u_1 and $\sigma = pu_1'$ respectively when the coefficient function p has a ratio of $\beta = p_2/p_1 = 10$. Figures 5.1 and 5.2 show the plots of the $\log h$ vs. $\log E_h$ data

h	$\|u_1 - u_h\|_\infty$	$\|u_1 - u_h\|_0$	$\|u_1 - u_h\|_1$
1/8	1.679×10^{-3}	6.790×10^{-4}	2.425×10^{-2}
1/16	4.723×10^{-4}	2.058×10^{-4}	1.262×10^{-2}
1/32	1.159×10^{-4}	4.621×10^{-5}	6.273×10^{-3}
1/64	2.894×10^{-5}	1.072×10^{-5}	3.112×10^{-3}
1/128	7.293×10^{-6}	2.765×10^{-6}	1.546×10^{-3}
1/256	1.931×10^{-6}	6.669×10^{-7}	7.752×10^{-4}
1/512	5.883×10^{-7}	2.279×10^{-7}	3.875×10^{-4}

Table 5.1: MLSIFEM error, $E_h = \|u_1 - u_h\|_s,\ s = \infty, 0, 1$ where $u_h \in \tilde{\mathcal{S}}_{h,1}$, $\sigma_h \in \mathcal{S}_{h,1}$ and $\beta = 10$.

h	$\|\sigma - \sigma_h\|_\infty$	$\|\sigma - \sigma_h\|_0$	$\|\sigma - \sigma_h\|_1$
1/8	9.711×10^{-3}	6.264×10^{-3}	6.479×10^{-2}
1/16	1.421×10^{-3}	5.961×10^{-4}	3.225×10^{-2}
1/32	4.594×10^{-4}	2.469×10^{-4}	1.613×10^{-2}
1/64	1.467×10^{-4}	9.284×10^{-5}	8.062×10^{-3}
1/128	3.243×10^{-5}	1.904×10^{-5}	4.031×10^{-3}
1/256	9.559×10^{-6}	6.187×10^{-6}	2.015×10^{-3}
1/512	1.015×10^{-6}	2.954×10^{-7}	1.008×10^{-3}

Table 5.2: MLSIFEM error, $E_h = \|\sigma - \sigma_h\|_s,\ s = \infty, 0, 1$ where $u_h \in \tilde{\mathcal{S}}_{h,1}$, $\sigma_h \in \mathcal{S}_{h,1}$ and $\beta = 10$.

using more data points than is shown in the tables. Using the data from the figures, it is seen that the order of the errors from the MLSIFEM for approximating the function u_1 are

$$\begin{aligned} E_{h,\infty} = \|u_1 - u_h\|_\infty &\approx 0.101\ h^{1.955}, \\ E_{h,0} = \|u_1 - u_h\|_0 &\approx 0.044\ h^{1.982}, \\ E_{h,1} = \|u_1 - u_h\|_1 &\approx 0.194\ h^{0.996}, \end{aligned} \tag{5.36}$$

while the orders for the errors for approximating the flux $\sigma = pu_1'$ are

$$\begin{aligned} E_{h,\infty} = \|\sigma - \sigma_h\|_\infty &\approx 0.603\ h^{2.047}, \\ E_{h,0} = \|\sigma - \sigma_h\|_0 &\approx 0.404\ h^{2.098}, \\ E_{h,1} = \|\sigma - \sigma_h\|_1 &\approx 0.517\ h^{1.000}. \end{aligned} \tag{5.37}$$

Both the approximations of u_1 and σ perform optimally in the sense that $E_{h,\infty}$ and $E_{h,0}$ have orders of $\mathcal{O}(h^2)$ and $E_{h,1}$ have orders of $\mathcal{O}(h)$ for both u_1 and σ. This definitely reinforces the idea that equal order elements do not have any problems using the MLSIFEM. Indeed, σ_h provides an approximation to σ with a higher order than that generated from a usual FE solution using a linear space by differentiation.

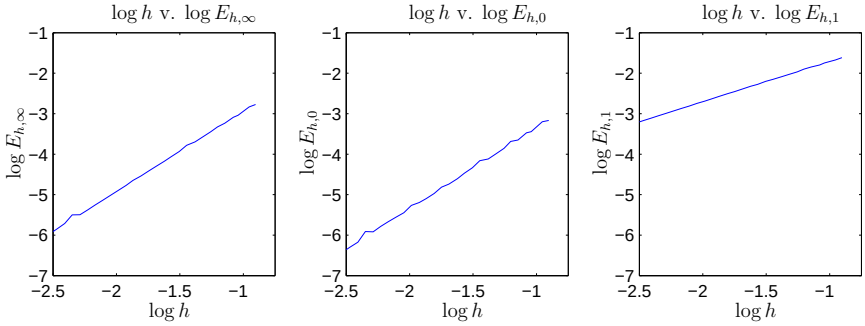


Figure 5.1: MLSIFEM error, $E_h = \|u_1 - u_h\|_s,\ s = \infty, 0, 1$ where $u_h \in \tilde{\mathcal{S}}_{h,1}$, $\sigma_h \in \mathcal{S}_{h,1}$ and $\beta = 10$.

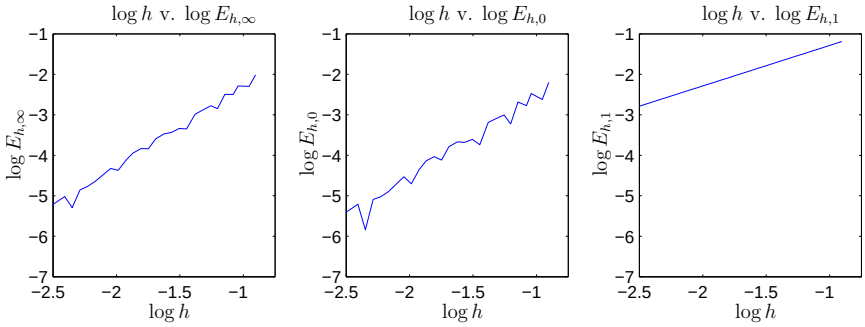


Figure 5.2: MLSIFEM error, $E_h = \|\sigma - \sigma_h\|_s,\ s = \infty, 0, 1$ where $u_h \in \tilde{\mathcal{S}}_{h,1}$, $\sigma_h \in \mathcal{S}_{h,1}$ and $\beta = 10$.

5.2.2 MLSIFEM: $u_h \in \bar{\mathcal{S}}_{h,2}$ and $\sigma_h \in \mathcal{S}_{h,1}$

Now we repeat the experiment from the previous section except that this time the quadratic IFE space $\bar{\mathcal{S}}_{h,2}(0,1)$ is used as the approximation space $\mathcal{V}_h$. The standard linear lagrangian fi-

nite element space $\mathcal{S}_{h,1}(0,1)$ is again used as $\mathcal{W}_h$. The errors in the MLSIFEM approximation errors of u_1 and $\sigma = pu_1'$ are shown in Tables 5.3 and 5.4 respectively when $\beta = 10$. Again

h	$\|u_1 - u_h\|_\infty$	$\|u_1 - u_h\|_0$	$\|u_1 - u_h\|_1$
1/8	8.400×10^{-5}	4.473×10^{-5}	4.364×10^{-4}
1/16	1.617×10^{-5}	1.070×10^{-5}	1.086×10^{-4}
1/32	3.819×10^{-6}	2.690×10^{-6}	2.694×10^{-5}
1/64	9.539×10^{-7}	6.768×10^{-7}	6.718×10^{-6}
1/128	2.410×10^{-7}	1.701×10^{-7}	1.684×10^{-6}
1/256	5.878×10^{-8}	4.242×10^{-8}	4.207×10^{-7}
1/512	1.464×10^{-8}	1.061×10^{-8}	1.052×10^{-7}

Table 5.3: MLSIFEM error, $E_h = \|u_1 - u_h\|_s,\ s = \infty, 0, 1$ where $u_h \in \bar{\mathcal{S}}_{h,2}$, $\sigma_h \in \mathcal{S}_{h,1}$ and $\beta = 10$.

h	$\|\sigma - \sigma_h\|_\infty$	$\|\sigma - \sigma_h\|_0$	$\|\sigma - \sigma_h\|_1$
1/8	3.077×10^{-3}	1.318×10^{-3}	6.446×10^{-2}
1/16	8.133×10^{-4}	3.338×10^{-4}	3.224×10^{-2}
1/32	2.075×10^{-4}	8.335×10^{-5}	1.612×10^{-2}
1/64	5.232×10^{-5}	2.079×10^{-5}	8.062×10^{-3}
1/128	1.312×10^{-5}	5.186×10^{-6}	4.031×10^{-3}
1/256	3.293×10^{-6}	1.298×10^{-6}	2.015×10^{-3}
1/512	8.244×10^{-7}	3.245×10^{-7}	1.008×10^{-3}

Table 5.4: MLSIFEM error, $E_h = \|\sigma - \sigma_h\|_s,\ s = \infty, 0, 1$ where $u_h \in \bar{\mathcal{S}}_{h,2}$, $\sigma_h \in \mathcal{S}_{h,1}$ and $\beta = 10$.

plots of the $\log h$ vs. $\log E_h$ data are shown in Figures 5.4 and 5.4 respectively. The trends of the $\log - \log$ plots are linear showing the stability of the MLSIFEM. This is somewhat significant since it now seems that the MLSIFEM works without any problems whether the approximation spaces have the same order (as in the previous section) or not. The slopes of the lines in the figures show that the orders of the errors for approximating u_1 using $u_h \in \bar{S}_h$ obey the relationships

$$\begin{aligned} E_{h,\infty} = \|u_1 - u_h\|_\infty &\approx 0.006\ h^{2.069}, \\ E_{h,0} = \|u_1 - u_h\|_0 &\approx 0.003\ h^{2.005}, \\ E_{h,1} = \|u_1 - u_h\|_1 &\approx 0.028\ h^{2.000}, \end{aligned} \tag{5.38}$$

while the orders for the errors of approximating the flux when using standard linear basis functions are

$$\begin{aligned} E_{h,\infty} = \|\sigma - \sigma_h\|_\infty &\approx 0.196\ h^{1.982}, \\ E_{h,0} = \|\sigma - \sigma_h\|_0 &\approx 0.084\ h^{1.998}, \\ E_{h,1} = \|\sigma - \sigma_h\|_1 &\approx 0.516\ h^{1.000}. \end{aligned} \tag{5.39}$$

As it turns out, the orders for $E_h(\sigma)$ in this case are optimal while the orders for $E_h(u_1)$ are not optimal. Although the MLSIFEM seems stable by looking at the graphs since the $\log - \log$ trends are nice and linear, the orders for $E_{h,\infty}(u_1)$ and $E_{h,0}(u_1)$ are $\mathcal{O}(h^2)$ which are a full order below what should be optimal. Actually this is not surprising since orders of the errors of u and σ are related. Jiang [30] indicates that the H^1 errors are of order $\mathcal{O}(h^q)$ where q is the smaller of the H^1 interpolation orders for the FE spaces used to approximate u and σ and hence

$$E_{h,1}(u_1) + E_{h,1}(\sigma) = \mathcal{O}(h^q). \tag{5.40}$$

Since $u_h \in \mathcal{S}_{h,1}(0,1)$ then $q = 1$ in this case. Then using the Aubin-Nitsche trick [8, 30], the L^2 errors for u_h and σ_h should be of order $\mathcal{O}(h^{q+1})$. The result of all of this is that the error $E_{h,0}(u_1)$ does perform as well as we expect when using a quadratic FE space and the Galerkin method. Some advantage is still gained, however, since σ_h gives a C^0 approximation of the flux, σ instead of the usual piecewise continuous approximation that comes with the Galerkin method. Using unequal order elements cannot be recommended

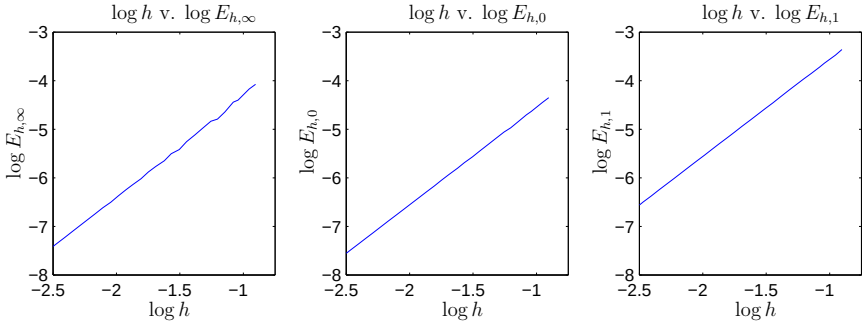


Figure 5.3: MLSIFEM error, $E_h = \|u_1 - u_h\|_s$, $s = \infty, 0, 1$ where $u_h \in \bar{\mathcal{S}}_{h,2}$, $\sigma_h \in \mathcal{S}_{h,1}$ and $\beta = 10$.

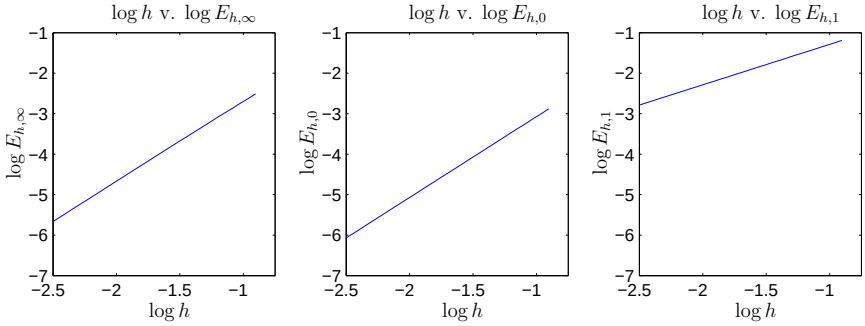


Figure 5.4: MLSIFEM error, $E_h = \|\sigma - \sigma_h\|_s$, $s = \infty, 0, 1$ where $u_h \in \bar{\mathcal{S}}_{h,2}$, $\sigma_h \in \mathcal{S}_{h,1}$ and $\beta = 10$.

when using MLSIFEM since this results in suboptimal results for the approximation of u_1. In fact, other than the increased order for $E_{h,1}(u_1)$, the quadratic/linear combination of $\mathcal{V}_h$ and $\mathcal{W}_h$ does not perform any better than the linear/linear case. Even further, if the quadratic IFE space $\bar{\mathcal{S}}_{h,2}(0,1)$ is going to be used for the approximation of u_1 then using the Galerkin finite element method as in Chapter 4 produces better results than the quadratic/linear MLSIFEM, and at less cost.

For example, if there are N elements in the partition, then the galerkin method with the quadratic IFE space $\bar{S}_h$ is solving a $(2N-1)\times(2N-1)$ system. On the other hand, the MLSIFEM with a quadratic/linear $\mathcal{V}_h/\mathcal{W}_h$ combination is required to solve a $(3N)\times(3N)$ system which is a little more than twice the size. For the extra computation that is required, the only benefit that is gained is a C^0 approximation of σ as compared to a piecewise continuous approximation of σ when using the Galerkin finite element method.

5.2.3 MLSIFEM: $u_h \in \bar{\mathcal{S}}_{h,2}$ and $\sigma_h \in \mathcal{S}_{h,2}$

Finally, we repeat the previous experiments using a quadratic/quadratic combination for $\mathcal{V}_h$ and $\mathcal{W}_h$ where the quadratic IFE space $\bar{\mathcal{S}}_{h,2}(0,1)$ is used for the former space and $\mathcal{S}_{h,2}(0,1)$ is used for the latter. The data from the MLSIFEM approximations for u_1 when $\beta = 10$ is shown in Table 5.5 while the corresponding data for σ is shown in Table 5.6. Plots of

h	$\|u_1-u_h\|_\infty$	$\|u_1-u_h\|_0$	$\|u_1-u_h\|_1$
1/8	2.457×10^{-5}	8.022×10^{-6}	4.041×10^{-4}
1/16	3.946×10^{-6}	1.101×10^{-6}	1.023×10^{-4}
1/32	5.674×10^{-7}	1.478×10^{-7}	2.521×10^{-5}
1/64	6.032×10^{-8}	1.641×10^{-8}	6.253×10^{-6}
1/128	8.806×10^{-9}	2.281×10^{-9}	1.562×10^{-6}
1/256	6.864×10^{-10}	2.358×10^{-10}	3.909×10^{-7}
1/512	1.156×10^{-10}	3.259×10^{-11}	9.769×10^{-8}

Table 5.5: MLSIFEM error, $E_h = \|u_1-u_h\|_s,\ s=\infty,0,1$ where $u_h \in \bar{\mathcal{S}}_{h,2}$, $\sigma_h \in \mathcal{S}_{h,2}$ and $\beta = 10$.

the $\log h$ vs. $\log E_h$ data for a larger set of partition sizes h are shown in Figures 5.5 and 5.6 for the approximations of u_1 and σ respectively. Again as in both the linear/linear and quadratic/linear cases, the $\log-\log$ data show nice linear trends. Calculating the slopes of these trends shows that the errors for approximating u_1 have the approximate order relationships of

$$\begin{aligned} E_{h,\infty} = \|u_1-u_h\|_\infty &\approx 0.015\ h^{3.000}, \\ E_{h,0} = \|u_1-u_h\|_0 &\approx 0.005\ h^{3.008}, \\ E_{h,1} = \|u_1-u_h\|_1 &\approx 0.025\ h^{1.999}, \end{aligned} \tag{5.41}$$

h	$\|\sigma - \sigma_h\|_\infty$	$\|\sigma - \sigma_h\|_0$	$\|\sigma - \sigma_h\|_1$
1/8	4.931×10^{-5}	2.233×10^{-5}	1.040×10^{-3}
1/16	9.784×10^{-6}	5.248×10^{-6}	2.602×10^{-4}
1/32	1.440×10^{-6}	8.455×10^{-7}	6.505×10^{-5}
1/64	1.440×10^{-7}	7.296×10^{-8}	1.626×10^{-5}
1/128	2.210×10^{-8}	1.275×10^{-8}	4.066×10^{-6}
1/256	1.369×10^{-9}	6.175×10^{-10}	1.016×10^{-6}
1/512	4.393×10^{-10}	2.555×10^{-10}	2.541×10^{-7}

Table 5.6: MLSIFEM error, $E_h = \|\sigma - \sigma_h\|_s,\ s = \infty, 0, 1$ where $u_h \in \bar{\mathcal{S}}_{h,2}$, $\sigma_h \in \mathcal{S}_{h,2}$ and $\beta = 10$.

while the errors for approximating σ using standard quadratic FE space $\mathcal{S}_{h,2}(0,1)$ have orders which obey the equations

$$\begin{aligned} E_{h,\infty} = \|\sigma - \sigma_h\|_\infty &\approx 0.035\ h^{2.985}, \\ E_{h,0} = \|\sigma - \sigma_h\|_0 &\approx 0.020\ h^{3.006}, \\ E_{h,1} = \|\sigma - \sigma_h\|_1 &\approx 0.067\ h^{2.000}. \end{aligned} \tag{5.42}$$

As in the linear/linear case, the orders shown here are optimal in all norms for both u_1 and σ. Consequently, equal order elements perform at least well as the standard Galerkin finite element method in approximating the function u_1. Additionally, a continuous approximation to $\sigma = pu_1'$ is generated whose errors have one full order better approximation capability than the corresponding errors from the Galerkin finite element method.

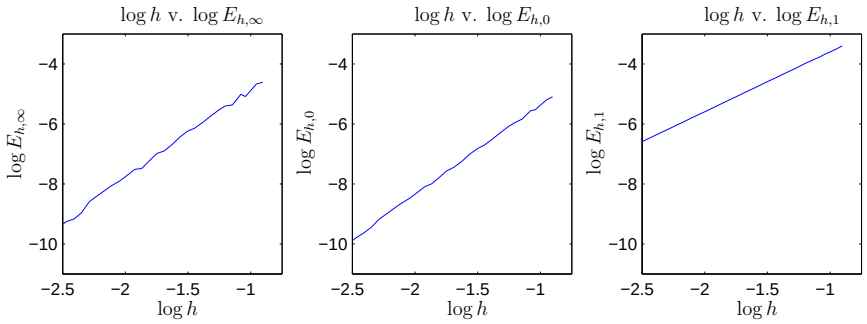


Figure 5.5: MLSIFEM error, $E_h = \|u_1 - u_h\|_s,\ s = \infty, 0, 1$ where $u_h \in \bar{\mathcal{S}}_{h,2}$, $\sigma_h \in \mathcal{S}_{h,2}$ and $\beta = 10$.

Overall, the MLSIFEM gives error results as expected based upon the interpolation results in Chapter 3 for the quadratic IFE space $\bar{\mathcal{S}}_{h,2}(0,1)$. We see, however, that computationally, using equal order IFE spaces for u_h and σ_h gives the best errors for the amount of effort that

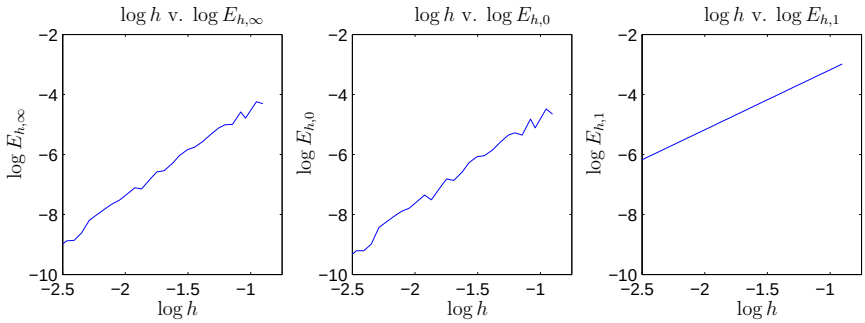


Figure 5.6: MLSIFEM error, $E_h = \|\sigma - \sigma_h\|_s,\ s = \infty, 0, 1$ where $u_h \in \bar{\mathcal{S}}_{h,2}$, $\sigma_h \in \mathcal{S}_{h,2}$ and $\beta = 10$.

is expended during the computation. For example, a quadratic/linear FE space pairing for u_h and σ_h does only slightly better than a linear/linear pairing of the spaces. Unfortunately for approximately the same results, the quadratic/linear pairing must solve a system with about one and a half times the unknowns of the linear/linear pairing.

Chapter 6

Inverse Elliptic Interface Problems

In this chapter, we use the IFE spaces developed in Chapter 2 to solve the interface inverse problem **(I)** numerically. Two important features of the IFE spaces make their use in solving problem **(I)** desirable: (i) they can use one uniform mesh in the iteration procedure for the solution of interface problems while the interface parameters are adjusted, and (ii) the dependence of the cost functional on the interface parameters is straightforward. The primary focus of this chapter is to employ the IFE spaces as tools to implement either an output least squares method or a mixed least squares method for the model interface problem.

6.1 Background

Recall our model elliptic interface problem: find u such that

$$\begin{aligned} -\nabla \cdot (p(\boldsymbol{x})\nabla u(\boldsymbol{x})) &= f(\boldsymbol{x}), \quad && \boldsymbol{x} \in \Omega, \\ u(\boldsymbol{x})\big|_{\partial\Omega} &= 0, && \boldsymbol{x} \in \partial\Omega, \end{aligned} \tag{6.1}$$

where p is a piecewise constant function as defined in (1.2), and recall **Problem (I)** which is an inverse problem related to the solution of (6.1):

Problem (I): Determine Γ so that

$$p(\boldsymbol{x}_j)\nabla u(\boldsymbol{x}_j) \cdot \boldsymbol{n} = g(\boldsymbol{x}_j), \text{ for } \boldsymbol{x}_j, \ j = 1, \ldots, m \in \partial\Omega, \tag{6.2}$$

where u is the solution of the elliptic interface problem (6.1).

In this statement of **Problem (I)**, $\boldsymbol{x}_j$ are a set of points that are on the boundary of the domain Ω where the flux $p\nabla u \cdot \boldsymbol{n}$ is required to match the values of some prescribed function g which is defined over the boundary. Alternatively, **Problem (I)** can be stated so that the flux must match the function g over the entire boundary. Then (6.2) becomes

$$p(\boldsymbol{x})\nabla u(\boldsymbol{x}) \cdot \boldsymbol{n} = g(\boldsymbol{x}), \text{ for } \boldsymbol{x} \in \partial\Omega. \tag{6.3}$$

We consider two methods for solving the inverse problem in this chapter: the Output Least Squares (OLS) method and the Least Squares Method with Mixed Equation Error (LSMEE). We assume that the interface is represented by a parameter or parameters so that determining the parameters allows us to determine the interface. For example, one way to represent the interface in two dimensions when $\Omega = (0,1) \times (0,1)$ is to use a parameter such as

$$\vec{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_q)^T, \tag{6.4}$$

so that Γ is represented by the curve

$$\boldsymbol{l}(t) = (x(t), y(t)) = \left(t, \sum_{i=1}^{q} \alpha_i \phi_i(t) \right), \tag{6.5}$$

where it is assumed that the parameter $\vec{\alpha}$ and the basis functions $\{\phi_i\},\ i = 1, \ldots, n$ satisfy constraints such that $0 \leq y(t) \leq 1$ for $t \in (0,1)$. Then the parameterization is a curve on the unit square.

Both least squares techniques revolve around formulating an appropriate cost functional and then finding the parameter, $\vec{\alpha}$, that minimizes the cost functional. The minimum of the chosen cost functional should correspond to the solution of the inverse problem or as close as possible (in a least squares sense). If **Problem (I)** is specified using boundary conditions 6.2 then the output least squares (OLS) cost functional can be defined by

$$\mathcal{J}_o(\vec{\alpha}) = \sum_{j=1}^{m} |p(\boldsymbol{x}_j; \vec{\alpha})\nabla u(\boldsymbol{x}_j; \vec{\alpha}) \cdot \boldsymbol{n} - g(\boldsymbol{x}_j)|^2, \tag{6.6}$$

where $\boldsymbol{x}_j,\ j = 1, \ldots, m\ \in \partial\Omega$ and $u(\boldsymbol{x}; \vec{\alpha})$ solves the interface problem (6.1) with the given parameter $\vec{\alpha}$. Alternatively, if the flux is to be matched at more than just a discrete set of points on the boundary according to (6.3), then the OLS cost function may be stated as

$$\tilde{\mathcal{J}}_o(\vec{\alpha}) = \int_{\partial\Omega} |p(\boldsymbol{x}, \vec{\alpha})\nabla u(\boldsymbol{x}, \vec{\alpha}) \cdot \boldsymbol{n} - g(\boldsymbol{x})|^2 ds. \tag{6.7}$$

Clearly if the value of the cost functional, $\mathcal{J}_o$, is zero then the inverse problem is solved.

The output least squares cost functional, $\mathcal{J}_o$, is typically difficult to deal with for a couple of reasons. First, the dependence of the cost function upon $u(\boldsymbol{x}; \vec{\alpha})$ requires that the elliptic

interface problem (6.1) must be solved in order to gain any information about $u(\boldsymbol{x};\vec{\alpha})$. Also, because of the way that $u(\boldsymbol{x};\vec{\alpha})$ is defined, it is often difficult, if not impossible, to take the derivative of this function with respect to the parameters that represent the interface. This prevents the use of more efficient iteration methods depending on gradients and Hessians to minimize the cost functional. However, the output least squares method is a valuable tool since it is straightforward formulation. In simple cases, a direct minimization strategy such as the golden section search or the multivariable DIRECT search algorithm of Jones, et. al. [27, 32, 49] can be utilized to determine the parameters where the cost functional is minimized without the need for the derivative to be calculated. However, this still requires that the interface problem be solved in the iteration, and either the Galerkin FEM or the least squares FEM for example are used to avoid re-partition in the iteration.

The least squares method with mixed equation error addresses the difficulties arising in implementing the output least squares method. Recall the mixed formulation of the elliptic interface problem that is used in Chapter 5.

$$\begin{aligned} p\nabla u - \boldsymbol{\sigma} &= 0, \text{ in } \Omega, \\ -\nabla\cdot\boldsymbol{\sigma} &= f, \text{ in } \Omega, \\ \nabla\times\boldsymbol{\sigma} &= 0, \text{ in } \Omega, \\ u &= 0, \text{ in } \partial\Omega, \\ \boldsymbol{n}\times\boldsymbol{\sigma} &= 0, \text{ in } \partial\Omega. \end{aligned} \tag{6.8}$$

Then the mixed equation error can be incorporated into a cost functional for **Problem (I)** as follows when boundary conditions are prescribed using (6.2):

$$\begin{aligned} \mathcal{J}_m(u,\boldsymbol{\sigma},\vec{\alpha}) = \; & \|p\nabla u - \boldsymbol{\sigma}\|_0^2 + \|\nabla\cdot\boldsymbol{\sigma} + f\|_0^2 + \|\nabla\times\boldsymbol{\sigma}\|_0^2 \\ & + \sum_{j=1}^{m} |p(\boldsymbol{x}_j;\vec{\alpha})\nabla u(\boldsymbol{x}_j)\cdot\boldsymbol{n} - g(\boldsymbol{x}_j)|^2 \, , \end{aligned} \tag{6.9}$$

where $\boldsymbol{x}_j \, , j = 1,\ldots,m \; \in \partial\Omega$. This is referred to as the Least Square with Mixed Equation Error (LSMEE) cost functional. Alternatively, the LSMEE cost functional can be written in terms of $\boldsymbol{\sigma}$ as

$$\begin{aligned} \mathcal{J}_m(u,\boldsymbol{\sigma},\vec{\alpha}) = \; & \|p\nabla u - \boldsymbol{\sigma}\|_0^2 + \|\nabla\cdot\boldsymbol{\sigma} + f\|_0^2 + \|\nabla\times\boldsymbol{\sigma}\|_0^2 \\ & + \sum_{j=1}^{m} |\boldsymbol{\sigma}(\boldsymbol{x}_j)\cdot\boldsymbol{n} - g(\boldsymbol{x}_j)|^2. \end{aligned} \tag{6.10}$$

Alternative forms of (6.9) and (6.10) based upon the boundary criteria (6.3) are easy to form by replacing the appropriate discrete boundary terms with integral terms. It should be noted that because of the $\|p\nabla u - \boldsymbol{\sigma}\|_0^2$ term that (6.9) and (6.10) have the same minima since $p\nabla u$ and $\boldsymbol{\sigma}$ are as close as possible in a least squares sense. This latter formulation (6.10) is the focus of the LSMEE methods discussed in this chapter.

Notice that the first three terms in the cost functional $\mathcal{J}_m$ are essentially the cost function terms used in the least squares FE formulation for the elliptic interface problem in Chapter 5. So if the first three terms of the cost functional are zero, then $(u, \boldsymbol{\sigma})$ satisfies (6.8) which is equivalent to the constraint that is imposed on the output least squares cost functional. The last term is just the cost functional from the output least squares method except that u and $\boldsymbol{\sigma}$ are now considered to be independent of the interface parameter $\vec{\alpha}$. Now the minimization of the cost functional is no longer a problem of constrained optimization as is the case with the output least squares method. Additionally, since u is now a variable of the cost functional, it becomes easier to find the gradient of the cost function so that gradient search techniques can be used to find the minimizer of the cost functional. The method for calculating the gradient of the cost functional uses methods similar to those used by Lin and others in [21, 43, 45]. As it turns out, being able to solve the elliptic interface problem using the MLSIFEM plays an important role in the calculation of the gradient of $\mathcal{J}_m(u, \boldsymbol{\sigma}, \vec{\alpha})$.

6.2 Inverse Problem in One Dimension

In one dimension, the elliptic interface problem (6.1) becomes

$$\begin{aligned} -\big(p(x;\alpha)u'(x,\alpha)\big)' &= f(x), \quad \text{for } x \in \Omega = (0,1), \\ u(0;\alpha) = u(1;\alpha) &= 0, \end{aligned} \tag{6.11}$$

where the coefficient p is defined by

$$p(x;\alpha) = \begin{cases} p_1, & \text{for } x < \alpha, \\ p_2, & \text{for } x \geq \alpha, \end{cases} \tag{6.12}$$

with p_1 and p_2 being positive constants. Recall that if u satisfies (6.11) then u also satisfies the interface conditions

$$u \in C(0,1) \text{ and } pu' \in C(0,1), \tag{6.13}$$

which implies that u and pu' satisfy the interface jump conditions

$$\begin{aligned} \big[u\big]_{x=\alpha} &= 0, \\ \big[pu'\big]_{x=\alpha} &= 0. \end{aligned} \tag{6.14}$$

The related inverse **Problem (I)** has the following form:

Problem (I-1): Find α such that

$$p(0;\alpha)u'(0;\alpha) = g_0 \text{ and } p(1;\alpha)u'(1;\alpha) = g_1, \tag{6.15}$$

where $u(x;\alpha)$ satisfies the elliptic interface problem (6.11) with the coefficient function $p(x;\alpha)$.

We have the following theorem about the non-uniqueness of a solution to **Problem (I-1)**.

Theorem 6.2.1. *Let u_1 be a $C^2(0,1)$ solution of the following initial value problem (IVP):*

$$\begin{aligned} -\big(p_1 u'(x)\big)' &= f(x),\ x \in (0,1), \\ u(0) &= 0, \\ p_1 u'(0) &= g_0, \end{aligned} \tag{6.16}$$

and let u_2 be a $C^2(0,1)$ solution of the following IVP:

$$\begin{aligned} -\big(p_2 u'(x)\big)' &= f(x),\ x \in (0,1), \\ u(1) &= 0, \\ p_2 u'(1) &= g_1. \end{aligned} \tag{6.17}$$

*In addition, let α_1 be a solution of **Problem (I-1)** such that the solution of (6.11) determined by α_1 is*

$$u(x;\alpha_1) = \begin{cases} u_1(x), & x < \alpha_1, \\ \\ u_2(x), & x \geq \alpha_1. \end{cases} \tag{6.18}$$

*If $u_1(\alpha_2) = u_2(\alpha_2)$ for $\alpha_2 \neq \alpha_1$ then α_2 is also a solution of **Problem (I-1)**. Additionally, the solution of (6.11) that is determined by α_2 is*

$$u(x;\alpha_2) = \begin{cases} u_1(x), & x < \alpha_2, \\ \\ u_2(x), & x \geq \alpha_2. \end{cases} \tag{6.19}$$

Proof. Since $u(x;\alpha_1)$ solves (6.11) and α_1 is a solution of **Problem (I-1)**, we have

$$p(0;\alpha_1)u'(0;\alpha_1) = g_0 \ \text{ and } \ p(1;\alpha_1)u'(1;\alpha_1) = g_1. \tag{6.20}$$

Then,

$$p(0;\alpha_2)u'(0;\alpha_2) = p_1 u_1'(0) = p(0;\alpha_1)u'(0;\alpha_1) = g_0, \tag{6.21}$$

and likewise

$$p(1;\alpha_2)u'(1;\alpha_2) = p_2 u_2'(1) = p(1;\alpha_1)u'(1;\alpha_1) = g_1. \tag{6.22}$$

These indicate that $u(x,\alpha_2)$ can match the data g_0, g_1 specified in the inverse **Problem (I-1)**. Also, since $u(0;\alpha_1) = u(1;\alpha_1) = 0$, it follows that $u_1(0) = 0$ and $u_2(0) = 0$ and hence $u(0;\alpha_2) = u(1;\alpha_2) = 0$. These imply that $u(x,\alpha_2)$ can also satisfy the boundary conditions of the interface problem (6.11).

By definition, we can see that $u(x,\alpha_2)$ satisfies the differential equation in (6.11), and $u(\alpha_2-,\alpha_2) = u_1(\alpha_2) = u_2(\alpha_2) = u(\alpha+,\alpha_2)$. What remains is to show that the flux continuity $\big[p\ u'\big]_{x=\alpha_2} = 0$ is also satisfied. Since $u_1(x)$ and $u_2(x)$ are solutions of (6.16) and

(6.17), respectively, both $-p_1u_1'(x)$ and $-p_2u_2'(x)$ are antiderivatives of the same function $f(x)$. Hence they may differ at most by a constant:

$$-p_1u_1'(x) = -p_2u_2'(x) + C, \tag{6.23}$$

for all $x \in (0,1)$. However, because $u(x;\alpha_1)$ is also a solution of (6.11), we must have $p_1u_1'(\alpha_1) = p_2u_2'(\alpha_1)$. This implies $C = 0$, and $p_1u_1'(x) = p_2u_2'(x)$ for all $x \in (0,1)$, and in particular, $p_1u_1'(\alpha_2) = p_2u_2'(\alpha_2)$. All of these lead to the conclusions that $u(x;\alpha_2)$ is a solution to the interface problem (6.11) with $\alpha = \alpha_2$ and α_2 is a solution of the inverse **Problem (I-1)**. ■

According to the theorem, if the solutions of the two IVP's, (6.16) and (6.17) share more than one point of intersection, then the inverse **Problem (I-1)** does not have a unique solution. On the other hand, the question of whether or not the solutions of these two IVP's intersect at all leads us to an existence theorem for **Problem (I-1)**.

Theorem 6.2.2. *If $u_1(x)$ is the unique solution of (6.16) and $u_2(x)$ is the unique solution of (6.17) and there exists a point $\alpha \in (0,1)$ such that $u_1(\alpha) = u_2(\alpha)$ then α is the solution of **Problem (I-1)**. In addition, the solution of (6.11) that is determined by α is*

$$u(x;\alpha) = \begin{cases} u_1(x) & x < \alpha, \\ u_2(x) & x \geq \alpha. \end{cases} \tag{6.24}$$

Proof. If $u_1(x)$ solves the IVP (6.16) and $u_2(x)$ solves the IVP (6.17) then the function $u(x;\alpha)$ defined by

$$u(x;\alpha) = \begin{cases} u_1(x) & x < \alpha, \\ u_2(x) & x \geq \alpha, \end{cases} \tag{6.25}$$

is a solution of the BVP (6.11) and hence α is a solution of **Problem (I-1)**. ■

This theorem gives us a sufficient condition for the existence of an exact solution to **Problem (I-1)**. This is because if α is a solution of **Problem (I-1)** then it certainly minimizes the least squares cost functional being used and thus it is also a solution in the least squares sense.

6.2.1 A Model Problem (I-1)

Now we consider a model problem to illustrate how these least squares methods can be applied to find the solution of the inverse **Problem (I-1)**. We look at **Problem (I-1)** when the right hand side of the elliptic interface problem (6.11) is $f(x) = e^x$ and the exact solution of **Problem (I-1)** is $\alpha^* = \pi^{-1} \approx 0.318309886$. The function that satisfies the elliptic

interface problem is $u_1(x)$ which is defined by (3.15) in Chapter 3. To avoid confusion, this function is simply referred to as $u(x)$ for the remainder of this section. Again, the coefficient function p has a ratio of $\beta = p_2/p_1 = 10$. Then by integrating and using the boundary conditions of (6.11) and interface conditions (6.14) to solve for constants, it is seen that the exact solution of (6.11) with the parameter $\alpha \in (0,1)$ is given by

$$u(x;\alpha) = \begin{cases} u_1(x;\alpha) = \dfrac{1}{-D(\alpha)}\big(e^x(\alpha p_2 + p_1) + \alpha(p_1 - p_2) - p_1 \\ \quad +x(e^{\alpha}(p_1 - p_2) + p_2 - ep_1)\big)\ , & \text{if } x < \alpha, \\ u_2(x;\alpha) = \dfrac{p_1}{p_2 D(\alpha)}\big(e^x p_1 - \alpha(e^x - e)(p_1 - p_2) \\ \quad +e^{\alpha}(p_1 - p_2)(x-1) + p_2(x-1) - ep_1 x\big)\ , & \text{if } x \geq \alpha, \end{cases} \tag{6.26}$$

where

$$D(\alpha) = p_1 + \alpha(p_2 - p_1). \tag{6.27}$$

Then the values that the flux is trying to be matched to are

$$\begin{aligned} g_0 &= -p_1 u_1'(0; 1/\pi) \approx -0.31740747230618, \\ g_1 &= p_2 u_2'(1; 1/\pi) \approx -1.40087435615287, \end{aligned} \tag{6.28}$$

and the exact OLS cost functional $\mathcal{J}_o$ is defined by

$$\mathcal{J}_o(\alpha) = |p_1 u_1'(0;\alpha) + g_0|^2 + |p_2 u_2'(0;\alpha) - g_1|^2, \tag{6.29}$$

where $p_1 = 1, p_2 = 10$.

According to Theorem 6.2.1 $\alpha^* = 1/\pi$ is not a unique solution of **Problem (I-1)**. This nonuniqueness of the solution occurs because $u_1(\bar{\alpha}; 1/\pi) = u_2(\bar{\alpha}; 1/\pi)$ with $\bar{\alpha} \approx 0.23240674424847$ and hence $\bar{\alpha}$ is also a solution of the inverse problem. This is confirmed by Figure 6.1 which shows the plot of the exact OLS cost functional $\mathcal{J}_o(\alpha)$ for $\alpha \in (0,1)$ and also more locally around $\alpha^* = 1/\pi$. It is clear from the plot on the right in Figure 6.1 that there are two minima which are associated with the cost function (6.29).

In the sections that follow, IFE spaces are used in the approximation of the exact solution, α^*, of **Problem (I-1)** using either the OLS cost functional or the LSMEE cost functional. In all cases, either the linear IFE space $\tilde{\mathcal{S}}_{h,1}(0,1)$ or the quadratic IFE space $\bar{\mathcal{S}}_{h,2}(0,1)$ are used in the approximation of u. On the other hand, in the mixed formulations, either the standard linear FE space $\mathcal{S}_{h,1}(0,1)$ or the standard quadratic FE space $\mathcal{S}_{h,2}(0,1)$ are used in the approximation of σ as is done in Chapter 5 for the MLSIFEM.

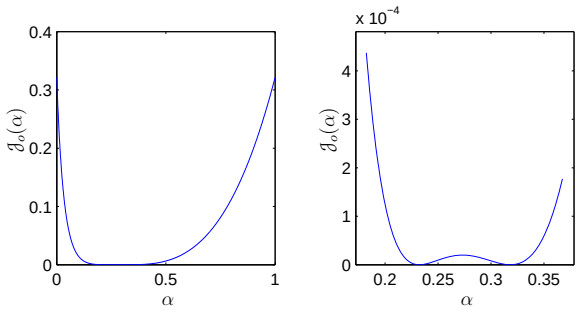


Figure 6.1: The output least squares cost functional $\mathcal{J}_o$ when $\alpha^* = 1/\pi$ is the exact solution of **Problem (I-1)** with $\beta = 10$.

6.2.2 Using an OLS Cost Functional to Solve Problem (I-1)

In this section, we try to minimize the OLS cost functional $\mathcal{J}_o(\alpha)$ that is introduced in the previous section. A typical approach to this problem is to take the derivative of the functional and then try to find at what α value the derivative is equal to zero using a Newton or quasi-Newton iteration. However, these methods require the ability to calculate the derivatives of the functional. It is possible that the derivative and then its second derivative can be approximated using difference quotients or other procedures, however, it is more straightforward to use a minimization technique for which just the evaluation of the function itself is sufficient. One such method is the golden section search method which requires only that the cost functional be calculated. However, each time that the functional is to be evaluated for a given α, either the elliptic interface problem (6.11) or its mixed formulation given by

$$\begin{aligned} p(x)u'(x) - \sigma(x) &= 0, \text{ if } x \in (0,1), \\ -\sigma'(x) &= f(x), \text{ if } x \in (0,1), \\ u(0) = u(1) &= 0, \end{aligned} \tag{6.30}$$

must be solved. Approximations of the solutions for (6.11) or (6.30) can of course be obtained using either the Galerkin IFEM or the MLSIFEM.

Here is an algorithm for the golden section search (or modified Fibonacci search) which uses only the evaluations of the function and hence no derivatives of the functional $\mathcal{J}_o$ need to be calculated [24, 25].

Algorithm 1. *Golden Section Search*

BEGIN
Given: a, b, tol, N
Step 1: let $\gamma = \frac{\sqrt{5}-1}{2}$ *and* $n = 1$.
Step 2: Find $m_1 = b - \gamma(b-a)$ *and* $m_2 = a + \gamma(b-a)$
Step 3: Calculate $\mathcal{J}_o(m_1)$ *and* $\mathcal{J}_o(m_2)$
Step 4: IF $\mathcal{J}_o(m_1) \leq \mathcal{J}_o(m_2)$ *THEN* $(a, b) \rightarrow (a, m_2)$ *ELSE* $(a, b) \rightarrow (m_1, b)$.
Step 5: IF $|a - b| < tol$ *OR* $n = N$ *THEN* $r = (m_1 + m_2)/2$ *END.*
Step 6: Let $n = n + 1$ *and GOTO Step 2.*
END

In this algorithm, a and b are the starting interval in which a minimum for $\mathcal{J}_o$ is sought. The search tolerance, tol, is how close the calculated minima, r, should be to the true minimum, α^* so that $|r - \alpha^*| < tol$. The maximum number of iterations that should be used by the algorithm is denoted by N. For the golden section search, the maximum number of iterations can be calculated precisely in terms of the starting interval and the desired tolerance [24, 25]:

$$N = \frac{\log\left(\frac{tol}{b-a}\right)}{\log \gamma}. \tag{6.31}$$

where the parameter γ is given by

$$\gamma = \frac{\sqrt{5}-1}{2} = \frac{2}{1+\sqrt{5}}\ . \tag{6.32}$$

Then using a search tolerance of $tol = 1 \times 10^{-10}$ and starting the search in the interval $[a, b] = [0.3, 0.35]$, the golden section search requires 49 iterations to return an approximation of the location of the minimum value. Although the golden section search appears to need two new function evaluations at each iteration, this is not the case. In fact, two function evaluations are only needed in the first iterative step. In subsequent iterative steps, one of the function evaluations is guaranteed to be either m_1 or m_2 from the prior evaluation. Thus after n iterations, the golden section search uses $n + 1$ function evaluations.

Applying an OLS Cost Functional with Galerkin IFEM

First, we use the Galerkin method to solve the interface problem (6.11) that is consider as a constraint upon the OLS cost functional that must be satisfied by u. Both the linear IFE space and the quadratic IFE space that uses extra continuity. Standard finite elements can be used with the Galerkin method, however, with each new calculation in the golden section search for the interface location α, a new mesh is introduced onto the domain. This is inconvenient at best.

In approximating the solution of the elliptic interface problem (6.11) using the Galerkin method, four gaussian quadrature nodes are used on each element. The starting search interval is $[a, b] = [0.3, 0.35]$ which is known to contain a minimum of the OLS cost functional since $\alpha^* = 1/\pi \in [0.3, 0.35]$ is the exact answer of **Problem (I-1)**. Then using the golden section search, 49 iterations are required to reach the desired tolerance of $tol = 1 \times 10^{-10}$. Note that this algorithm requires 2 evaluations of $\mathcal{J}_o$ in the first iteration of the algorithm and then 1 new evaluation of $\mathcal{J}_o$ at each iteration which results in 50 evaluations of the OLS cost functional $\mathcal{J}_o$. Using the IFE Galerkin method, this means that the elliptic interface problem needs to be solved 50 different times in order to approximate the value of the solution. However, the partition used in the Galerkin method for the interface problem are kept the same because of the use of IFE spaces.

The OLS approximation to the solution of inverse **Problem (I-1)** generated by using the Linear IFE space with the Galerkin method are shown in Table 6.1 while the results gener-

h	α_h	$E_h = \lvert\alpha_h - \alpha^*\rvert$	$\mathcal{J}_o(\alpha_h)$
1/181	0.30000000007	1.8310×10^{-2}	5.2928×10^{-5}
1/215	0.30009243597	1.8217×10^{-2}	3.7333×10^{-5}
1/256	0.30365955874	1.4650×10^{-2}	2.6339×10^{-5}
1/304	0.30635026936	1.1960×10^{-2}	1.8681×10^{-5}
1/362	0.30850369417	9.8062×10^{-3}	1.3177×10^{-5}
1/431	0.31022469919	8.0852×10^{-3}	9.2967×10^{-6}
1/512	0.31160150004	6.7084×10^{-3}	6.5886×10^{-6}

Table 6.1: Golden Section Search Errors using OLS Cost Functional with Galerkin IFEM $u_h \in \tilde{\mathcal{S}}_{h,1}$.

ated by using the Quadratic IFE space $\bar{\mathcal{S}}_{h,2}$ are shown in Table 6.2. Each table shows the approximate interface location, α_h, that is the minima of the OLS cost functional $\mathcal{J}_o(m_2)$ found by using a partition of size h. The errors are then given to indicate how close the approximate location is to the exact answer $\alpha^* = 1/\pi$. The Linear IFE space $\tilde{\mathcal{S}}_{h,1}(0, 1)$ does not perform well at all and requires a meshsize at least as small as $h = 1/181$ in order to return any useful results. The reason for this is that the cost function which is generated by using approximations of u from the Galerkin IFEM using only a linear IFE space does not begin to display two minima until h becomes smaller than about $h = 1/181$. This is probably due to the large error that results in $pu' = \sigma$ using the linear IFE space. In fact, the error in approximating pu' are of order $\mathcal{O}(h)$. A local maximum occurs between the two local minimums in the OLS cost function (see Figure 6.1) where the functional $\mathcal{J}_o$ has a value about 1×10^{-5}. If the IFE space used does not have errors below that level, then the errors act as a kind of regularization causing the two local minima to be hidden. Then not surprisingly, the order of the error in the approximate location of the minima (for $h < 1/181$)

h	α_h	$E_h = \lvert\alpha_h - \alpha^*\rvert$	$\mathcal{J}_o(\alpha_h)$
1/32	0.31717179662	1.1381×10^{-3}	9.2569×10^{-9}
1/45	0.31773660415	5.7328×10^{-4}	2.4050×10^{-9}
1/64	0.31802693002	2.8296×10^{-4}	5.9466×10^{-10}
1/91	0.31817000681	1.3988×10^{-4}	1.4667×10^{-10}
1/128	0.31823919056	7.0696×10^{-5}	3.7676×10^{-11}
1/181	0.31827453211	3.5354×10^{-5}	9.4607×10^{-12}
1/256	0.31829220957	1.7677×10^{-5}	2.3708×10^{-12}
1/362	0.31830104636	8.8398×10^{-6}	5.9414×10^{-13}
1/512	0.31830546476	4.4214×10^{-6}	1.4867×10^{-13}

Table 6.2: Golden Section Search Errors using OLS Cost Functional with Galerkin IFEM and $u_h \in \bar{\mathcal{S}}_{h,2}$.

is $E_h(\alpha_h) \approx 4.43521h^{1.03752}$, or only $\mathcal{O}(h)$, which agrees with the H^1-norm approximation capability of the linear IFE space.

Using the quadratic IFE space $\bar{\mathcal{S}}_{h,2}(0,1)$ to approximate the solution u to the interface problem for each α does a little better job because the data in Table 6.2 indicates that the order in the OLS solution of the inverse problem is $E_h(\alpha_h) \approx 1.16657h^{2.00128}$. This order also agrees with the approximation capability of the quadratic IFE space $\bar{\mathcal{S}}_{h,2}$ that is demonstrated in Chapters 3 through 5. Figure 6.2 shows the $\log h$ vs. $\log E_h$ plots for the OLS solution to the inverse problem using both the linear and quadratic IFE spaces. Especially noticeable in the plot for the linear IFE space is its behavior when $h > 1/181$. The order for the quadratic IFE approximations is found by using all of the data in the plot, while the order for the linear IFE space approximations is obtained by using only the portion of the plot that shows the error decreasing as h decreases. Because of this obvious deficiency in the behavior of the linear IFE space, it is more advantageous to use the quadratic IFE space $\bar{\mathcal{S}}_{h,2}$. This is true even though the system of equations used to solve for $u_h \in \bar{\mathcal{S}}_{h,2}$ has twice as many unknowns as the linear case. Consider if $h = 1/181$. Although the quadratic IFE space requires a system with 361 unknowns to be solved, it not only does much better than the linear IFE space but it also out performs the linear IFE space when it uses a similar number of unknowns at $h = 1/362$.

Applying an OLS Cost Functional with MLSIFEM

Next we apply the golden section search method while using the MLSIFEM to approximate the function u and the flux σ. The advantage here is that the order of approximation of the flux σ should be better than that obtained by differentiating the Galerkin IFE solution. Despite using the MLSIFEM, the number of evaluations of $\mathcal{J}_o$ that the golden section search

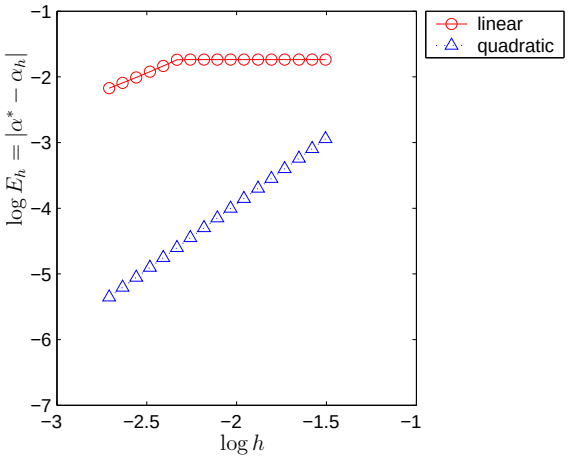


Figure 6.2: Output Least Squares Errors using Galerkin IFEM with Golden Section Search

requires is still going to be 49 with the tolerance of $tol = 1 \times 10^{-10}$. However, we expect this approach to produce more accurate results since the approximation of the flux is more accurate.

Table 6.3 shows the approximate minima of the functional when a linear IFE space is used for u and the standard linear FE space is used for σ. The order of the errors in approximating the location of the minima is $E \approx 1.80072h^{1.89678}$. Tables 6.4 and 6.5 show the approximations

h	α_h	$E_h = \lvert\alpha_h - \alpha\rvert$	$\mathcal{J}_o(\alpha_h)$
1/32	0.31595985967	2.3500×10^{-3}	9.7752×10^{-9}
1/45	0.31660226636	1.7076×10^{-3}	2.4998×10^{-9}
1/64	0.31738810459	9.2178×10^{-4}	6.1102×10^{-10}
1/91	0.31809685959	2.1303×10^{-4}	1.4949×10^{-10}
1/128	0.31808999619	2.1989×10^{-4}	3.8190×10^{-11}
1/181	0.31818604752	1.2384×10^{-4}	9.5517×10^{-12}
1/256	0.31824661691	6.3269×10^{-5}	2.3869×10^{-12}
1/362	0.31828537939	2.4507×10^{-5}	5.9699×10^{-13}
1/512	0.31830421187	5.6743×10^{-6}	1.4917×10^{-13}

Table 6.3: Golden Section Search Errors using Output Least Squares with linear/linear MLSIFEM for u_h and σ_h

using golden section search with quadratic-linear and quadratic-quadratic finite element spaces being used for u and σ. Similarly, the order of the errors in approximating the interface at $\alpha^* = 1/\pi$ using the quadratic IFE space $\bar{\mathcal{S}}_{h,2}$ for u_h and standard linear FE space

h	α_h	$E_h = \lvert\alpha_h - \alpha\rvert$	$\mathcal{J}_o(\alpha_h)$
1/32	0.31750712119	8.0276×10^{-4}	5.1233×10^{-23}
1/45	0.31790446869	4.0542×10^{-4}	3.1152×10^{-23}
1/64	0.31810945323	2.0043×10^{-4}	6.6677×10^{-24}
1/91	0.31821050281	9.9383×10^{-5}	1.0304×10^{-23}
1/128	0.31825964057	5.0246×10^{-5}	5.0135×10^{-23}
1/181	0.31828479022	2.5096×10^{-5}	4.6567×10^{-23}
1/256	0.31829735230	1.2534×10^{-5}	1.6031×10^{-23}
1/362	0.31830362160	6.2646×10^{-6}	1.3953×10^{-23}
1/512	0.31830675288	3.1333×10^{-6}	1.7580×10^{-22}

Table 6.4: Golden Section Search Errors for Output Least Squares using quadratic/linear MLSIFEM for u_h and σ_h

h	α_h	$E_h = \lvert\alpha_h - \alpha\rvert$	$\mathcal{J}_o(\alpha_h)$
1/32	0.31831574466	5.8585×10^{-6}	1.2622×10^{-26}
1/45	0.31831158981	1.7036×10^{-6}	7.7981×10^{-24}
1/64	0.31831034538	4.5920×10^{-7}	5.3036×10^{-23}
1/91	0.31830980574	8.0443×10^{-8}	5.8071×10^{-25}
1/128	0.31830979803	8.8150×10^{-8}	1.0753×10^{-23}
1/181	0.31830986804	1.8145×10^{-8}	6.0074×10^{-23}
1/256	0.31830988680	6.1764×10^{-10}	6.5439×10^{-23}
1/362	0.31830989007	3.8831×10^{-9}	4.0973×10^{-23}
1/512	0.31830988808	1.8968×10^{-9}	2.7091×10^{-21}

Table 6.5: Golden Section Search Errors for Output Least Squares using quadratic/quadratic MLSIFEM for u_h and σ_h

for σ_h is $E_h \approx 0.84147h^{2.00461}$. Likewise, the order when a quadratic-quadratic combination for u_h and σ_h are used is $E_h \approx 0.17484h^{3.09123}$. The orders are obtained by doing linear regression on the $\log h$ vs. $\log E_h$ errors which are shown in Figure 6.3. The orders all agree with the approximation capabilities of the respective spaces. In particular, it is not surprising that the quadratic-linear pairing results in errors of order $\mathcal{O}(h^2)$ since linear basis functions are used to approximate $\sigma(0)$ and $\sigma(1)$ in the output least squares cost functional and the L^2 approximation capability of linear elements is $\mathcal{O}(h^2)$.

Although the expectation is that using MLSIFEM in conjunction with the OLS cost functional is better than when the Galerkin IFEM is used, it is not clear that this is the case. For example, when a meshsize of $h = 1/64$ is used, using $\bar{\mathcal{S}}_{h,2}$ with the Galerkin IFEM to approximate pu_h' out performs the equivalent linear/linear pairing of the MLSIFEM. Both of these methods require an approximately equal number of unknowns to be solved for. The real advantage of the MLSIFEM method is not seen until a quadratic/quadratic pairing

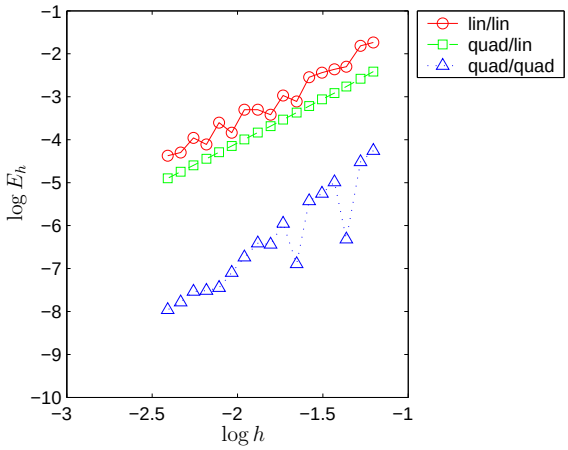


Figure 6.3: Output Least Squares Errors using MLSIFEM with Golden Section Search

with $u_h \in \bar{\mathcal{S}}_{h,2}$ and $\sigma_h \in \mathcal{S}_{h,2}$ is used. Then not only are the errors better when using the MLSIFEM with the OLS cost functional but the order of the errors are one order better as well. The disadvantage is that the number of unknowns is doubled to gain this advantage in accuracy.

6.2.3 Using a LSMEE Cost Functional to Solve Problem (I-1)

Next we investigate the Least Squares with Mixed Equation Error (LSMEE) method. While a golden section search approach can minimize the OLS cost functional in order to find an approximate solution to the inverse **Problem (I-1)**, more efficient minimization methods using a fewer number of function evaluations for a specified error tolerance can be used in conjunction with the LSMEE cost functional. The reason for this is that the LSMEE approach allows us to efficiently calculate the derivative of the cost functional with respect to the interface; hence faster gradient type algorithms can be employed to minimize the involved cost functional.

Specifically, the LSMEE cost functional in one dimension is

$$\begin{aligned} \mathcal{J}_m(u,\sigma,\alpha) \;=\; & \|f+\sigma'\|_0^2 + \|\sigma - p\ u'\|_0^2 \\ & +|\sigma(0)+g_0|^2 + |\sigma(1)-g_1|^2. \end{aligned} \tag{6.33}$$

where g_0 and g_1 are defined in (6.28).

Drawing upon standard calculus of variation techniques, it is necessary for the first variation of the cost functional $\mathcal{J}_m$ to be equal to zero at a the point where a minimum is attained.

However, the first variation of $\mathcal{J}_m$ can be written as

$$\nabla \mathcal{J}_m(u, \sigma, \alpha) \cdot (\delta u, \delta \sigma, \delta \alpha), \tag{6.34}$$

where $\nabla \mathcal{J}_m(u, \sigma, \alpha)$ is the gradient of $\mathcal{J}_m$ and δu, $\delta \sigma$ and $\delta \alpha$ are admissible variations of u, σ and α. Since these variations are arbitrary, then the requirement that the first variation be equal to zero leads to the gradient being equal to zero. We write the gradient of $\mathcal{J}_m$ as

$$\nabla \mathcal{J}_m(u, \sigma, \alpha) = \begin{pmatrix} \nabla_u \mathcal{J}_m(u, \sigma, \alpha) \\ \nabla_\sigma \mathcal{J}_m(u, \sigma, \alpha) \\ \nabla_\alpha \mathcal{J}_m(u, \sigma, \alpha) \end{pmatrix} , \tag{6.35}$$

where we use the notations $\nabla_u \mathcal{J}_m$, $\nabla_\sigma \mathcal{J}_m$ and $\nabla_\alpha \mathcal{J}_m$ to refer to the different components of the gradient. Since a necessary condition to have a minimum is that the gradient $\nabla \mathcal{J}_m = 0$, then this means that $\nabla_u \mathcal{J}_m$ and $\nabla_\sigma \mathcal{J}_m$ must of course be equal to zero. However, these two components of the gradient are almost identical to the first variations of the cost functional (5.25) that are used in Chapter 5 to motivate the MLSIFEM. The only difference comes in the matrix $K_{2,2}$ where there are contributions from the flux terms at the boundary in $\mathcal{J}_m$. For example, if we let α be fixed, then setting $\nabla_u \mathcal{J}_m = 0$ and $\nabla_\sigma \mathcal{J}_m = 0$ we get the following variational problem: find $(u, \sigma) \in \mathcal{V} \times \mathcal{W}$ such that

$$\begin{aligned} &\int_0^1 (pu' - \sigma) p\nu' \, dx = 0 \text{ for all } \nu \in \mathcal{V}, \\ &\int_0^1 \big\{ (pu' - \sigma)(-\xi) + (-\sigma' - f)(-\xi') \big\} dx \\ &\qquad + \xi(0)(\sigma(0) + g_0) + \xi(1)(\sigma(1) - g_1) = 0 \text{ for all } \xi \in \mathcal{W}, \end{aligned} \tag{6.36}$$

where $\mathcal{V}$ and $\mathcal{W}$ are defined in (5.23) in Chapter 5 . The only difference between these equations and (5.26) from Chapter 5 are the terms that arise from matching the flux at the endpoints of the domain. These terms only serve to make $K_{2,2}$ more diagonally dominant and hence do not destroy the nonsingularity of the coefficient matrix from (5.30). This observation brings up an interesting technique for finding the gradient of $\mathcal{J}_m$ in a Gauss-Seidel type of approach. It seems very natural to first find where the first variations of $\mathcal{J}_m$ with respect to u and σ are equal to zero since this amounts to finding the solution of the elliptic interface problem (6.11).

The derivative of $\mathcal{J}_m$ with respect to α can also be found directly. To do this, we first introduce the finite element discretizations for u and σ:

$$u_h = \sum_{i=1}^{N} U_i \phi_i \quad \text{and} \quad \sigma_h = \sum_{j=1}^{M} S_j \psi_j \ , \tag{6.37}$$

where $\{\phi_i\}$ $, i = 1, \ldots, N$ form the basis of an IFE space $\mathcal{V}_h \subset \mathcal{V}$ and $\{\psi_j\}$ $, j = 1, \ldots, M$ form the basis of a standard FE space $\mathcal{W}_h \subset \mathcal{W}$. Because u and σ are variables in the

cost function $\mathcal{J}_m$, then it follows that the coefficients $\vec{U}$ and $\vec{S}$ of u_h and σ_h are considered independent of α. Likewise, the basis functions $\psi_j,\ j = 1, \ldots, M$ are independent of α since they are constructed from a finite element partition which does not depend upon α. Consequently, it is seen that σ_h does not depend upon the parameter α and so

$$\frac{\partial}{\partial \alpha}(\sigma_h) = 0. \tag{6.38}$$

On the other hand, the space $\mathcal{V}_h$ that contains u_h does depend upon the parameter α since the basis functions are required to satisfy the interface conditions (6.14). In particular, the only basis functions of $\mathcal{V}_h$ which do depend upon α are those which are non-zero on an interface element. Hence we have that

$$\frac{\partial}{\partial \alpha}(u_h) = \sum_{i=1}^{N} U_i \frac{\partial}{\partial \alpha}(\phi_i(x)) \ , \tag{6.39}$$

Then taking the derivative of the cost function becomes very straightforward. The discretized cost function is given by

$$\begin{aligned} \mathcal{J}_m(u_h, \sigma_h, \alpha) &= \int_0^1 (pu_h' - \sigma_h)^2 dx + \int_0^1 (\sigma_h' + f)^2 dx \\ &\quad + |\sigma_h(0) + g_0|^2 + |\sigma_h(1) - g_1|^2 \ , \end{aligned} \tag{6.40}$$

Then taking derivatives with respect to α we get

$$\begin{aligned} \frac{d\mathcal{J}_m}{d\alpha}(u_h, \sigma_h, \alpha) &= \int_0^1 2\frac{\partial}{\partial \alpha}(pu_h')(pu_h' - \sigma_h) dx \\ &\quad + (pu_h' - \sigma_h)\big|_{x=\alpha-} - (pu_h' - \sigma_h)\big|_{x=\alpha+} \\ &\quad + (\sigma_h' + f)\big|_{x=\alpha-} - (\sigma_h' + f)\big|_{x=\alpha+} \ , \end{aligned} \tag{6.41}$$

and

$$\begin{aligned} \frac{d^2\mathcal{J}_m}{d\alpha^2} &= \int_0^1 \left\{ 2\left[\frac{\partial}{\partial \alpha}(pu' - \sigma)\right]^2 + 2\frac{\partial^2}{\partial \alpha^2}(pu' - \sigma)(pu' - \sigma) \right\} dx \\ &\quad - \frac{\partial}{\partial \alpha}\left(\left[pu_h' - \sigma_h\right]_{x=\alpha} \right) - \frac{\partial}{\partial \alpha}\left(\left[\sigma_h' + f\right]_{x=\alpha} \right) \ , \end{aligned} \tag{6.42}$$

where the derivatives of $pu_h'(x)$ are non-zero only when x is in interface element.

There are a couple of important notes to make here. First, the terms that are evaluated at $\alpha+$ and $\alpha-$ typically cancel out if α occurs in the interior of an element since $pu_h' - \sigma_h$ and $\sigma_h' + f$ are continuous on each element. Then the only times that these terms need to be considered is when α is one of the vertices of an element. In this situation, however, the immersed basis functions ϕ_i become the standard basis functions since they are consistent

with the standard FE basis functions (see Theorems 2.1.3 and 2.2.7). Because of this if α is an endpoint of the interface element, then the integral terms are zero.

Second, the $m-$th derivative of pu_h' is given by

$$\frac{\partial^m}{\partial \alpha^m}(pu_h') = \sum_{i=1}^{N} \left\{ U_i \frac{\partial^m}{\partial \alpha^m}(p\phi_i'(x)) \right\}, \tag{6.43}$$

where m is a non-negative integer. One may notice that on an interface element, p and u_h', are typically discontinuous. However, the product pu_h' is continuous and in fact, C^∞ with respect to α. For example, if $\phi_{k,i}(x)$ are the local linear IFE basis functions on the interface element e_k then the derivatives of $p\phi_{k,i}'(x),\ i = 1, 2$ are

$$\frac{\partial}{\partial \alpha}\big(p\phi_{k,1}'(x)\big) = p_1 p_2 K^{-2}(p_2 - p_1), \tag{6.44}$$

$$\frac{\partial}{\partial \alpha}\big(p\phi_{k,2}'(x)\big) = -p_1 p_2 K^{-2}(p_2 - p_1), \tag{6.45}$$

$$\frac{\partial^2}{\partial \alpha^2}\big(p\phi_{k,1}'(x)\big) = -2p_1 p_2 K^{-3}(p_2 - p_1)^2, \tag{6.46}$$

$$\frac{\partial^2}{\partial \alpha^2}\big(p\phi_{k,2}'(x)\big) = 2p_1 p_2 K^{-3}(p_2 - p_1)^2, \tag{6.47}$$

where $K = p_1(t_{k,2} - \alpha) + p_2(\alpha - t_{k,1})$. On the other hand when $t_{k,1} < \alpha < t_{k,2}$ and $\phi_{k,i} \in \bar{\mathcal{S}}_{h,2}$ then the first two derivatives of $p\bar{\phi}_{k,i}'(x),\ i = 1, 2, 3$ are given by

$$\begin{aligned} \frac{\partial}{\partial \alpha}\big(p\bar{\phi}_{k,1}'(x)\big) \ &= \ D_1^{-2}\big(p_1^3 p_2(p_1 - p_2)(t_{k,2} - t_{k,3})^2(2\alpha \\ &\quad -t_{k,2} - t_{k,3})(-2x + t_{k,2} + t_{k,3})\big)\ , \end{aligned} \tag{6.48}$$

$$\begin{aligned} \frac{\partial}{\partial \alpha}\big(p\bar{\phi}_{k,2}'(x)\big) \ &= \ D_1^{-2}\big(p_1^3 p_2(p_1 - p_2)(t_{k,2} - t_{k,3})(2x - \\ &\quad t_{k,2} - t_{k,3})(A\alpha - p_2 t_{k,1}^2 + p_1 t_{k,3}^2 \\ &\quad +2\alpha(p_2 t_{k,1} - p_1 t_{k,3}))\big)\ , \end{aligned} \tag{6.49}$$

$$\begin{aligned} \frac{\partial}{\partial \alpha}\big(p\bar{\phi}_{k,3}'(x)\big) \ &= \ D_1^{-2}\big(p_1^2 p_2(p_2 - p_1)(t_{k,2} - t_{k,3})(-2x \\ &\quad +t_{k,2} + t_{k,3})(A\alpha - p_2 t_{k,1}^2 + p_1 t_{k,2}^2 \\ &\quad +2\alpha(p_2 t_{k,1} - p_1 t_{k,2}))\big)\ , \end{aligned} \tag{6.50}$$

$$\begin{aligned} \frac{\partial^2}{\partial \alpha^2}\big(p\bar{\phi}_{k,1}'(x)\big) \ &= \ D_1^{-3}\big\{2p_1^4 p_2(p_1 - p_2)(t_{k,2} - t_{k,3})^3(2x \\ &\quad -t_{k,2} - t_{k,3})\big(3A(\alpha + t_{k,2} + t_{k,3}) \\ &\quad +p_1(t_{k,2}^2 + t_{k,2}t_{k,3} + t_{k,3}^2) \\ &\quad -p_2(t_{k,1}^2 - t_{k,1}(t_{k,2} + t_{k,3}) + (t_{k,2} + t_{k,3})^2)\big)\big\}\ , \end{aligned} \tag{6.51}$$

$$\begin{aligned}\frac{\partial^2}{\partial\alpha^2}\big(p\bar{\phi}'_{k,2}(x)\big) \;=\; & D_1^{-3}\big\{p_1^3p_2(p_1-p_2)(t_{k,2}-t_{k,3})^2(2x- \\ & t_{k,2}-t_{k,3})\big(-A^2\alpha+p_2^2t_{k,1}^3+p_1^2t_{k,3}^3 \\ & +3A\alpha(-p_2t_{k,1}+p_1t_{k,3})-p_1p_2(t_{k,3}^2(t_{k,2}+t_{k,3}) \\ & -t_{k,1}t_{k,3}(2t_{k,2}+t_{k,3})+t_{k,1}^2(t_{k,2}+2t_{k,3})) \\ & -3\alpha(p_2^2t_{k,1}^2+p_1^2t_{k,3}^2-p_1p_2(t_{k,1}^2+t_{k,3}^2)))\big\}\;,\end{aligned} \tag{6.52}$$

$$\begin{aligned}\frac{\partial^2}{\partial\alpha^2}\big(p\bar{\phi}'_{k,3}(x)\big) \;=\; & D_1^{-3}\big\{2p_1^3p_2(p_1-p_2)(t_{k,2}-t_{k,3})^2(2x-t_{k,2}-t_{k,3})\big(A^2\alpha \\ & -p_2^2t_{k,1}^3-p_1^2t_{k,2}^3+3A\alpha(-p_2t_{k,1}+p_1t_{k,2}) \\ & +3\alpha(p_2^2t_{k,1}^2+p_1^2t_{k,2}^2-p_1p_2(t_{k,1}^2+t_{k,2}^2)) \\ & +p_1p_2(t_{k,2}^2(t_{k,2}+t_{k,3})+t_{k,1}^2(2t_{k,2}+t_{k,3})-t_{k,1}t_{k,2}(t_{k,2}+2t_{k,3})))\big\}\;,\end{aligned} \tag{6.53}$$

where $A=\alpha(p_1-p_2)$ and

$$D_1=p_1(t_{k,2}-t_{k,3})(A(\alpha-t_{k,2}-t_{k,3})+p_2t_{k,1}(t_{k,1}-t_{k,2}-t_{k,3})+p_1t_{k,2}^2t_{k,3}^2. \tag{6.54}$$

Then while holding u and σ fixed, the derivatives of $\mathcal{J}_m(u_h,\sigma_h,\alpha)$ with respect to α can be calculated by using the above formulas in (6.41) and (6.42). Figures 6.4 and 6.5 show plots of the first and second derivatives of $\mathcal{J}_m(u_h,\sigma_h,\alpha)$ with respect to α for $\alpha\in(0,1)$ as well as a closer look at the derivatives in a region around between $\alpha=0.2$ and $\alpha=0.4$. These

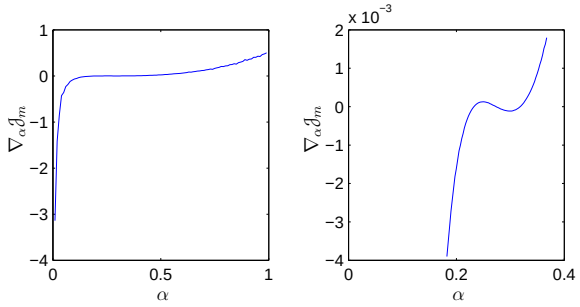


Figure 6.4: The first derivative of $\mathcal{J}_m$ with respect to α (with u and σ determined by $\nabla_u\mathcal{J}_m=0$ and $\nabla_\sigma\mathcal{J}_m=0$).

derivatives are calculated using u_h and σ_h which are found by first setting the first variations of $\mathcal{J}_m(u_h,\sigma_h,\alpha)$ with respect to u and σ equal to zero with $h=1/64$. Then we substitute u_h and σ_h into (6.41) and (6.41) to generate the derivatives. The quadratic IFE space $\bar{\mathcal{S}}_{h,2}$ is used for the approximations of u_h and the standard quadratic FE space is used for the approximations of σ_h.

Now that derivatives of $\mathcal{J}_m(u_h,\sigma_h,\alpha)$ with respect to α are obtained, we can apply a Gauss-Seidel type of iteration to minimize the cost functional. The process is to first fix α and find u_h and σ_h by their first variations. Then while u_h and σ_h are fixed, we can solve for the new

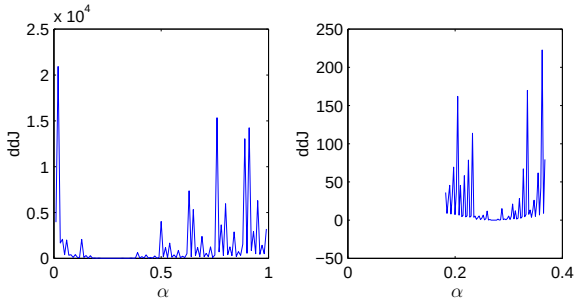


Figure 6.5: The second derivative of $\mathcal{J}_m$ with respect to α (with u and σ determined by $\nabla_u \mathcal{J}_m = 0$ and $\nabla_\sigma \mathcal{J}_m = 0$).

approximation to α by solving $\nabla_\alpha J(u_h, \sigma_h, \alpha) = 0$. The process is then repeated by first finding approximates to u and σ and then finding approximation to α again and again.

The main benefit here is that the mixed least squares method allows us to efficiently compute all the derivative components of $\nabla \mathcal{J}_m(u_h, \sigma_h, \alpha)$ which in turn allows more efficient minimization iteration methods to be used. This is rather difficult, if not impossible, to do with standard finite elements in such a straightforward way since each time the cost functional or its derivative needs to be calculated, the finite element partition needs to be recalculated. Additionally, this means that the coefficients of u_h and σ_h are changing in a uncertain way due to the new partition since the location of the unknowns is moving. Making u and σ independent of the parameter α then taking the derivative in the LSMEE method alleviates this difficulty.

Applying LSMEE Cost Functional with Bisection Method

This idea is summarized in the following algorithm [15]:

Algorithm 2. *(Gauss-Seidel Bisection)*

BEGIN
Given: a, b, tol, N.
Step 1: Use $\alpha = a$ *to find* $u_h^{(a)}$ *and* $\sigma_h^{(a)}$ *using (6.36). Use* $\alpha = b$ *to find* $u_h^{(b)}$ *and* $\sigma_h^{(b)}$ *using (6.36).*
Step 2: Calculate $\nabla_\alpha \mathcal{J}_m^{(a)}$ *and* $\nabla_\alpha \mathcal{J}_m^{(b)}$. *IF NOT*$(\nabla_\alpha \mathcal{J}_m^{(a)} \nabla_\alpha \mathcal{J}_m^{(b)} < 0)$ *THEN END*
Step 3: Let $m = (a+b)/2$.
Step 4: Use $\alpha = m$ *to find* $u_h^{(m)}$ *and* $\sigma_h^{(m)}$.
Step 5: Calculate $\nabla \mathcal{J}_m^{(m)}$.
Step 6: IF $\nabla_\alpha \mathcal{J}_m^{(a)} \nabla_\alpha \mathcal{J}_m^{(m)} < 0$ *THEN* $(a, b) \to (a, m)$ *ELSE* $(a, b) \to (m, b)$
Step 7: IF $n = N$ *THEN* $r = m$ *END*
Step 8: GOTO Step 3
END

Here we use the notation $\nabla_\alpha \mathcal{J}_m^{(a)} = \nabla_\alpha \mathcal{J}_m(u_h^{(a)}, \sigma_h^{(a)}, a)$ for example. If the desired tolerance is *tol* then the bisection method needs N iterations where

$$N = \frac{\ln(b-a) - \ln tol}{\ln 2}. \tag{6.55}$$

(see [15, 25]). For N iterations, the bisection method requires $N+2$ function evaluations. This is a significant improvement over the golden section search. For example, if $tol = 1 \times 10^{-10}$ and the starting interval is $[0.3, 0.35]$, then the minimization procedure here needs a maximum of 29 iterations. Less than this may be used if Step 6 is modified to also test for whether or not m is actually a root where $\nabla_\alpha \mathcal{J}_m^{(m)} = 0$. If this is true then the iteration should be stopped.

Now we apply Algorithm 2 to solve the model inverse problem. Results for using u_h approximated by the linear IFE space and σ_h approximated by the standard linear FE space are given in Table 6.6. Likewise, the errors in approximating the solution $\alpha^* = 1/\pi$ the model inverse problem with the bisection method when using a quadratic-linear combination for u_h and σ_h in the MLSIFEM are shown in Table 6.7, and Table 6.8 shows errors when the bisection method is used in conjunction with quadratic-quadratic spaces for u_h and σ_h in the MLSIFEM.

The order of the approximations to α^* using a linear-linear combination for the approximation spaces of u_h and σ_h is $E_h \approx 5.08402h^{2.09676}$. Unlike the output least squares method, this error is $\mathcal{O}(h^2)$ since the flux is now being approximated directly by σ_h. Similarly, the order of the approximation to α^* using a quadratic-linear pairing of the finite element spaces results in an order of $E_h \approx 0.82142h^{1.99998}$. Finally, the order of the approximation that is generated by the quadratic-quadratic pairing for u_h and σ_h is $E_h \approx 0.10304h^{2.99226}$. All three

h	α_h	$E_h = \lvert\alpha_h - \alpha\rvert$	$\mathcal{J}_m(u_h, \sigma_h, \alpha_h)$	Evaluations of $\mathcal{J}_m$
1/64	0.31740906741	9.0082×10^{-4}	1.2998×10^{-4}	30
1/91	0.31810926218	2.0062×10^{-4}	6.4293×10^{-5}	30
1/128	0.31809576359	2.1412×10^{-4}	3.2496×10^{-5}	30
1/181	0.31818884425	1.2104×10^{-4}	1.6252×10^{-5}	29
1/256	0.31824799255	6.1894×10^{-5}	8.1241×10^{-6}	28
1/362	0.31828605477	2.3831×10^{-5}	4.0629×10^{-6}	30
1/512	0.31830456275	5.3234×10^{-6}	2.0310×10^{-6}	30

Table 6.6: Errors for LSMEE: linear u_h and linear σ_h with bisection

h	α_h	$E_h = \lvert\alpha_h - \alpha\rvert$	$\mathcal{J}_m(u_h, \sigma_h, \alpha_h)$	Evaluations of $\mathcal{J}_m$
1/64	0.31810945319	2.0043×10^{-4}	6.4991×10^{-5}	31
1/91	0.31821050290	9.9383×10^{-5}	3.2147×10^{-5}	30
1/128	0.31825964041	5.0246×10^{-5}	1.6248×10^{-5}	29
1/181	0.31828479022	2.5096×10^{-5}	8.1258×10^{-6}	27
1/256	0.31829735208	1.2534×10^{-5}	4.0620×10^{-6}	30
1/362	0.31830362184	6.2643×10^{-6}	2.0315×10^{-6}	31
1/512	0.31830675220	3.1340×10^{-6}	1.0155×10^{-6}	31

Table 6.7: Errors for LSMEE: quadratic u_h and linear σ_h with bisection

h	α_h	$E_h = \lvert\alpha_h - \alpha\rvert$	$\mathcal{J}_m(u_h, \sigma_h, \alpha_h)$	Evaluations of $\mathcal{J}_m$
1/64	0.31831034534	4.5915×10^{-7}	5.2890×10^{-10}	29
1/91	0.31830980573	8.0455×10^{-8}	1.2940×10^{-10}	30
1/128	0.31830979800	8.8185×10^{-8}	3.3057×10^{-11}	31
1/181	0.31830986803	1.8150×10^{-8}	8.2678×10^{-12}	31
1/256	0.31830988694	7.5603×10^{-10}	2.0661×10^{-12}	30
1/362	0.31830989029	4.1088×10^{-9}	5.1674×10^{-13}	30
1/512	0.31830988536	8.2722×10^{-10}	1.2913×10^{-13}	31

Table 6.8: Errors for LSMEE: quadratic u_h and quadratic σ_h with bisection

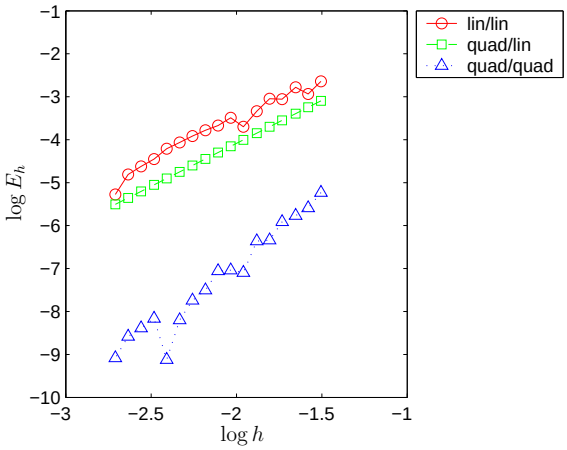


Figure 6.6: Errors for LSMEE: using Bisection method

of these orders agree nicely with the approximation capabilities of the involved IFE spaces. Plots of the $\log h$ vs. $\log E_h = |\alpha_h - \alpha^*|$ for the different combinations of basis spaces are shown in Figure 6.6 when the bisection method is used.

These plots show that the quadratic-quadratic pairing consistently performs the best. The quadratic-linear combination, however, does not perform much better than the linear-linear combination. It seems that the trend with a quadratic-linear mixed least squares approach seems to be more linear but this seems to be the only obvious benefit. In terms of calculation, the quadratic-linear is solving a system to find u_h and σ_h that has about $3N$ unknowns compared with about $2N$ unknowns in the linear-linear case where there are N elements in the finite element partition.

Applying LSMEE Cost functional with Secant Method

In this section, we incorporate the secant method with the Gauss-Seidel iteration idea in the LSMEE method to solve the inverse **Problem (I-1)**. Since $\nabla_{\alpha\alpha}\mathcal{J}_m = \dfrac{d^2\mathcal{J}_m}{d\alpha^2}$ has values which are too erratic, it seems not to be appropriate to use a Newton type method to solve the gradient equations. See Figures 6.4 and 6.5 for plots of the $\nabla_\alpha \mathcal{J}_m$ and $\nabla_{\alpha\alpha}\mathcal{J}_m$ respectively. These plots are generated using a finite element mesh of size $h = 1/64$, quadratic IFE functions for u_h and standard quadratic FE functions for σ_h. The instability of the second derivative is probably due to the nature of the sensitivities of the immersed basis functions to the location of α in an interface element. On the other hand, behavior of $\nabla_\alpha \mathcal{J}_m$ is quite nice, and consequently the super-linear secant type iterative methods turn out to be a better choice.

h	α_h	$E_h = \lvert\alpha_h - \alpha\rvert$	$\mathcal{J}_m(u_h, \sigma_h, \alpha_h)$	Evaluations of $\mathcal{J}_m$
1/64	0.31740906745	9.0082×10^{-4}	1.2998×10^{-4}	9
1/91	0.31810926211	2.0062×10^{-4}	6.4293×10^{-5}	11
1/128	0.31809576359	2.1412×10^{-4}	3.2496×10^{-5}	9
1/181	0.31818884431	1.2104×10^{-4}	1.6252×10^{-5}	10
1/256	0.31824799252	6.1894×10^{-5}	8.1241×10^{-6}	10
1/362	0.31828605474	2.3831×10^{-5}	4.0629×10^{-6}	10
1/512	0.31830456274	5.3234×10^{-6}	2.0310×10^{-6}	10

Table 6.9: Errors for LSMEE: linear u_h and linear σ_h with secant method.

The calculation of the secant method is very similar to what is done to implement the bisection method. The following algorithm shows how the secant method may be implemented [15, 24].

Algorithm 3. *(Secant Method)*

BEGIN
Given: $\alpha^{(0)}, \alpha^{(1)}, tol, N$.
Step 1: Use $\alpha^{(0)}$ *to find* $u_h^{(0)}$ *and* $\sigma_h^{(0)}$ *using (6.36). Calculate* $\nabla_\alpha \mathcal{J}_m^{(0)}$.
Use $\alpha^{(1)}$ *to find* $u_h^{(1)}$ *and* $\sigma_h^{(1)}$ *using (6.36). Calculate* $\nabla_\alpha \mathcal{J}_m^{(1)}$.
Let $i = 1$.
Step 2: Calculate h *where*

$$h = -\frac{\nabla_\alpha \mathcal{J}_m^{(i)}(\alpha^{(i)} - \alpha^{(i-1)})}{\nabla_\alpha \mathcal{J}_m^{(i)} - \nabla_\alpha \mathcal{J}_m^{(i-1)}}. \tag{6.56}$$

Step 3: Find $\alpha^{(i+1)} = \alpha^{(i)} + h$.
Step 4: Use $\alpha^{(i+1)}$ *to find* $u_h^{(i+1)}$ *and* $\sigma_h^{(i+1)}$ *using (6.36). Calculate* $\nabla_\alpha \mathcal{J}_m^{(i+1)}$.
Step 5: IF $n = N$ *OR* $|\nabla_\alpha \mathcal{J}_m^{(i+1)}| < tol$ *THEN END*
Step 6: Let i=i+1. GOTO Step 2.
END

Tables 6.9 through 6.11 show the errors in the approximation of the α^* for the various pairings of an IFE space for u_h and a standard FE space for σ_h. In each table, the number of evaluations of $\mathcal{J}_m$ that the secant method requires to converge in each case is included. It is obvious that using the secant method resulted in a drastic reduction in the number of iterations that are required for the bisection method. Additionally, the secant method only requires one new function evaluation in each iteration. Then comparatively with the bisection method, the secant method results about a third of the function evaluations being

h	α_h	$E_h = \lvert\alpha_h - \alpha\rvert$	$\mathcal{J}_m(u_h, \sigma_h, \alpha_h)$	Evaluations of $\mathcal{J}_m$
1/64	0.31810945317	2.0043×10^{-4}	6.4991×10^{-5}	10
1/91	0.31821050281	9.9383×10^{-5}	3.2147×10^{-5}	10
1/128	0.31825964044	5.0246×10^{-5}	1.6248×10^{-5}	10
1/181	0.31828479023	2.5096×10^{-5}	8.1258×10^{-6}	10
1/256	0.31829735209	1.2534×10^{-5}	4.0620×10^{-6}	10
1/362	0.31830362179	6.2644×10^{-6}	2.0315×10^{-6}	10
1/512	0.31830675219	3.1340×10^{-6}	1.0155×10^{-6}	10

Table 6.10: Errors for LSMEE: quadratic u_h and linear σ_h with secant method.

h	α_h	$E_h = \lvert\alpha_h - \alpha\rvert$	$\mathcal{J}_m(u_h, \sigma_h, \alpha_h)$	Evaluations of $\mathcal{J}_m$
1/64	0.31831034532	4.5914×10^{-7}	5.2890×10^{-10}	10
1/91	0.31830980571	8.0470×10^{-8}	1.2940×10^{-10}	10
1/128	0.31830979798	8.8208×10^{-8}	3.3057×10^{-11}	10
1/181	0.31830986800	1.8186×10^{-8}	8.2678×10^{-12}	10
1/256	0.31830988695	7.6507×10^{-10}	2.0661×10^{-12}	10
1/362	0.31830989020	4.0195×10^{-9}	5.1674×10^{-13}	10
1/512	0.31830988533	8.5859×10^{-10}	1.2913×10^{-13}	10

Table 6.11: Errors for LSMEE: quadratic u_h and quadratic σ_h with secant method.

used to converge to the desired tolerance. This is expected from the secant method since it usually has a super-linear convergence. The orders of the errors in approximating α^* with the different combinations of immersed and standard FE spaces are displayed in Table 6.12. These orders are very similar to those we obtain in the previous sections when the MLSIFE method is used in conjunction with the output least squares cost functional and when the LSMEE cost functional is used with the bisection method. Similarly, Figure 6.7 shows the

u_h	σ_h	Approximate Order
linear	linear	$E \approx 5.08402h^{2.09676} = \mathcal{O}(h^2)$
quadratic	linear	$E \approx 0.82140h^{1.99998} = \mathcal{O}(h^2)$
quadratic	quadratic	$E \approx 0.10006h^{2.98637} = \mathcal{O}(h^3)$

Table 6.12: Orders of $E_h = \lvert\alpha_h - \alpha^*\rvert$ using LSMEE cost functional with the secant method

plots of $\log h$ vs. $\log E_h$ for the various pairings of IFE and FE spaces for u_h and σ_h.

As in the case above in which the LSMEE cost functional with bisection method is used to solve the inverse problem, again, the quadratic IFE space $\bar{\mathcal{S}}_{h,2}$ in combination with the

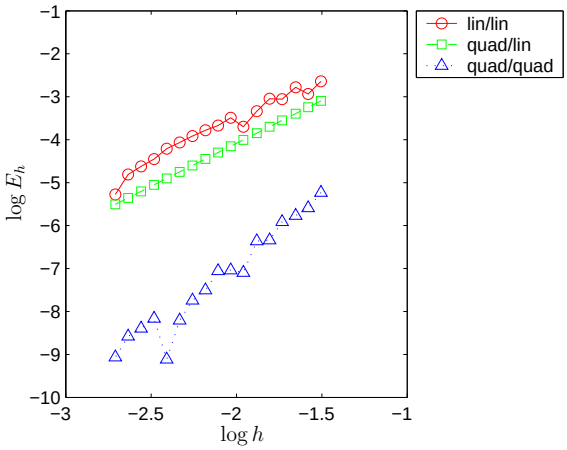


Figure 6.7: Errors for LSMEE: using Secant method

standard quadratic FE space performs the best here in terms of order and in terms of magnitude of error. Also, the quadratic-linear pairing performs slightly better than the linear-linear pairing although probably not well enough to justify its use since it is the same order as the linear-linear pairing and is computationally more expensive.

6.2.4 Summary

The main emphasis of this section is to use the immersed finite element spaces to apply various numerical strategies for finding the minimum of either the output least squares cost functional $\mathcal{J}_o$ or the least squares with equation error cost functional $\mathcal{J}_m$. The orders of the errors in the approximations of the location of α^* agree in all cases with the approximation capabilities of the finite element spaces involved. The worst orders are generated by the output least squares method using just the Galerkin finite element method with IFE spaces to implement the golden section search. These orders are of $\mathcal{O}(h)$ for the linear IFE space and $\mathcal{O}(h^2)$ for the quadratic IFE space $\bar{\mathcal{S}}_{h,2}$.

The quadratic IFE space $\bar{\mathcal{S}}_{h,2}$ really stands out in the numerical results presented here. The most telling result however is that the linear IFE space does not perform well at all in the OLS method using the Galerkin IFEM to implement the golden section search. Two improvements are found for the bad results that are exhibited by the linear space. First, quadratic elements work well and do not seem to have many problems at the larger h values. For instance, at $h = 1/64$, the quadratic IFE space with the OLS cost functional and the Galerkin IFEM produces reasonable results while even at $h = 1/128$ the linear space can still not produce meaningful results using approximately the same number of unknowns. Second,

a mixed formulation using a least squares approach also serves to improve upon the results based upon the linear IFE space $\tilde{\mathcal{S}}_{h,1}$.

When the LSIFEM is used in conjunction with the output least squares and when the LSMEE method is used with either the bisection or secant methods, the order of the approximations agree with the orders for approximating the flux variable σ. Also, using a LSMEE formulation allows components of the gradient of the cost functional to be efficiently calculated so that either the bisection or secant methods can be employed to find an approximation to the solution α^* to the inverse problem with a much smaller number of function evaluations.

Table 6.13 shows a comparison of the different methods we use to solve **Problem (I-1)** when the meshsize is $h = 1/64$. Additionally the approximate orders are given for each method.

Type of Cost Function	$\mathcal{V}_h$	$\mathcal{W}_h$	Minimizing Technique	$E_h = \lvert\alpha_h - \alpha^*\rvert$	Function Evals.	No. of Variables	Order
OLS w/IFEM	$\mathcal{S}_{h,2}$	n/a	golden section	2.8296×10^{-4}	53	125	$\mathcal{O}(h^2)$
OLS w/LSIFEM	$\tilde{\mathcal{S}}_{h,1}$	$\mathcal{S}_{h,1}$	golden section	9.2178×10^{-4}	53	126	$\mathcal{O}(h^2)$
OLS w/LSIFEM	$\mathcal{S}_{h,2}$	$\mathcal{S}_{h,1}$	golden section	2.0043×10^{-4}	53	190	$\mathcal{O}(h^2)$
OLS w/LSIFEM	$\mathcal{S}_{h,2}$	$\mathcal{S}_{h,2}$	golden section	4.5920×10^{-7}	53	252	$\mathcal{O}(h^3)$
LSMEE	$\tilde{\mathcal{S}}_{h,1}$	$\mathcal{S}_{h,1}$	bisection	9.0082×10^{-4}	30	126	$\mathcal{O}(h^2)$
LSMEE	$\mathcal{S}_{h,2}$	$\mathcal{S}_{h,1}$	bisection	2.0043×10^{-4}	31	190	$\mathcal{O}(h^2)$
LSMEE	$\mathcal{S}_{h,2}$	$\mathcal{S}_{h,2}$	bisection	4.5915×10^{-7}	29	252	$\mathcal{O}(h^3)$
LSMEE	$\tilde{\mathcal{S}}_{h,1}$	$\mathcal{S}_{h,1}$	secant	9.0082×10^{-4}	9	126	$\mathcal{O}(h^2)$
LSMEE	$\mathcal{S}_{h,2}$	$\mathcal{S}_{h,1}$	secant	2.0043×10^{-4}	10	190	$\mathcal{O}(h^2)$
LSMEE	$\mathcal{S}_{h,2}$	$\mathcal{S}_{h,2}$	secant	4.5914×10^{-7}	10	252	$\mathcal{O}(h^3)$

Table 6.13: Summary of the methods for solving the inverse **Problem (I-1)** using a meshsize of $h = 1/64$.

Table 6.14 shows a comparison of the CPU times that different methods need when a meshsize of $h = 1/256$ is used. The same information is shown in Table 6.15 except, the number of unknowns in each method is kept about equal. CPU times are obtained using Matlab 6.1 on an Intel Pentium 4 computer. The processor speed is 2.4 Gigahertz with hyperthreading and the computer has a gigabyte of RAM. The CPU times easily show the benefit of being able to take the derivative of the cost functional as the secant method is more efficient than the other methods of comparable accuracy.

Type of Cost Function	$\mathcal{V}_h$	$\mathcal{W}_h$	Minimizing Technique	No. of Variables	CPU Time (seconds)
OLS w. IFEM	$\tilde{\mathcal{S}}_{h,1}$	n/a	Golden Section	255	16.42
OLS w. IFEM	$\bar{\mathcal{S}}_{h,2}$	n/a	Golden Section	511	15.94
OLS w. MLSIFEM	$\tilde{\mathcal{S}}_{h,1}$	$\mathcal{S}_{h,1}$	Golden Section	512	12.28
OLS w. MLSIFEM	$\bar{\mathcal{S}}_{h,2}$	$\mathcal{S}_{h,1}$	Golden Section	766	21.67
OLS w. MLSIFEM	$\bar{\mathcal{S}}_{h,2}$	$\mathcal{S}_{h,2}$	Golden Section	1020	26.54
LSMEE	$\tilde{\mathcal{S}}_{h,1}$	$\mathcal{S}_{h,1}$	Bisection	512	11.41
LSMEE	$\bar{\mathcal{S}}_{h,2}$	$\mathcal{S}_{h,1}$	Bisection	766	19.27
LSMEE	$\mathcal{S}_{h,2}$	$\mathcal{S}_{h,2}$	Bisection	1020	23.98
LSMEE	$\tilde{\mathcal{S}}_{h,1}$	$\mathcal{S}_{h,1}$	Secant	512	4.366
LSMEE	$\bar{\mathcal{S}}_{h,2}$	$\mathcal{S}_{h,1}$	Secant	766	6.725
LSMEE	$\mathcal{S}_{h,2}$	$\mathcal{S}_{h,2}$	Secant	1020	8.425

Table 6.14: Summary of CPU times for various methods of solving **Problem (I-1)** using $h = 1/256$.

Type of Cost Function	$\mathcal{V}_h$	$\mathcal{W}_h$	Minimizing Technique	Meshsize h	No. of Variables	CPU Time (seconds)
OLS w. IFEM	$\tilde{\mathcal{S}}_{h,1}$	n/a	Golden Section	1/512	511	16.42
OLS w. IFEM	$\bar{\mathcal{S}}_{h,2}$	n/a	Golden Section	1/256	509	15.94
OLS w. MLSIFEM	$\tilde{\mathcal{S}}_{h,1}$	$\mathcal{S}_{h,1}$	Golden Section	1/256	512	12.28
OLS w. MLSIFEM	$\bar{\mathcal{S}}_{h,2}$	$\mathcal{S}_{h,1}$	Golden Section	1/171	511	14.49
OLS w. MLSIFEM	$\bar{\mathcal{S}}_{h,2}$	$\mathcal{S}_{h,2}$	Golden Section	1/128	508	13.43
LSMEE	$\tilde{\mathcal{S}}_{h,1}$	$\mathcal{S}_{h,1}$	Bisection	1/256	512	11.41
LSMEE	$\mathcal{S}_{h,2}$	$\mathcal{S}_{h,1}$	Bisection	1/171	511	11.35
LSMEE	$\mathcal{S}_{h,2}$	$\mathcal{S}_{h,2}$	Bisection	1/128	508	12.67
LSMEE	$\tilde{\mathcal{S}}_{h,1}$	$\mathcal{S}_{h,1}$	Secant	1/256	512	4.366
LSMEE	$\mathcal{S}_{h,2}$	$\mathcal{S}_{h,1}$	Secant	1/171	511	4.625
LSMEE	$\mathcal{S}_{h,2}$	$\mathcal{S}_{h,2}$	Secant	1/128	508	4.394

Table 6.15: Summary of CPU times for various methods of solving **Problem (I-1)** using approximately the same number of unknowns.

Chapter 7

Conclusions

The focus of this thesis is to develop a class of immersed finite element spaces which is suitable for use in solving both forward and inverse elliptic interface problems in $\Omega \subset \mathbb{R}^n$ with $n = 1$ or 2 that have a piecewise constant coefficient. Solutions of elliptic interface problems are required to satisfy the interface requirements that u and its flux are continuous on the domain Ω. Immersed finite element spaces are finite element spaces which are constructed so that the basis functions can satisfy these interface conditions. Defining FE basis functions in this way allows the partition upon which the IFE spaces are constructed to be independent of the interface thereby immersing it in the partition of the domain.

Quadratic IFE spaces are not uniquely determined by the interface conditions and the usual nodal value specifications of the basis functions. By imposing an extra condition, quadratic IFE spaces are uniquely determined. Three quadratic IFE spaces and their related bi-quadratic IFE spaces are formed so that the basis functions satisfy the appropriate interface conditions and so that the spaces are uniquely determined. The first IFE spaces, $\tilde{\mathcal{S}}_{h,2}(0,1)$ and $\tilde{\mathcal{S}}_{h,2,\square}\big((0,1)^2\big)$, are constructed using a hierarchical approach based upon lower degree immersed interpolating functions. The second IFE spaces, $\bar{\mathcal{S}}_{h,2}(0,1)$ and $\bar{\mathcal{S}}_{h,2,\square}\big((0,1)^2\big)$, are constructed by applying an extra continuity requirement upon the basis functions. Specifically, the derivative of the flux is required to also be continuous on the domain Ω. Finally, IFE spaces using a local refinement technique of Y. Lin, $\hat{\mathcal{S}}_{h,2}(0,1)$ and $\hat{\mathcal{S}}_{h,2,\square}\big((0,1)^2\big)$, are constructed.

The interpolation properties of these spaces are numerically tested by using them to interpolate a set of appropriate testing functions. The spaces using a hierarchical construction and those using an extra continuity requirement have mixed results depending upon the continuity of the derivative of the flux. For example, the IFE space $\bar{\mathcal{S}}_{h,2}(0,1)$ performs optimally in interpolating the function $u_1(x)$ which has a differentiable flux. The third IFE space based on a local refinement technique performs optimally regardless of which testing function is

used. For the most part, the approximation results we obtain using the Galerkin IFEM and the MLSIFEM agree with the interpolation results.

Finally, the IFE space $\bar{\mathcal{S}}_{h,2}(0,1)$ is used to solve inverse **Problem (I-1)** using an OLS cost functional and a LSMEE cost functional. The best results are obtained using the LSMEE cost functional since the derivative of the cost functional is easy to obtain. Hence, a quasi-newton method such as the secant method is employed to solve the problem. In the OLS method, the derivative of the cost function is not readily available and so the golden section search algorithm is employed to solve the problem. The best results are found when the approximate functions u_h and σ_h are in the quadratic spaces $\bar{\mathcal{S}}_{h,2}(0,1)$ and $\mathcal{S}_{h,2}(0,1)$ respectively.

There are a few main results that arise from this thesis. The most obvious result is that IFE spaces can be used efficiently with moving interface problems such as inverse **Problem (I-1)** presented in Chapter 6. Also, IFE spaces work well in conjunction with a mixed formulation such as the MLSIFEM. The advantage of this is that a continuous representation of the flux is obtained where normal finite element approximations only allow for a piecewise continuous approximation of the flux. Lastly, quadratic IFE spaces provide better approximation capability than linear IFE spaces. This feature of quadratic IFE spaces is shown when they are used to solve the inverse **Problem (I-1)**. For example, linear IFE spaces by themselves with the OLS cost functional provide very poor results while the quadratic IFE space $\bar{\mathcal{S}}_{h,2}(0,1)$ perform acceptably.

Some questions are not answered by this thesis. There is a need to develop quadratic and biquadratic IFE spaces that are based on more general interfaces in 2D. One way to construct these IFE spaces is to use a hierarchical approach which uses linear immersed interpolating functions that are constructed on triangular elements. Extra continuity requirements also have several possible implementations for the formation of quadratic and biquadratic IFE spaces using linear and rectangular elements respectively. Also, analysis of the interpolation errors and the approximation errors obtained using the Galerkin IFEM and the MLSIFEM needs to be performed.

Bibliography

[1] R. A. Adams. *Sobolev Spaces*. Academic Press, 1975.

[2] A.K. Aziz and I. Babuška. Survey lectures on the mathematical foundations of the finite element method. In A.K. Aziz, editor, *The Mathematical Foundation of the Finite Element Method with Applications to Partial Differential Equations*, pages 3–363. Academic Press, 1972.

[3] H. T. Banks and K. Kunisch. *Estimation Techniques for Distributed Parameter Systems*. Birkhäuser, Boston, 1989.

[4] J. Baumeister. *Stable Solution of Inverse Problems*. Advanced Lectures in Mathematics. Friedr. Vieweg and Sohn, 1987.

[5] J. V. Beck, B. Blackwell, and St. Clair, Jr., C. R. *Inverse Heat Conduction: Ill-posed Problems*. John Wiley and Sons, Inc., 1985.

[6] G. Birkhoff and R. E. Lynch. *Numerical Solution of Elliptic Problems*. SIAM Studies in Applied Mathematics. SIAM, Philadelphia, 1984.

[7] P. Bochev and M. Gunzburger. Finite element methods of least-squares type. *SIAM Review*, 40(5):789–837, 1998.

[8] D. Braess. *Finite Elements: Theory, fast solvers, and applications in solid mechanics*. Cambridge University Press, 1997.

[9] S. C. Brenner and L. R. Scott. *The Mathematical Theory of Finite Element Methods*, volume 15 of *Texts in Applied Mathematics*. Springer, 1994.

[10] F. Brezzi and M. Fortin. *Mixed and Hybrid Finite Element Methods*. Springer-Verlag, 1991.

[11] M. Burger. A level set method for inverse problems. *Inverse Problems*, 17:1327–1355, 2001.

[12] Z. Cai, R. Lazarov, T.A. Manteuffel, and S.F. McCormick. First-order system least squares for second-order partial differential equations: part i. *SIAM Journal on Numerical Analysis*, 31(6):1785–1799, 1994.

[13] Z. Cai, T.A. Manteuffel, and S.F. McCormick. First-order system least squares for second-order partial differential equations: Part ii. *SIAM Journal of Numerical Analysis*, 34(2):425 – 454, 1997.

[14] T. Chan, B. Heimsund, T. Nilssen, and X. Tai. Level set methods and augmented lagrangian for a parameter identification problem. *preprint*, 2002.

[15] W. Cheney and D. Kinkaid. *Numerical Analysis: Mathematics of Scientific Computing*. Brooks/Cole Publishing Company, New York, 2nd edition, 1996.

[16] A. Cherkaev. *Variational Methods for Structural Optimization*, volume 140 of *Applied Mathematical Sciences*. Springer, New York, 2000.

[17] P.G. Ciarlet. *The finite element method for elliptic problems*. North-Holland, Amsterdam, 1978.

[18] O. Dorn, E. L. Miller, and C. M. Rappaport. A shape reconstruction method for electromagnetic tomography using adjoint fields and level sets. *Inverse Problems*, 16:1119–1156, 2000.

[19] R. Ewing, R. Falk, and T. Lin. Inverse and ill-posed problems in reservoir simulation. In H.W. Engl and C.W. Groetsch, editors, *Inverse and Ill-Posed Problems*, pages 483–497. Academic Press, Inc., 1987.

[20] R. Ewing, Z. Li, T. Lin, and Y. Lin. The immersed finite volume element method for the elliptic interface problems. *Math. Comput. Simulation*, 50(1-4):63–76, 1999. Modelling '98 (Prague).

[21] R. Ewing, T. Lin, and Y. Lin. A mixed least-squares method for an inverse problem of a nonlinear beam equation. *Inverse Problems*, 15(1):19–32, 1999. Conference on Inverse Problems, Control and Shape Optimization (Carthage, 1998).

[22] R. Ewing, T. Lin, and Y. Lin. On the accuracy of the finite volume element method based on piecewise linear polynomials. *SIAM Journal of Numerical Analysis*, 39(6):1865–1888, 2002.

[23] G. Fletcher. *Introduction to Mathematical Methods in Physics*. Wm. C. Brown Publishers, 1994.

[24] R. Fletcher. *Practical Methods of Optimization*, volume 1 of *Unconstrained Optimization*. John Wiley and Sons, New York, 1980.

[25] P. Gill, W. Murray, and M. Wright. *Practical Optimization*. Academic Press, Inc., New York, 1981.

[26] J. Haslinger, K.-H. Hoffmann, and M. Kočvara. Control/fictitious domain method for solving optimal shape design problems. *RAIRO Modél. Math. Anal. Numér.*, 27(2):157–182, 1993.

[27] J. He, L. Watson, and et al. Dynamic data structures for a direct search algorithm. *Computational Optimization and Applications*, 25, 2002.

[28] V. Isakov. *Inverse Problems for Partial Differential Equations*, volume 127 of *Applied Mathematical Sciences*. Springer, 1998.

[29] K. Ito, K. Kunisch, and Z. Li. Level-set function approach to an inverse interface problem. *Inverse Problems*, 17:1225 – 1242, 2001.

[30] B. Jiang. *The Least-Squares Finite Element Method: Theory and Applications in Computational Fluid Dynamics and Electromagnetics*. Springer, 1998.

[31] C. Johnson. *Numerical Solution of Partial Differential Equations by the Finite Element Method*. Cambridge University Press, 1987.

[32] D. R. Jones. The direct global optimization algorithm. In *Encyclopedia of Optimization, Volume 1*, pages 431–440. Kluwer Academic, 2001.

[33] S. Kesavan. *Topics in Functional Analysis and Applications*. John Wiley and Sons, 1989.

[34] I. Knowles. Descent methods for inverse problems. *Nonlinear Analysis*, 47:3235 – 3245, 2001.

[35] K. Kobayashi, K. Onishi, and Y. Ohura. On identifying dirichlet condition for 2d laplace equation by bem. *Engineering Analysis with Boundary Elements*, 17:223–230, 1996.

[36] S. Kobayashi and N. Nishimura. Determination of cracks having arbitrary shapes with the boundary integral equation method. *Engineering Analysis with Boundary Elements*, 15:189–195, 1994.

[37] M. Křížek and P. Neittaanmäki. *Mathematical and Numerical Modelling in Electrical Engineering: Theory and Applications*, volume 1 of *Mathematical Modelling: Theory and Applications*. Kluwer Academic Publishers, 1996.

[38] K. Kurpisz and A.J. Nowak. *Inverse Thermal Problems*. Computational Mechanics Publications, 1995.

[39] L.P. Lebedev, I.I. Vorovich, and G.M.L. Gladwell. *Functional Analysis: Applications in Mechanics and Inverse Problems*, volume 41 of *Solid Mechanics and Its Applications*. Kluwer Academic Publishers, 1996.

[40] Z. Li. The immersed interface method using a finite element formulation. *Applied Numerical Mathematics*, 27:253–267, 1998.

[41] Z. Li, T. Lin, Y. Lin, and R. Rogers. An immersed finite element space and its approximation capability. *Numerical Methods for Partial Differential Equations*, 2003. (to appear).

[42] Z. Li, T. Lin, and X. Wu. New cartesian grid methods for interface problems using finite element formulation. *Numerische Mathematik*, 96(1):61–98, 2003.

[43] T. Lin. A C^0 finite element method for an inverse problem. *Appl. Math. Comput.*, 90(2-3):253–284, 1998.

[44] T. Lin, Y. Lin, R. Rogers, and L.M. Ryan. A rectangular immersed finite element method for interface problems. In *Advances in Computation: Theory and Practice*, volume Vol. 7, pages 107–114. Nova Science Publishers, Inc., 2001.

[45] T. Lin and E. Ramirez. A numerical method for parameter identification of a boundary value problem. *Applicable Analysis*, 69(3-4):349–379, 1998.

[46] B. D. Lowe and W. Rundell. The determination of a coefficient in an elliptic equation with average flux data. *Journal of Applied and Computational Mathematics*, 70:173–187, 1996.

[47] V.A. Morozov. *Methods for Solving Incorrectly Posed Problems*. Springer-Verlag, 1984.

[48] Y. Ohura, K. Onishi, and Q. Wang. On identifying dirichlet condition for underdetermined problem of the laplace equation by bem. *Engineering Analysis with BEM*, 26:87–93, 2002.

[49] C. D. Perttunen, D. R. Jones, and B. E. Stuckman. Liptschitzian optimization without the lipschitz constant. *Journal of Optimization Theory and Applications*, 79(1):157–181, 1993.

[50] O. Pironneau. *Optimal shape design for elliptic systems*. Springer Series in Computational Physics. Springer-Verlag, New York, 1984.

[51] E. Ramirez. *Finite Element Methods for Parameter Identification Problem of Linear and Nonlinear Steady-state Diffusion Equations*. PhD thesis, Virginia Polytechnic Institute and State University, 1997.

[52] M. Renardy and R. C. Rogers. *An Introduction to Partial Differential Equations*. Springer, 1993.

[53] D. P. Seraphim, R. Lasky, and C. Li. *Principles of Electronic Packaging*. McGraw-Hill Series in Materials Science and Electrical Engineering. McGraw-Hill, 1989.

[54] K. Singh and M. Tanaka. Dual reciprocity boundary element analysis of inverse heat conduction problems. *Computer Methods in Applied Mechanics and Engineering*, 190:5283–5295, 2001.

[55] W. Yeh. Review of parameter identification procedures in groundwater hydrology: the inverse problem. *Water Resource Research*, 22:95–108, 1986.

Vita

Brian David Camp was born in Hartford, Connecticut in 1972. Upon graduating from Simsbury High School in Simsbury, Connecticut in 1990, Brian matriculated to Virginia Polytechnic Institute and State University where he began to study engineering. Armed with a B.S. in Chemical Engineering in 1995, Brian moved to Rhode Island to pursue a career in plastics engineering in the employment of Toray Plastics America, Inc., which specializes in producing thin films. This career, however, was unfulfilling to Brian, so in 1997 he returned to Virginia Polytechnic Institute and State University to study Mathematics so that one day he could teach the subject to others. In 1999 Brian received his M.S. in Mathematics and in 2003 Brian received his Ph.D. in Mathematics.