Efficiency-based *hp*-refinement for finite element methods

by

Lei Tang

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Mathematics
in
Applied Mathematics

Waterloo, Ontario, Canada, 2007

AUTHOR'S DECLARATION FOR ELECTRONIC SUBMISSION OF A THESIS

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public

Abstract

Two efficiency-based grid refinement strategies are investigated for adaptive finite element solution of partial differential equations. In each refinement step, the elements are ordered in terms of decreasing local error, and the optimal fraction of elements to be refined is determined based on efficiency measures that take both error reduction and work into account. The goal is to reach a pre-specified bound on the global error with a minimal amount of work. Two efficiency measures are discussed, work times error and accuracy per computational cost. The resulting refinement strategies are first compared for a one-dimensional model problem that may have a singularity. Modified versions of the efficiency strategies are proposed for the singular case, and the resulting adaptive methods are compared with a threshold-based refinement strategy. Next, the efficiency strategies are applied to the case of hp- refinement for the one-dimensional model problem. The use of the efficiency-based refinement strategies is then explored for problems with spatial dimension greater than one. The work times error strategy is inefficient when the spatial dimension, d, is larger than the finite element order, p, but the accuracy per computational cost strategy provides an efficient refinement mechanism for any combination of d and p.

Acknowledgements

First, I would like to thank my supervisor Hans De Sterck for all of his help and support. It has been a great pleasure to work with you, and I am grateful for all you have taught me. Thank you for introducing me to scientific computation, for your patience and your support through out all my study. Thank you very much!

I would like to thank Josh Nolting, Prof. Tom Manteuffel, Prof. Steve McCormick and Prof. John Ruge in Applied Mathematics, University of Colorado, for answering me questions regarding AMG and FEM. In particular, thank you Prof. T. Manteuffel and Prof. S. McCormick, for your discussion regarding the error convergence of FEM in singular case. Thank you, Prof. J. Ruge, for providing me FOSPACK for 2D FOSLS simulations, and for your patience in answering questions regarding FOSPACK. Thank you, Josh, for implementation of the WEE and ACE strategies in FOSPACK.

I would also like to thank Prof. Lilia Krivodonova and Prof. Stephen Vavasis, for reading this thesis. Your comments are much appreciated.

Thank you to the Applied Mathematics Department for all of their financial support in the past two years, and to the Waterloo Graduate Studies Office for their travel support in my attending the 2007 Copper Mountain Conference on multigrid methods.

Thank you, Prof. David Siegel and Prof. Justin Wan, for teaching me functional analysis and sparse matrices, and especially for all your help in my PhD applications.

Thank you to all my friends in the department who are dear to me. Thank you for your encouragement and kindness.

Lastly and most importantly, thank you, my parents, for all your love and support. I would not be where I am today without you.

Contents

List of Figures

Chapter 1

Introduction

Finite element methods (FEM) are an attractive class of methods for the numerical solution of partial differential equations (PDEs). They are motivated by variational formulations of a PDE boundary value problem in a certain properly chosen solution space U. By partitioning the problem domain Ω into a union of sub-domains, finite dimensional subspaces $U_h \subset U$, are easily constructed. Then, in a general setting, finite element methods seek approximate solutions u_h by solving variational problems in U_h. Finite element methods possess many features that are absent from other numerical PDE methods. For example, in general settings, the finite element approximation u_h is a projection of u into the finite dimensional subspace, satisfying certain minimization properties. This implies that techniques from functional analysis and interpolation theory can be used directly for the analysis of error convergence. Also, finite element methods can deal with irregular domains and unstructured grids. Furthermore, the resulting linear systems are usually easy to solve. In the literature, it is shown that the linear systems of many PDE problems discretized with finite element methods can be solved by multigrid efficiently. Moreover, the most important feature of FEM is that adaptive grid refinement based on an error estimator can be applied to reduce the error in an efficient fashion when a sharp, easily computed local a posteriori error estimator is available.

Adaptive finite element methods are being used extensively as powerful tools for approximating solutions of partial differential equations (PDEs) in a variety of application fields, see, e.g., [1, 2, 3, 12, 18]. Various strategies have been developed for adaptive refinement and for generating mesh sequences with roughly optimal error convergence, i.e., highly accurate mesh sequences. For example, one can consider threshold-based refinement strategies which refine a properly chosen fixed fraction of elements on each refinement level, or strategies which refine elements with local error greater than a fixed fraction of the maximum local error (cf [9, 12]). However, in these approaches the total work to generate the mesh sequence is barely considered. There are examples which show that, if the solution contains highly singular points, threshold-based strategies would only refine the singular elements multiple times. Note that in order to evaluate the a-posteriori error estimator, we generally require solving the linear system after each refinement. Even if optimal linear solvers, such as multigrid, are used, large amount of work may be required. This issue motivates us to develop new strategies by taking into account both work and error reduction, in order to generate a highly accurate mesh sequence efficiently.

This thesis develops and investigates the behaviour of two efficiency-based grid refinement strategies for adaptive finite element solution of PDEs. In each refinement step, the elements are ordered in terms of decreasing local error, and the optimal fraction of elements to be refined in the current step is determined based on efficiency measures that take both error reduction and work into account. The goal is to reach a pre-specified bound on the global error with a minimal amount of work. It is assumed that optimal solvers are used for the discrete linear systems, and that the computational work for solving these systems is, thus, proportional to the number of degrees of freedom (DOF). Two efficiency measures are discussed. The first efficiency measure is 'work times error' efficiency (WEE), which was originally proposed in [11]. A second measure proposed in this thesis is called 'accuracy per computational cost' efficiency (ACE).

This thesis is divided into four parts. First, preliminary background in Sobolev spaces and finite element methods are reviewed in Chapter 2. In Chapter 3, assumptions on adaptive refinements are presented and the WEE and ACE refinement strategies are described. Then, in Chapter 4, we apply WEE and ACE strategies to a 1D model problem using the standard Galerkin finite element method. Numerical results of h- and hp-refinement for both smooth and singular cases are presented. Comparisons with general threshold based strategies are discussed. Finally, in Chapter 5, WEE and ACE are applied for a 2D problem using first-order least-squares (FOSLS) finite element methods.

Chapter 2

Preliminary background

In this chapter, concepts of finite element methods are reviewed. Background material on Sobolev spaces is introduced in section 2.1, along with the notation used throughout this thesis. Then, in section 2.2, properties of finite element methods are briefly are briefly discussed. Error estimate of FEM are especially addressed. This chapter is devoted to provide theoretical background for proposing our refinement strategies in next chapters. Readers who are familiar with Sobolev spaces and FEM may skip this introductory part.

2.1 Function spaces

In this section, we present the basic definitions and properties of Sobolev spaces which are used throughout this thesis. For more details on Sobolev spaces, one can refer to [5].

2.1.1 Domains and boundaries

Let $\Omega \subset \mathcal{R}^d$ be an open, connected bounded domain with boundary $\Gamma = \partial\Omega$. For example, for $d = 1$, $\Omega = (a, b)$ is an open interval and $\Gamma = \{a, b\}$ consisting of the end points.

Definition 2.1 (Lipschitz boundary)**.** *If $d \geq 2$, we say that Γ is a Lipschitz boundary, if there exists a finite open cover $O_1, O_2, ..., O_m$ of Γ such that for $j = 1, 2, ..., m$.*

a) $\Gamma \cup O_j$ is the graph of a Lipschitz function g_j and

b) $\Omega \cup O_j$ is on one side of this graph.

2.1.2 Spaces of continuous functions

Definition 2.2 (Multi-index). *A multi-index, α, is a d-tuple of non-negative integers, α_i. The length of α is given by*

$$|\alpha| = \sum_{i=1}^{d} \alpha_i. \tag{2.1}$$

Then we denote by $D^\alpha u$ the partial derivative

$$\frac{\partial^{|\alpha|} u}{\partial x_1^{\alpha_1} \partial x_2^{\alpha_2} ... \partial x_d^{\alpha_d}}.$$

Let $\Omega \subset \mathcal{R}^d$ be open and bounded. We will use the following spaces of continuous functions.

Definition 2.3.

$$\begin{aligned}
&C(\Omega) = \{u : \Omega \to \mathcal{R} | u \textit{ is continuous}\} \\
&C(\overline{\Omega}) = \{u \in C(\Omega) | u \textit{ is uniformly continuous}\} \\
&C^k(\Omega) = \{u : \Omega \to \mathcal{R} | D^\alpha u \textit{ is continuous for all} |\alpha| \le k\} \\
&C^k(\overline{\Omega}) = \{u \in C^k(\Omega) | D^\alpha u \textit{ is uniformly continuous for all } |\alpha| \le k\} \\
&C_0(\Omega), C_0^k(\Omega) \textit{ denote the sets of functions in } C(\Omega), C^k(\Omega) \textit{ with compact support in } \Omega
\end{aligned} \tag{2.2}$$

Definition 2.4 (Hölder norm). *(i) If $u : \Omega \to \mathcal{R}$ is bounded and continuous, we write*

$$||u||_{C(\overline{\Omega})} := \sup_{x \in \Omega} |u(x)|. \tag{2.3}$$

(ii) The β^{th}-Hölder semi-norm of $u : \Omega \to \mathcal{R}$ is

$$[u]_{C^{0,\beta}(\Omega)} := \sup_{x \ne y \in \Omega} \left\{ \frac{|u(x) - u(y)|}{|x - y|^\beta} \right\}, \tag{2.4}$$

and the β^{th}-Hölder norm is

$$||u||_{C^{0,\beta}(\Omega)} := ||u||_{C(\overline{\Omega})} + [u]_{C^{0,\beta}(\Omega)}. \tag{2.5}$$

Definition 2.5 (Hölder spaces)**.** *The Hölder space*

$$C^{k,\beta}(\Omega)$$

consists of all functions $u \in C^k(\overline{\Omega})$ *for which the norm*

$$||u||_{C^{k,\beta}(\Omega)} := \sum_{|\alpha|\leq k} ||D^\alpha u||_{C(\overline{\Omega})} + \sum_{|\alpha|=k} [D^\alpha u]_{C^{0,\beta}(\Omega)} \tag{2.6}$$

is finite.

As a function space, the Hölder space $C^{k,\beta}(\Omega)$ is a Banach space.

2.1.3 Sobolev spaces

Definition 2.6 (Locally integrable functions)**.** *Given a domain* Ω *(not necessarily open), the set of locally integrable functions is denoted by*

$$L^1_{loc}(\Omega) := \left\{ f \ : \ f \in L^1(K) \qquad \forall \textit{ compact } K \subset \textit{ interior } \Omega \right\}. \tag{2.7}$$

Definition 2.7 (Weak derivative)**.** *We say a function* $f \in L^1_{loc}(\Omega)$ *has a weak derivative,* $D^\alpha_w f$*, provided there exists a function* $g \in L^1_{loc}(\Omega)$ *such that*

$$\int_\Omega g(x)\phi(x)\,dx = (-1)^{|\alpha|} \int_\Omega f(x) D^\alpha \phi(x)\,dx \qquad \forall \phi \in C^\infty_0(\Omega). \tag{2.8}$$

If such g *exists, define* $D^\alpha_w f = g$*.*

A weak derivative of u, $D^\alpha_w u$, if it exists, is uniquely defined up to a set of measure zero. Moreover, if $u \in C^{|\alpha|}(\Omega)$, then the weak derivative $D^\alpha_w u$ is given by the classical derivative $D^\alpha u$ up to a set of measure zero. As a consequence, we will ignore the difference between $D^\alpha_w u$ and $D^\alpha u$ from now on.

Now we can use the notation of weak derivative to define Sobolev spaces.

Definition 2.8 (Sobolev spaces $W^k_p(\Omega)$)**.** *Let* k *be a non-negative integer, and let* $f \in L^1_{loc}(\Omega)$*. Suppose that the weak derivative* $D^\alpha f$ *exist for all* $|\alpha| \leq k$*. Define the Sobolev norm*

$$||f||_{W^k_p(\Omega)} := \left(\sum_{|\alpha|\leq k} \int_\Omega |D^\alpha f|^p \right)^{1/p} \tag{2.9}$$

in the case $1 \leq p < \infty$*, and in the case* $p = \infty$

$$||f||_{W^k_\infty(\Omega)} := \max_{|\alpha| \leq k} ||D^\alpha f||_{L^\infty(\Omega)}. \tag{2.10}$$

In either case, define the Sobolev spaces via

$$W^k_p(\Omega) := \left\{ f \in L^1_{loc}(\Omega) \ : \ ||f||_{W^k_p(\Omega)} < \infty \right\}. \tag{2.11}$$

The Sobolev space $W^k_p(\Omega)$ is a Banach space. Throughout this thesis, mainly the case $k = 2$ will be used. As usual, $W^k_2(\Omega)$ is denote by $H^k(\Omega)$, and we further have

Theorem 2.1. $H^k(\Omega)$ *is a Hilbert space with inner product*

$$(f, g)_{H^k(\Omega)} = \sum_{|\alpha| \leq k} \int_\Omega D^\alpha f \, D^\alpha g \, dx \qquad \forall f, g \in H^k(\Omega). \tag{2.12}$$

Likewise, we define the Sobolev semi-norms.

Definition 2.9 (Sobolev semi-norm)**.** *For* k *a non-negative integer and* $f \in W^k_p(\Omega)$*, let*

$$|f|_{W^k_p(\Omega)} = \left(\sum_{|\alpha| = k} \int_\Omega |D^\alpha f|^p \, dx \right)^{1/p} \tag{2.13}$$

in the case $1 \leq p < \infty$*, and in the case* $p = \infty$

$$|f|_{W^k_\infty(\Omega)} = \max_{|\alpha| = k} ||D^\alpha f||_{L^\infty(\Omega)}. \tag{2.14}$$

The following theorem relates Sobolev spaces to continuous function spaces. It will be used to determine the smoothness of finite element subspaces in the following chapters. For more information on embedding theory in Sobolev spaces, please refer to [5], Chapter 5.

Theorem 2.2 (The Sobolev imbedding theorem)**.** *[5] Let* Ω *be a bounded domain in* $\mathcal{R}^d$ *with Lipschitz boundary. Let* j *and* m *be non-negative integers and let* p *satisfy* $1 \leq p < \infty$*. Suppose* $mp > d > (m-1)p$*. Then*

$$W^{j+m}_p(\Omega) \hookrightarrow C^{j,\beta}(\overline{\Omega}) \qquad 0 < \beta \leq m - \frac{d}{p}. \tag{2.15}$$

Here $W_p^{j+m}(\Omega) \hookrightarrow C^{j,\beta}(\overline{\Omega})$ denotes that $W_p^{j+m}(\Omega)$ can be imbedded in $C^{j,\beta}(\overline{\Omega})$, i.e.,

(i) $W_p^{j+m}(\Omega) \subset C^{j,\beta}(\overline{\Omega})$, and

(ii) there exists a constant C such that

$$||u||_{C^{j,\beta}(\overline{\Omega})} \leq C||u||_{W_p^{j+m}(\Omega)} \qquad \forall u \in W_p^{j+m}(\Omega). \tag{2.16}$$

2.1.4 Traces: Sobolev spaces on boundaries

For a boundary value problem posed in Sobolev spaces, it is nontrivial to interpret the boundary condition. In $\mathcal{R}^1$, the Sobolev inequality (2.16) implies that $H^k(\Omega) \hookrightarrow C^{k+1,\frac{1}{2}}(\overline{\Omega})$. It follows that for any $u \in H^k(\Omega)$, u is well defined at boundary points. However, in higher dimensional spaces, e.g, $d \geq 2$, $H^1(\Omega)$ contains unbounded functions. Thus, we can not interpret boundary conditions in a pointwise sense. The following theorem is used to interpret a restriction of Sobolev-class functions on the boundary.

Theorem 2.3 (Trace Theorem)**.** *[20] Assume Ω is bounded with Lipschitz boundary Γ. Then there exists a bounded linear operator*

$$T : W_p^1(\Omega) \to L^p(\Gamma)$$

such that

$$Tu = u|_\Gamma \text{ if } u \in W_p^1(\Omega) \cap C(\overline{\Omega})$$

and

$$||Tu||_{L^p(\Gamma)} \leq C||u||_{W_p^1(\Omega)},$$

for each $u \in W^{1,p}(\Omega)$, with the constant C depending only on p and Ω.

Definition 2.10. *We call Tu the trace of u on Γ.*

Theorem 2.4. *The range space of $T(H^1(\Omega))$ is a proper and dense subspace of $L^2(\Gamma)$, called $H^{1/2}(\Gamma)$. For any $v \in H^{1/2}(\Gamma)$, define the norm*

$$||v||_{1/2,\Gamma} = \inf_{u \in H^1(\Omega), Tu = v} ||u||_{H^1(\Omega)}. \tag{2.17}$$

Then, $H^{1/2}(\Gamma)$ is a Hilbert space.

When Γ is sufficiently smooth, $H^{m-\frac{1}{2}}(\Omega)$ can be similarly defined as the trace space of $H^m(\Omega)$. For more details, one can refer to [5].

The Sobolev space $H_0^k(\Omega)$ is defined as the completion of $C_0^\infty(\Omega)$ with the norm $||\cdot||_{H^k(\Omega)}$. In the literature, under certain assumptions, it has been shown that $u \in H_0^k(\Omega)$ if and only if $Tu = 0$. Thus, we define

Definition 2.11.

$$H_0^k(\Omega) = \{u \in H^k(\Omega) : u|_\Gamma = 0\},$$

where $u|_\Gamma = 0$ *in the trace sense.*

2.1.5 $H(\mathbf{div})$ and $H(\mathbf{curl})$ spaces

In an addition to the usual Sobolev spaces, the following two spaces are often considered in FOSLS finite element methods, which will be used in Chapter 5. For more details, one can refer to [20].

We define

$$\begin{aligned} H(\text{div};\Omega) &= \{\mathbf{u} \in (L^2(\Omega))^d : \nabla \cdot \mathbf{u} \in L^2(\Omega)\}, \\ H(\text{curl};\Omega) &= \{\mathbf{u} \in (L^2(\Omega))^d : \nabla \times \mathbf{u} \in (L^2(\Omega))^{2d-3}\}, \end{aligned} \tag{2.18}$$

which are Hilbert spaces with norms

$$\begin{aligned} ||\mathbf{u}||^2_{H(\text{div};\Omega)} &= ||\mathbf{u}||^2_{L^2(\Omega)} + ||\nabla \cdot u||^2_{L^2(\Omega)}, \\ ||\mathbf{u}||^2_{H(\text{curl};\Omega)} &= ||\mathbf{u}||^2_{L^2(\Omega)} + ||\nabla \times u||^2_{L^2(\Omega)}. \end{aligned} \tag{2.19}$$

2.2 PDE problem and finite element method

In this section, basic properties of finite element methods are introduced in three parts. First, it is shown how a boundary value problem can be cast into a variational problem using model problems. Then, requirements a variational problem must satisfy in order to be well-posed are presented. Lastly, error estimates are briefly discussed. For more information about FEM, one can refer to [3, 4].

2.2.1 Variational formulation of PDE problem

Let Ω be a bounded domain in $\mathcal{R}^d$, and consider a linear PDE boundary value problem written abstractly as

$$\begin{aligned} Lu &= f \text{ in } \Omega, \\ Bu &= g \text{ on } \partial\Omega. \end{aligned} \tag{2.20}$$

The variational problem is obtained by associating (2.20) with a bilinear form $B(u, v)$ and a linear functional $F(v)$. In a proper variational problem, both $B(u, v)$ and $F(v)$ must be defined on properly chosen normed linear spaces U and V, i.e. $B : U \times V \to \mathcal{R}$, $F : V \to \mathcal{R}$. Then the variational formulation (or weak form) of BVP (2.20) is defined as follows:

$$\text{find } u \in U \text{ such that } B(u, v) = F(v) \qquad \forall v \in V. \tag{2.21}$$

Here V is called the test space and U the trial space.

To derive this associated variational problem, one can consider the Galerkin method, that is, if u solves (2.20), then $(Lu, v)_{L^2(\Omega)} = (f, v)_{L^2(\Omega)}$ for all test functions v in a properly chosen space. By integrating by parts, we may obtain a bilinear form which allows the generalized solution u to be less smooth. We illustrate the procedure by the following model problem.

Consider the boundary value problem

$$\begin{cases} -\nabla \cdot \underline{A}\, \nabla u + \mathbf{b}(x) \cdot \nabla u + c(x)u = f(x) & \text{in } \Omega, \\ u = 0 & \text{on } \partial\Omega, \end{cases} \tag{2.22}$$

for $f \in L^2(\Omega)$, where $\underline{A}$ is a $d \times d$ symmetric matrix with entries in $L^2(\Omega)$, $\mathbf{b}(x) \in (L^2(\Omega))^d$, and $\mathbf{n}$ is the outward unit vector normal to the boundary. Suppose u solves (2.22), then we have

$$(-\nabla \cdot \underline{A}\, \nabla u + \mathbf{b}(x) \cdot \nabla u + c(x)u\, , v)_{L^2(\Omega)} = (f, v)_{L^2(\Omega)} \qquad \forall v \in H^1_0(\Omega). \tag{2.23}$$

Integrating by parts, and noting that $v = 0$ on the boundary, we obtain

$$(\underline{A}\, \nabla u\, , \nabla v)_{(L^2(\Omega))^d} + (\mathbf{b}(x) \cdot \nabla u + c(x)u\, , v)_{L^2(\Omega)} = (f, v)_{L^2(\Omega)} \qquad \forall v \in H^1_0(\Omega). \tag{2.24}$$

It follows that the associated variational problem of the boundary value problem (2.22) is to find $u \in H_0^1(\Omega)$ such that (2.24) holds. Inhomogeneous boundary conditions are easily treated. For example, suppose $u = g$ on $\partial\Omega$. For simplicity, assume that g is defined on all of Ω with $g \in H^1(\Omega)$. Then the variational problem of (2.22) is as follows: find $u \in H^1(\Omega)$ such that $u - g \in H_0^1(\Omega)$ and such that $B(u, v) = (f, v)_{0,\Omega}$, $\forall v \in H_0^1(\Omega)$.

One can also consider a Ritz-Rayleigh method, in which the variational problem is obtained by minimizing a functional over a certain trial space U, e.g., FOSLS formulations, which are discussed in Chapter 4.

Generally, the FEM is a projection method based on a variational formulation of a BVP. It generates an approximate solution in a finite dimensional subspace, U_h, of U. An approximate solution u_h of the following form is sought:

$$u_h(x) = \sum_{i=1}^{N} c_i \psi_i(x), \tag{2.25}$$

where $\psi_i(x) \in U$, $i = 1, 2, ...N$ are N linearly independent functions in U, and a_i are real numbers. The set U_h of all functions u_h of the form (2.25) is a linear space of dimension N contained in U. Let $V_h \subset V$ also be a subspace of dimension N. Then the discrete weak form of (2.21) is defined by

$$\text{find } u_h \in U_h \text{ such that } B(u_h, v) = F(v) \qquad \forall v \in V_h. \tag{2.26}$$

This problem is equivalent to solving an $N \times N$ linear system. We can write

$$u_h = \sum_{i=1}^{N} c_i \psi_i,$$

and

$$v = \sum_{i=1}^{N} a_i \phi_i.$$

Hence

$$\sum_{i=1}^{N} \sum_{j=1}^{N} a_i B(\psi_j, \phi_i) c_j = \sum_{i=1}^{N} a_i F(\phi_i) \qquad \forall \mathbf{a} = (a_1, a_2, ..., a_N)^T \in \mathcal{R}^N.$$

Therefore

$$\underline{K}\mathbf{c} = \mathbf{q}, \tag{2.27}$$

where the stiffness matrix $\underline{K}$ is an $N \times N$ matrix with entries $k_{ij} = B(\psi_j, \phi_i)$, and $q_i = F(\phi_i)$. In order for (2.27) to have a unique solution, we require $\underline{K}$ to be nonsingular. This implies that the bilinear form $B(\cdot, \cdot)$ must satisfy certain conditions, which are discussed in the next section.

2.2.2 Well-posed variational problems

We are seeking conditions that a variational problem must satisfy in order to be well-posed, i.e., there exists a unique solution solving the variational problem. A minimal requirement on $F(v)$ is continuity: there exists a constant $C_F > 0$ such that, for all $v \in V$,

$$|F(v)| \leq C_F ||v||_V. \tag{2.28}$$

If (2.28) holds, $F(\cdot)$ is called a continuous linear functional on V. Denote by V' the set of all such functionals, with norm given by

$$||F||_{V'} = \sup_{v \neq 0} \frac{|F(v)|}{||v||_V}. \tag{2.29}$$

It follows that V' is a Banach space, and it is called the dual space of V, see [7]. As usual, we use the notation $< F, v >= F(v)$ for all $F \in V'$, $v \in V$.

We define continuity and coercivity of a bilinear form $B(\cdot, \cdot)$ as follows:

Definition 2.12 (Continuity and coercivity)**.** *A bilinear form $B(\cdot, \cdot)$ on a normed linear space X is said to be continuous if $\exists$ $C_1 > 0$ such that*

$$|B(u, v)| \leq C_1 ||u||_X ||v||_X \qquad \forall u, v \in X, \tag{2.30}$$

and coercive on $V \subset X$ if $\exists$ $C_2 > 0$ such that

$$B(v, v) \geq C_2 ||v||_X^2 \qquad \forall v \in V. \tag{2.31}$$

First, consider the well-posedness of variational problems in a Hilbert space. Let H be a Hilbert space with inner product $(\cdot, \cdot)$. For any continuous functional F on a Hilbert space, we have the following well-known theorem:

Theorem 2.5 (Riesz Representation Theorem)**.** *Any continuous linear functional F on a Hilbert space H can be represented uniquely as*

$$F(v) = (u, v) \tag{2.32}$$

for some $u \in H$. Furthermore, we have

$$||F||_{H'} = ||u||_H. \tag{2.33}$$

One can refer to [4] for a proof.

Assume that the bilinear form $B(\cdot, \cdot)$ is symmetric on H, i.e., $B(u, v) = B(v, u)$ for all $u, v \in H$. If $U = V$, a closed subspace of H, and the bilinear form $B(\cdot, \cdot)$ is continuous and coercive, then $(V, B(\cdot, \cdot))$ is also a Hilbert space. Applying the Riesz Representation Theorem, and noting that $F \in V'$, there exists a unique solution $u \in V$ solving (2.21).

In general, suppose that the following three conditions are valid.

$$\begin{cases} (1) & (H, (\cdot, \cdot)) \text{ is a Hilbert space.} \\ (2) & V \text{ is a closed subspace of H.} \\ (3) & B(\cdot, \cdot) \text{ is a continuous, symmetric bilinear form that is coercive on V.} \end{cases} \tag{2.34}$$

Then the variational problem:

$$\text{given } F \in V', \text{ find } u \in V \text{ such that } B(u, v) = F(v) \qquad \forall v \in V, \tag{2.35}$$

is well-posed. Likewise, under conditions (2.34), the approximation problem:

$$\begin{aligned} &\text{given a finite-dimensional subspace } V_h \subset V \text{ and } F \in V', \\ &\text{find } u_h \in V_h \text{ such that} \\ &\qquad\qquad B(u_h, v) = F(v) \qquad \forall v \in V_h, \end{aligned} \tag{2.36}$$

is also well-posed.

The bilinear form is not always symmetric, e.g., the bilinear form (2.24) of the model problem (2.22) is not symmetric on $H_0^1(\Omega)$. For a nonsymmetric bilinear form on a Hilbert space, we have the following theorem:

Theorem 2.6 (Lax-Milgram)**.** *Given a Hilbert space $(V,(\cdot,\cdot))$, a continuous, coercive bilinear form $B(\cdot,\cdot)$ and a continuous functional $F \in V'$, there exists a unique $u \in V$ such that*

$$B(u,v) = F(v) \qquad \forall v \in V.$$

Proof. **Step 1**: Let $u \in V$ be arbitrary, fixed. Then the continuity of $B(\cdot,\cdot)$ implies that

$$T_u(v) := B(u,v) \in V'.$$

Applying the Riesz Representation Theorem, there exists a unique $u^* \in V$ such that $(u^*,v) = T_u(v) = B(u,v)$ for all $v \in V$. Let $u^* = \Phi(u)$. Then $\Phi : V \to V$ is linear:

$$\begin{aligned}(\Phi(a_1u_1 + a_2u_2), v) &= B(a_1u_1 + a_2u_2, v)\\ &= a_1B(u_1,v) + a_2B(u_2,v)\\ &= a_1(\Phi(u_1),v) + a_2(\Phi(u_2),v)\\ &= (a_1\Phi(u_1) + a_2\Phi(u_2), v) \qquad \forall v \in V.\end{aligned}$$

Moreover, Φ is continuous and bounded below:

$$||\Phi(u)||_V^2 = (\Phi(u),\Phi(u)) = B(u,\Phi(u)) \le C_1||\Phi(u)||_V||u||_V \Rightarrow ||\Phi(u)||_V \le C_1||u||_V,$$

and

$$C_2||u||_V^2 \le B(u,u) = (u,\Phi(u)) \le ||u||_V||\Phi(u)||_V \Rightarrow C_2||u||_V \le ||\Phi(u)||_V.$$

Step 2: We show that the range of Φ, $\mathcal{R}(\Phi)$, is a closed subspace of V.

Obviously, $\mathcal{R}(\Phi)$ is a subspace of V since Φ is a linear operator defined from V to V. We claim that $\mathcal{R}(\Phi)$ is closed in V. To see this, let $\{\Phi(u_n)\}_{n=1}^{\infty}$ be a Cauchy sequence in V, i.e.,

$$\exists\ z \in V \text{ s.t. } \lim_{n\to\infty} ||\Phi(u_n) - z||_V = 0. \tag{2.37}$$

From step 1, we have

$$||\Phi(u_n) - \Phi(u_m)||_V = ||\Phi(u_n - u_m)||_V \ge C_1||u_n - u_m||_V.$$

It follows that $\{u_n\}_{n=1}^{\infty}$ is also a Cauchy sequence in V, i.e.,

$$\exists\ u \in V \text{ s.t. } \lim_{n\to\infty} ||u_n - u||_V = 0.$$

This implies that

$$\lim_{n\to\infty} ||\Phi(u_n) - \Phi(u)||_V = \lim_{n\to\infty} ||\Phi(u_n - u)||_V \leq \lim_{n\to\infty} C_2||u_n - u||_V = 0.$$

It follows that $z = \Phi(u) \in \mathcal{R}(\Phi)$. Thus, $\mathcal{R}(\Phi)$ is closed.

Step 3: We show that $\mathcal{R}(\Phi) = V$.

Since $\mathcal{R}(\Phi)$ is a closed subspace of V, we have $V = \mathcal{R}(\Phi) \oplus \mathcal{R}(\Phi)^{\perp}$. If $\mathcal{R}(\Phi) \neq V$, then $\mathcal{R}(\Phi)^{\perp} \neq \{0\}$. For any $0 \neq v \in \mathcal{R}(\Phi)^{\perp}$, $(\Phi(u), v) = 0$ holds for all $u \in V$. Putting $u = v$, we have $C_2||v||^2 \leq B(v, v) = (\Phi(v), v) = 0$. Therefore $v = 0$, a contradiction.

Step 4: From step 1, Φ is bounded below, i.e., $||\Phi(u)||_V \geq C_2||u||_V$. It has a bounded inverse $\Phi^{-1} : V \to V$. For any given functional $F \in V'$, by the Riesz Representation Theorem, there exists a unique $p \in V$ s.t. $F(v) = (v, p)$ for all $v \in V$. Let $u = \Phi^{-1}(p)$. Then u is the unique solution such that $B(u, v) = F(v)$ for all $v \in V$.

□

In general, suppose that the following five conditions are valid.

$$\begin{cases} (1) \quad (H, (\cdot,\cdot)) \text{ is a Hilbert space.} \\ (2) \quad V \text{ is a closed subspace of } H. \\ (3) \quad B(\cdot,\cdot) \text{ is a bilinear form on } V, \text{ not necessarily symmetric.} \\ (4) \quad B(\cdot,\cdot) \text{ is continuous on V.} \\ (5) \quad B(\cdot,\cdot) \text{ is coercive on V.} \end{cases} \tag{2.38}$$

Applying the Lax-Milgram Theorem, the variational problem (2.35) and the approximation problem (2.36) are well-posed.

Next, consider that the trial space U and the test space V are not the same space, e.g., the mixed variational formulation of the Stokes equation for steady flow of a viscous fluid in $\mathcal{R}^d$ results in bilinear forms with $U = H(\text{div})$ and $V = \left(L^2(\Omega)\right)^d$. The following theorem addresses the existence and uniqueness in this case by assuming that U and V are reflexive Banach spaces. A Banach space X is called reflexive if its second dual X'' equals X. For more information about the second dual of a Banach space, one can refer to [7]. It is obvious that every Hilbert space is reflexive.

Theorem 2.7 (inf-sup condition). *[9] Let U, V be real reflexive Banach spaces with norms $||\cdot||_U$ and $||\cdot||_V$, respectively. Let further $B(\cdot,\cdot): U \times V \to \mathcal{R}$ be a bilinear form such that there exist $0 < C_1$, $C_2 < \infty$ with*

$$\begin{aligned} &(1) \qquad |B(u,v)| \le C_1 ||u||_U ||v||_V \qquad \forall u \in U, v \in V, \\ &(2) \qquad \inf_{0\neq u\in U} \sup_{0\neq v\in V} \frac{|B(u,v)|}{||u||_U\,||v||_V} \ge C_2, \\ &(3) \qquad \sup_{u\in U} B(u,v) > 0 \qquad \forall 0 \neq v \in V. \end{aligned} \tag{2.39}$$

Then for every continuous linear functional $F \in V'$ there exists a unique $u_0 \in U$ such that

$$B(u_0, v) = F(v) \qquad \forall v \in V.$$

Proof. **Step 1**: For any arbitrary, fixed $u \in U$, define $T_u(v) := B(u,v)$. Then $T_u \in V'$. By the definition of V', there exists a unique $z \in V'$ such that $< z, v >= T_u(v) = B(u,v)$ for all $v \in V'$.

Denote by $z = \Phi(u)$. Then $\Phi : u \to z$ is continuous and linear:

$$\begin{aligned} \Phi : U \to V',\ ||\Phi||_{U\to V'} &= \sup_{0\neq u\in U} \frac{||\Phi(u)||_{V'}}{||u||_U} \\ &= \sup_{0\neq u\in U} \frac{\sup_{0\neq v\in V} \frac{|<\Phi(u),v>|}{||v||_V}}{||u||_U} \\ &= \sup_{0\neq u\in U} \frac{\sup_{0\neq v\in V} \frac{|B(u,v)|}{||v||_V}}{||u||_U} \\ &\le \sup_{0\neq u\in U} \frac{\sup_{0\neq v\in V} \frac{C_1||u||_U||v||_V}{||v||_V}}{||u||_U} \\ &= C_1. \end{aligned}$$

Step 2: We show that the range of Φ, $\mathcal{R}(\Phi)$ is closed in V'. Condition (2.39) gives

$$||\Phi(u)||_{V'} = \sup_{0\neq v\in V} \frac{| < \Phi(u), v > |}{||v||_V} = \sup_{0\neq v\in V} \frac{|B(u,v)|}{||v||_V} \ge C_2 ||u||_U.$$

Let $\{u_n\}_{n=1}^{\infty} \subset U$ such that $\{\Phi(u_n)\}_{n=1}^{\infty}$ is a Cauchy sequence in V'. Then $\{u_n\}_{n=1}^{\infty}$ is a Cauchy sequence in U. Hence $\mathcal{R}(\Phi)$ is closed in V'. Also, it is easy to see that the null space of Φ, $\mathcal{N}(\Phi) = \Phi^{-1}(\{0\})$ is trivial, i.e. $\mathcal{N}(\Phi) = \{0\}$. This implies that Φ is injective.

Step 3: We prove that $\mathcal{R}(\Phi) = V'$.

Assuming not, then $\mathcal{R}(\Phi) = \overline{\mathcal{R}(\Phi)}^{||\cdot||_{V'}} \neq V'$. Taking $0 \neq \psi \in \mathcal{R}(\Phi)^{\perp}$, by the Hahn-Banach theorem, see [7], there exists $v_0 \neq 0 \in V''$ such that $< v_0, \phi >= 0$ for all $\phi = \Phi(u) \in \mathcal{R}(\Phi)$ but $< v_0, \psi >= 1$. Since V is reflexive, $v_0 \in V$ and hence

$$0 =< \phi, v_0 >=< \Phi(u), v_0 >= B(u, v_0) \qquad \forall u \in U,$$

a contradiction to (2.39) (3). Hence $\mathcal{R}(\Phi) = V'$.

Step 4: By steps 2 and 3, Φ is a bijection; thus for any $F \in V'$, there exists a unique $u_0 \in U$ s.t. $\Phi(u_0) = F$. That is $F(v) =< \Phi(u_0), v >= B(u_0, v)$ for all $v \in V$.

□

Likewise, conditions for the well-posedness of an approximation problem are as follows:

Theorem 2.8 (discrete inf-sup condition)**.** *Let $B : U \times V \rightarrow \mathcal{R}$ be continuous, i.e.,*

$$|B(u, v)| \leq C_1||u||_U||v||_V \qquad \forall u \in U, v \in V.$$

Let $U_h \subset U$, $V_h \subset V$ be subspaces of dimension N such that the inf-sup conditions hold, i.e.,

$$\inf_{0\neq u\in U_h} \sup_{0\neq v\in V_h} \frac{|B(u, v)|}{||u||_U\,||v||_V} \geq C_2(U_h, V_h) > 0. \tag{2.40}$$

Then for every continuous linear functional $F \in V'$ there exists a unique $u_h \in U_h$ such that

$$B(u_h, v) = F(v) \qquad \forall v \in V_h.$$

Proof. Both U_h and V_h are reflexive Banach spaces since they are finite dimensional. Again, define $\Phi : U_h \rightarrow V_h'$ such that $< \Phi(u), v >= B(u, v)$ for all $v \in V_h$. Then Φ is continuous, linear and injective. The dual space, V_h', of V_h, has the same dimension as V_h since V_h is finite dimensional. Then U_h and V_h' are of dimension N, and the injection $\Phi : U_h \rightarrow V_h'$ is also surjective. This completes the proof. □

2.2.3 Error estimates for finite element methods

Let u be the solution to the variational problem and u_h be the solution to the approximation problem. Assume that the error $e = u - u_h$ is evaluated by $||u - u_h||_U$. We first consider the symmetric variational problem in a Hilbert space with $U = V$. Suppose that conditions (2.34) hold. Let u and u_h be solutions to problem (2.35) and (2.36), respectively. Then $u - u_h$ is $B(\cdot, \cdot)$-orthogonal to the space V_h, i.e.,

$$B(u - u_h, v) = 0 \qquad \forall v \in V_h. \tag{2.41}$$

Define $||v||_B = B(v, v)$ for all $v \in V$. We have the following minimization property:

$$||u - u_h||_B = \min_{v \in V_h} ||u - v||_B. \tag{2.42}$$

Next, consider the nonsymmetric problem in a Hilbert space with $U = V$. The following theorem holds.

Theorem 2.9 (Céa). *Suppose that conditions (2.38) hold and that u solves (2.35). For the finite element approximation problem (2.36) we have*

$$||u - u_h||_V \leq \frac{C_1}{C_2} \min_{v \in V_h} ||u - v||_V, \tag{2.43}$$

where C_1 is the continuity constant and C_2 is the coercivity constant of $B(\cdot, \cdot)$ on V.

Proof. Since $B(u, v) = F(v)$ and $B(u_h, v) = F(v)$ for all $v \in V_h$, we have

$$B(u - u_h, v) = 0 \qquad \forall v \in V_h.$$

For all $v \in V_h$,

$$\begin{aligned} C_2 ||u - u_h||_V^2 &\leq B(u - u_h, u - u_h) \\ &= B(u - u_h, u - v) + B(u - u_h, v - u_h) \\ &= B(u - u_h, u - v) \qquad (\text{ since } v - u_h \in V_h) \\ &\leq C_1 ||u - u_h||_V ||u - v||_V. \end{aligned}$$

Hence,

$$||u - u_h||_V \leq \frac{C_1}{C_2} ||u - v||_V \qquad \forall v \in V_h.$$

Therefore

$$\begin{aligned} ||u - u_h||_V &\leq \inf_{v \in V_h} \frac{C_1}{C_2} ||u - v||_V \\ &= \frac{C_1}{C_1} \min_{v \in V_h} ||u - v||_V. \qquad (\text{since } V_h \text{ is closed}) \end{aligned}$$

□

At last, we address the case where $U \neq V$.

Theorem 2.10. *[3, 9] Suppose that the conditions in theorem 2.3 and theorem 2.4 hold. Let u and u_h be the solution of the variational problem and the approximation problem, respectively. We have*

$$||u - u_h||_U \leq \left(1 + \frac{C_1}{C_2(U_h, V_h)}\right) \min_{w \in U_h} ||u - w||_U. \tag{2.44}$$

Proof. By triangle inequality, we have

$$\begin{aligned} ||u - u_h||_U &= ||u - w + w - u_h||_U \\ &\leq ||u - w||_U + ||u_h - w||_U \qquad \forall w \in U_h. \end{aligned}$$

Recall that u_h satisfies

$$B(u, v) = B(u_h, v) = F(v) \qquad \forall v \in V_h.$$

By discrete inf-sup condition (2.40) on $U_h \times V_h$ and the continuity condition on $U \times V$, we have

$$\begin{aligned} C_2(U_h, V_h)||u_h - w||_U &\leq \sup_{0 \neq v \in V_h} \frac{B(u_h - w, v)}{||v||_V} \\ &= \sup_{0 \neq v \in V_h} \frac{B(u - w, v)}{||v||_V} \qquad \forall v \in V_h \\ &\leq \sup_{0 \neq v \in V_h} \frac{C_1||u - w||_U||v||_V}{||v||_V} \\ &= C_1||u - w||_U. \end{aligned}$$

Hence

$$||u - u_h||_U \leq \left(1 + \frac{C_1}{C_2(U_h, V_h)}\right) ||u - w||_U \qquad \forall w \in U_h.$$

Therefore

$$||u - u_h||_U \leq \left(1 + \frac{C_1}{C_2(U_h, V_h)}\right) \min_{w \in U_h} ||u - w||_U. \qquad \text{(since } U_h \text{ is closed)}$$

□

Denote by m a nonnegative integer. For many boundary value problems, the trial space U is usually chosen to be a subspace of the Sobolev space $H^m(\Omega)$. Moreover, the finite dimensional trial space U_h usually consists of piecewise polynomials defined on a partition of Ω. To give a qualitative analysis of the error $||u-u_h||_{H^m(\Omega)}$, we need polynomial approximation theory in Sobolev spaces. First of all, some definitions (cf. [4]) are presented:

Definition 2.13 (Finite Element)**.** *Let*

(i) $K \subset \mathcal{R}^d$ be a domain with piecewise smooth boundary (the element domain),

(ii) $\mathcal{P}$ be a finite-dimensional space of functions on K (the shape functions) and,

(iii) $\mathcal{N} = \{N_1, N_2, ..., N_k\}$ be a basis for $\mathcal{P}'$ (the nodal variables).

Then $(K, \mathcal{P}, \mathcal{N})$ is called a finite element.

Definition 2.14 (Nodal basis)**.** *Let $(K, \mathcal{P}, \mathcal{N})$ be a finite element, and let $\{\psi_1, \psi_2, ..., \psi_k\}$ be a basis for $\mathcal{P}$ dual to $\mathcal{N}$ $(N_i(\psi_j) = \delta_{ij})$. This set is called the nodal basis for $\mathcal{P}$.*

Example 2.1 (the 1D Lagrange element)**.** *Let $K = [a, b]$ and $\mathcal{P}_k =$the set of all polynomials on K of degree less or equal than k. Let $\mathcal{N} = \{N_0, N_1, ..., N_k\}$, where $N_i(v) = v(a + \frac{(b-a)i}{k})$ for all $v \in \mathcal{P}_k$ and $i = 0, 1, ..., k$. Then $\{K, \mathcal{P}_k, \mathcal{N}\}$ is a finite element.*

Example 2.2 (the 2D Linear Lagrange triangular element)**.** *Let K be a triangle in $\mathcal{R}^2$ with three vertices z_1, z_2 and z_3. Let $\mathcal{P} = \mathcal{P}_1$. Let $\mathcal{N} = \{N_1, N_2, N_3\}$, where $N_i(v) = v(z_i)$ for all $v \in \mathcal{P}_1$ and $i = 1, 2, 3$. Then $\{K, \mathcal{P}_1, \mathcal{N}\}$ is a finite element.*

Example 2.3 (the 2D bilinear Lagrange rectangular element)**.** *In $\mathcal{R}^2$, define*

$$\mathcal{Q}_k := \left\{ \sum_j p_j(x) q_j(y) \; : \; p_j, \; q_j \textit{ are polynomials of degree } \leq k \right\}.$$

Let K be any rectangle with vertices z_1, z_2, z_3 and z_4. Let $\mathcal{N} = \{N_i, i = 1, 2, 3, 4\}$, where $N_i(v) = v(z_i)$ for all $v \in \mathcal{Q}_1$ and $i = 1, 2, 3, 4$. Then $\{K, \mathcal{Q}_1, \mathcal{N}\}$ forms a finite element.

After defining the finite element, we need to partition the domain Ω into a mesh $\mathcal{T}$ in order to construct the finite element subspace. The mesh $\mathcal{T}$ is defined as:

Definition 2.15 (Mesh)**.** *A mesh (or subdivision) $\mathcal{T}$ of a domain Ω is a finite collection of open sets $\{K_i\}$ such that*

(1) *$K_i \cap K_j = \emptyset$ if $i \neq j$ and*

(2) $\bigcup \overline{K}_i = \overline{\Omega}$.

Assume that each element K in the mesh is equipped with some type of shape functions, $\mathcal{P}$, and nodal variables, $\mathcal{N}$, such that $(K, \mathcal{P}, \mathcal{N})$ forms a finite element. Then a finite element subspace U_h can be easily constructed. Usually, shape functions, $\mathcal{P}$, consist of polynomials. Then the finite element subspace U_h consists of piecewise polynomials. Next, we define the local interpolant and global interpolant of a given function.

Definition 2.16 (Interpolant)**.** *Given a finite element $(K, \mathcal{P}, \mathcal{N})$, let the set $\{\psi_1, \psi_2, ..., \psi_n\} \subseteq \mathcal{P}$ be the basis dual to $\mathcal{N}$. If v is a function for which all $N_i \in \mathcal{N}$, $i = 1, 2, ..., n$, are defined, then define the local interpolant by*

$$\mathcal{I}_K v := \sum_{i=1}^{n} N_i(v)\psi_i. \tag{2.45}$$

Definition 2.17 (Global interpolant)**.** *Suppose Ω is a domain with a mesh $\mathcal{T}$. Assume each element K in the mesh is equipped with some type of shape functions, $\mathcal{P}$, and nodal variables, $\mathcal{N}$, such that $(K, \mathcal{P}, \mathcal{N})$ forms a finite element. Let m be the order of highest partial derivative involved in the nodal variables. For $f \in C^m(\overline{\Omega})$, the global interpolant is defined by*

$$\mathcal{I}_\mathcal{T} f|_{K_i} = \mathcal{I}_{K_i} f \tag{2.46}$$

for all $K_i \in \mathcal{T}$.

Definition 2.18 (C^ℓ finite element space)**.** *We say that an interpolant has continuity order m if $\mathcal{I}_\mathcal{T} f \in C^\ell$. The space, $V_\mathcal{T} = \{\mathcal{I}_\mathcal{T} f\}$, is said to be a C^ℓ finite element space.*

After presenting these definitions, we can qualitatively give an error bound for $||u - \mathcal{I}_\mathcal{T} u||_U$. Then an error bound for $||u - u_h||_U$ is automatically obtained since $||u - u_h||_U \leq C||u - \mathcal{I}_\mathcal{T} u||_U$ by using the Céa Theorem or Theorem 2.6, respectively. We proceed by the following steps.

Step 1. Approximation property of local interpolant

For any bounded region $K \subset \mathcal{R}^d$, let $\hat{K} = \{(1/\text{diam}(K))x : x \in K\}$, where $\text{diam}(K)$ is the diameter of K. For simplicity, here we assume that K is convex and the origin lies in K such that $\text{diam}(\hat{K}) = 1$. Again, let $\mathcal{P}_k$ be the set of polynomials in d variables of degree less than or equal to k. For any given function $v \in H^m(K)$, the local interpolant $\mathcal{I}_K v$ has the following approximate property.

Theorem 2.11. *Let $(K, \mathcal{P}, \mathcal{N})$ be a finite element satisfying*

(1) K is a convex polygonal domain,

(2) $\mathcal{P}_{m-1} \subseteq \mathcal{P} \subseteq W^m_\infty(K)$ and

(3) $\mathcal{N} \subseteq (C^\ell(\overline{K}))'$.

Suppose $m - \ell - d/2 > 0$. Then for $0 \leq i \leq m$ and $v \in H^m(K)$ we have

$$|v - \mathcal{I}_K v|_{H^i(K)} \leq C_{m,d,\gamma(K),\sigma(\hat{K})}(diam(K))^{m-i}|v|_{H^m(K)}. \tag{2.47}$$

Here the constant $C_{m,d,\gamma(K),\sigma(\hat{K})}$ depends only on $m, d, \gamma(K)$ and $\sigma(\hat{K})$, where $\gamma(K)$ is given by

$$\gamma(K) = \frac{diam(K)}{\sup\{\rho : \text{ a ball of radius } \rho \text{ is contained in } K\}},$$

and $\sigma\{\hat{K}\}$ is the operator norm of $\mathcal{I}_{\hat{K}} : C^\ell\left(\overline{\hat{K}}\right) \to H^m(\hat{K})$, which is bounded.

Proof. It can be shown (cf theorem 4.4.4 in [4]) that a similar approximation property holds for more general case, in which the element K is a bounded polygonal domain, not necessarily convex, star-shaped w.r.t. some ball, and the interpolation error is evaluated by the W^i_p semi-norm for $1 \leq p \leq \infty$. Note that $H^i = W^i_2$, and convex domain K is always star-shaped w.r.t. any ball contained in K. Therefore, (2.47) holds. □

Remark The constant C increases w.r.t. $\gamma(K)$. Therefore, small $\gamma(K)$ is preferred.

Step 2. Approximation property of global interpolant

Consider polygonal domain Ω and mesh $\mathcal{T} = \cup T$ with given shape functions and nodal variable duals. The approximation property of the global interpolant can be derived by taking the sum of local interpolants and using Theorem 2.7. However, some conditions must be satisfied in order to obtain a uniform bound for $C_{m,d,\gamma(K),\sigma(\hat{T})}$ for all $T \in \mathcal{T}$. To see this, let $(K, \mathcal{P}, \mathcal{N})$ be a reference element, satisfying the conditions in Theorem 2.7. First, we define affine equivalence between finite elements.

Definition 2.19 (Affine equivalence)**.** *Let $(K, \mathcal{P}, \mathcal{N})$ be a finite element and let $F(x) = \underline{A}x + b$ (with $\underline{A}$ nonsingular) be an affine map. The finite element $(\widehat{K}, \widehat{\mathcal{P}}, \widehat{\mathcal{N}})$ is affine equivalent to $(K, \mathcal{P}, \mathcal{N})$ if*

(i) $F(K) = \widehat{K}$,

(ii) $F^* \widehat{\mathcal{P}} = \mathcal{P}$,

(iii) $F_* \mathcal{N} = \widehat{\mathcal{N}}$,
where F^* *is defined by* $F^*(\hat{f}) := \hat{f} \circ F$ *and* F_* *is defined by* $(F_* N)(\hat{f}) := N(F^*(\hat{f}))$.

For adaptive refinement, a uniform bound for $C_{m,d,\gamma(K),\sigma(\hat{T})}$ needs to be derived for a family of meshes. This requires that the family of meshes must be nondegenerate.

Definition 2.20. *Let* Ω *be a given domain and mesh* $\mathcal{T} = \cup T$.

(1) Define the meshwidth

$$h(\mathcal{T}) = \max_{T \in \mathcal{T}} diam(T).$$

For each element T, *define*

$$\rho(T) = \sup\{\rho : \text{ a ball of radius } \rho \text{ is contained in } T\}.$$

(2) A family of meshes $\{\mathcal{T}\}_{j=1}^{\infty}$ *is said to be nondegenerate if there exists a constant* κ, *independent of* j, *such that*

$$\sup_{T \in \mathcal{T}_j} \frac{diam(T)}{\rho(T)} \leq \kappa < \infty, \qquad j = 1, 2, ... \tag{2.48}$$

Now suppose that all element are affine equivalent to the reference element K. And let the family of meshes be nondegenerate. Then it can be shown that the constant in Theorem 2.7 can be uniformly bounded, see ([4]). At last, we obtain the following Theorem for the global interpolant.

Theorem 2.12. *[4] Let* $\{\mathcal{T}\}_{j=1}^{\infty}$ *be a nondegenerate family of meshes of a polyhedral domain* Ω *in* $\mathcal{R}^d$. *Let* $(K, \mathcal{P}, \mathcal{N})$ *be a reference element, satisfying the conditions of Theorem 2.7 for some* ℓ *and* m. *For all* $T \in \mathcal{T}_j$, $j = 1, 2, ...$ *let* $(T, \mathcal{P}_T, \mathcal{N}_T)$ *be the affine-equivalent element. Then there exists a positive constant* C *depending on the reference element,* d, m *and the number* κ *in (2.48) such that for* $0 \leq i \leq m$,

$$||v - \mathcal{I}_{\mathcal{T}_j} v||_{H^i(\Omega)} \leq C \left(h(\mathcal{T}_j)\right)^{m-i} |v|_{H^m(\Omega)} \qquad j = 1, 2, ... \tag{2.49}$$

Proof. To prove (2.49), observe that for all $j = 1, 2, \ldots$

$$\begin{aligned}\sum_{T\in\mathcal{T}_j} ||v - \mathcal{I}_{\mathcal{T}_j} v||^2_{H^i(T)} &\leq \sum_{T\in\mathcal{T}_j} C^2_{m,d,\gamma(T),\sigma(\hat{T})} (\mathrm{diam}(T))^{2(m-i)} |v|^2_{H^m(T)} \\ &\leq \sum_{T\in\mathcal{T}_j} C^2 (h(\mathcal{T}_j))^{2(m-i)} |v|^2_{H^m(T)} \\ &\leq C^2 (h(\mathcal{T}_j))^{2(m-i)} \sum_{T\in\mathcal{T}_j} |v|^2_{H^m(T)} \\ &= C^2 (h(\mathcal{T}_j))^{2(m-i)} |v|^2_{H^m(\Omega)},\end{aligned}$$

where C is the uniform bound of $C_{m,d,\gamma(T),\sigma(\hat{T})}$ for all elements T, which only depends on the reference element, d, m and the number κ in (2.48). □

Lastly, suppose the variational problem (2.21) is well posed on U and V, with the trial space, U, a subspace of Sobolev space $H^i(\Omega)$. Let the family of meshes $\{\mathcal{T}_j\}_{j=1}^{\infty}$ satisfy the conditions of Theorem 2.8. Let the discrete variational problem also be well-posed. Moreover, assume that $u \in H^m(\Omega)$. Then by using Theorem 2.5 or Theorem 2.6, respectively, and Theorem 2.8, we have

$$||u - u_h||_{H^i(\Omega)} \leq C(h(\mathcal{T}_j))^{m-i} ||u||_{H^m(\Omega)} \qquad j = 1, 2..., \; i \leq m \tag{2.50}$$

where the constant C depends on the reference element, d, m, the number κ in (2.48), the continuity constant and the coercivity constant.

To end this section, consider that the solution u is not smooth enough, i.e., $u \in H^s(\Omega)$ where s is not necessarily an integer. We briefly discuss how to extend the error estimate (2.50) to fractional order Sobolev spaces by using operator-interpolation theory. For θ a real number in the interval $(0, 1)$, and i a non-negative integer, define the fractional order Sobolev norm:

$$||u||^2_{H^{i+\theta}(\Omega)} := ||u||^2_{H^i(\Omega)} + \sum_{|\alpha|=i} \int_\Omega \int_\Omega \frac{|D^\alpha(x) - D^\alpha(y)|^2}{|x-y|^{d+2\theta}}. \tag{2.51}$$

We first define Banach spaces that interpolate between two given Banach spaces, B_0 and B_1. Then we show that the fractional order Sobolev space can be defined as the interpolated space between two integer order Sobolev spaces, and give the equivalence between the interpolated operator norm and the norm defined in (2.51). Lastly, finite element convergence estimates are presented for fractional order Sobolev spaces.

Given two Banach spaces, B_0 and B_1. For simplicity, assume that $B_1 \subset B_0$. For example, $B_0 = H^1(\Omega)$ and $B_1 = H^2(\Omega)$. For any $u \in B_0$ and $t > 0$, define

$$K(t,u) := \inf_{v \in B_1} (||u - v||_{B_0} + t||v||_{B_1}).$$

K measures how well u can be approximated by B_1. For $0 < \theta < 1$ and $1 \leq p < \infty$, define a norm

$$||u||_{[B_0,B_1]_{\theta,p}} := \left(\int_0^\infty t^{-p\theta} K(t,u)^p \frac{dt}{t} \right)^{1/p}. \tag{2.52}$$

The set

$$[B_0, B_1]_{\theta,p} = B_{\theta,p} = \left\{ u \in B_0 : ||u||_{[B_0,B_1]_{\theta,p} < \infty} \right\} \tag{2.53}$$

forms a Banach space with norm (2.52), see [8]. The key result of operator-interpolation theory is as follow:

Theorem 2.13. *Suppose that A_i and B_i are two pairs of Banach spaces as above, and that T is a linear operator that maps A_i to B_i ($i = 0, 1$). Then T maps $A_{\theta,p}$ to $B_{\theta,p}$. Moreover,*

$$||T||_{A_{\theta,p} \to B_{\theta,p}} \leq ||T||^{1-\theta}_{A_0 \to B_0} ||T||^{\theta}_{A_1 \to B_1}. \tag{2.54}$$

Proof. Let $M_i := ||T||_{A_i \to B_i}$. For any $v \in A_1$,

$$\begin{aligned} K_B(t, Tu) &\leq ||Tu - Tv||_{B_0} + t||Tv||_{B_1} \\ &\leq M_0 ||u - v||_{A_0} + tM_1 ||v||_{A_0} \\ &\leq M_0 (||u - v||_{A_0} + tM_1/M_0 ||v||_{A_0}). \end{aligned}$$

Taking the infimum over $v \in A_1$, we have

$$K_B(t, Tu) \leq M_0 K_A(tM_1/M_0, u).$$

Integrating this inequality, we have

$$\begin{aligned}
||Tu||_{B_{\theta,p}} &= \left(\int_0^\infty t^{-p\theta} K_B(t, Tu)^p \frac{dt}{t} \right)^{1/p} \\
&\leq \left(\int_0^\infty t^{-p\theta} \left(M_0 K_A(tM_1/M_0, u) \right)^p \frac{dt}{t} \right)^{1/p} \\
&= \left(\int_0^\infty M_1^{p\theta} M_0^{p-p\theta} s^{-p\theta} \left(K_A(s, u) \right)^p \frac{ds}{s} \right)^{1/p} \\
&= M_0^{1-\theta} M_1^{\theta} \left(\int_0^\infty s^{-p\theta} \left(K_A(s, u) \right)^p \frac{ds}{s} \right)^{1/p} \\
&= M_0^{1-\theta} M_1^{\theta} ||u||_{A_{\theta,p}}.
\end{aligned}$$

□

Consider the space, $[H^i(\Omega), H^{i+1}(\Omega)]_{\theta,2}$, where i is a nonnegative integer. Define the fractional order Sobolev space as

$$[H^i(\Omega), H^{i+1}(\Omega)]_{\theta,2} = H^{i+\theta}(\Omega).$$

It can be shown that if Ω has Lipschitz boundary, then the norms (2.51) and (2.52) are equivalent.

Lastly, consider a situation in which u is less smooth such that the global interpolant $\mathcal{I}u$ is not well defined. We might not be able to conclude anything regarding the error convergence as (2.50). However, using space-interpolation theory, the following bound can be derived.

Theorem 2.14. *Let $i < s$, $s = \lfloor s \rfloor + \theta$, $\theta \in (0, 1)$. Suppose that (2.50) hold for $m = \lfloor s \rfloor, \lfloor s \rfloor + 1$. Then for any $u \in H^s(\Omega)$,*

$$||u - u_h||_{H^i(\Omega)} \leq Ch^{s-i} ||u||_{H^s(\Omega)}. \tag{2.55}$$

Proof. Define an operator, T, that maps u to the error $u - u_h$:

$$Tu := u - u_h.$$

Estimate (2.50) implies that T maps $H^i(\Omega)$ to $H^m(\Omega)$, $m = \lfloor s \rfloor, \lfloor s \rfloor + 1$, with

$$||T||_{H^i(\Omega)\to H^m(\Omega)} \leq c_2 h^{m-i}.$$

Thus, by setting Banach-space interpolation: $A_0 = H^{\lfloor s \rfloor}(\Omega)$, $A_1 = H^{\lfloor s \rfloor+1}(\Omega)$, and $B_0 = B_1 = H^i(\Omega)$, we conclude that

$$\begin{aligned} ||T||_{H^s(\Omega)\to H^i(\Omega)} &\leq Ch^{\lfloor s \rfloor(1-\theta)} h^{(\lfloor s \rfloor+1)\theta} \\ &= Ch^s. \end{aligned}$$

This is equivalent to (2.55). □

Chapter 3

Efficiency-based refinement strategies

In this chapter, adaptive refinement and two efficiency-based refinement strategies are described. Study of efficiency-based refinement strategies is the main contribution of this thesis. First, in section 3.1, assumptions on PDE problems, finite element methods, error estimators, refinement processes and linear solvers are considered. Then the basic idea of efficiency-based refinement strategies is discussed. By taking into account both work and error reduction, refinement decisions are made by optimizing a certain efficiency measure. Lastly, in section 3.2, two refinement strategies are presented: the 'work times error' (WEE) and 'accuracy per computational cost' (ACE) strategies.

3.1 Assumptions on FEM, adaptive refinement process, and linear solver

First, consider using a FEM to solve the BVP problem (2.20). Assume that the associated variational problem is well-posed in trial space U and test space V. Let $\mathcal{T}$ be a partition of the domain, Ω, into finite elements, i.e., $\overline{\Omega} = \bigcup_{T_i \in \mathcal{T}} \overline{T}_i$, with given shape functions $\mathcal{P}_{T_i}$ and nodal variable $\mathcal{N}_{T_i}$. Here we only consider the case that the shape functions are polynomials. For any element T_i, denote by p_i the maximum polynomial order of all shape functions in $\mathcal{P}_{T_i}$. Then we say element T_i has order p_i. Let U_h be the finite element subspace. It is obvious that U_h consists of piecewise polynomials. Denote by m a nonnegative integer, and assume that $||\cdot||_U$ and $||\cdot||_{H^m(\Omega)}$ are equivalent norms. For any element T_i with order p_i, by Theorem 2.11, under certain conditions, we can assume that the following estimate holds

$$||u - \mathcal{I}_{T_i} u||_{H^m(T_i)} \leq c_i \ (\mathrm{diam}(T_i))^{p_i+1-m} ||u||_{H^{p_i+1}(T_i)}. \tag{3.1}$$

Here the constant c_i depends on p_i, d and the element T_i. Let u_h be the finite element approximation, and assume the discrete variational problem is also well posed such that we have the following error estimate:

$$\begin{aligned} ||u - u_h||_{H^m(\Omega)} &\approx ||u - u_h||_U \\ &\leq C||u - \mathcal{I}_\mathcal{T} u||_U \\ &\approx C||u - \mathcal{I}_\mathcal{T} u||_{H^m(\Omega)} \\ &\leq C \sum_{T_i \in \mathcal{T}} c_i \ (\mathrm{diam}(T_i))^{p_i+1-m} ||u||_{H^{p_i+1}(T_i)}, \end{aligned} \tag{3.2}$$

where C is the constant in Theorem 2.9 or Theorem 2.10, respectively. Moreover, consider that all elements have the same order p. Suppose that all elements are affine equivalent, and the mesh is nondegenerate. Using Theorem 2.12 and Theorem 2.14, the following error estimate holds

$$||u - u_h||_{H^m(\Omega)} \leq C(h(\mathcal{T}))^{s-m} ||u||_{H^s(\Omega)}, \tag{3.3}$$

where $0 \leq m < s \leq p+1$, with s a real number. Further assume that we obtain a locally sharp a-posteriori error estimate $E(u_h, f)$ that is equivalent to $||u - u_h||_{H^m(\Omega)}$. The associated error functional is given by $\mathcal{F}(u_h, f) = E^2(u_h, f)$. For example, the L^2 functional is a natural a-posteriori error estimate for First-Order System Least Squares (FOSLS) finite element methods, and equivalence to the H^1 norm has been proved for several relevant second-order PDE systems of elliptic type [11, 13, 14, 15]. The local value of the error, E, on element T_i is denoted by ϵ_i. Assume that the local value of the error also has similar asymptotic behaviour as (3.2), i.e., assume that for $u \in H^{s_i}(T_i)$ with $m < s_i \leq p_i + 1$

$$\epsilon_i \approx ||u - u_h||_{H^m(T_i)} \leq c_i \ (\mathrm{diam}(T_i))^{s_i - m} ||u||_{H^{s_i}(T_i)}. \tag{3.4}$$

Remark. Note that if we assume element wise that $||u - u_h||_{H^m(T_i)} \leq C||u - v||_{H^m(T_i)}$ for all $v \in \mathcal{P}_{T_i}$, then we could choose v to be the local interpolant and obtain (3.4) when

s_i is an integer, and use operator-interpolation theory to obtain the result for less smooth cases. However, we only know that the minimization property holds for $||u - u_h||_{H^m(\Omega)}$ over the whole domain Ω. This does not necessarily imply that the same minimization property holds element wise. Here, we assume that our locally sharp error estimator has the asymptotic behaviour which is described above.

Next, consider an adaptive *hp* refinement process of the following form. The refinement process starts on the coarse grid, $\mathcal{T}_0$, and proceeds through levels $\ell = 1, 2, \dots, L$ until the error measure, $E_\ell(u_h, f)$, has a value less than a given bound. In each step, some elements may be refined in h by splitting them equally in each dimension, resulting in 2^d subelements, and some elements may be refined in p by doubling the element order. This refinement process generates a sequence of meshes, $\{\mathcal{T}_j\}_{j=1}^{L}$, with meshwidth $h_0 \geq h_1 \geq \dots \geq h_L$, and a sequence of finite element subspaces satisfying $U_{h_0} \subset U_{h_1} \subset \dots \subset U_{h_L}$.

It is assumed that optimal solvers, *e.g.*, algebraic multigrid (AMG) (cf. [10]), are used for the discrete linear systems. Denote by K_ℓ the number of V-cycles required to converge to the desired error, C_0 the work units to setup a V-cycle, and C_1 the work units per V-cycle. Let ρ be the multigrid convergence factor. The number of V-cycles is given by

$$K_\ell = \lceil \frac{\log(\mathcal{F}_{\ell+1}/\mathcal{F}_\ell)}{\log(\rho)} \rceil. \tag{3.5}$$

Let N_ℓ be the number of DOF on level ℓ. The computational work for solving these systems on level ℓ is, thus, assumed to be:

$$W_\ell = (C_0 + C_1\, K_\ell)\, N_\ell. \tag{3.6}$$

For simplicity, in the following discussion, we assume the computational work to be a fixed constant times the number of DOF, i.e.,

$$W_\ell = c\, N_\ell. \tag{3.7}$$

We allow the domain to contain singularities, i.e., points (or lines or surfaces) in whose neighbourhood the full convergence order of the finite element method cannot be attained due to lack of smoothness of the solution. For simplicity, assume that those singularities can only be located at coarse-level grid points (or grid edges or faces), and that their power and location are known. If this information is not known in advance, the location and strength of singularities can be estimated by monitoring reduction rates of local error functionals during a few steps of initial uniform refinement.

3.2 Two refinement strategies: WEE and ACE

The decision of which elements to refine is based on the information provided by the local error estimator, and by heuristics that may take into account predicted error reduction and work. In particular, we consider strategies where the elements are ordered in terms of decreasing local error, such that elements with larger error are considered for refinement first. Standard threshold-based approaches then may refine, for example, a fixed fraction of the elements in every step, or a fixed fraction of the total error functional. Our goal is to reach a pre-specified bound on the global error, $E_\ell(u_h, f)$, with a minimal amount of total work, $\sum_{\ell=1}^{L} W_\ell$. Finding this optimal grid sequence may be difficult, even if we restrict the process to h-refinement alone. Hence, we turn to seeking nearly optimal solutions by using heuristics of greedy type. We consider refinement heuristics that determine the fraction of elements to be refined based on optimizing an efficiency measure in every step. We expect that a desirable grid sequence needs to be a high accuracy sequence, *i.e.*, a grid sequence for which the error, $E_\ell(N_\ell)$, decreases with nearly optimal order as a function of the number of DOF, N_ℓ, on grid level ℓ.

On each level, order the elements such that the local error, ϵ_j, satisfies $\epsilon_1 \geq \ldots \geq \epsilon_{N_\ell}$. With $r \in (0, 1]$ denoting the to-be-determined fraction of elements that will be refined, let $f(r) \in [0, 1]$ be the fraction of the total error functional in the refinement region, $\gamma(r) \in [0, 1]$ the predicted functional reduction, and $\eta(r) \in [1, 2^d]$ the ratio of the number of DOF on level $\ell + 1$ and level ℓ, *i.e.*, $N_{\ell+1} = \eta(r)\, N_\ell$. The first refinement strategy, 'work times error' efficiency (WEE), was initially proposed in [11]. Here, the fraction, r, of elements to be refined on the current level is determined by minimizing the following efficiency measure:

$$\text{work} \times \text{error reduction} = \eta(r)\, \sqrt{\gamma(r)}, \tag{3.8}$$

i.e.,

$$r_{\text{opt}} = \arg \min_{r \in (0,1]} \eta(r) \sqrt{\gamma(r)}. \tag{3.9}$$

The motivation for this heuristic is as follows: more work on the current level is justified when it results in increased error reduction that offsets the extra work. While this choice does not guarantee that a globally optimal grid sequence is obtained, this local optimization in each step results in an overall strategy of greedy type, which can be expected to lead to a reasonable approximation to the optimal grid sequence.

A second strategy, 'accuracy per computational cost' efficiency (ACE) was suggested by John Ruge . We define the predicted effective functional reduction factor

$$\gamma(r)^{\text{eff}} = \gamma(r)^{1/\eta(r)}. \tag{3.10}$$

The fraction, r, of elements to be refined on the current level is determined by minimizing this effective reduction factor, which is the same as minimizing $\log(\gamma(r)^{\text{eff}})$, *i.e.*,

$$r_{\text{opt}} = \arg\min_{r\in(0,1]} \frac{\log(\gamma(r))}{\eta(r)}. \tag{3.11}$$

The effective functional reduction factor, $\gamma(r)^{\text{eff}}$, measures the functional reduction per unit work. Indeed, compare two hypothetical error-reducing processes with functional reduction factors γ_1 and γ_2, and work proportional to η_1 and η_2. Assume that process 2 requires double the work of process 1, $\eta_2 = 2\,\eta_1$. Then the two processes would be equally effective when $\gamma_2 = \gamma_1^2$, because process 1 could be applied twice to obtain the same error reduction as process 2, using the same total amount of work as process 2. Minimizing the effective functional reduction in every step, thus, chooses the fraction, r, of elements to be refined by locally minimizing the functional reduction per unit work.

Both the strategies of minimizing work times error reduction, and minimizing the effective functional reduction factor, are ways for optimizing the efficiency of the refinement process at each level. Hence, we call the two proposed refinement strategies efficiency-based.

The predicted functional reduction ratio, $\gamma(r)$, and element growth ratio, $\eta(r)$, can be determined as follows for the case of h-refinement in $\mathcal{R}^d$.

The element growth ratio, $\eta(r)$, can be determined easily. We have N_ℓ elements on level ℓ. Of these, rN_ℓ are refined into 2^d new elements each, while $(1-r)N_\ell$ elements are left unrefined. Thus, the number of elements on level $\ell+1$ is $N_{\ell+1} = (1-r)N_\ell + 2^d\,rN_\ell = (1-r+2^d\,r)N_\ell$. This yields

$$\eta(r) = 1 - r + 2^d\,r. \tag{3.12}$$

The predicted functional reduction factor, $\gamma(r)$, depends on the error estimate and the smoothness of the solution. As mentioned above, we consider the case that the error estimate is equivalent to the H^1 norm of $u - u_h$, i.e., $\mathcal{F}(u_h, f) \approx ||u - u_h||^2_{H^1(\Omega)}$ and $\epsilon_i^2 \approx ||u - u_h||^2_{H^1(T_i)}$.

Let h_i be the diameter of element T_i. For elements T_i in which the solution is smooth (at least in $H^{p_i+1}(\tau_i)$ if order p_i elements are used), we have

$$\begin{aligned} \epsilon_i^2 &\leq C_i h_i^{2p_i} ||u||^2_{H^{p_i+1}(T_i)} \\ &\leq C_i M_{p_i+1} h_i^{2p} h_i^d. \end{aligned} \tag{3.13}$$

Here, we can take $M_{p_i+1} = \| \sum_{|\alpha|=0}^{p_i+1} (D^\alpha u)^2 \|_{\infty, T_i}$, such that $||u||^2_{H^{p_i+1}(T_i)} \leq M_{p_i+1}\, h_i^d$. If τ_i is split equally in each dimension, we have 2^d new elements, $T_{i,1}, T_{i,2}, ..., T_{i,2^d}$ with the same order p_i. Then we can assume that

$$\frac{\epsilon_{i,1}^2 + \epsilon_{i,2}^2 + ... + \epsilon_{i,2^d}^2}{\epsilon_i^2} \approx \left(\frac{1}{2}\right)^{2p}. \tag{3.14}$$

However, if u is less smooth in some element T_i, i.e., if we can only assume that $u \in H^{s_i}(T_i)$ with $s_i \in \mathcal{R},\ s_i < p_i + 1$, then we have

$$\epsilon_i^2 \leq C_i h_i^{2(s_i-1)} ||u||^2_{H^{s_i}(T_i)}. \tag{3.15}$$

For simplicity, we only consider the highly singular case here, for which $s_i << p_i + 1$. If, again, T_i is split in two in each dimension, assuming only element $T_{i,1}$ contains the singularity, then $\epsilon_{i,1} >> \epsilon_{i,j}$ for all $j = 2, 3, ..., 2^d$ and $||u||_{H^{s_i}(T_{i,1})} \approx ||u||_{H^{s_i}(T_i)}$. We then obtain that

$$\frac{\epsilon_{i,1}^2 + \epsilon_{i,2}^2 + ... + \epsilon_{i,2^d}^2}{\epsilon_i^2} \approx \frac{\epsilon_{i,1}^2}{\epsilon_i^2} \approx \left(\frac{1}{2}\right)^{2(s_i-1)}. \tag{3.16}$$

Suppose the solution is sufficiently smooth in the whole domain. For h refinement with uniformly fixed element order p, the refinement process described above generates a nondegenerate mesh sequence. Then the predicted functional reduction factor, $\gamma(r)$, can be obtained as follows. We apply (3.14) to the elements that are refined. A fraction $1 - f(r)$ of elements do not get refined, and so we assume that their errors are not reduced. This results in

$$\gamma(r) = 1 - f(r) + \left(\frac{1}{2}\right)^{2p} f(r). \tag{3.17}$$

It is cumbersome to give a general expression for the singular case. However, assuming that we know the power and location of the singularities in advance, one can easily compute $\gamma(r)$ using (3.14) and (3.16).

For p-refinement, the corresponding error reduction ratios are discussed in the next chapter in the context of a 1D model problem.

Chapter 4

Numerical Performance in 1D

In this chapter, we study the performance of the proposed efficiency-based refinement strategies for a standard model problem in 1D [3, 12]:

$$u'' = \alpha(\alpha - 1)x^{\alpha-2}, u(0) = 0, u(1) = 1, \tag{4.1}$$

with exact solution given by $u = x^\alpha$.

While the efficiency-based refinement strategies can be applied to various types of finite element methods and associated error estimates, we choose to illustrate the refinement strategies for model problem (4.1) using standard Galerkin finite element methods. We choose this model problem because asymptotically optimal h and hp finite element grids have been developed for them [3, 12], which can be used as a point of comparison for the refinement strategies to be presented. In addition, there is a x^α-type singularity at $x = 0$ by choosing certain values for α. The refinement strategies presented can be equally applied to other finite element methods, as is illustrated in next chapter, where we present results for a 2D problem using the FOSLS finite element method [14, 11].

This chapter is organized as follows. In section 4.1, properties of hp-finite elements in one dimension are discussed. The approximation properties for p-refinement are especially addressed. Then, in section 4.2, properties of the 1D model problem and the finite element method applied are considered. In section 4.3, the performance of the WEE and ACE h-refinement strategies for the 1D model problem are analyzed. Modified WEE and ACE refinement strategies for the singular case are considered in Section 4.4. Lastly, in section

4.5, efficiency-based *hp*-refinement strategies are discussed and illustrated for the model problem.

4.1 *hp*-finite elements in 1D

We choose a standard Galerkin finite element method to solve model problem (4.1). First, we construct a finite dimensional subspace $S^{\mathbf{p},m}(\Omega,\mathcal{T})$ of $H^m(\Omega)$, which consists of piecewise polynomials as follows. Let $\mathcal{T}$ be a partition of $\Omega = (a,b)$ into $N(\mathcal{T})$ open, disjoint subintervals T_i, i.e., $\mathcal{T} = \cup_i T_i$, $T_i = (x_{i-1}, x_i)$, $a = x_0 < x_1 < ... < x_{N(\mathcal{T})} = b$. Denote by $h_i = x_i - x_{i-1}$ the width of element T_i, and $h = \max_{1\leq i\leq N} h_i$ the meshwidth of $\mathcal{T}$.

Each element $T_i \in \mathcal{T}$ can be mapped onto $\widehat{T} = (-1,1)$, the reference element. Denote the map by Q_i, i.e.,

$$\begin{aligned} &T_i = Q_i(\widehat{T}), \\ &x = Q_i(\xi) = \frac{1}{2}(1-\xi)x_{i-1} + \frac{1}{2}(1+\xi)x_i, \qquad \xi \in \widehat{T}, \\ &\xi = Q_i^{-1}(x) = \frac{2x - x_i - x_{i-1}}{x_i - x_{i-1}}, \qquad x \in T_i. \end{aligned} \tag{4.2}$$

The space $S^{\mathbf{p},m}(\Omega,\mathcal{T})$ is defined as follows:

Definition 4.1. *Let $\Omega = (a,b)$ be an interval, $\mathcal{T}$ be a mesh and $\boldsymbol{p} = (p_1, p_2, ..., p_N)$ a vector of polynomial degrees p_i (the degree vector), and let $m \geq 0$ be an integer. Let S^p denote the set of polynomials of degree p on the reference element $\widehat{T}$. Then*

$$S^{\boldsymbol{p},m}(\Omega,\mathcal{T}) = \left\{ u \in H^m(\Omega): \ u|_{T_i} = s_i(Q_i^{-1}(x)), \ s_i \in S^{p_i}(\widehat{T}) \right\}. \tag{4.3}$$

Since Q_i is linear, $u \in S^{\mathbf{p},m}(\Omega,\mathcal{T})$ implies that on $T_i \in \mathcal{T}$, u is a polynomial of degree p_i. We define further

Definition 4.2. *Let $\Omega = (a,b)$ be an interval, $\mathcal{T}$ be a mesh and $\boldsymbol{p} = (p_1, p_2, ..., p_N)$ a vector of polynomial degree p_i (the degree vector), and let $m \geq 0$ be an integer. Let S^p denote the set of polynomials of degree p on the reference element $\widehat{T}$. Then*

$$S_0^{\boldsymbol{p},m}(\Omega,\mathcal{T}) = \left\{ u \in S^{\boldsymbol{p},m}(\Omega,\mathcal{T}): \ u(a) = u(b) = 0 \right\}. \tag{4.4}$$

Denote by $\psi_i,\ i = 1, 2, ..., p+1$ the basis functions (shape functions) of S^p. The selection of shape functions depends on the degree m of smoothness of $S^{\mathbf{p},m}(\Omega, \mathcal{T})$. Here, we only consider the case where $m = 1$.

Shape functions

If $m = 1$, we select the set

$$\begin{aligned} \psi_1(\xi) &= \frac{1-\xi}{2},\ \psi_2 = \frac{1+\xi}{2}, \\ \psi_i(\xi) &= \sqrt{\frac{2i-3}{2}} \int_{-1}^{\xi} L_{i-2}(t)\, dt, \qquad 3 \le i \le p+1. \end{aligned} \tag{4.5}$$

Here, $L_i(x)$ are Legendre polynomials defined on $\widehat{T} = (-1, 1)$. The first Legendre polynomials are given by

$$\begin{aligned} L_0(x) &= 1, \\ L_1(x) &= x, \\ L_2(x) &= (3x^2 - 1)/2, \\ L_3(x) &= (5x^3 - 3x)/2, \\ L_4(x) &= (35x^4 - 30x^2 + 3)/8. \end{aligned} \tag{4.6}$$

They satisfy the Legendre differential equation

$$\big((1 - x^2)L_i'(x)\big)' + i(i + 1)L_i(x) = 0 \text{ in } \widehat{T}. \tag{4.7}$$

Further, we have orthogonality

$$\int_{-1}^{1} L_i(x)L_j(x)\, dx = \begin{cases} \frac{2}{2i+1} & \text{for } i = j, \\ 0 & \text{otherwise .} \end{cases} \tag{4.8}$$

Moreover, $\{L_i(x)\}_{i=0}^{\infty}$ is a basis of $L^2(\widehat{T})$, i.e., any function $u \in L^2(\widehat{T})$ can be expanded into Legendre series

$$u(x) = \sum_{i=0}^{\infty} a_i L_i(x). \tag{4.9}$$

where "=" is understood in the sense that

$$\lim_{p\to\infty} ||u - \sum_{i=0}^{p} a_i L_i||_{L^2(\widehat{T})} = 0. \tag{4.10}$$

Multiplying (4.9) by $L_j(x)$, integrating over $\widehat{T}$ and referring to (4.8) we find

$$a_i = \frac{2i+1}{2} \int_{-1}^{1} u(x) L_i(x)\, dx. \tag{4.11}$$

Lastly, the truncated Legendre expansion gives the best approximation in S^p. Let $Q_p(x)$ be any polynomial of degree $\leq p$ in $\widehat{T}$. For any $u \in L^2(\widehat{T})$, we have

$$\min_{Q_p \in S^p} ||u - Q_p||_{L^2(\widehat{T})} = ||u - \sum_{i=0}^{p} a_i L_i||_{L^2(\widehat{T})}. \tag{4.12}$$

Now, after selecting shape functions on S^p, we need to define nodal variables $\mathcal{N}_p = \{N_1, N_2, ..., N_{p+1}\}$ such that $\{\widehat{T}, S^p, \mathcal{N}_p\}$ is a finite element.

Nodal Variables

If $m = 1$, we select the set

$$\begin{aligned} &N_1(v) = v(-1),\ N_2(v) = v(1), \\ &N_i(v) = \sqrt{\frac{2i-3}{2}} \int_{-1}^{1} v'(s) L_{i-2}(s)\, ds, \qquad 3 \leq i \leq p+1 \qquad \forall v \in S^p. \end{aligned} \tag{4.13}$$

Then for any $v \in S^p$, it can be shown that v can be written as

$$v(x) = \sum_{i=0}^{p} N_i(v) \psi_i(x). \tag{4.14}$$

Proof. For any $v \in S^p$, we have

$$\begin{aligned}
v(x) &= v(-1) + \int_{-1}^{x} v'(t)\,dt \\
&= v(-1) + \int_{-1}^{x} \sum_{i=0}^{p-1} \left(\frac{2i+1}{2} \int_{-1}^{1} v'(s) L_i(s)\,ds \right) L_i(t)\,dt \\
&= v(-1) + \frac{1}{2}(v(1) - v(-1))(x+1) + \sum_{i=1}^{p-1} \int_{-1}^{x} \left(\frac{2i+1}{2} \int_{-1}^{1} v'(s) L_i(s)\,ds \right) L_i(t)\,dt \\
&= v(-1)\frac{1-x}{2} + v(1)\frac{1+x}{2} + \sum_{i=3}^{p+1} \sqrt{\frac{2i-3}{2}} \left(\int_{-1}^{1} v'(s) L_{i-2}(s)\,ds \right) \sqrt{\frac{2i-3}{2}} \int_{-1}^{x} L_{i-2}(t)\,dt \\
&= \sum_{i=0}^{p} N_i(v)\psi_i(x).
\end{aligned}$$

□

Remark. The reason why we choose (4.5) as shape functions for S^p is that

$$\int_{-1}^{1} \psi_i' \psi_j' d\xi = \delta_{i,j},\ \ i,j \geq 3.$$

The weak form of (4.1) contains product of first derivatives of ψ_i, and this choice makes the resulting matrix sparse. One can also choose the standard 1D Lagrange elements described in Chapter 2. These two types of elements are essentially equivalent, and thus give the same approximate solution u_h.

The standard shape functions $\psi_j(\xi)$ and corresponding nodal variables N_j induce, via the mapping Q_i, element shape functions on T_i.

Definition 4.3. *Let $\{\psi_j\}_{j=1}^{p_i+1}$ be the set of standard shape functions (of order m and degree p_i). Then the corresponding element shape functions of order ℓ and degree p_i are defined by*

$$\nu_j^{[i]}(x) := \psi_j(Q_i^{-1}(x)),\ \ x \in T_i,\ \ j = 1, 2, ..., p_i + 1. \tag{4.15}$$

And the corresponding nodal variables are given by

$$n_j^{[i]}(v(x)) = N_j(v(Q_i(\xi))),\ \ j = 1, 2, ..., p_i + 1. \tag{4.16}$$

After setting up basis functions and nodal variables elementwise, basis functions of the space $S^{\mathbf{p},m}(\Omega,\mathcal{T})$ can be deduced by extending all element shape functions to Ω. Here, we only consider the case where $m = 1$. For any element T_i, we distinguish external element shape functions $\nu_j^{[i]}(x), j = 1,2$ and internal element shape functions $\nu_j^{[i]}(x), j = 3,...,p_i+1$. It is easy to see that nodal variables corresponding to external element shape functions give function values at nodes x_{i-1} and x_i. Since in one dimensional space, by the Sobolev Imbedding Theorem, $H^1(\Omega) \hookrightarrow C^{0,\frac{1}{2}}(\overline{\Omega})$, it follows that each extension of $\nu_j^{[i]}(x)$ must be continuous. To achieve this, the corresponding nodal element shape functions of neighboring elements must be joined at node x_i, $i = 1,2,...,N(\mathcal{T})-1$. Every internal shape function can be extended by zero to Ω to obtain a basis function. Thus, the dimension of $S^{\mathbf{p},1}(\Omega,\mathcal{T})$ can be given by

$$\dim(S^{\mathbf{p},1}(\Omega,\mathcal{T})) = \sum_{i=1}^{N(\mathcal{T})}(p_i+1) - (N(\mathcal{T})-1). \tag{4.17}$$

Approximation properties of $S^{\mathbf{p},1}(\Omega,\mathcal{T})$

Assuming that $u \in H^{p_i+1}(T_i)$, Theorem 2.11 implies that the local interpolant $\mathcal{I}_{T_j}u$ satisfies:

$$||u' - (\mathcal{I}_{T_i}u)'||_{L^2(T_i)} \leq c_i\,(h_i)^{p_i}||u||_{H^{p_i+1}(T_i)}.$$

Here c_i depends on p_i. More precisely, it is shown in [3] that the following sharp estimate holds:

Theorem 4.1 (Approximation properties of $S^{\mathbf{p},1}(\Omega,\mathcal{T})$)**.** *Assume that $u \in H^{p_i+1}(T_i)$ for $T_i \in \mathcal{T}$. Let $\mathcal{I}_{T_i}u$ be the local interpolant of u. Then $\mathcal{I}_{T_i}u$ satisfies:*

$$||u' - (\mathcal{I}_{T_i}u)'||_{L^2(T_i)} \leq \frac{1}{2^{p_i}\sqrt{(2p_i)!}}(h_i)^{p_i}||u||_{H^{p_i+1}(T_i)}. \tag{4.18}$$

This estimate can be applied to predict the error reduction for p-refinement in one dimension if it is known in advance that solution u is smooth enough, e.g., (4.18) is used in section 4.5 for hp-version WEE and ACE strategies.

4.2 Model problem

Next, we formulate problem (4.1) into a variational problem using standard Galerkin method. Let $\Omega = (0,1)$, and let $f(x) = \alpha(\alpha-1)x^{\alpha-2}$. Define $H_D^1 = \{u \in H^1(\Omega) : u(0) = 0,\ u(1) = 1\}$. Multiplying both sides of (4.1) by any given function $v \in H_0^1(\Omega)$ and integrating by parts gives the variational problem:

$$\begin{aligned}&\text{find } u \in H_D^1 \text{ such that}\\ &-\int_0^1 u'v'\,dx = \int_0^1 fv\,dx \qquad \forall v \in H_0^1(\Omega).\end{aligned} \tag{4.19}$$

However, H_D^1 is not a linear subspace of $H^1(\Omega)$ and the well-posedness theory of section 2.2.2 does not apply. To treat this case properly, let us assume that a function g_D (in general not unique) is defined on all of $\Omega = (0,1)$, with $g_D \in H^1(\Omega)$ and $g_D|_{\partial\Omega} = u$. Then we can write

$$u = w + g_D. \tag{4.20}$$

Substituting into (4.19), we obtain an equivalent variational problem as

$$\begin{aligned}&\text{find } w \in H_0^1(\Omega) \text{ such that}\\ &-\int_0^1 w'v'\,dx = \int_0^1 (fv + g_D'v')\,dx \qquad \forall v \in H_0^1(\Omega).\end{aligned} \tag{4.21}$$

It is easy to see that the right hand side of (4.21) defines a continuous linear functional on $H_0^1(\Omega)$. The continuity of the bilinear form on $H_0^1(\Omega) \times H_0^1(\Omega)$ is a direct consequence of the Cauchy-Schwarz inequality. Since a Poincaré inequality holds for all $v \in H_0^1(\Omega)$, i.e., there exists a constant c such that $||v||_{L^2(\Omega)} \leq c||v'||_{L^2(\Omega)}$ for all $v \in H_0^1(\Omega)$, the coercivity follows. Therefore, problem (4.21) admits a unique solution w by the Lax-Milgram Theorem. It follows that problem (4.19) has a unique solution given by $u = w + g_D$.

Let $S_D^{\mathbf{p},1}(\Omega, \mathcal{T}) = S^{\mathbf{p},1}(\Omega, \mathcal{T}) \cap H_D^1$. A finite element approximation u_h is obtained as usual:

$$\begin{aligned}&\text{find } u_h \in S_D^{\mathbf{p},1}(\Omega, \mathcal{T}) \text{ such that}\\ &-\int_0^1 u_h'v'\,dx = \int_0^1 fv\,dx \qquad \forall v \in S_0^{\mathbf{p},1}(\Omega, \mathcal{T}).\end{aligned} \tag{4.22}$$

Proving the well-posedness of (4.22) involves, typically, the use of an interpolant, $\mathcal{I}^h g_D$, of the Dirichelet data into $S^{\mathbf{p},1}(\Omega, \mathcal{T})$. Again, for any $u_h \in S_D^{\mathbf{p},1}(\Omega, \mathcal{T})$, let $w_h = u_h - \mathcal{I}^h g_D$ such that $w_h \in S_0^{\mathbf{p},1}(\Omega, \mathcal{T})$ (note that here we can not let $w_h = u_h - g_D$ since $u_h - g_D$ may not be in space $S_0^{\mathbf{p},1}(\Omega, \mathcal{T})$). We have

$$\text{find } w_h \in S_0^{\mathbf{p},1}(\Omega, \mathcal{T}) \text{ such that}$$
$$-\int_0^1 w_h' v' \, dx = \int_0^1 fv + (\mathcal{I}^h g_D)' v' \, dx \qquad \forall v \in S_0^{\mathbf{p},1}(\Omega, \mathcal{T}). \tag{4.23}$$

Noting that $|\cdot|_{H^1(\Omega)}$ and $||\cdot||_{H^1(\Omega)}$ are equivalent norms in $H_0^1(\Omega)$, problem (4.23) admits a unique solution w_h satisfying

$$|w - w_h|_{H^1(\Omega)} = \min_{v \in S_0^{\mathbf{p},1}(\Omega, \mathcal{T})} |w - v|_{H^1(\Omega)}. \tag{4.24}$$

This implies

$$|u - g_D - (u_h - \mathcal{I}^h g_D)|_{H^1(\Omega)} = \min_{v \in S_0^{\mathbf{p},1}(\Omega, \mathcal{T})} |u - g_D - v|_{H^1(\Omega)}. \tag{4.25}$$

Here, we can choose $g_D = u(1)\nu_2^{[N(\mathcal{T})]}$, i.e., the function value of u at $x = 1$ multiplied by the corresponding nodal element shape function. Then $g_D = \mathcal{I}^h g_D$ since $g_D \in S^{\mathbf{p},1}(\Omega, \mathcal{T})$. And $S_D^{\mathbf{p},1}(\Omega, \mathcal{T}) = S_0^{\mathbf{p},1}(\Omega, \mathcal{T}) + g_D$. From (4.25), we have

$$|u - u_h|_{H^1(\Omega)} = \min_{z \in S_D^{\mathbf{p},1}(\Omega, \mathcal{T})} |u - z|_{H^1(\Omega)}. \tag{4.26}$$

For our model problem (4.1), it turns out that $u(x_i) - u_h(x_i) = 0$ at each grid point, [3, 12]. In addition, together with (4.26), the finite element approximation can be obtained easily, namely, $u_h = \mathcal{I}^h u$. Let the error be estimated by the H^1 seminorm of the actual error, $e = u - u_h$, i.e., $\mathcal{F}(u_h, f) = ||u' - u_h'||^2_{L^2(\Omega)}$ and $\epsilon_j^2 = ||u' - u_h'||^2_{L^2(\tau_j)}$. The actual error will be used as the error estimator throughout this chapter for the study of efficiency-based refinement strategies. However, in real practise, the exact solution is usually unknown. In this case, an equivalent locally sharp a posteriori error estimate needs to be derived. For more details, one can refer to [3].

The smoothness of the exact solution of our model problem (4.1), $u(x) = x^\alpha$, depends on α. First, α must be greater than $\frac{1}{2}$ such that $u \in H^1(\Omega)$. Otherwise, the variational problem (4.19) can not be solved in $H^1(\Omega)$. In addition, we have

$$x^\alpha \in H^{k+\theta}((0,1)) \qquad \forall k < k+\theta \leq \alpha + \frac{1}{2} < k+1 \tag{4.27}$$

with k an nonnegative integer and $\theta \in (0,1)$.

One can refer to [3] for proof. This implies that $u \in H^{\alpha+\frac{1}{2}}(T_1)$, and is infinitely smooth in elements T_i, $i = 2, 3, ..., N(\mathcal{T})$. Therefore, using (4.26), letting $z = \mathcal{I}^h u$ (actually, here $u_h = \mathcal{I}^h u$), and using Theorem 4.1, following asymptotic behaviour of the error holds for our model problem [3]:

$$\begin{aligned} \epsilon_1 &\leq ||u' - (\mathcal{I}^h u)'||_{L^2(T_1)} \leq C(\alpha) \frac{h_1^{\min(p_1, \alpha - \frac{1}{2})}}{p_1^{2\alpha-1}} ||u||_{H^{\min(p_1,\alpha-\frac{1}{2})+1}(T_1)}, \\ \epsilon_i &\leq ||u' - (\mathcal{I}^h u)'||_{L^2(T_i)} \leq \frac{1}{2^{p_i}\sqrt{(2p_i)!}} (h_i)^{p_i} ||u||_{H^{p_i+1}(T_i)}, \qquad i = 2, ..., N(\mathcal{T}). \end{aligned} \tag{4.28}$$

Then, for h-refinement, we can apply local error reduction ratio (3.16) for element T_1 and (3.14) for element T_i, $i = 2, ..., N(\mathcal{T})$ to predict the global error functional reduction $\gamma(r)$. Since (4.28) also shows how error bounds depend on p, error reduction w.r.t. p-refinement can be predicted.

Remark 1. Asymptotic behaviours of the error in (4.28) for the singular element is shown by using space-interpolation theory (cf [3]). It only holds for x^α-type singularity. However, error estimate for smooth elements holds generally in 1D.

Remark 2. Throughout his chapter, we say that the solution u is smooth in element T_i provided that $u \in H^{p_i+1}(T_i)$. For our model problem, consider that a uniform polynomial order p is used, then u is smooth if and only if $\alpha > p + \frac{1}{2}$.

4.3 Performance of the WEE and ACE h-refinement strategies in 1D

In this section, we apply the WEE and the ACE h-refinement strategies to our 1D model problem (4.1) with polynomial order $p = 1$. On each level ℓ, each element is allowed to be refined at most once. Numerical results are presented for both smooth and singular cases. We use the terminology and notation described in section 3.2.

4.3.1 Performance of WEE and ACE for smooth solutions

We first consider the nonsingular case and choose $\alpha > 3/2$ such that $u \in H^2((0,1))$. It follows that using (3.17) the predicted functional reduction factor, $\gamma(r)$, is given by

$$\gamma(r) = 1 - \frac{3}{4} f(r). \tag{4.29}$$

Note that, for a given error bound, our ultimate goal is to choose a grid sequence that minimizes the total work, $\sum_{\ell=1}^{L} \mathcal{W}_\ell$, which is the same as minimizing $\sum_{\ell=1}^{L} N_\ell$, based on our assumption that the work is proportional to N_ℓ. For a given error bound, the number of elements on final grid N_L is determined by the convergence rate of the global error w.r.t. the DOF, which in fact is determined by the refinement strategy. For our model problem, it has been shown in [12] that the rate of convergence is never better than $(Np)^{-p}$, where N is the number of elements and p is the degree of the polynomial.

Theorem 4.2. *[12] Let $E = (\sum \epsilon_i^2)^{\frac{1}{2}}$. Then there is a constant, $C = C(\alpha, p) > 0$, for any grid $\{0 = x_0 < x_1 < ... < x_N = 1\}$, such that*

$$E \geq C\,(Np)^{-p}. \tag{4.30}$$

For our example problem, an asymptotically optimal final grid, called a radical grid, is described in [3, 12]:

$$x_j = (j/N)^{(p+1/2)/(\alpha-1/2)},\ j = 0, ..., N. \tag{4.31}$$

This grid is optimal in the sense that, in the limit of large N, it results in the smallest error as a function of the number of DOF. If the WEE or the ACE strategy results in a grid sequence with approximately optimal convergence rate of the global error w.r.t. DOF, then the number of elements on the final grid must be close to the optimal number of elements, which only depends on the given error bound. Because we want to minimize work, it follows that, among the methods with approximately optimal convergence rate, the methods for which the sequence $\{N_\ell\}$ increases fast are preferable. Large refinements are, thus, advantageous.

We compare the numerical results of the WEE strategy and the ACE strategy, and the radical grid, for $\alpha = 2.1$ and $p = 1$ in Fig. 4.1 to Fig. 4.6. In the numerical results, we carry out the refinement process until $E_L(u_h, f) \leq 2e{-}5$ on final grid level L.

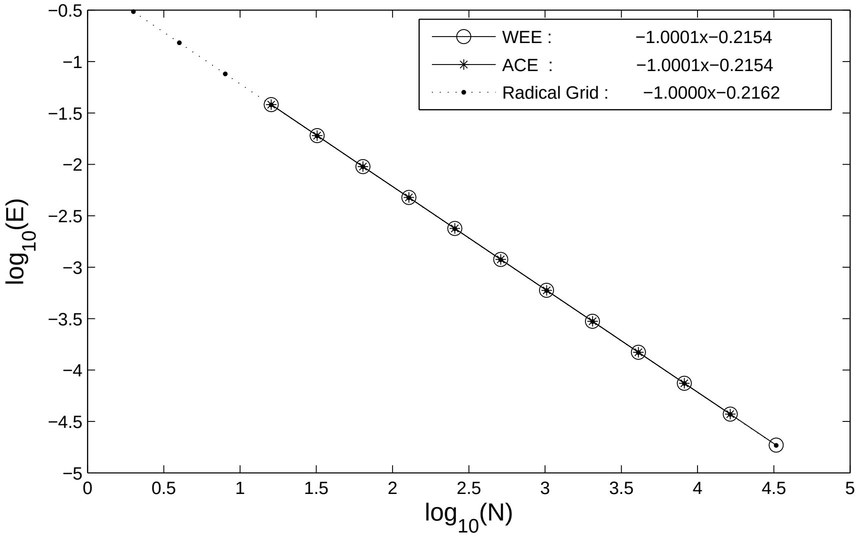
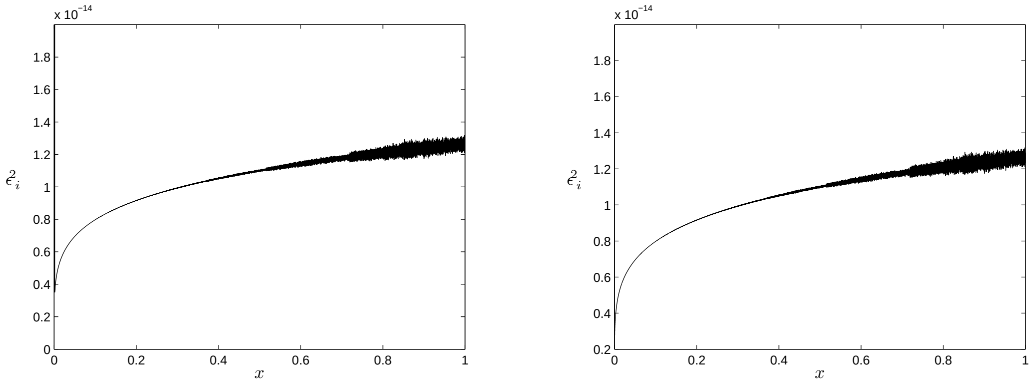


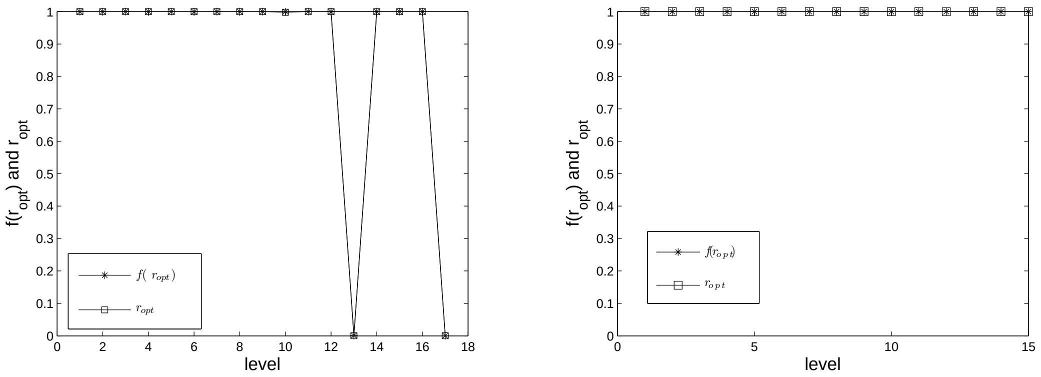
Figure 4.1: Error versus DOF, $\alpha = 2.1$ (no singularity), $p = 1$.



(a) WEE: $N_L = 32,741$, $E_L = 1.859e-5$, $L = 18$, total work $= 102,313$

(b) ACE: $N_L = 32,760$, $E_L = 1.858e-5$, $L = 16$, total work $= 65,520$

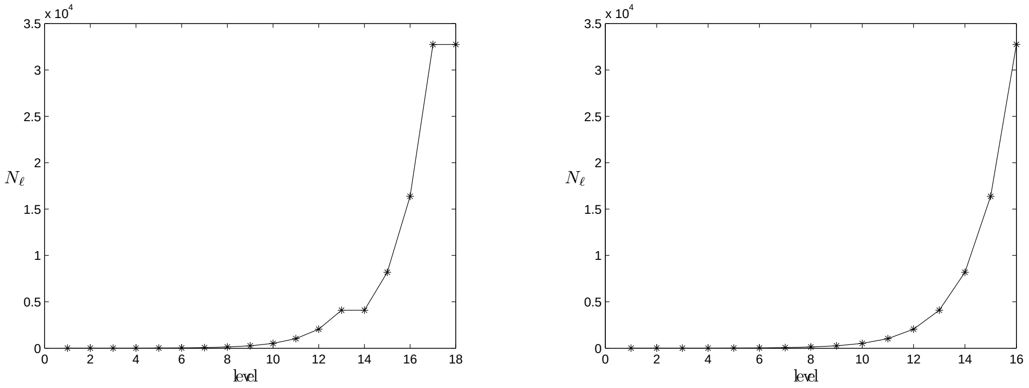
Figure 4.2: Local error functional, ϵ_i^2, versus grid location on the final grid, $\alpha = 2.1$ (no singularity), $p = 1$.



(a) WEE

(b) ACE

Figure 4.3: Refined fraction of error functional, $f(r_{\text{opt}})$, versus level, ℓ, and refined fraction of elements, r_{opt}, versus level, ℓ, $\alpha = 2.1$ (no singularity), $p = 1$.



(a) WEE

(b) ACE

Figure 4.4: Number of elements, N_ℓ, versus level, ℓ, $\alpha = 2.1$ (no singularity), $p = 1$.

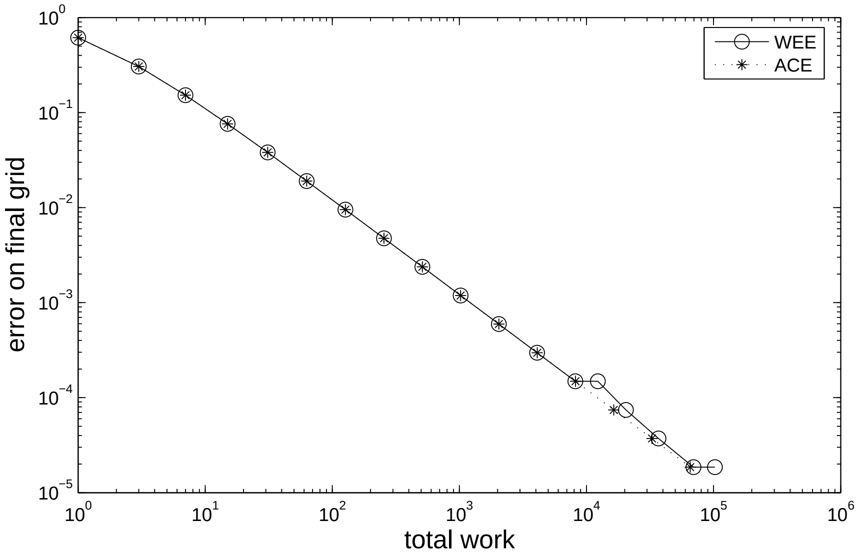
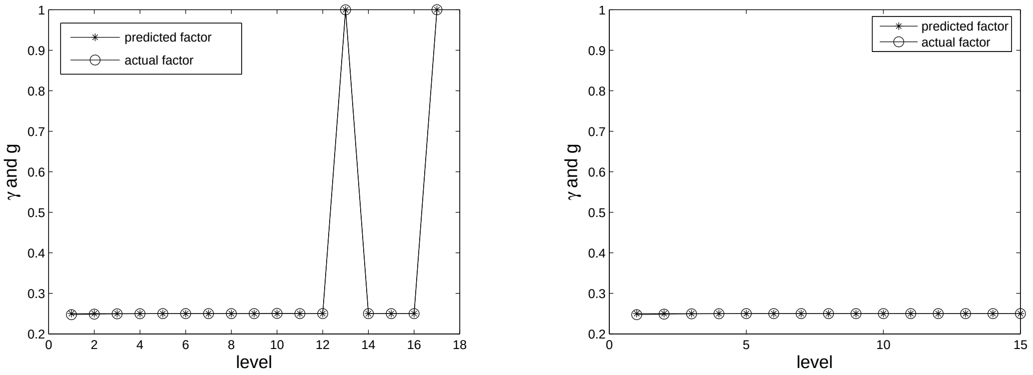


Figure 4.5: Final error, E_L, versus total work, $\sum_{\ell=1}^{L} N_\ell$, $\alpha = 2.1$ (no singularity), $p = 1$.



(a) WEE

(b) ACE

Figure 4.6: Predicted functional reduction factor, $\gamma(r_{opt})$, and actual functional reduction factor, g, versus level, ℓ, $\alpha = 2.1$ (no singularity), $p = 1$.

From Fig. 4.1, it can be observed that both strategies result in a highly accurate grid sequence. Thus, for a given error bound, the difference in the number of elements on the final grid is very small. This can be verified on Fig. 4.2. Fig. 4.3 and Fig. 4.4 show that the ACE strategy is slightly more efficient than the WEE strategy for our model problem in the smooth case. There are two small refinements in the WEE refinement process, while there are no small refinements for the ACE strategy. It follows that for a given error bound on the final grid, the WEE strategy may require slightly more total work than the ACE strategy, see Fig. 4.5. Fig. 4.2 shows that, for both strategies, the local errors in all elements tend to be equally distributed. This explains why the values of $f(r_{opt})$ and r_{opt} are close in Fig. 4.3. From Fig. 4.6 one can see that the predicted reduction factor $\gamma(r_{opt})$ is very accurate. This suggests a modification of the refinement process that can be considered to increase performance: one does not need to solve the linear systems until the new level is refined enough to have a significant number of additional elements in it. In this way complexity is never a problem, and we can still have a highly accurate grid sequence.

4.3.2 Performance of WEE and ACE for singular solutions

Next, we consider a singular example: let $\alpha = 0.6$, so that $u \in H^{1.1}((0,1))$. In the numerical results, we carry out the refinement process until $E_L(u_h, f) \le 7e{-}4$ on final grid level L. For $p = 1$, the error reduction in the element that contains $x = 0$ can be approximately given by $\left(\frac{1}{2}\right)^{0.2}$, see (3.16, 4.28). The predicted reduction factor $\gamma(r)$ is given by

$$\gamma(r) = 1 - \frac{3}{4} f(r) + \left[\left(\frac{1}{2}\right)^{0.2} - \frac{1}{4} \right] f(\frac{1}{N_\ell}). \tag{4.32}$$

Here, we assume that the local error in the element that contains $x = 0$ is always the largest.

The numerical results in Fig. 4.7 to Fig. 4.12 show that the two refinement strategies fail for this singular case. Fig. 4.7 shows that the WEE strategy results in a highly accurate grid sequence, while the ACE strategy becomes inaccurate by comparison with the radical grid. For both strategies, the local error in the first element, which contains the singularity, is always the largest, see Fig. 4.8. Hence, it is refined by the WEE and the ACE in every step.

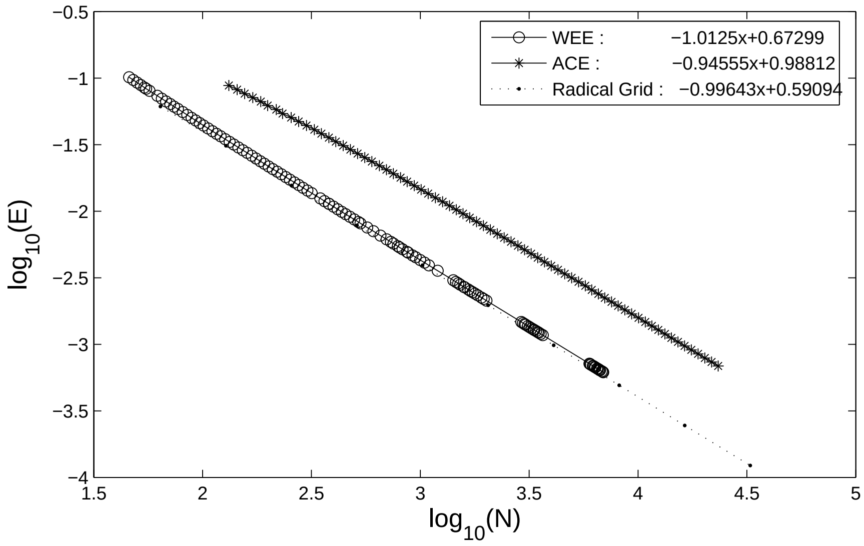
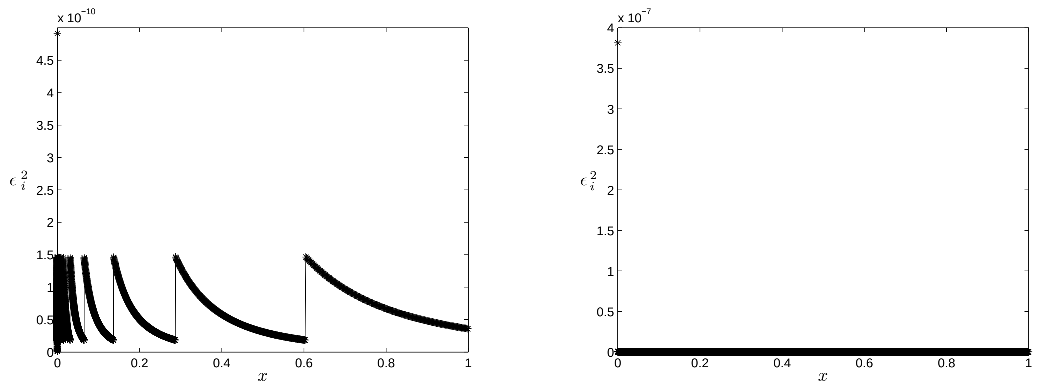


Figure 4.7: Error versus DOF, $\alpha = 0.6$ (singular case), $p = 1$.

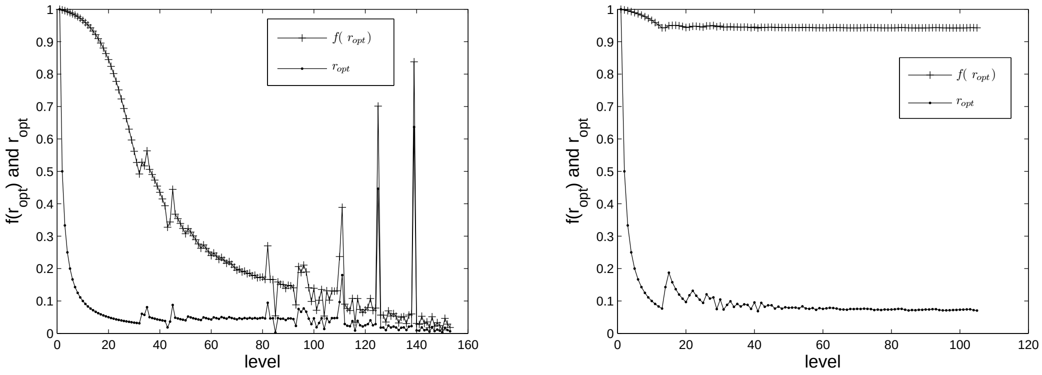


(a) WEE: $N_L = 6,925$, $E_L = 6.169e-4$, $L = 154$, total work $= 192,775$

(b) ACE: $N_L = 24,986$, $E_L = 6.411e-4$, $L = 106$, total work $= 365,420$

Figure 4.8: Local error functional, ϵ_i^2, versus grid location on the final grid, $\alpha = 0.6$ (singular case), $p = 1$.

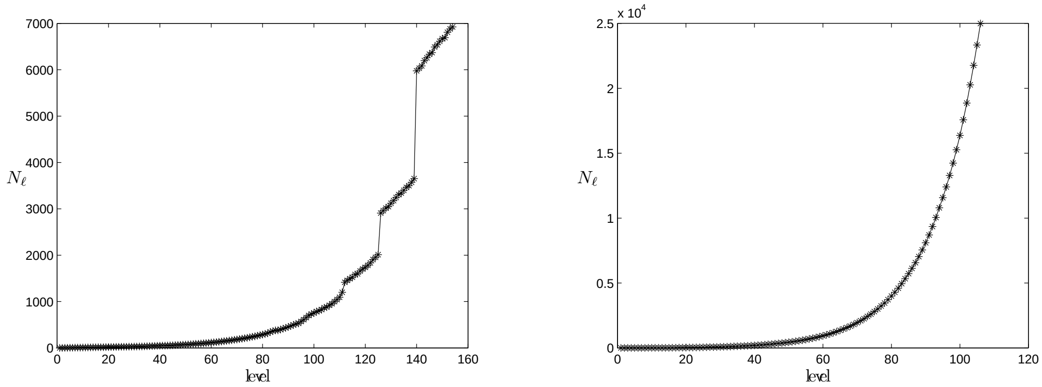
This also confirms that the predicted reduction factor can be given by (4.32). The WEE strategy generates a grid sequence with local errors being nearly equally distributed, but the



(a) WEE

(b) ACE

Figure 4.9: Refined fraction of error functional, $f(r_{\text{opt}})$, versus level, ℓ, and refined fraction of elements, r_{opt}, versus level ℓ, $\alpha = 0.6$ (singular case), $p = 1$.



(a) WEE

(b) ACE

Figure 4.10: Number of elements, N_ℓ, versus level, ℓ, $\alpha = 0.6$ (singular case), $p = 1$.

ACE strategy does not: more than 90% of the global error accumulates in only 10% of the elements; see Fig. 4.8 and Fig. 4.9. Most refinement steps of the WEE strategy are small refinements: only the first element (possibly with a few other elements) is continuously being refined (see Figs. 4.9 and 4.10). This implies that the number of elements increases slowly as a function of refinement level. It follows that the total work is very large. The ACE strategy does choose a refinement region with large fraction of the error in it. However,

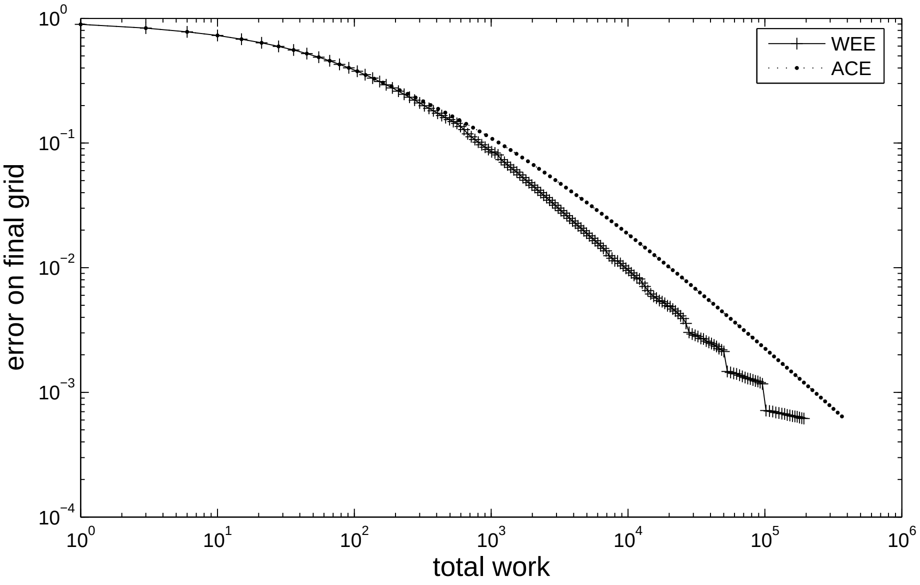
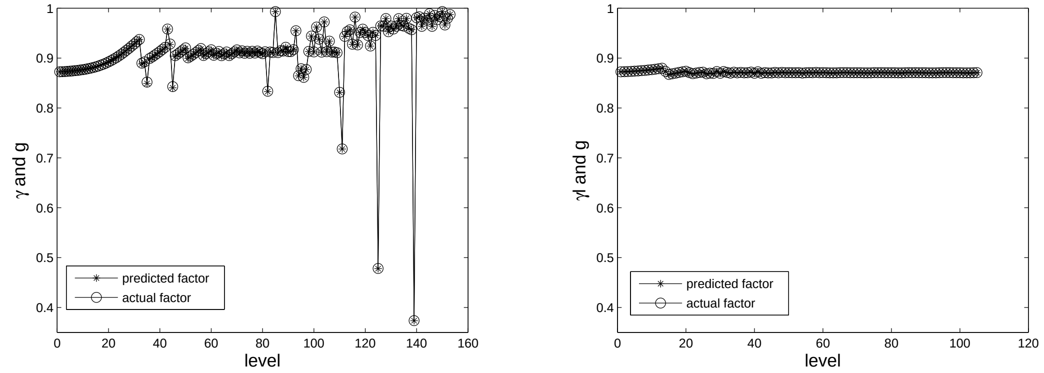


Figure 4.11: Final error, E_L, versus total work, $\sum_{\ell=1}^{L} N_\ell$, $\alpha = 0.6$ (singular case), $p = 1$.



(a) WEE

(b) ACE

Figure 4.12: Predicted functional reduction factor, $\gamma_\ell(r_{opt})$, and actual functional reduction factor, g_ℓ, versus level, ℓ, $\alpha = 0.6$ (singular case), $p = 1$.

this large fraction of error is only contained in a few elements. As a result, only a small fraction of elements are refined. Thus, the required total work is still large; see Figs. 4.10

and 4.11. Compared to the nonsingular case (Fig. 4.5), the slope of the error versus total work plot in Fig. 4.11 is much less steep, especially in the initial phase of the refinement process. The predicted reduction factors for both strategies are accurate, see Fig. 4.12. This suggests that we can make the same modification as for the smooth case to increase performance: one can wait on solving the linear systems until the number of elements has increased sufficiently. In this way, one can assure that complexity is never a problem, but calculating and minimizing the WEE and ACE functions many times may be costly as well. In conclusion, for the highly singular case, the WEE strategy results in an accurate grid sequence but is not efficient due to too many small refinements; the ACE strategy is worse than the WEE strategy in this case, because the grid sequence is not accurate and many small refinements are performed.

4.4 Modified WEE and ACE for singular solutions in 1D

In this section, modifications of WEE and ACE for singular solutions are considered. First, putting a graded grid is proposed for improving numerical performance for singular solutions. Next, we apply modified WEE and ACE strategies to our model problem. Then comparisons between our proposed efficiency-based strategies and traditional threshold based strategies are shown. Finally, we give numerical results for higher order p.

4.4.1 Modified WEE and ACE h-refinement strategies

The inefficiency of the WEE and ACE strategies for the highly singular solution is due to many steps of small refinement for the singular elements. Therefore, we attempt to avoid these steps by using a geometrically graded grid starting from the singular point, with the aim of saving work while attempting to keep the grid sequence accurate.

As was discussed before, we assume that singularities can only be located at coarse-level grid points, and that we know the location and the power of the singularities in advance. We propose to do graded-grid refinement for elements containing a singularity, in such a way that we obtain the same error reduction factor as in elements in which the solution is smooth. For example, for a singularity located at a domain boundary, the element at the boundary is split in two, and then, within the same refinement step, the new element at the singularity is repeatedly split in two again, until the predicted error reduction factor matches the desired error reduction. We modify the predicted functional reduction factor, $\gamma(r)$, and the work increase ratio, $\eta(r)$, accordingly. We expect the correspondingly modified WEE and ACE strategies (MWEE and MACE) to generate a highly accurate grid sequence in an efficient way. This results in the following modified efficiency-based refinement strategies:

Modified WEE and ACE

1) Order the elements such that the local error, ϵ_j, satisfies $\epsilon_1 \geq \epsilon_2 \geq ... \geq \epsilon_{N_\ell}$.

2) Perform graded grid refinement for elements containing a singularity, i.e., if $u \in H^{s_j}(\tau_j)$, then graded grid refinement with m_j levels is used for any τ_j that needs to be refined, with m_j satisfying

$$\left(\frac{1}{2}\right)^{2m_j(s_j-1)} \approx \left(\frac{1}{2}\right)^{2p} \Rightarrow m_j = \lceil \frac{p}{s_j - 1} \rceil.$$

Note that we assume here that the error in the first, singular new element dominates the sum of the errors in the other new elements of the graded grid. This is a good approximation for a strong singularity. For elements in which the solution is smooth, single refinement is performed: $m_j = 1$. Let k_j be the number of new elements after τ_j is refined: $k_j = m_j + 1$.

3) The predicted functional reduction factor, $\gamma(r)$, and the work increase ratio, $\eta(r)$, are given by

$$\begin{aligned} \eta(r) &= 1 - r + \frac{\sum_{j \leq rN_\ell} k_j}{N_\ell}, \\ \gamma(r) &= 1 - f(r) + \left(\frac{1}{2}\right)^{2p} f(r). \end{aligned} \tag{4.33}$$

4) Find the optimal r defined in (3.9) for the MWEE strategy, and in (3.11) for the MACE strategy.

5) Repeat.

4.4.2 Performance of the modified WEE and ACE for singular solutions

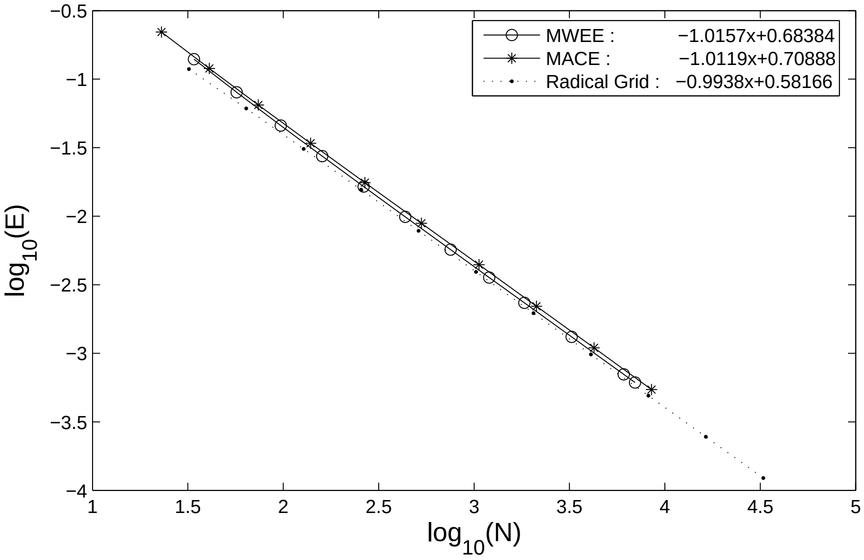
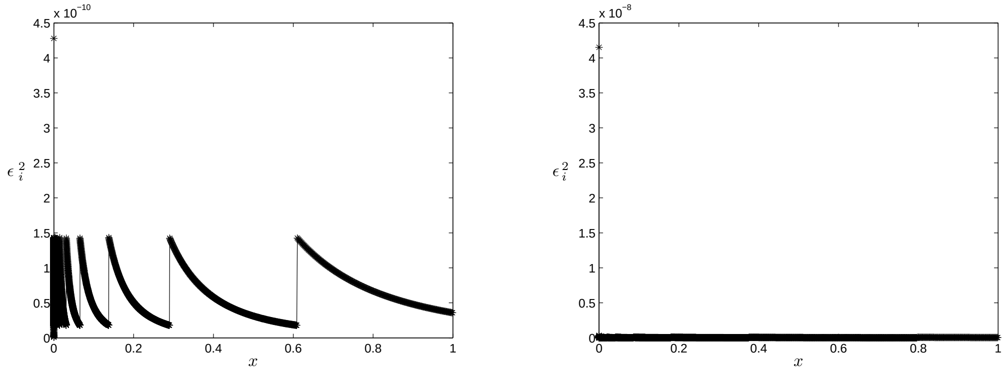


Figure 4.13: Error versus DOF, $\alpha = 0.6$ (singular case), $p = 1$.

We again choose $\alpha = 0.6$ and $p = 1$ for our example problem. There is a singularity at $x = 0$, with error reduction factor bound $(\frac{1}{2})^{0.2}$. Therefore, for the element that contains $x = 0$, we use $11-$graded refinement $(m = \lceil \frac{1}{0.1} \rceil)$. Numerical results are shown in Figs. 4.13-4.18.

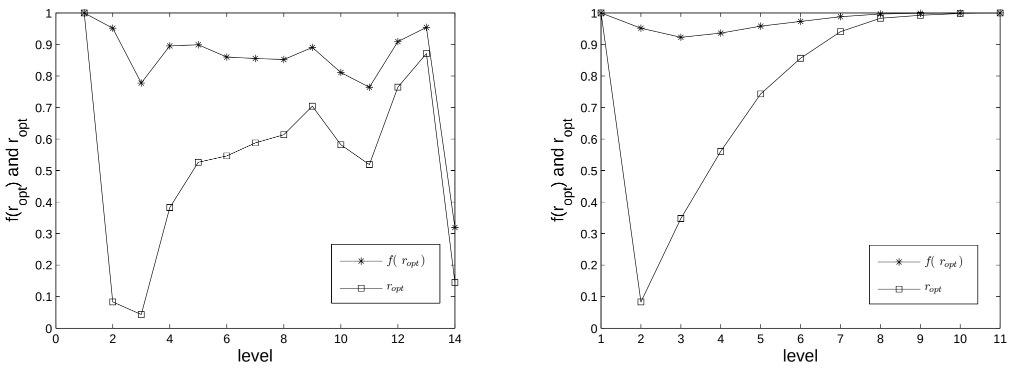
By comparing the numerical results for the modified strategies with the results for the original methods, we see the following. Both the MWEE and MACE strategies results in highly accurate grid sequences: the convergence rate is very close to the optimal rate (Fig. 4.13). Local error functionals on the final MWEE grid are more equally distributed than for the MACE grid. For the MWEE strategy, the local error functional in the singular element is only 3 times larger than in the smooth elements. However, for the MACE strategy, that ratio is as large as $1,000$ (Fig. 4.14). For the MWEE strategy, the number of elements, N_ℓ, increases much faster than for the WEE strategy, which reduces the work considerably (Fig. 4.15). However, there still exist a few small refinement steps. For the



(a) MWEE: $N_L = 6,975$, $E_L = 6.125e-4$, $L = 15$, total work $= 21,176$

(b) MACE: $N_L = 8,517$, $E_L = 5.443e-4$, $L = 12$, total work $= 17,044$

Figure 4.14: Local error functional, ϵ_i^2, versus grid location on the final grid, $\alpha = 0.6$ (singular case), $p = 1$.

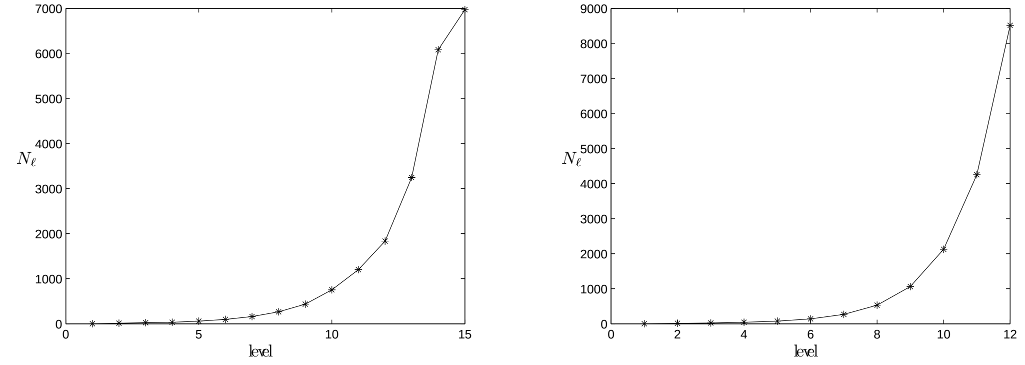


(a) MWEE

(b) MACE

Figure 4.15: Refined fraction of error functional, $f(r_{\text{opt}})$, versus level, ℓ, and refined fraction of elements, r_{opt}, versus level, ℓ, $\alpha = 0.6$ (singular case), $p = 1$.

MACE strategy, it seems that the strategy tends to do uniform refinement after several initial steps (Fig. 4.15(b)). Similar to the smooth solution case, the MWEE strategy may need slightly more work to reach the same error bound than the MACE strategy due to a few steps of small refinement (Fig. 4.17). However, since the MWEE strategy is slightly



(a) MWEE (b) MACE

Figure 4.16: Number of elements, N_ℓ, versus level, ℓ, $\alpha = 0.6$ (singular case), $p = 1$.

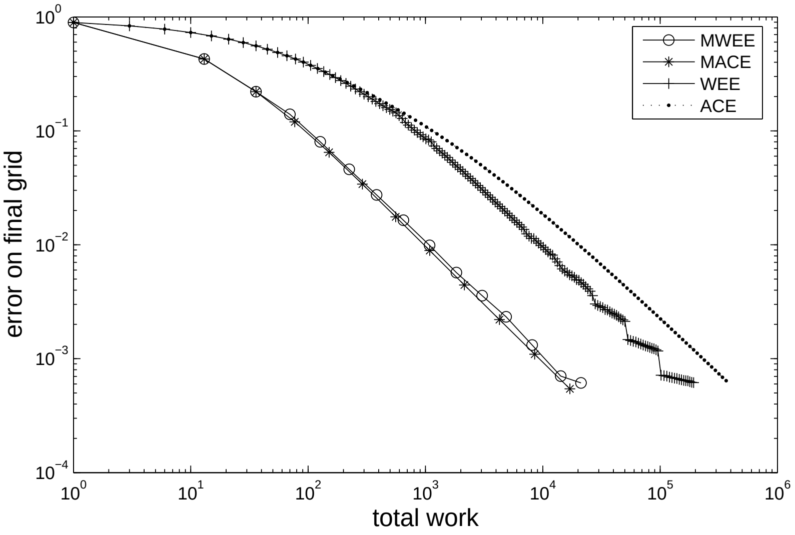
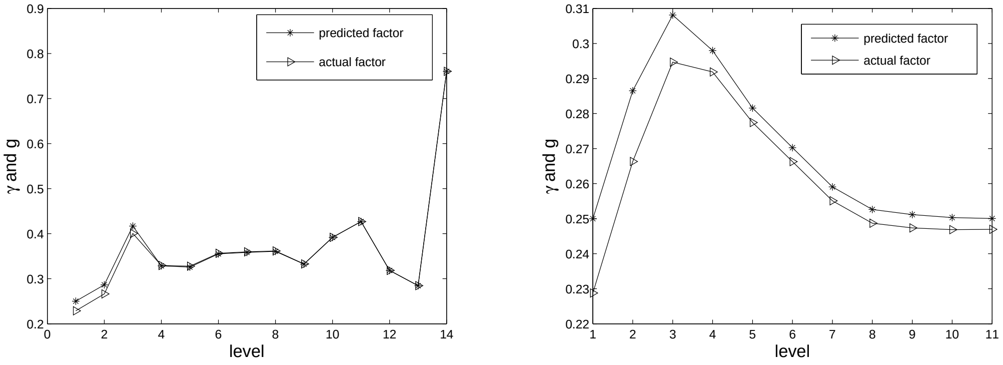


Figure 4.17: Final error, E_L, versus total work, $\sum_{\ell=1}^{L} N_\ell$, $\alpha = 0.6$ (singular case), $p = 1$.

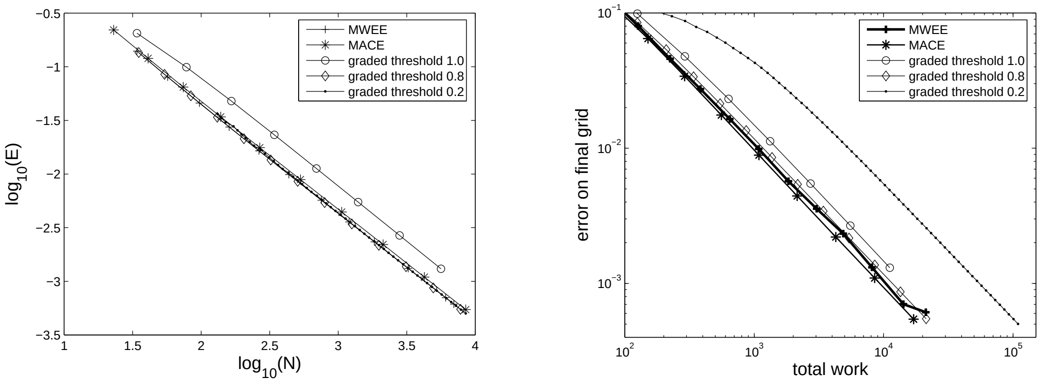
more accurate, the difference is very small. Again, the predicted functional reduction factors are good approximations of the actual factors for both strategies (Fig. 4.18).



(a) MWEE

(b) MACE

Figure 4.18: Predicted functional reduction factor, $\gamma(r_{opt})$, and actual functional reduction factor, g, versus level, ℓ, $\alpha = 0.6$ (singular case), $p = 1$.



(a)

(b)

Figure 4.19: : (a) Error versus DOF. (b) Final error, E_L, versus total work, $\sum_{\ell=1}^{L} N_\ell$. (Both $\alpha = 0.6$ (singular case), $p = 1$.)

4.4.3 Comparison with threshold-based refinement strategy

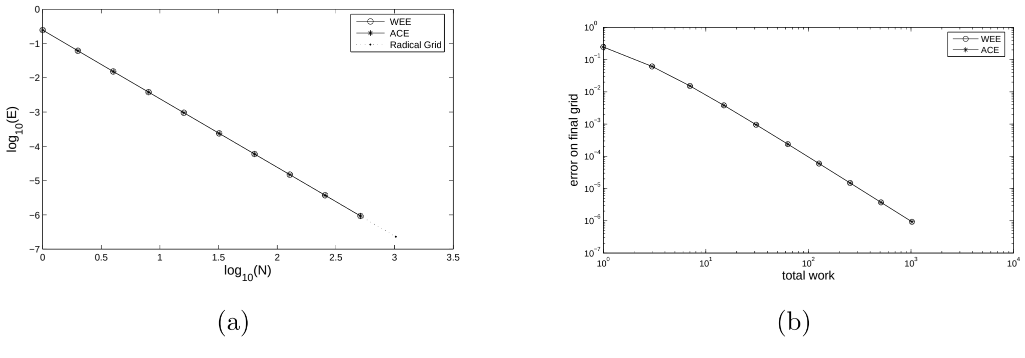
It is instructive to compare the MWEE and the MACE strategies with the threshold-based refinement strategy that chooses to refine a fixed fraction of the error functional on each level, i.e., $f_\ell(r) \equiv \theta$. The same graded grid refinement strategy is used for the elements that contain a singularity. We find the following for our example problem.

If we choose to refine a fixed fraction of the global error that is too small (less than the average of $f(r_{\text{opt}})$ in the modified efficiency-based strategies), *e.g.*, $\theta = 0.2$ in Fig. 4.19, then the resulting grid sequence is almost of optimal accuracy, but the total work increases significantly since N_ℓ increases slowly. A threshold value that is too large (larger than the average of $f_\ell(r_{\text{opt}})$ in the modified efficiency-based strategies), *e.g.*, $\theta = 1.0$ in Fig. 4.19, makes the number of elements, $\{N_\ell\}_{\ell=1}^{L}$, increase faster, but the large threshold results in a less accurate grid sequence. This implies that more total work is required to reach the same error bound. A threshold value that is close to the average of $f(r_{\text{opt}})$ in the modified efficiency-based strategies, namely, $\theta = 0.8$ in Fig. 4.19, results in a refinement process that performs similar to the efficiency-based refinement processes.

In conclusion, the efficiency-based refinement strategies automatically and adaptively choose a nearly optimal fraction of the error to be refined. As a result, they generate nearly optimal grid sequences in an efficient way, and there is no need to determine the optimal value of a threshold parameter.

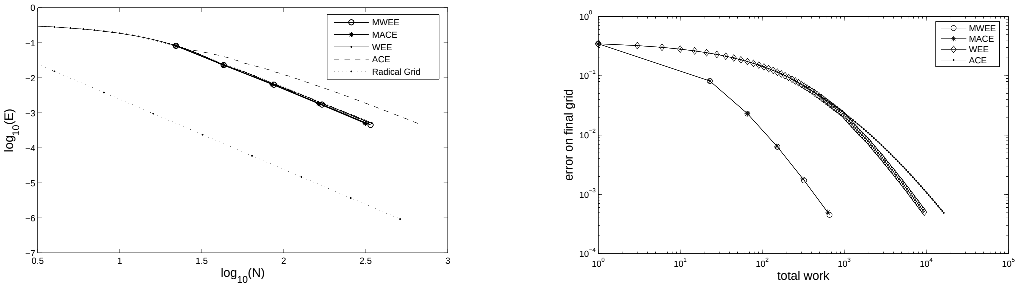
4.4.4 Results for $p = 2$

In this section, we briefly illustrate how the (M)WEE and (M)ACE strategies perform for finite element polynomial order $p = 2$.



(a) (b)

Figure 4.20: Efficiency-based refinement strategies for a smooth problem with $p = 2$ ($\alpha = 3.1$). (a) Error versus DOF. (b) Final error, E_L, versus total work, $\sum_{\ell=1}^{L} N_\ell$.



(a) (b)

Figure 4.21: Efficiency-based refinement strategies for a singular problem with $p = 2$ ($\alpha = 0.6$). (a) Error versus DOF. (b) Final error, E_L, versus total work, $\sum_{\ell=1}^{L} N_\ell$.

First, consider a smooth case with $\alpha = 3.1$, such that $u \in H^3$ and $u \notin H^4$. Error versus DOF and total work are plotted for WEE and ACE in Fig. 4.20. Both strategies lead to global refinement in every step for this example, and produce a sequence of grids that are very close to optimal radical grids.

Fig. 4.21 shows results for a highly singular case, with $\alpha = 0.6$, such that $u \in H^1$ and $u \notin H^2$. WEE and ACE produce small refinements, but this is remedied by the MWEE and MACE strategies, resulting, as before, in much less work for the modified strategies. It has to be noted, however, that the MWEE and MACE grids contain many more elements than optimal graded grids. This is probably due to the fact that the singularity is very strong for $\alpha = 0.6$ and $p = 2$, such that a geometrically graded grid with a grading factor of $\frac{1}{2}$ does not decrease the grid size fast enough in the vicinity of the singularity. Nevertheless, we can conclude that, within the constraint of refinement based on splitting cells in two, the MWEE and MACE strategies lead to an efficient refinement process.

4.5 Efficiency-based *hp*-refinement strategies in 1D

Assuming that we know a good approximation for the *p*-refinement error reduction factor for each element, we can apply the efficiency-based refinement strategies to *hp*-refinement processes. As we discussed before, the difficulty for *hp*-refinement strategies lies in predicting the *p*-error reduction ratio correctly. For our simple model problem, two approximations for the p-error reduction ratio are considered. One is obtained in [12], and is a sharp approximation specific to our model problem. The other one is obtained by using the minimization property of the finite element solution and the interpolation error, namely, the second part of formula (4.28). This p-ratio may be less sharp but more general, and may be used for other BVPs.

In this section, we first describe the hp-version of the WEE and ACE refinement strategies, with modified versions corresponding to singular cases. Then both strategies are applied to our model problem for a singular case. Numerical results are presented and compared with an optimal geometric hp-grid.

4.5.1 hp-version of the (M)WEE and (M)ACE refinement strategies

Consider an hp-finite element method for our simple example problem (4.1). Again, let $\mathcal{T} = \{0 = x_0 < x_1 < ... < x_N = 1\}$ be the grid and let $\mathbf{p} = \{p_1, p_2, ..., p_N\}$ be the degrees of the polynomials in the elements. Let $u_h \in S_D^{\mathbf{p},1}(\Omega, \mathcal{T})$ be the Galerkin finite element solution of (4.1) and $\epsilon_i^2(p_i) = ||u' - u_h'||_{0,T_i}^2$ the local error functional in element $T_i = [x_{i-1}, x_i]$ with polynomial of degree p_i. Then we have the following theorem:

Theorem 4.3. *[12] Let $\epsilon_i^2(p_i)$ be the local error of the finite element solution of problem (4.1), and let*

$$\tau_i = [x_{i-1}, x_i],\ \theta_i = \frac{\sqrt{x_i} - \sqrt{x_{i-1}}}{\sqrt{x_i} + \sqrt{x_{i-1}}}.$$

Then

$$\epsilon_1^2(p_1) \approx \frac{h_1^{2\alpha-1}}{p_1^{4\alpha-2}}. \tag{4.34}$$

If θ_i $(2 \le i \le N(\mathcal{T}))$ is not close to 1, then

$$\epsilon_i^2(p_i) \approx \left\{ h_i^{\alpha-1/2} \left(\frac{1-\theta_i^2}{2\theta_i} \right)^{\alpha-1} \frac{\theta_i^{p_i}}{p_i^{\alpha}} \right\}^2. \tag{4.35}$$

Since higher order polynomials do not provide significant improvement in error reduction in the singular element, we only consider h-refinement for the first element, which contains the singularity. Then we have the error functional reduction factor bound $(\frac{1}{2})^{2\alpha-1}$ as in (3.16) and (4.28). For an element T_i that does not contain a singularity, note that θ_j is small, and again we obtain the same h-reduction factor bound, $(\frac{1}{2})^{2p_j}$, as before (see (3.14) and (4.28). Moreover, if we double the degree of the polynomial p_i, we obtain the p-reduction factor bound as follows:

$$\frac{\epsilon_i(2p_i)}{\epsilon_i(p_i)} \approx \left(\frac{\theta_i^{p_i}}{2^\alpha}\right)^2 . \tag{4.36}$$

We can then develop an hp-version of the MWEE strategy as follows:

***hp*-version MWEE and MACE**

1) Order the elements such that the local error, ϵ_j, satisfies $\epsilon_1 \geq \epsilon_2 \geq ... \geq \epsilon_{N_\ell}$.

2) Let $p_{\max}$ be the maximal polynomial order to be used in the refinement process. Three types of refinement are used, depending on the element. We use a graded grid with $p = 1$ for the elements containing a singularity, in such a way that the predicted error-reduction factor attains $\frac{1}{4}$. (Note that a target reduction factor of up to $\frac{1}{2^{p_{max}}}$ could be used, but we choose $\frac{1}{4}$ for simplicity in our numerical tests.) For elements without a singularity, p-refinement (doubling p) is used if the solution is locally smooth enough (which, in general, can be detected *a posteriori* by comparing predicted and observed error-functional reduction ratios) and $p < p_{\max}$. Otherwise, $h-$refinement is used and the degree p is inherited by both sub-elements. As before, we assume that the work of solving the linear systems is proportional to the number of DOF. Then, doubling p or splitting the element into two elements with order p has the same computational complexity.

3) Calculate the MWEE or MACE efficiency functions and find the optimal fraction of elements to be refined, r_{opt}.

4) Refine elements τ_j, $1 \leq j \leq r_{opt}N_\ell$.

5) Repeat.

For a general problem different from (4.1), it may be difficult to find a sharp approximation formula for the error reduction in the case of p-refinement. Hence we are interested in seeking a more general but possibly less sharp p-error reduction factor. Recall that for elements T_i in which the solution is smooth (at least in $H^{p_i+1}(T_i)$ if order p_i elements are used), we can use formula (4.28) to get

$$\epsilon_i^2(p_i) \leq \left(\frac{h_i}{2}\right)^{2p_i} \frac{1}{(2p_i)!}||u||^2_{H^{p_i+1}(T_i)}. \tag{4.37}$$

Assuming that $\frac{1}{(2p_i)!}||u||^2_{H^{p_i+1}(T_i)} \leq M$, where M is a constant, we obtain the following general p-error reduction factor

$$\frac{\epsilon_i^2(2p_i)}{\epsilon_i^2(p_i)} \approx \left(\frac{h_i}{2}\right)^{2p_i}, \tag{4.38}$$

for elements T_i that do not contain a singularity.

Just as in the case of h-refinement, we seek some kind of optimal grid for comparison. Suppose the locations of the grid points are given by

$$x_i = q^{N(\mathcal{T})-i}, \quad 0 < q < 1, \; j = 1, 2, ..., N(\mathcal{T}). \tag{4.39}$$

Let $\theta_i = \theta = \frac{1-\sqrt{q}}{1+\sqrt{q}}, \forall i : 1 \leq i \leq N(\mathcal{T})$. It was shown in [12] that the optimal degree distribution of $\mathbf{p}$ for these grid locations tends to a linear distribution with slope

$$s_o = (\alpha - 1/2)\frac{\log q}{\log \theta}. \tag{4.40}$$

Furthermore, the optimal geometric grid factor q and linear slope s_o combination is given by

$$q_{opt} = (\sqrt{2}-1)^2, \; s_{opt} = 2\alpha - 1. \tag{4.41}$$

4.5.2 Numerical results and comparisons

We apply the hp-version MWEE and MACE strategies with the two $p-$refinement reduction factors given by (4.36) and (4.38) to our model problem 4.1 with $\alpha = 0.6$, and compare the numerical results with the optimal geometric grid with $q = q_{opt}$ and $q = \frac{1}{2}$; see Figs. 3.22 and 3.23. In the numerical results, we carry out the refinement process until $E_L(u_h, f) \leq 5e{-}3$ on final grid level L.

Observe that the hp-finite element methods result in much faster error convergence rates than the h-finite element method. Both the hp-MACE and hp-MWEE strategies

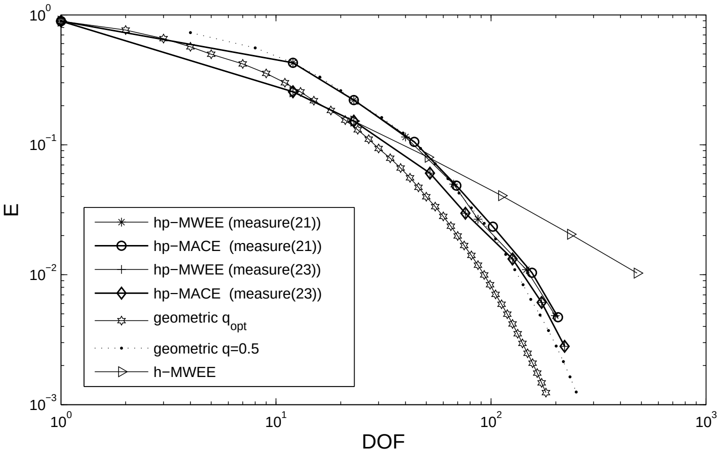


Figure 4.22: Error versus DOF, $\alpha = 0.6$ (singular case), $p = 1$.

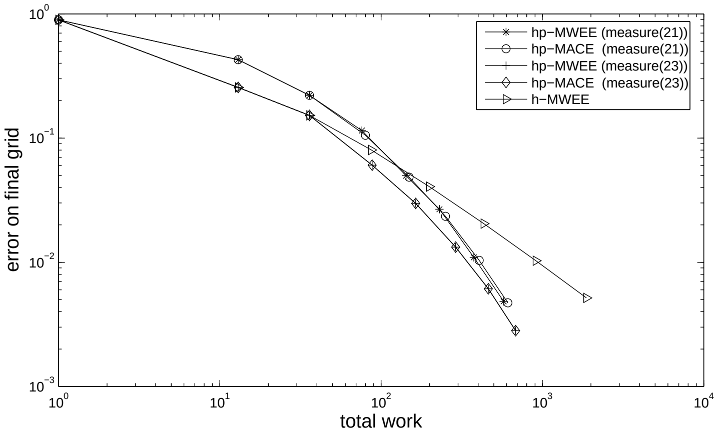


Figure 4.23: Final error, E_L, versus total work, $\sum_{\ell=1}^{L} N_\ell$, $\alpha = 0.6$ (singular case), $p = 1$.

result in a highly accurate grid sequence with rate-of-error convergence very close to the geometrical grid with grading number $q = 0.5$. Also, the refinement process is efficient, *i.e.*, the number of DOF increases fast w.r.t. the refinement level. Surprisingly, $hp-$refinement strategies using the more general, but less accurate, error reduction factor (4.38), result in a better grid sequence than with the more accurate Babuška factor, (4.36). The results are even better than the optimal geometric grid sequence when the number of DOF is small.

Note that for all elements which do not contain a singularity, the general p-reduction factor is always less than the h-reduction factor. It follows that p-refinement is always considered first for such elements before reaching $p_{\max}$. However, why using the general factor results in a better grid than using the more accurate factor is still not clear. More work needs to be done to verify whether the general factor (4.38) works well for more general problems. Furthermore, the convergence rate of the optimal geometric grid is faster than the MWEE and MACE and the geometrical grid with $q = 0.5$. Unlike the case of the h-version, the difference is noticeable here.

Chapter 5

2D results: FOSLS Finite Element Methods

In this chapter, we explore the use of the proposed efficiency-based refinement strategies in two spatial dimensions. In these initial considerations, we only discuss problems with sufficiently smooth solutions.

To illustrate the broad applicability of our refinement strategies, we solve a 2D Poisson equation using a first-order system least-squares (FOSLS) finite element method, rather than the Galerkin method that was used for the 1D test problems. We choose FOSLS mainly because FOSLS possesses one advantageous feature that other FEMs do not possess: FOSLS functionals provide an easily computable, and locally sharp a-posteriori error estimate that can be used for adaptive refinement.

This chapter is organized as follows. In section 5.1, concepts of FOSLS finite element methods are introduced. We review the FOSLS methodology, and discuss how FOSLS functionals can be used in adaptive refinement. Then, in section 5.2, FOSLS for second-order elliptic PDEs is specifically discussed. Lastly, in section 5.3, we apply the proposed efficiency-based refinement strategies to a 2D Poisson equation and present numerical results.

5.1 First-Order System Least Squares (FOSLS) finite element methods

5.1.1 The FOSLS methodology

We briefly review the FOSLS methodology here. For details, one can refer to [13, 14, 15].

We start with a (typically second order elliptic type) PDE, or a system of PDEs together with appropriate boundary conditions:

$$\begin{cases} \mathbf{Lw} = \mathbf{f} \text{ in } \Omega \subset \mathcal{R}^d, d = 2, 3, \\ \mathbf{Rw} = \mathbf{g} \text{ on } \Gamma = \partial\Omega, \end{cases} \tag{5.1}$$

where $\mathcal{L}$ is a linear differential operator and $\mathcal{R}$ is a linear boundary operator.

By introducing new variables, the FOSLS methodology yields a system of first-order (linear) PDEs,

$$\begin{cases} L_i \mathbf{u} = f_i, \ i = 1, 2, ..., M, \text{ in } \Omega, \\ R_i \mathbf{u} = g_i, \ i = 1, 2, ..., N, \text{ on } \Gamma, \end{cases} \tag{5.2}$$

which is equivalent to the original problem (5.1). The resulting FOSLS L^2-functional is the scaled sum of L^2-norms of the residuals of system (5.2):

$$\mathcal{G}(\mathbf{u}, \mathbf{f}) = \sum_{i=1}^{M} ||a_i(L_i\mathbf{u} - f_i)||^2_{L^2(\Omega)} + \sum_{i=1}^{N} ||b_i(R_i\mathbf{u} - g_i)||^2_{L^2(\Gamma)}. \tag{5.3}$$

Here a_i and b_i are weight functions to improve the FOSLS functional convergence, e.g., see [16]. For simplicity, we can assume $a_i = b_i = 1$. We only consider the L^2 case here, even though the L^2 norm can be replaced by other suitable norms, e.g., the H^{-1} norm, see [17], or the $H^{\frac{1}{2}}$ norm for boundary terms, see [15]. The FOSLS minimization problem is

$$\mathbf{u} = \arg \min_{\mathbf{v} \in \mathbf{W}} \mathcal{G}(\mathbf{v}, \mathbf{f}), \tag{5.4}$$

where $\mathbf{W}$ is a proper Hilbert space with suitable norm $||| \cdot |||$, often (equivalent to) a product of H^1 spaces. To derive the weak form, assuming that $\mathbf{u}$ solves (5.4), we have for any function $\mathbf{v} \in \mathbf{W}$,

$$\frac{d\mathcal{G}(\mathbf{u} + \lambda\mathbf{v}, \mathbf{f})}{d\lambda}|_{\lambda=0} = 0. \tag{5.5}$$

Therefore

$$\sum_{i=1}^{N} 2\left(a_i(L_i\mathbf{u}-f_i), a_iL_i\mathbf{v}\right)_{0,\Omega} + \sum_{i=1}^{M} 2\left(b_i(R_i\mathbf{u}-g_i), b_iR_i\mathbf{v}\right)_{0,\Gamma} = 0 \qquad \forall\mathbf{v}\in\mathbf{W}. \tag{5.6}$$

Hence, define

$$\mathcal{B}(\mathbf{u},\mathbf{v}) = \sum_{i=1}^{N}\left(a_iL_i\mathbf{u}, a_iL_i\mathbf{v}\right)_{0,\Omega} + \sum_{i=1}^{M}\left(b_iR_i\mathbf{u}, b_iR_i\mathbf{v}\right)_{0,\Gamma}, \tag{5.7}$$

and

$$\mathcal{F}(\mathbf{v}) = \sum_{i=1}^{N}\left(a_if_i, a_iL_i\mathbf{v}\right)_{0,\Omega} + \sum_{i=1}^{M}\left(b_ig_i, b_iR_i\mathbf{v}\right)_{0,\Gamma}. \tag{5.8}$$

The weak form of the minimization problem (5.4) is

$$\begin{aligned} &\text{find } \mathbf{u}\in\mathbf{W} \text{ such that} \\ &\qquad\qquad \mathcal{B}(\mathbf{u},\mathbf{v}) = \mathcal{F}(\mathbf{v}) \qquad \forall\mathbf{v}\in\mathbf{W}. \end{aligned} \tag{5.9}$$

Assume that the bilinear form $\mathcal{B}(\cdot,\cdot)$ is continuous and coercive in $\mathbf{W}$, i.e., there exist constants C_1 and C_2 such that

$$\begin{aligned} &\mathcal{B}(\mathbf{u},\mathbf{v}) \le C_1|||\mathbf{u}|||\ |||\mathbf{v}|||, \\ &C_2|||\mathbf{u}|||^2 \le \mathcal{B}(\mathbf{u},\mathbf{u}), \qquad \forall\mathbf{u},\mathbf{v}\in\mathbf{W}. \end{aligned} \tag{5.10}$$

Then, by the Lax-Milgram theorem (Theorem 2.6), the weak problem (5.9) is well posed.

The inclusion of the boundary residual in (5.3) allows the use of minimization spaces $\mathbf{W}$ that are not constrained to satisfy the boundary condition, i.e., such conditions are enforced weakly through the variational principle. This is advantageous whenever the boundary condition is difficult to satisfy computationally and represents a beneficial feature of least-squares based methods. If, on the other hand, the boundary condition can be easily imposed, one can consider (5.3) with the boundary term omitted. Then, the functions belonging to the space $\mathbf{W}$ should be required to satisfy the boundary condition, i.e., the boundary condition is enforced strongly or directly on candidate minimizers. In this case, the admissible set may not be a linear space. In order to formulate the well-posedness, one can extend the boundary data to all of Ω by introducing $\mathbf{g}_D$. By introducing a new variable $\mathbf{w} = \mathbf{u} - \mathbf{g}_D$, one can formulate an equivalent FOSLS problem in a subspace of $\mathbf{W}$ with zero boundary conditions. However, as we discussed in Chapter 3, the discrete weak problem involves using an interpolant of $\mathbf{g}_D$. Hence, the use of $\mathcal{I}^h\mathbf{g}_D$ may affect the convergence of the finite element solution, especially in higher dimensional spaces, see (4.25).

Denote by $\mathbf{W}_h$ a finite dimensional subspace (often consisting of piecewise polynomials). The corresponding discrete variational problem is given

$$\begin{aligned}&\text{find } \mathbf{u}^h \in \mathbf{W}^h \text{ such that}\\ &\qquad\qquad \mathcal{B}(\mathbf{u}^h, \mathbf{v}) = \mathcal{F}(\mathbf{v}) \qquad \forall \mathbf{v} \in \mathbf{W}^h.\end{aligned} \tag{5.11}$$

Together with (5.9), this implies that

$$|||\mathbf{u} - \mathbf{u}^h||| \le C \min_{\mathbf{v}\in\mathbf{W}^h} |||\mathbf{u} - \mathbf{v}||| \tag{5.12}$$

5.1.2 Local FOSLS functionals

The FOSLS functional is a sum of integrals and, hence, can be evaluated over any subdomains of Ω. Again, let $\mathcal{T} = \cup T$ be a partition of Ω. We call

$$\mathcal{G}_T(\mathbf{u}, \mathbf{v}) = \sum_{i=1}^{M} ||a_i(L_i\mathbf{u} - f_i)||^2_{L^2(T)} + \sum_{i=1}^{N} ||b_i(R_i\mathbf{u} - g_i)||^2_{L^2(\partial T)}. \tag{5.13}$$

the local FOSLS functional. Obviously, the FOSLS functional can be written as the sum of all local functionals over all elements:

$$\mathcal{G}(\mathbf{u}, \mathbf{f}) = \sum_{T\in\mathcal{T}} \mathcal{G}_T(\mathbf{u}, \mathbf{f}). \tag{5.14}$$

Here, we discuss how local functionals can be used for adaptive refinement. We first consider $\mathcal{G}_T(\mathbf{u}, \mathbf{v})$ as a candidate a-posteriori error estimate. It has been suggested that the value of the local least-squares functional on an element in a given mesh can be used as an a-posteriori error estimate, see [11, 18]. For any given function $\mathbf{u}^h$ in the finite-dimensional space $\mathbf{W}^h \in \mathbf{W}$, the coercivity bound implies that

$$\mathcal{G}(\mathbf{u}^h, \mathbf{f}) = \mathcal{G}(\mathbf{u}^h, L\mathbf{u}) = \mathcal{G}(\mathbf{u}^h - \mathbf{u}, 0) = \mathcal{B}(\mathbf{u}^h - \mathbf{u}, \mathbf{u}^h - \mathbf{u}) \ge C_2|||\mathbf{u}^h - \mathbf{u}|||. \tag{5.15}$$

This implies

$$|||\mathbf{u}^h - \mathbf{u}|||^2 \leq \frac{1}{C_2} \sum_{T \in \mathcal{T}} \mathcal{G}_T(\mathbf{u}^h, \mathbf{f}). \tag{5.16}$$

If all of the local error estimates $\mathcal{G}(\mathbf{u}^h, \mathbf{f})$ are small, then the global error is also small.

In the literature, there are many examples (cf [4, 13, 14, 15]) showing that a proof analogous to the continuity bound can be used to establish a similar bound on the local functional, i.e., we have

$$\mathcal{G}_T(\mathbf{u}^h, \mathbf{f}) = \mathcal{B}_T(\mathbf{u}^h - \mathbf{u}, \mathbf{u}^h - \mathbf{u}) \leq C_1 |||\mathbf{u}^h - \mathbf{u}|||_T^2 \tag{5.17}$$

for any T, which implies

$$|||\mathbf{u}^h - \mathbf{u}|||_T^2 \geq \frac{1}{C_1} \mathcal{G}_T(\mathbf{u}^h, \mathbf{f}) \tag{5.18}$$

for any $T \in \mathcal{T}$.

Bounds (5.16) and (5.18) indicate that the local FOSLS functional $\mathcal{G}_T(\mathbf{u}^h, \mathbf{f})$ can be used as an a-posteriori error estimate. In fact, these two bounds do not depend on how the approximation $\mathbf{u}^h \in \mathbf{W}^h$ is obtained. For standard Galerkin methods, bounds of this type usually depend on the fact that $\mathbf{u}^h$ must be a discrete finite element solution, and moreover they can be very tedious to derive, see [3].

We show how the local FOSLS functional can be used to obtain a local a-priori error estimate in order to provide an (fine) initial mesh for adaptive refinement.

For a given element T, define the estimate

$$\overline{\eta}_T := \sqrt{\min_{\mathbf{v}^h \in \mathbf{W}^h} \mathcal{G}_T(\mathbf{v}^h, \mathbf{f})}, \tag{5.19}$$

which is entirely local to T, and thus can be evaluated efficiently in parallel, see [11]. Then, for any partition $\mathcal{T}$ that contains T, we have

$$\overline{\eta}_T \leq \mathcal{G}_T(\mathbf{v}^h, \mathbf{f}), \qquad \forall \mathbf{v}^h \in \mathbf{W}^h_{\mathcal{T}}. \tag{5.20}$$

Together with (5.18), this implies that

$$\overline{\eta}_T \leq \mathcal{G}_T(\mathbf{u}^h, \mathbf{f}) \leq C_1 |||\mathbf{u}^h - \mathbf{u}|||_T^2 \tag{5.21}$$

for any partition $\mathcal{T}$ with associated finite element space $\mathbf{W}_{\mathcal{T}}^h$ and finite element approximation $\mathbf{u}^h$. A global a-priori lower bound for the value of the FOSLS functional for the given partition $\mathcal{T}$ can then be calculated:

$$\sum_{T\in\mathcal{T}} \overline{\eta}_T^2 \leq \mathcal{G}(\mathbf{u}^h, \mathbf{f}). \tag{5.22}$$

Thus, for a given initial grid $\mathcal{T}_0$, one can use the indicator in (5.19) to refine $\mathcal{T}_0$ adaptively until the refined grid $\mathcal{T}_1$ satisfies $\sum_{T\in\mathcal{T}} \overline{\eta}_T^2$ is less than an acceptable bound. Then, one can solve the associated discrete variational problem and use local FOSLS functionals $\mathcal{G}_T(\mathbf{u}^h, \mathbf{f})$ as an a-posteriori error estimate for further refinement.

5.2 Introduction to FOSLS for second-order PDEs

Before we illustrate applicability of our refinement strategies using FOSLS, we study FOSLS for solving second-order elliptic partial differential equations. For details, one can refer to [13, 14].

Assume Ω is a bounded, open, connected domain in $\mathcal{R}^d$ ($d = 2$ or 3) with Lipschitz boundary $\partial\Omega$. Consider the following second-order elliptic BVP:

$$\begin{cases} -\nabla\cdot(\underline{A}\nabla p) + \mathbf{X}p = f, & \text{in } \Omega, \\ p = 0, & \text{on } \Gamma_D, \\ \mathbf{n}\cdot\underline{A}\nabla p = 0, & \text{on } \Gamma_N, \end{cases} \tag{5.23}$$

where $\underline{A}(x)$ is a $d\times d$ symmetric matrix with entries in $L^\infty(\Omega)$, $\mathbf{X}$ is an at most first-order linear differential operator, $\Gamma_D \cap \Gamma_N = \Gamma$ is a partitioning of the boundary of Ω, and $\mathbf{n}$ is the outward unit vector normal to the boundary. $\underline{A}$ is assumed to be uniformly symmetric positive definite and scaled appropriately: there exist positive constants

$$0 < \lambda \leq 1 \leq \Lambda \tag{5.24}$$

such that

$$\lambda\xi^T\xi \leq \xi^T\underline{A}\xi \leq \Lambda\xi^T\xi \tag{5.25}$$

for all $\xi \in \mathcal{R}^d$ and almost all $x \in \overline{\Omega}$.

Following the FOSLS methodology, the flux variable is introduced as

$$\mathbf{u} = \underline{A}\nabla p. \tag{5.26}$$

Then problem (5.23) can be rewritten as a first-order system of PDEs as follows:

$$\begin{cases} \mathbf{u} - \underline{A}\nabla p = \mathbf{0}, & \text{in } \Omega, \\ -\nabla \cdot \mathbf{u} + \mathbf{X}p = f, & \text{in } \Omega, \\ p = 0, & \text{on } \Gamma_D, \\ \mathbf{n} \cdot \mathbf{u} = 0, & \text{on } \Gamma_N. \end{cases} \tag{5.27}$$

Under appropriate assumptions on Γ_D and $\mathbf{X}$, the associated weak form of the system (5.27) is uniquely solvable in $H^1(\Omega)$ for any $f \in H^{-1}(\Omega)$ or uniquely solvable in $H^1(\Omega)/\mathcal{R}$ (cf. [20]) iff f satisfies the compatibility condition $\int_\Omega f = 0$ for the case when $\Gamma_D = \emptyset$.

Define the subspaces

$$\mathbf{W}_0(\text{div};\Omega) = \{\mathbf{v} \in H(\text{div};\Omega) : \mathbf{n} \cdot \mathbf{v} = 0 \text{ on } \Gamma_N\},$$
$$V = \{q \in H^1(\Omega) : q = 0 \text{ on } \Gamma_D\}.$$

Consider the FOSLS functional of system (5.27):

$$\mathbf{G}_{\text{grad-div}}(\mathbf{v}, q; f) = ||\mathbf{v} - \underline{A}\nabla q||^2_{(L^2(\Omega))^2} + || - \nabla \cdot \mathbf{v} + \mathbf{X}q - f||^2_{L^2(\Omega)} \tag{5.28}$$

for $(\mathbf{v}, q) \in \mathbf{W}_0(\text{div};\Omega) \times V$. Under a certain assumption (cf.[13]), it is shown that $\mathbf{G}_{\text{grad-div}}(\mathbf{v}, q; 0)$ is equivalent to the $H(\text{div};\Omega) \times H^1(\Omega)$ norm on $\mathbf{W}_0(\text{div};\Omega) \times V$. This implies optimal convergence for finite element subspaces of $H(\text{div},\Omega) \times H^1$. However, since the associated bilinear form is not elliptic w.r.t. the $(H^1(\Omega))^{d+1}$ norm, additive multigrid algorithms applied to the discrete functionals are not optimally convergent.

This functional is modified by adding a compatible *curl* constraint and imposing additional boundary conditions on the first-order system (5.27), e.g., see [14]. Let $\nabla\times$ denote the curl operator. Note that if $\mathbf{u}$ is sufficiently smooth, then the properly scaled solution $\underline{A}^{-1}\mathbf{u}$ is curl free, i.e., $\nabla \times (\underline{A}^{-1}\mathbf{u}) = \mathbf{0}$, and the homogeneous Dirichlet boundary condition on Γ_D implies that the tangential flux satisfies

$$\mathbf{n} \times (\underline{A}^{-1}\mathbf{u}) = 0, \qquad \text{on } \Gamma_D. \tag{5.29}$$

The augmented system for (5.27) is then

$$\begin{cases} \mathbf{u} - \underline{A}\nabla p = \mathbf{0}, & \text{in } \Omega, \\ -\nabla \cdot \mathbf{u} + \mathbf{X}p = f, & \text{in } \Omega, \\ \nabla \times \underline{A}^{-1}\mathbf{u} = \mathbf{0}, & \text{in } \Omega, \\ p = 0, & \text{on } \Gamma_D, \\ \mathbf{n} \times \mathbf{u} = 0, & \text{on } \Gamma_N, \\ \mathbf{n} \times (\underline{A}^{-1}\mathbf{u}) = 0, & \text{on } \Gamma_D. \end{cases} \tag{5.30}$$

The associated FOSLS functional is given by

$$\begin{aligned} \mathcal{G}_{\text{grad-div-curl}}(\mathbf{v}, q; f) = ||\mathbf{v} - \underline{A}\nabla q||^2_{(L^2(\Omega))^2} + || - \nabla \cdot \mathbf{v} + \mathbf{X}q - f||^2_{L^2(\Omega)} \\ + ||\nabla \times (\underline{A}^{-1}\mathbf{v})||^2_{(L^2(\Omega))^{2d-3}}. \end{aligned} \tag{5.31}$$

Let

$$H(\text{curl } \underline{A}; \Omega) = \left\{ \mathbf{v} \in (L^2(\Omega))^d : \nabla \times (\underline{A}^{-1}\mathbf{v}) \in (L^2(\Omega))^{2d-3} \right\}. \tag{5.32}$$

This is a Hilbert space with norm

$$||\mathbf{v}||_{H(\text{curl } \underline{A};\Omega)} = \left(||\mathbf{v}||^2_{(L^2(\Omega))^d} + ||\nabla \times (\underline{A}^{-1}\mathbf{v})||^2_{(L^2(\Omega))^{2d-3}} \right)^{1/2}. \tag{5.33}$$

Define the subspaces

$$\mathbf{W}_0(\text{curl } \underline{A}; \Omega) = \left\{ \mathbf{v} \in H(\text{curl } \underline{A}; \Omega) : \mathbf{n} \times (\underline{A}^{-1}\mathbf{v}) = \mathbf{0} \text{ on } \Gamma_D \right\},$$

and

$$\mathbf{W} = \mathbf{W}_0(\text{div}; \Omega) \cap \mathbf{W}_0(\text{curl } \underline{A}; \Omega).$$

Then it follows that $\mathcal{G}_{\text{grad-div-curl}}(\mathbf{v}, q; 0)$ is equivalent to the $(H(\text{div}; \Omega) \cap H(\text{curl } \underline{A}; \Omega)) \times H^1(\Omega)$ for all $(\mathbf{v}, p) \in \mathbf{W} \times V$. Moreover, in [14], the functional $\mathcal{G}_{\text{grad-div-curl}}(\mathbf{v}, q; 0)$ is equivalent to the $H^1(\Omega)^{d+1}$ norm on $\mathbf{W} \times V$ under some additional hypotheses on $\underline{A}$ and Ω, see [14].

Let $\mathcal{T}_h$ be a regular partition of Ω into finite elements. Assume two finite element subspaces $\mathbf{W}_h \subset \mathbf{W}$ and $V_h \subset V$ are defined on $\mathcal{T}$. Let $(\mathbf{u}_h, p_h)$ be the finite element solutions. Suppose the conditions in Theorem 2.12 are satisfied such that the approximation properties of the finite element interpolant hold. Then by the Céa Theorem, the general error estimate for finite element methods holds. More precisely, suppose that $p \in H^{r+1}(\Omega)$ and $\mathbf{u} \in H^{k+1}(\Omega)^2$, where r is the polynomial order of V_h and k is the polynomial order of $\mathcal{W}_h$.Let $s = \min(k, r)$. Then

$$||p - p_h||_{H^1(\Omega)} + ||\mathbf{u} - \mathbf{u}_h||_{H^1(\Omega)^d} \leq Ch^s(||p||_{H^{s+1}(\Omega)} + ||\mathbf{u}||_{H^{s+1}(\Omega)^d}). \tag{5.34}$$

Also, one can show that if $\mathbf{u}$ and p belong to fractional Sobolev spaces, the estimates in Theorem 2.14 also hold.

5.3 Numerical performance in 2D

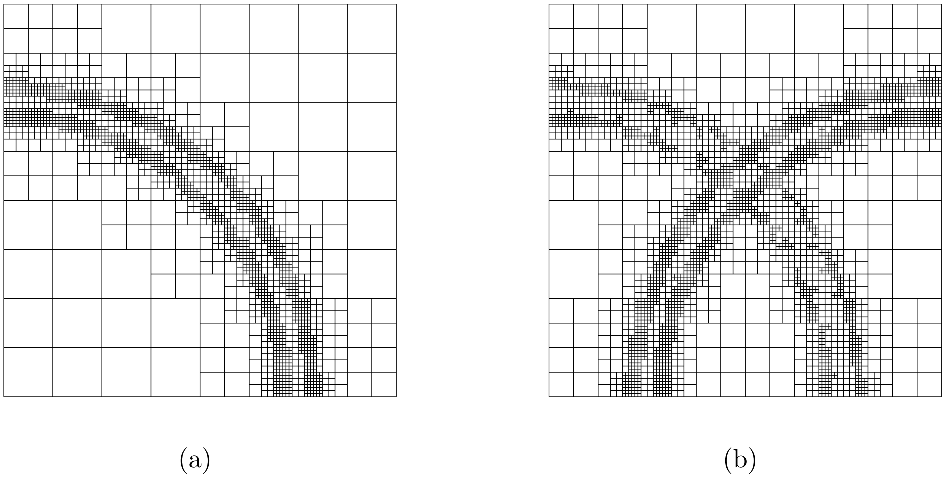


Figure 5.1: Adaptively refined grids using the ACE refinement strategy for 2D problems with $p = 2$. (a) Single arc on a unit square domain. (b) Double arc on a unit square domain.

The following 2D finite element problem is considered to illustrate the efficiency-based refinement strategies. We solve the Poisson equation with inhomogeneous Dirichlet boundary condition

$$\begin{cases} -\Delta p = f & \text{in} \quad \Omega, \\ p = g & \text{on} \quad \partial\Omega, \\ \Omega = (0,1) \times (0,1), \end{cases} \tag{5.35}$$

with the right-hand side f and boundary conditions g chosen such that the solution is given by

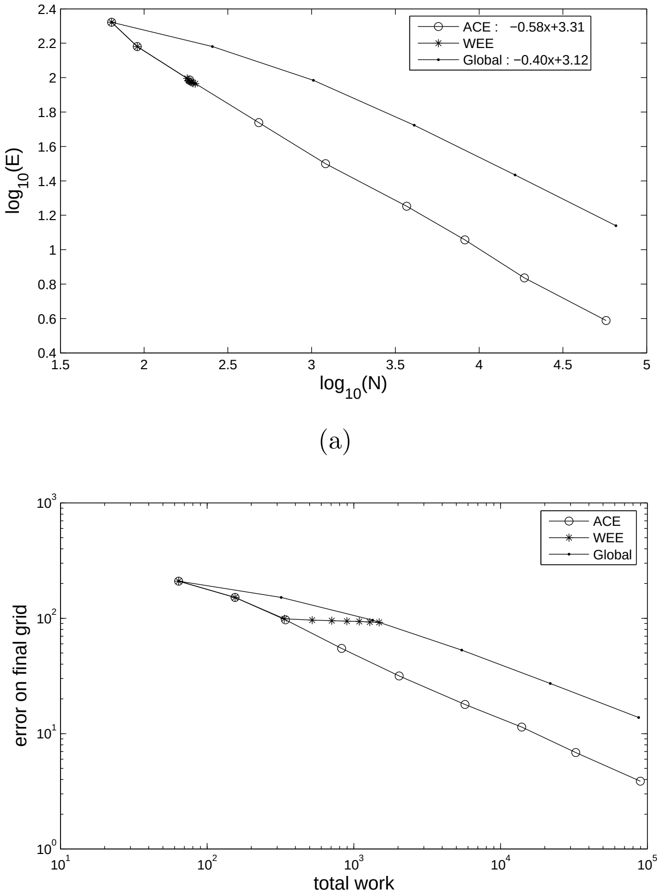
$$p(r,\theta) = \begin{cases} 1 & r \leq r_0, \\ h(r) & r_0 \leq r \leq r_1, \\ 0 & r_1 \leq r. \end{cases} \tag{5.36}$$

Here, (r,θ) are the usual polar coordinates and $h(r)$ is the unique polynomial of degree five such that $p \in C^2(\Omega)$. We choose $r_0 = 0.7$ and $r_1 = 0.8$. The solution of this test problem takes on the unit value in the lower left corner of the domain, and is zero elsewhere, except for a steep gradient in the thin strip $0.7 \leq r \leq 0.8$. Fig. 5.1(a) shows the grid obtained after several refinement steps for this model problem.

BVP (5.35) is rewritten as a first-order system BVP

$$\begin{cases} -\nabla \cdot U = f \quad \text{in} \quad \Omega, \\ \qquad U = \nabla p \\ \nabla \times U = 0 \\ \qquad p = g \quad \text{on} \quad \partial\Omega, \\ \quad \vec{\tau} \cdot U = \dfrac{\partial g}{\partial \tau} \\ \qquad \Omega = (0,1) \times (0,1), \end{cases} \tag{5.37}$$

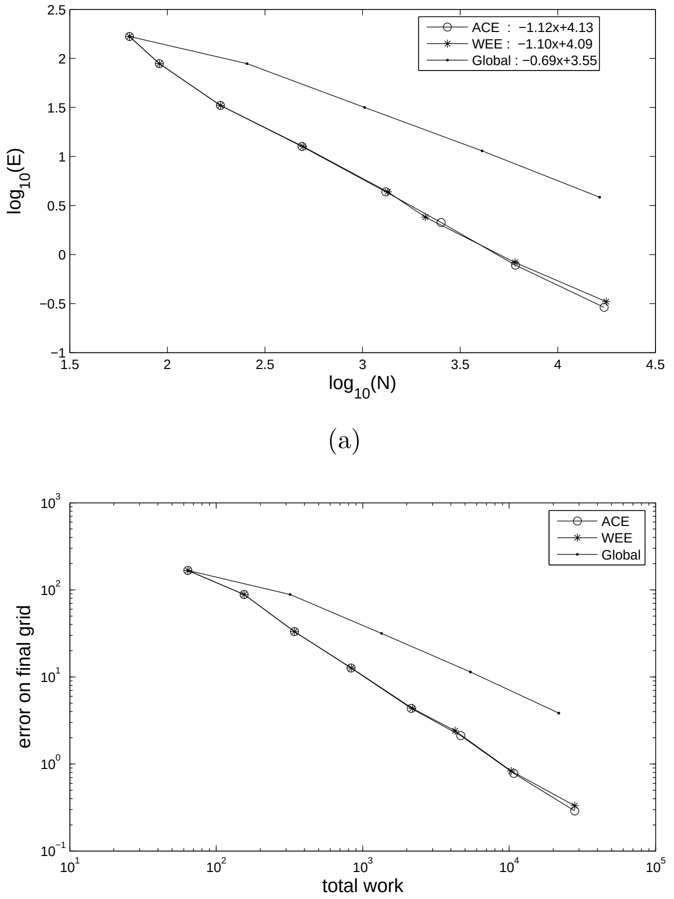
where U is a vector of auxiliary unknowns, and $\vec{\tau}$ is the unit vector tangent to $\partial\Omega$. The FOSLS error estimator is given by $\mathcal{F}(p_h, U_h; f) = \|\nabla \cdot U_h + f\|^2_{L^2(\Omega)} + \|U_h - \nabla p_h\|^2_{(L^2(\Omega))^2} + \|\nabla \times U_h\|^2_{L^2(\Omega)}$. We treat the inhomogeneous boundary conditions strongly, i.e., we extend g to all of Ω such that $g \in H^2(\Omega)$ and transform the BVP into a BVP with homogeneous boundary condition. This can be done since $g \in H^{3/2}(\Omega)$. The same argument as in chapter 3 shows that the FOSLS variational problem and discrete problem are well posed. We also have $\mathcal{G}(p_h, U_h; f) \approx ||p - p_h||^2_{H^1(\Omega)} + \|U - U_h\|^2_{(H^1(\Omega))^2}$. As discussed before, for any element T, we choose $\epsilon_T = G(p_h, U_h; f)$. And we assume that $\epsilon_T \approx ||p-p_h||^2_{H^1(T)} + \|U - U_h\|^2_{(H^1(T))^2}$ holds such that all assumptions on ϵ_T in Chapter 3 hold.



(a)

(b)

Figure 5.2: Efficiency-based refinement strategies for the 2D model problem with $p = 1$. (a) Error versus DOF. (b) Final error, E_L, versus total work, $\sum_{\ell=1}^{L} N_\ell$.



(a)

(b)

Figure 5.3: Efficiency-based refinement strategies for the 2D model problem with $p = 2$. (a) Error versus DOF. (b) Final error, E_L, versus total work, $\sum_{\ell=1}^{L} N_\ell$.

It should be noted here that the WEE measure may be problematic in dimensions higher than one. This can be seen as follows. The WEE measure determines r_{opt} by minimizing $M_{WEE} \equiv \eta(r)\sqrt{\gamma(r)}$ over $r \in [1/N, 1]$. For smooth solutions, $\eta(1/N) \approx 1$ and $\gamma(1/N) \approx 1$, such that $M_{WEE}(1/N) \approx 1$. For $r = 1$, however, it can be observed that $\eta(1) = 2^d$ and $\gamma(1) = (\frac{1}{2})^{2p}$, such that $M_{WEE}(1) = 2^{d-p}$. This means that $M_{WEE} > 1$ when $d > p$. $M_{WEE}(r)$ is often a very smooth function, so r_{opt} is likely to be close to $1/N$ when $d > p$, resulting in small refinements, which are inefficient. We, thus, expect that the

WEE strategy may not be efficient when $d > p$. We investigate this issue in the numerical results presented below. Also, it can be noted that this problem does not occur for the ACE strategy.

We present numerical results (obtained by Josh Nolting using FOSPACK) for the 2D model problem using C^0 elements with $p = 1$ and $p = 2$ in Figs. 5.2 and 5.3, respectively. The figures show error versus DOF and total work for the WEE and ACE refinement strategies, compared with global refinement in every step.

For $p = 1$, the ACE strategy results in an efficient algorithm, but, as expected, the WEE strategy produces many small refinement steps for this case where $d > p$, and is, thus, not efficient (Fig. 5.2). Fig. 5.3 shows that, for $p = 2$ ($d = p$), both the ACE and WEE strategies produce an efficient refinement process.

Fig. 5.1(b) shows the resulting grid when the ACE strategy is applied to a slightly more complicated test problem, in which two circular steps are superimposed ($u = 1$ in the lower left corner, $u = 2$ in the lower right corner, $u = 3$ where the two steps overlap, and $u = 0$ in the top part of the domain). The adaptive refinement process adequately captures the error generated at the steep gradients.

Remark. One can see there exist hanging nodes in Fig. 5.1. We give constrained values to the hanging nodes by using interpolation of appropriate free nodes such that the approximate solutions are C^0. It is shown in literature that the general error convergence result holds for grids with hanging nodes under certain conditions, e.g., see [3].

Chapter 6

Conclusions

Two efficiency-based adaptive refinement strategies for finite element methods, WEE and ACE, were discussed. The two strategies take both error reduction and work into account. The two strategies were first compared for a 1D model problem. For the case of h-refinement with smooth solutions, the efficiency-based strategies generate a highly accurate grid sequence and an efficient refinement process. However, for singular solutions, the refinement process becomes inefficient due to many steps of small refinements. Use of a graded grid for elements with a singularity leads to significant improvement. For both the WEE and ACE strategies, this modification saves a lot of work, and also results in a highly accurate grid sequence. For the hp-refinement case, similar conclusions are obtained. However, for general problems, the difficulty here may lie in how to find a good approximation for the p-error reduction factor. Application to problems with spatial dimension larger than one shows that the WEE strategy is inefficient when the dimension, d, is larger than the finite element order, p. The ACE strategy, however, produces an efficient refinement process for any combination of d and p.

Future work will include application of these grid refinement strategies to problems with singularities in multiple spatial dimensions. Also, an idea to be explored in the future is to enhance the refinement strategies by allowing double or triple refinement for some elements, and determining, in each step, the optimal number of elements to be refined once, twice and thrice. More realistic measures for computational work must be considered, that may, for instance, take into account matrix assembly costs and multigrid convergence factors, and their dependence on the finite element order and the spatial dimension of the problem.

Another topic of interest is the parallelization of the efficiency-based refinement strategies. Binning strategies need to be considered in order to reduce the work for minimizing the efficiency measures, and to reduce the communication between processors [11]. Also, load balancing issues are important for parallel adaptive methods (see, e.g., [19]). After initial solution of a coarse level problem on a single processor, the domain may be partitioned such that each parallel processor receives a subdomain with approximately the same amount of error. This may be a fruitful strategy for load balancing in that, as the grid becomes finer, the optimal refinement approaches global refinement, which requires minimal load balancing. This will be explored in future research.

Bibliography

[1] U. Ruede, Mathematical and Computational Techniques for Multilevel Adaptive Methods, Volume 13 of Frontiers in Applied Mathematics, SIAM, Philadelphia (1993).

[2] R. Verfuerth, A Review of a Posteriori Error Estimation and Adaptive Mesh-Refinement Techniques, Teubner Verlag and J. Wiley, Stuttgart, (1996).

[3] C. Schwab, p- and hp-Finite Element Methods, Clarendon press, Oxford (1998).

[4] S. C. Brenner and L. R. Scott, The Mathematical Theory Of Finite Element Methods, Springer-Verlag, New York (1996).

[5] R. A. Adams, Sobolev Spaces, Academic Press, New York (1975).

[6] L. C. Evans, Partial Differential Equations, American Mathematical Society, Providence, (1998).

[7] W. Rudin, Functional Analysis, Second Edition, McGraw-Hill, (1991).

[8] J. Bergh and J. Löfstrom, Interpolation Spaces, and Introduction. Springer-Verlag, Berlin, (1976).

[9] I. Babuška, Error-Bounds For Finite Element Method, *Numerische Mathematik* 16, pp. 322-333 (1971).

[10] W. L. Briggs, V. E. Henson, and S. F. McCormick, A Multigrid Tutorial, SIAM, PA (1999)

[11] M. Berndt, T. A. Manteuffel and S. F. McCormick, Local Error Estimates And Adaptive Refinement for First-Order System Least Squares (FOSLS), *E.T.N.A.* 6, pp. 35-43 (1997).

[12] W. Gui and I. Babuška, The *h*, *p* and *hp* Versions Of The Finite Element Method in 1 Dimension, Parts I, II, III. *Numerische Mathematik* 49, pp. 577-683 (1986).

[13] Z. Cai, R. Lazarov, T. A. Manteuffel, and S. F. McCormick, First-Order System Least Squares For Second-Order Partial Differential Equations. I, *SIAM J. Numer. Anal.* 31, pp. 1785–1799 (1994).

[14] Z. Cai, T. A. Manteuffel, and S. F. McCormick, First-Order System Least Squares For Second-Order Partial Differential Equations. II, *SIAM J. Numer. Anal.* 34, pp. 425–454 (1997).

[15] P.B. Bochev and M.D. Gunzburger, Finite Element Methods of Least-Squares Type, *SIAM Review* 40, pp. 789–837 (1998).

[16] E. Lee, T. A. Manteuffel, and C. R. Westphal, Weighted-Norm First-Order System Least Squares (FOSLS) For Problems with Corner Singularities, *SIAM J. Numer. Anal.* 44, pp. 1974–1996 (2006).

[17] J. Bramble, R. Lazarov, and J. Pasciak, A Least-Squares Approach Based On A Discrete Minus One Inner Product First Order Systems, *Math. Comp.* 66, pp. 935–955 (1997).

[18] G. F. Carey and B. N. Jiang, Adaptive Refinement For Least Squares Finite Elements With Element-by-element Conjugate Gradient Solution, *Int. J. Numer. Meth. Engr.* 24, pp. 569–580 (1987).

[19] R. E. Bank and M. J. Holst, A New Paradigm For Parallel Adaptive Meshing Algorithms, *SIAM Review* 45, pp. 292–323 (2003).

[20] V. Girault and P. A. Raviart, Finite Element Approximation Of The Navier-Stokes Equations, Springer-Verlag, Berlin Heidelberg (1979).

[21] H. De Sterck, T. Manteuffel, S. McCormick, J. Nolting, J. Ruge, and L. Tang, Efficiency-based h- and hp-refinement Strategies For Finite Element Methods, J. Num. Lin. Alg. Appl., submitted (2007)