Acta Numerica (2017), *pp.* 1–000

Hierarchical Bases and the Finite Element Method

Randolph E. Bank*

The choice of basis functions for a finite element space has important consequences in the practical implementation of the finite element method. A traditional choice is the nodal or Lagrange basis. Many of the computational advantages of this basis derive from the property of compact support enjoyed by the basis functions. Here we study a second choice, the hierarchical basis, and examine its application to some specialized computations in finite element analysis. In particular, we examine the computation of a posteriori error estimates using hierarchical basis functions, and multilevel iterative methods for solving large sparse linear systems of finite element equations.

CONTENTS

1. Introduction

In this work we present a brief introduction to hierarchical bases, and the important part they play in contemporary finite element calculations. In particular, we examine their role in a posteriori error estimation, and in the

* Department of Mathematics, University of California at San Diego, La Jolla, CA 92093.
The work of this author was supported by the Office of Naval Research under contract N00014-89J-1440.

formulation of iterative methods for solving the large sparse sets of linear equations arising from the finite element discretization.

Our goal is that the development should be largely self-contained, but at the same time accessible and interesting to a broad range of mathematicians and engineers. We focus on the simple model problem of a self-adjoint, positive definite, elliptic equation. For this simple problem, the usefulness of hierarchical bases is already readily apparent, but we are able to avoid some of the more complicated technical hurdles that arise in the analysis of more general situations.

A posteriori error estimates play an important role in two related aspects of finite element calculations. First, such estimates provide the user of a finite element code with valuable information about the overall accuracy and reliability of the calculation. Second, since most a posteriori error estimates are computed locally, they also contain significant information about the distribution of error among individual elements, which can form the basis of adaptive procedures such as local mesh refinement. Space considerations prevent us from exploring these two topics in depth, and we will limit our discussion here to the error estimation procedure itself.

Hierarchical basis iterative methods have enjoyed a fair degree of popularity as elliptic solvers. These methods are closely related to the classical multigrid V-cycle and the BPX methods. Hierarchical basis methods typically have a growth in condition number of order k^2, where k is the number of levels*. This is in contrast to multigrid and BPX methods, where the generalized condition number is usually bounded independent of the number of unknowns. Although the rate of convergence is less than optimal, hierarchical basis methods offer several important advantages. First, classical multigrid methods require a sequence of subspaces of geometrically increasing dimension, having work estimates per cycle proportional to the number of unknowns. Such a sequence is sometimes difficult to achieve if adaptive local mesh refinement is used. Hierarchical basis methods, on the other hand, require work per cycle proportional to the number of unknowns for any distribution of unknowns among levels. Second, the analysis of classical multigrid methods often relies on global properties of the mesh and solution (e.g. quasiuniformity of the meshes, $\mathcal{H}^2$ regularity of the solution), whereas analysis of hierarchical basis methods relies mainly on local properties of the mesh (e.g. shape regularity of the triangulation). This yields a method which is very robust over a broad range of problems.

Our analysis of a posteriori error estimates and hierarchical basis iterative methods is based on so-called strengthened Cauchy-Schwartz inequalities. The basic inequality for two levels, along with some other important

* This result is for two space dimensions. For three space dimensions the growth is much faster, like $N^{1/3}$, where N is the number of unknowns

properties of the hierarchical basis decomposition, is presented in Section 3. In Section 4 we use these results to analyze a posteriori error estimates, while in Section 5 we analyze basic two-level iterative methods. In Section 6, we develop a suite of strengthened Cauchy-Schwartz inequalities for k-level hierarchical decompositions, which are then used in Section 7 to analyze multilevel hierarchical basis iterations.

Notation is often a matter of personal preference and provokes considerable debate. We have chosen to use a mixture of the function space notation typical in the mathematical analysis of finite element methods, and matrix-vector notation, which is often most useful when considering questions of practical implementation. We switch freely and frequently between these two types of notation, using that which we believe affords the clearest statement of a particular result. Some important results are presented using both types of notation.

2. Preliminaries

For background on finite element discretizations, we refer the reader to Aziz and Babuška (1972) [3], Brenner and Scott (1994) [20], and Ciarlet (1980) [21]. For simplicity, we will consider the solution of the self-adjoint elliptic partial differential equation

$$-\nabla(a\nabla u) + bu = f \tag{2.1}$$

in a polygonal region $\Omega \subset I\!R^2$, with the homogeneous Neumann boundary conditions

$$a\nabla u \cdot n = 0 \tag{2.2}$$

on $\partial\Omega$, where n is the outward pointing unit normal. Most of our results apply with small modification to the case of Dirichlet boundary conditions $u = 0$ on $\partial\Omega$. We assume that $a(x)$, $b(x)$ are smooth functions satisfying $0 < \underline{a} \leq a(x) \leq \bar{a}$ and $0 < \underline{b} \leq b(x) \leq \bar{b}$ for $x \in \Omega$. The requirement that $\underline{b} > 0$ rather than $\underline{b} \geq 0$ is mainly for convenience.

The $\mathcal{L}^2(\Omega)$ inner product $(\cdot, \cdot)$ is defined by

$$(u, v) = \int_\Omega uv \, \mathrm{d}x$$

and the corresponding norm

$$\|u\|^2 = (u, u) = \int_\Omega u^2 \, \mathrm{d}x.$$

Let $\mathcal{H} = \mathcal{H}^1(\Omega)$ denote the usual Sobolev space equipped with the norm

$$\|u\|_1^2 = \|\nabla u\|^2 + \|u\|^2 = \int_\Omega |\nabla u|^2 + u^2 \, \mathrm{d}x,$$

where $|\cdot|$ denotes the Euclidean norm on $I\!R^2$. The energy inner product

$a(\cdot,\cdot)$ is defined by

$$a(u,v) = \int_\Omega a\nabla u \cdot \nabla v + buv \, \mathrm{d}x, \tag{2.3}$$

for $u, v \in \mathcal{H}$. For $u \in \mathcal{H}$, we define the energy norm $|||u|||$ by

$$|||u|||^2 = a(u,u).$$

This norm is comparable to the $\mathcal{H}^1$ norm in the sense that there exist positive constants c_1 and c_2, depending on a and b, such that

$$c_1 |||u||| \le \|u\|_1 \le c_2 |||u|||.$$

The weak form of the elliptic boundary value problem (2.1)-(2.2) is as follows: find $u \in \mathcal{H}$ such that

$$a(u,v) = (f,v) \tag{2.4}$$

for all $v \in \mathcal{H}$.

Let $\mathcal{T}$ be a triangulation of the region Ω. While the results presented here do not depend on the uniformity or quasiuniformity of the triangulation, many of the constants depend on the shape regularity of the mesh. Let h_t denote the diameter of triangle $t \in \mathcal{T}$, and let d_t denote the diameter of the inscribed circle for t. We assume there exists a positive constant δ_0 such that

$$h_t \delta_0 \le d_t \tag{2.5}$$

for all $t \in \mathcal{T}$. Later, when we consider sequences or families of triangulations, the constant δ_0 will be assumed to be uniform over all triangulations considered. While a shape regularity condition like (2.5) does not imply a *globally* quasiuniform triangulation, it does imply a *local* quasiuniformity for the mesh.

Many of the constants in our estimates depend only on the local variation of the functions a and b; thus we define

$$\alpha_0 = \max_{t\in\mathcal{T}} \frac{\max_{x\in t} a(x)}{\min_{x\in t} a(x)}, \quad \text{and} \quad \beta_0 = \max_{t\in\mathcal{T}} \frac{\max_{x\in t} b(x)}{\min_{x\in t} b(x)}.$$

The fact that our estimates have only a local dependence on the coefficients can be very important in practice. For example, suppose a is piecewise constant, varying by orders of magnitude over the region Ω. If the jumps in a are aligned with edges of the triangulation, then our estimates will be independent of a ($\alpha_0 = 1$), irrespective of the magnitudes of the jumps.

Let $\mathcal{M}$ be an N-dimensional finite element subspace of $\mathcal{H}$, consisting of continuous piecewise polynomials with respect to the triangulation $\mathcal{T}$. We will be more specific about requirements for $\mathcal{M}$ later. The finite element approximation $u_h \in \mathcal{M}$ satisfies

$$a(u_h,v) = (f,v) \tag{2.6}$$

for all $v \in \mathcal{M}$. From (2.4) and (2.6), it is easy to see that the finite element solution is the *best approximation* of u with respect to the energy norm

$$|||u - u_h||| = \inf_{v \in \mathcal{M}} |||u - v|||.$$

Let ϕ_i $1 \leq i \leq N$ be a basis for $\mathcal{M}$. Then (2.6) can be transformed to the linear system of equations

$$AU = F \tag{2.7}$$

where

$$A_{ij} = a(\phi_j, \phi_i), \qquad F_i = (f, \phi_i), \qquad \text{and} \qquad u_h = \sum_{i=1}^{N} U_i \phi_i.$$

The matrix A is typically large, sparse, symmetric, and positive definite. We note that

$$\|x\|_A^2 \equiv x^t A x = |||\chi|||^2,$$

where

$$\chi = \sum_{i=1}^{N} x_i \phi_i.$$

Thus the A-norm of a vector in $\mathbb{R}^N$ is equivalent to the energy norm of the corresponding finite element function.

At the computational level, many aspects of implementation of the finite element method are carried out on an elementwise basis. For example, the *stiffness matrix* A is typically computed as the sum of element stiffness matrices, in which integration is restricted to a single element $t \in \mathcal{T}$. The element stiffness matrix is usually computed by first mapping t to a fixed *reference element* t_r, and then computing the relevant integrals on the reference element. Because such mappings play an important role in our analysis, we begin by considering them in some detail.

Let $\mathcal{S}$ denote the set of triangles t satisfying $h_t = 1$, $\delta_0 \leq d_t/h_t$ and one vertex at the origin. Roughly speaking, the set $\mathcal{S}$ characterizes all shape regular triangles of diameter one. We will denote a particular triangle $t_r \in \mathcal{S}$ as the reference triangle. The reference triangle t_r can be mapped to any other triangle $t \in \mathcal{S}$ using a simple linear transformation (which can be represented as a 2×2 matrix). Shape regularity of the triangles in $\mathcal{S}$ implies that such transformations are well conditioned, with condition numbers depending only on the constant δ_0.

Let $\mathcal{A}$ denote the set of linear transformations mapping the reference triangle t_r to $t \in \mathcal{S}$. Since the triangles in the triangulation $\mathcal{T}$ are all shape regular, any triangle $t \in \mathcal{T}$ can be generated by a simple scaling and translation of an element $\hat{t} \in \mathcal{S}$. Thus the reference element t_r can be mapped to t using a linear transformation from the set $\mathcal{A}$ followed by a simple scaling and translation.

We now suppose that the finite element space $\mathcal{M}$ has the direct sum hierarchical decomposition $\mathcal{M} = \mathcal{V} \oplus \mathcal{W}$. Thus for $u \in \mathcal{M}$ we have the unique decomposition $u = v + w$, where $v \in \mathcal{V}$ and $w \in \mathcal{W}$. Let $\mathcal{V}_t$ and $\mathcal{W}_t$ denote the restrictions of $\mathcal{V}$ and $\mathcal{W}$ to each triangle $t \in \mathcal{T}$, and write $u_t = v_t + w_t$. Often, $\mathcal{V}_t$ and $\mathcal{W}_t$ will be polynomial spaces (as opposed to *piecewise* polynomial spaces), being restricted to a single element. Let $\mathcal{V}_r$ and $\mathcal{W}_r$ denote reference spaces of polynomials defined with respect to the reference triangle t_r. We require that the finite element space $\mathcal{M} = \mathcal{V} \oplus \mathcal{W}$ satisfy the following assumptions for all $t \in \mathcal{T}$:

A1. If $u_t = c$ is constant then $w_t = 0$ and $v_t = c$.
A2. The mapping from t_r to t, consisting of a linear mapping from $\mathcal{A}$ followed by simple scaling and translation induces maps from $\mathcal{V}_r$ onto $\mathcal{V}_t$ and $\mathcal{W}_r$ onto $\mathcal{W}_t$.

These conditions are very weak and are satisfied by many common finite element spaces, although sometimes with a nonstandard choice of basis functions. For example, consider the spaces of continuous piecewise polynomials of degree $p > 1$. For this choice, we let $\mathcal{V}$ be the space of continuous piecewise *linear* polynomials and $\mathcal{W}$ be the space of piecewise polynomials of degree p which are zero at the vertices of the triangulation $\mathcal{T}$. A basis for $\mathcal{V}$ is just the usual nodal basis for the space of continuous piecewise linear polynomials. A basis for $\mathcal{W}$ consists of all the nodal basis functions for the continuous piecewise polynomials of degree p *except* those associated with the triangle vertices. For example, for $p = 2$, $\mathcal{W}$ consists of the span of the quadratic "bump functions" associated with edge midpoints in the triangulation. This is called the hierarchical basis for the piecewise quadratic polynomial space, in contrast to the usual nodal basis, and is often employed in practice in the p-version of the finite element method. It is typically the case that the dimension of the space $\mathcal{W}$ is larger than that of $\mathcal{V}$. In this example, the space $\mathcal{V}$ has a dimension of approximately N/p^2, or about $\dim \mathcal{M}/4$ for the case $p = 2$, and an increasingly smaller fraction as p increases.

We now consider a decomposition of the form $\mathcal{M} = \mathcal{V} \oplus \mathcal{W}$ for the case of continuous piecewise linear polynomials. In this case, we imagine that the triangulation $\mathcal{T} \equiv \mathcal{T}_f$, which we will call the fine grid, arose from the refinement of a coarse grid triangulation $\mathcal{T}_c$. For example, we can consider the case of uniform refinement, in which each triangle $t \in \mathcal{T}_c$ is refined into four similar triangles in $\mathcal{T}$ by pairwise connecting the midpoints of the edges of t. In this case the space $\mathcal{V} \equiv \mathcal{M}_c$ is just the space of continuous piecewise linear polynomials associated with the coarse mesh, while $\mathcal{W}$ consists of the span of the fine grid nodal basis functions associated with vertices in $\mathcal{T}$ which are not in $\mathcal{T}_c$. If uniform refinement is used, then the space $\mathcal{V}$

has a dimension of approximately $N/4$ while the dimension of $\mathcal{W}$ will be approximately $3N/4$. For iterative methods, it is important in practice that the dimension of the space $\mathcal{V}$ be as small as conveniently possible. In this vein, we note that the hierarchical decomposition of $\mathcal{M}$ can be recursively applied to the space $\mathcal{V}$, assuming that $\mathcal{T}_c$ arose from the refinement of an even coarser triangulation. This anticipates the k-level iterations discussed in later sections.

Let $\mathcal{M} = \mathcal{V} \oplus \mathcal{W}$. Let dim $\mathcal{V} = N_{\mathcal{V}}$ and dim $\mathcal{W} = N_{\mathcal{W}} = N - N_{\mathcal{V}}$, and let $\{\phi_i\}_{i=1}^{N_{\mathcal{V}}}$ be a basis for $\mathcal{V}$ and $\{\phi_i\}_{N_{\mathcal{V}}+1}^{N}$ be a basis for $\mathcal{W}$. This induces a natural block 2×2 partitioning of the linear system of (2.7) as

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} U_1 \\ U_2 \end{bmatrix} = \begin{bmatrix} F_1 \\ F_2 \end{bmatrix} \tag{2.8}$$

where A_{11} is of order $N_{\mathcal{V}}$, and A_{22} is of order $N_{\mathcal{W}}$.

We note that if the vector $U \in I\!R^N$ corresponds to the finite element function $u = v + w \in \mathcal{M}$, then

$$U_1^t A_{11} U_1 = |\!|\!| v |\!|\!|^2, \qquad U_2^t A_{22} U_2 = |\!|\!| w |\!|\!|^2, \qquad \text{and} \qquad U_1^t A_{12} U_2 = a(v, w).$$

3. Fundamental Two-Level Estimates

In this section we develop some of the mathematical properties of the hierarchical basis. Chief among these properties is the so-called strengthened Cauchy inequality. One interesting feature of this strengthened Cauchy inequality is that it is a local property of the hierarchical basis: that is, it is true for the hierarchical decomposition corresponding to individual elements in the mesh as well as on the space as a whole. As a result, the constant in the strengthened Cauchy inequality does not depend strongly on such things as global regularity of solutions, the shape of the domain, quasiuniformity of the mesh, global variation of coefficient functions, and other properties that typically appear in the mathematical analysis of finite element methods. By the same reasoning, it is not surprising that the constant in the strengthened Cauchy inequality does depend on local properties like the shape regularity of the elements.

Our analysis of the strengthened Cauchy inequality in this section is taken from Bank and Dupont (1980) [10], but see also Eijkhout and Vassilevski (1991) [26]. We begin our analysis with a preliminary technical lemma.

Lemma 1 Let $(\cdot,\cdot)$ and $\langle\cdot,\cdot\rangle$ denote two inner products defined on a vector space X. Let $\|\cdot\|$ and $|\cdot|$ denote the corresponding norms. Suppose that there exist positive constants $\underline{\lambda}$ and $\bar{\lambda}$ such that

$$0 < \underline{\lambda} \le \frac{(z,z)}{\langle z,z\rangle} \le \bar{\lambda}, \tag{3.1}$$

for all nonzero $z \in X$. For any nonzero $x, y \in X$, let

$$\beta = \frac{(x,y)}{\|x\|\,\|y\|} \quad \text{and} \quad \gamma = \frac{\langle x,y\rangle}{|x|\,|y|}. \tag{3.2}$$

Then

$$1 - \beta^2 \geq \mathcal{K}^{-2}(1-\gamma^2) \tag{3.3}$$

where $\mathcal{K} = \bar{\lambda}/\underline{\lambda}$.

Proof. Lemma 1 states that if two inner products give rise to norms that are comparable as in (3.1), then the angles measured by those inner products must also be comparable. Without loss we can assume $|x| = |y| = 1$. Then from (3.1), we have

$$\begin{aligned}
1-\beta^2 &= (1-\beta)(1+\beta) \\
&= \frac{1}{4}\left\|\frac{x}{\|x\|} + \frac{y}{\|y\|}\right\|^2 \left\|\frac{x}{\|x\|} - \frac{y}{\|y\|}\right\|^2 \\
&= \frac{1}{4\|x\|^4}\|x+\theta y\|^2\|x-\theta y\|^2 \\
&\geq \frac{\underline{\lambda}^2}{4\|x\|^4}|x+\theta y|^2|x-\theta y|^2,
\end{aligned}$$

where $\theta = \|x\|/\|y\|$. Since

$$|x \pm \theta y|^2 = 1 + \theta^2 \pm 2\theta\gamma,$$

we have

$$\begin{aligned}
1-\beta^2 &\geq \frac{\underline{\lambda}^2}{\|x\|^4}\left\{\theta^2(1-\gamma^2) + \frac{1}{4}(1-\theta^2)^2\right\} \\
&\geq \frac{\underline{\lambda}^2\theta^2}{\|x\|^4}(1-\gamma^2) \\
&= \frac{\underline{\lambda}^2}{\|x\|^2\|y\|^2}(1-\gamma^2) \\
&\geq \mathcal{K}^{-2}(1-\gamma^2).
\end{aligned}$$

□

We now state the main Lemma of this section, the strengthened Cauchy inequality.

Lemma 2 Let $\mathcal{M} = \mathcal{V} \oplus \mathcal{W}$ satisfy the assumptions A1 and A2 above. Then there exists a number $\gamma = \gamma(\alpha_0, \beta_0, \delta_0, \mathcal{V}_r, \mathcal{W}_r) \in [0, 1)$ such that

$$|a(v,w)| \leq \gamma\,\|v\|\,\|w\| \tag{3.4}$$

for all $v \in \mathcal{V}$ and all $w \in \mathcal{W}$.

Proof. This proof is done in detail, as many later proofs follow a similar pattern. The first step is to reduce (3.4) to an element-by-element estimate. In particular, suppose that for each $t \in \mathcal{T}$,

$$|a(v,w)_t| \leq \gamma_t |||v|||_t |||w|||_t, \tag{3.5}$$

where

$$a(v,w)_t = \int_t a\nabla v \cdot \nabla w + bvw \, \mathrm{d}x$$

is the restriction of $a(\cdot,\cdot)$ to t, and $|||\cdot|||_t$ is the corresponding norm. Then

$$\begin{aligned} |a(v,w)| &= \left| \sum_t a(v,w)_t \right| \\ &\leq \sum_t |a(v,w)_t| \\ &\leq \sum_t \gamma_t |||v|||_t |||w|||_t \\ &\leq \gamma \left(\sum_t |||v|||_t^2 \right)^{1/2} \left(\sum_t |||w|||_t^2 \right)^{1/2} \\ &= \gamma \, |||v||| \, |||w|||, \end{aligned}$$

where

$$\gamma = \max_{t \in \mathcal{T}} \gamma_t.$$

Thus, if we can show (3.5), then (3.4) follows.

To prove (3.5), we derive the pair of inequalities

$$|a(v,w)_{1,t}| \leq \gamma_{1,t} |||v|||_{1,t} |||w|||_{1,t}, \tag{3.6}$$

$$|a(v,w)_{0,t}| \leq \gamma_{0,t} |||v|||_{0,t} |||w|||_{0,t}, \tag{3.7}$$

where

$$a(v,w)_{1,t} = \int_t a\nabla v \cdot \nabla w \, \mathrm{d}x, \qquad a(v,w)_{0,t} = \int_t bvw \, \mathrm{d}x,$$

and $|||\cdot|||_{i,t}$, $i = 0, 1$, are the corresponding (semi) norms. If (3.6)-(3.7) hold, then for

$$\gamma_t = \max(\gamma_{0,t}, \gamma_{1,t}),$$

we have

$$\begin{aligned} a(v,w)_t^2 &= \left(a(v,w)_{0,t} + a(v,w)_{1,t}\right)^2 \\ &\leq \gamma_t^2 \left(|||v|||_{0,t} |||w|||_{0,t} + |||v|||_{1,t} |||w|||_{1,t}\right)^2 \\ &\leq \gamma_t^2 \left(|||v|||_{0,t}^2 + |||v|||_{1,t}^2\right) \left(|||w|||_{0,t}^2 + |||w|||_{1,t}^2\right) \\ &= \gamma_t^2 |||v|||_t^2 |||w|||_t^2. \end{aligned}$$

We now restrict attention to (3.6); the proof of (3.7) follows a similar pattern. We note that $\|\cdot\|_{1,t}$ defines a strong norm of $\mathcal{W}_t$, but only a seminorm on $\mathcal{V}_t$, since $\mathcal{V}_t$ contains the constant function, and $\|c\|_{1,t} = 0$ for any constant c. It is sufficient to show (3.6) only for the subspace $\tilde{\mathcal{V}}_t = \{v \in \mathcal{V}_t | \int_t v \ \mathrm{d}x = 0\}$, whose elements have average value zero. For any $v \in \mathcal{V}_t$ let $c = |t|^{-1} \int_t v \ \mathrm{d}x$, and note $v - c \in \tilde{\mathcal{V}}_t$. Then

$$a(v,w)_{1,t} = a(v-c,w)_{1,t} \quad \text{and} \quad a(v,v)_{1,t} = a(v-c,v-c)_{1,t}$$

for any $w \in \mathcal{W}_t$. Thus we need show (3.6) only for $v \in \tilde{\mathcal{V}}_t$ and $w \in \mathcal{W}_t$ and note that $\|\cdot\|_{1,t}$ is a strong norm on the space $\tilde{\mathcal{V}}_t \oplus \mathcal{W}_t$.

A simple homogeneity argument now shows that $\gamma_{1,t}$ does not depend on the size of the element h_t. Making the change of variable

$$\hat{x} = \frac{x - x_0}{h_t},$$

where x_0 is any vertex of t, (3.6) becomes

$$\left| \int_{\hat{t}} \hat{a} \nabla \hat{v} \cdot \nabla \hat{w} \ \mathrm{d}\hat{x} \right| \leq \gamma_{1,t} \left(\int_{\hat{t}} \hat{a} |\nabla \hat{v}|^2 \ \mathrm{d}\hat{x} \right)^{1/2} \left(\int_{\hat{t}} \hat{a} |\nabla \hat{w}|^2 \ \mathrm{d}\hat{x} \right)^{1/2}, \tag{3.8}$$

where $\hat{t} \in \mathcal{S}$ is the image of t under the change of variables, $\hat{v}(\hat{x}) = v(x)$, $\hat{w}(\hat{x}) = w(x)$, and $\hat{a}(\hat{x}) = a(x)$. In view of (3.8), we can restrict our attention to the set of triangles $\mathcal{S}$, the set of linear mappings $\mathcal{A}$, and the reference spaces $\mathcal{V}_r$ and $\mathcal{W}_r$.

Let $J \in \mathcal{A}$ be the linear mapping that takes the reference triangle t_r to $\hat{t}$. Then we have

$$\int_{\hat{t}} \hat{a} \nabla \hat{v} \cdot \nabla \hat{w} \ \mathrm{d}\hat{x} = |\det J| \int_{t_r} \tilde{a} (J^{-t} \nabla \tilde{v}) \cdot (J^{-t} \nabla \tilde{w}) \ \mathrm{d}\tilde{x}. \tag{3.9}$$

The right-hand side of (3.9) defines an inner product on the reference triangle t_r. A second inner product is given by the right-hand side of (3.9) with $\tilde{a} \equiv 1$ and $J = I$:

$$\langle v, w \rangle = \int_{t_r} \nabla v \cdot \nabla w \ \mathrm{d}\tilde{x}.$$

Since $\hat{t} \in \mathcal{S}$, there is a positive constant $C = C(\delta_0)$ such that, for all $z \in \tilde{\mathcal{V}}_r \oplus \mathcal{W}_r$,

$$\underline{a}_t C^{-1} \leq \frac{|\det J| \int_{t_r} \tilde{a} |J^{-t} \nabla z|^2 \ \mathrm{d}\tilde{x}}{\int_{t_r} |\nabla \tilde{z}|^2 \ \mathrm{d}\tilde{x}} \leq C \bar{a}_t. \tag{3.10}$$

Here $\underline{a}_t \leq a \leq \bar{a}_t$ for $x \in t$, and $\tilde{\mathcal{V}}_r = \left\{ v \in \mathcal{V}_r | \int_{t_r} v \ \mathrm{d}\tilde{x} = 0 \right\}$. Lemma 1 now tells us that angles measured by these two inner products are comparable.

The last step of the proof is to note that for $v \in \tilde{\mathcal{V}}_r$ and $w \in \mathcal{W}_r$, there exists $\gamma_r = \gamma_r(\mathcal{V}_r, \mathcal{W}_r)$, $0 \le \gamma_r < 1$ for which

$$\left| \int_{t_r} \nabla v \cdot \nabla w \, \mathrm{d}\tilde{x} \right| \le \gamma_r \left(\int_{t_r} |\nabla v|^2 \, \mathrm{d}\tilde{x} \right)^{1/2} \left(\int_{t_r} |\nabla w|^2 \, \mathrm{d}\tilde{x} \right)^{1/2}. \tag{3.11}$$

Estimate (3.11) follows since $\tilde{\mathcal{V}}_r$ and $\mathcal{W}_r$ are linearly independent subspaces, so there must be a nonzero angle between them. Through the use of Lemma 1, it follows that $0 \le \gamma_{1,t}(\alpha_0, \delta_0, \mathcal{V}_r, \mathcal{W}_r) < 1$. The estimate $0 \le \gamma_{0,t}(\beta_0, \delta_0, \mathcal{V}_r, \mathcal{W}_r) < 1$ follows by similar reasoning, except that the reduction to $\tilde{\mathcal{V}}_t$ is unnecessary. □

Analysis of methods employing hierarchical bases is often framed in terms of bounds of certain interpolation operators between fine and coarse spaces. See for example Borneman and Yserentant (1993) [15], Bramble (1993) [17], Oswald (1994) [33], Xu (1989) and (1992) [37] [38], and Yserentant (1992) [40]. In the present context, the fine space is $\mathcal{M}$ while the coarse space is $\mathcal{V}$. The following lemma shows that this approach is entirely equivalent to the use of strengthened Cauchy inequalities.

Lemma 3 Suppose $\mathcal{M} = \mathcal{V} \oplus \mathcal{W}$, and let $\mathcal{I}$ denote the interpolation operator defined as follows: if $u = v + w \in \mathcal{M}$, $v \in \mathcal{V}$, and $w \in \mathcal{W}$, then $\mathcal{I}(u) = v$. Then

$$|\!|\!|\mathcal{I}(u)|\!|\!| \le C |\!|\!|u|\!|\!| \tag{3.12}$$

if and only if

$$|a(v, w)| \le \gamma |\!|\!|v|\!|\!| \, |\!|\!|w|\!|\!| \tag{3.13}$$

for $\gamma < 1$ and for all $v \in \mathcal{V}$ and $w \in \mathcal{W}$.

Proof. First, we assume (3.13) in order to prove (3.12). Let $u = v + w$, $v \in \mathcal{V}$, $w \in \mathcal{W}$. Then

$$\begin{aligned} |\!|\!|u|\!|\!|^2 &= a(v + w, v + w) \\ &= |\!|\!|v|\!|\!|^2 + |\!|\!|w|\!|\!|^2 + 2a(v, w) \\ &\ge |\!|\!|v|\!|\!|^2 + |\!|\!|w|\!|\!|^2 - 2\gamma |\!|\!|v|\!|\!| \, |\!|\!|w|\!|\!| \\ &\ge (1 - \gamma^2) |\!|\!|v|\!|\!|^2. \end{aligned}$$

Therefore

$$|\!|\!|\mathcal{I}(u)|\!|\!| \le (1 - \gamma^2)^{-1/2} |\!|\!|u|\!|\!|.$$

Now we assume (3.12) to show (3.13). It suffices to take $|\!|\!|v|\!|\!| = |\!|\!|w|\!|\!| = 1$. Then, from (3.12)

$$|\!|\!|v - w|\!|\!| \ge \frac{1}{C} |\!|\!|v|\!|\!| = \frac{1}{C}.$$

Thus,

$$a(v,w) = \frac{1}{2}\left(|||v|||^2 + |||w|||^2 - |||v-w|||^2 \right) \le 1 - \frac{1}{2C^2}.$$

□

The last result in this section is related to the space $\mathcal{W}$. The functions in $\mathcal{W}$ are necessarily quite oscillatory, since by assumption $\mathcal{V}$ contains local constants. Indeed, typically $\mathcal{V}$ contains the larger space of local linear functions, although it has not been necessary to assume this. The solution of equations using the space $\mathcal{W}$ should be quite simple, because on such an oscillatory space, an elliptic differential operator behaves very much like a large multiple of the identity.

To make this more precise, suppose that there is a basis for the reference space $\mathcal{W}_r$ whose elements are mapped onto the computational basis functions $\{\phi_j\}_{j=1}^{N_r}$ for $\mathcal{W}_t$ by the affine mapping of t_r onto t. This is a very natural assumption for the case of nodal finite elements, and is typically exploited in practical computations in algorithms for the assembly of the stiffness matrix and right-hand side. With this additional assumption, we have the following lemma:

Lemma 4 Suppose $\{\phi_j\}_{j=1}^{N_\mathcal{W}}$ is the basis for $\mathcal{W}$ and let

$$w = \sum_{j=1}^{N_\mathcal{W}} w_j \phi_j(x,y).$$

Then there exist finite positive constants $\underline{\mu}$ and $\bar{\mu}$, depending only on α_0, β_0, and δ_0, such that

$$\underline{\mu}\, |||w|||^2 \le \sum_{j=1}^{N_\mathcal{W}} w_j^2 |||\phi_j|||^2 \le \bar{\mu}\, |||w|||^2. \tag{3.14}$$

Proof. The proof follows the pattern of Lemma 2, so we will provide only a short sketch here. One first shows it is sufficient to prove

$$\underline{\mu}_t\, |||w|||_t^2 \le \sum_{j=1}^{N_r} w_j^2 |||\phi_j|||_t^2 \le \bar{\mu}_t\, |||w|||_t^2,$$

and set $\underline{\mu} = \min_t \underline{\mu}_t$ and $\bar{\mu} = \max_t \bar{\mu}_t$. (We have been a bit sloppy in our use of subscripts on w_j and ϕ_j in order to avoid more complicated notation.) We then reduce this to showing the pair of inequalities

$$\underline{\mu}_{0,t}\, |||w|||_{0,t}^2 \le \sum_{j=1}^{N_r} w_j^2 |||\phi_j|||_{0,t}^2 \le \bar{\mu}_{0,t}\, |||w|||_{0,t}^2,$$

and

$$\underline{\mu}_{1,t} \, \|w\|_{1,t}^2 \le \sum_{j=1}^{N_r} w_j^2 \|\phi_j\|_{1,t}^2 \le \bar{\mu}_{1,t} \, \|w\|_{1,t}^2,$$

with $\underline{\mu}_t = \min\{\underline{\mu}_{0,t}, \underline{\mu}_{1,t}\}$ and $\bar{\mu}_t = \max\{\bar{\mu}_{0,t}, \bar{\mu}_{1,t}\}$.

A change of variable as in (3.8), mapping $t \in \mathcal{T}$ to an element $\hat{t} \in \mathcal{S}$ proves that $\underline{\mu}$ and $\bar{\mu}$ are independent of h_t. Finally, changing variables as in (3.9) and using equivalence of norms as in (3.10)-(3.11) yields the result. $\square$

We now apply Lemmas 2 and 4 to several finite element spaces having hierarchical decompositions. Much of our analysis of these examples comes from the work of Maitre and Musy (1982) [31]. See also Braess (1981) [16]. In these examples, we will compute the constants $\gamma_{1,t}$, $\underline{\mu}_{1,t}$, and $\bar{\mu}_{1,t}$ for the case $a = 1$, illustrating the effect of shape regularity on the estimates. Let t be a triangle with vertices ν_i, edges ϵ_i, and angles θ_i, $1 \le i \le 3$.

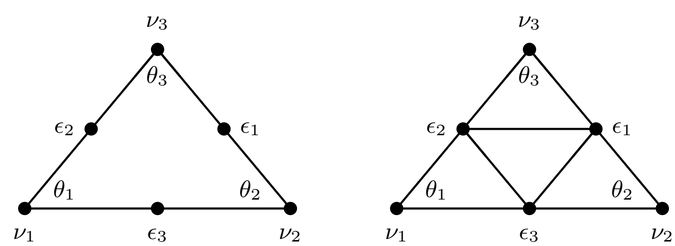

Fig. 1. Quadratic element (left) and piecewise linear element (right).

In our first example, we consider the space of continuous piecewise quadratic finite elements, illustrated on the left in Figure 1. Let ϕ_i $1 \le i \le 3$ denote the linear basis functions for t. Then $\mathcal{V}_t = \langle \phi_i \rangle_{i=1}^3$. The space $\mathcal{W}_t$ is composed of the quadratic bump functions $\mathcal{W}_t = \langle \psi_i \rangle_{i=1}^3$, where $\psi_i = 4\phi_j\phi_k$, and (i, j, k) is a cyclic permutation of $(1, 2, 3)$.

In the second example, we consider the space of continuous piecewise linear polynomials on a refined mesh, illustrated in Figure 1 on the right. Here $\mathcal{V}_t$ contains the linear polynomials on the coarse mesh element t; $\mathcal{V}_t = \langle \phi_i \rangle_{i=1}^3$, with ϕ_i defined as in the first example. The space $\mathcal{W}_t$ contains the continuous piecewise polynomials on the fine grid that are zero at the vertices of t. Thus $\mathcal{W}_t = \langle \hat{\phi}_i \rangle_{i=1}^3$, where $\hat{\phi}_i$ is the standard nodal piecewise linear basis function associated with the midpoint of edge ϵ_i of triangle t.

By direct computation, we can establish the relation

$$-|t| \nabla\phi_j \cdot \nabla\phi_k = \frac{1}{2} \cot\theta_i = \frac{1}{2} L_i.$$

Let

$$A = \begin{bmatrix} L_2 + L_3 & -L_3 & -L_2 \\ -L_3 & L_3 + L_1 & -L_1 \\ -L_2 & -L_1 & L_1 + L_2 \end{bmatrix}, \tag{3.15}$$

and

$$D = \begin{bmatrix} L_1 & 0 & 0 \\ 0 & L_2 & 0 \\ 0 & 0 & L_3 \end{bmatrix}. \tag{3.16}$$

Then the element stiffness matrix for the quadratic hierarchical basis can be shown to be

$$M_Q = \begin{bmatrix} A/2 & -2A/3 \\ -2A/3 & 4(A+D)/3 \end{bmatrix}. \tag{3.17}$$

We know that

$$\begin{aligned} \gamma_{1,t} &= \max\{a(v,w) : \|v\| = \|w\| = 1\} \\ &= \max\{2x^t Ay/3 : x^t Ax = 2,\ y^t(A+D)y = 3/4\} \end{aligned}$$

Standard algebraic manipulations yield

$$\gamma_{1,t}^2 = \frac{2}{3}(1 - \lambda_{\min}),$$

where $\lambda_{\min}$ is the smallest eigenvalue of the generalized eigenvalue problem

$$Dx = \lambda(A+D)x. \tag{3.18}$$

By direct computation and the use of various trigonometric identities, in particular $L_1L_2 + L_2L_3 + L_3L_1 = 1$, we can compute

$$\det\{D - \lambda(A+D)\} = 2(p-s)\lambda^3 + 3(s-p)\lambda^2 - s\lambda + p = 0,$$

where

$$\begin{aligned} p &= L_1L_2L_3, \\ s &= L_1 + L_2 + L_3. \end{aligned}$$

The corresponding eigenvalues are $\lambda = 1$ and $\lambda = (1 \pm \sqrt{4c-3})/4$, where

$$c = \cos^2\theta_1 + \cos^2\theta_2 + \cos^2\theta_3,$$

and

$$\frac{p}{s} = \frac{1-c}{3-c}.$$

Thus

$$\gamma_{1,t}^2 = \frac{3 + \sqrt{4c-3}}{6}. \tag{3.19}$$

For the second example, the element stiffness matrix for the piecewise linear hierarchical basis is given by

$$M_L = \begin{bmatrix} A/2 & -A \\ -A & A+D \end{bmatrix}. \tag{3.20}$$

We see that repeating the arguments for the first example leads to the same values for $\gamma_{1,t}$ but scaled by $\sqrt{3}/2$, that is

$$\gamma_{1,t}^2 = \frac{3+\sqrt{4c-3}}{8}. \tag{3.21}$$

We now turn to the bounds for $\underline{\mu}$ and $\bar{\mu}$ of Lemma 4. These may be expressed in terms of the largest and smallest eigenvalues in the generalized eigenvalue problem

$$(A+D)x = s\lambda x, \tag{3.22}$$

so that

$$\det\{A+D-s\lambda I\} = s^3(1-\lambda)^3 - s(s^2-2)(1-\lambda) - 2p = 0.$$

One can easily write down the analytic solutions of this cubic equation in terms of p and s, but there is no major simplification as in the case of $\gamma_{1,t}$. The bounds for the case of the piecewise linear hierarchical basis are given by $\underline{\mu}_{1,t} = \lambda_{\min}$ and $\bar{\mu}_{1,t} = \lambda_{\max}$. Those for the quadratic case are a simple scaling by 4/3; $\underline{\mu}_{1,t} = 4\lambda_{\min}/3$ and $\bar{\mu}_{1,t} = 4\lambda_{\max}/3$.

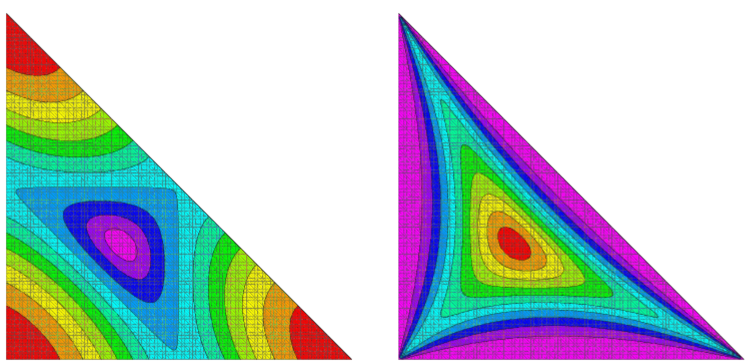

Fig. 2. The contour map for $\gamma_{1,t}$ (left) and for $\underline{\mu}_{1,t}/\bar{\mu}_{1,t}$ (right).

In Figure 2, we have plotted $\gamma_{1,t}$ and the ratio $\kappa_t^{-1} = \underline{\mu}_{1,t}/\bar{\mu}_{1,t}$ as a function of $0 \le \theta_1 \le \pi$ and $0 \le \theta_2 \le \pi - \theta_1$, with $\theta_3 = \pi - \theta_1 - \theta_2$. For the case of quadratic elements, the smallest value $\gamma_{1,t} = 1/\sqrt{2}$ occurs for an equilateral triangle, while the largest value $\gamma_{1,t} = 1$ occurs for the degenerate cases $\theta_i = \theta_j = 0, \theta_k = \pi$. For the case of piecewise linear elements, one should scale all values of $\gamma_{1,t}$ by $\sqrt{3}/2$; for this case $\gamma_{1,t} < 1$, even in the degenerate cases.

It is the ratio $\kappa = \bar{\mu}/\underline{\mu}$ that plays a central role in our later analysis. However, we plot the reciprocal to confine the ratio to the interval $[0,1]$. Here the largest value occurs again for the equilateral triangle, where $\kappa_t^{-1} = 1/4$, while $\kappa_t = 0$ whenever $\theta_i = 0$, $1 \le i \le 3$. A special case occurs in the corners of the domain where the function κ_t^{-1} is discontinuous. For example, if one approaches the origin along the edge $\theta_1 = 0$, then the limiting cubic equation is $(1-\lambda)^3 - (1-\lambda) = 0$, with a corresponding $\kappa_t^{-1} = 0$. However, if we approach along, say, the line $\theta_1 = \theta_2 = \delta$, then the limiting cubic is $(2/3 - \lambda)(\lambda^2 - 7\lambda/3 + 4/9) = 0$, and $\kappa_{1,t}^{-1} = (7 - \sqrt{33})/(7 + \sqrt{33}) > 0$.

4. A Posteriori Error Estimates

A posteriori error estimates are now widely used in the solution of partial differential equations. A recent survey of the field is given by Verfürth (1995) [35], which contains a good bibliography on the subject. See also Ainsworth and Oden (1992 and 1993) [1] [2], Babuška and Gui (1986) [4], Babuška and Rheinboldt (1978a) and (1978b) [6] [7], Bank and Weiser (1985) [14], Weiser (1981) [36], Zienkiewicz *et al* (1982) [41], and the book edited by Babuška *et al* (1986) [5]. Our discussion here is motivated by Bank and Smith (1993) [13].

A posteriori error estimates provide useful indications of the accuracy of a calculation and also provide the basis of adaptive local mesh refinement or local order refinement schemes. For example, if one has solved a problem for a given p, corresponding to a finite element space $\mathcal{M}$, one can enrich the space to, say, order $p+1$ by adding certain hierarchical basis functions to the set of basis functions already used for $\mathcal{M}$. If $\bar{\mathcal{M}}$ is the new space, then we have the hierarchical decomposition

$$\bar{\mathcal{M}} = \mathcal{M} \oplus \mathcal{W},$$

where $\mathcal{W}$ is the subspace spanned by the additional basis functions.

If we resolve the problem with the space $\bar{\mathcal{M}}$ using the hierarchical basis, one expects intuitively that the component of the new solution lying in $\mathcal{M}$ will change very little from the previous calculation. Therefore, the component lying in $\mathcal{W}$ should be a good approximation to the error for the solution on the original space $\mathcal{M}$.

In fact, for our error estimate, we simply solve an (approximate) problem in the space $\mathcal{W}$ rather than $\bar{\mathcal{M}}$ to estimate the error. Let $\bar{u}_h \in \bar{\mathcal{M}}$ be the finite element solution on the enriched space satisfying

$$a(\bar{u}_h, v) = (f, v) \tag{4.1}$$

for all $v \in \bar{\mathcal{M}}$, and

$$|\!|\!| u - \bar{u}_h |\!|\!| = \inf_{v \in \bar{\mathcal{M}}} |\!|\!| u - v |\!|\!|. \tag{4.2}$$

Although we don't explicitly compute $\bar{u}_h$, it enters into our theoretical analysis of the a posteriori error estimate for $u - u_h$. In particular, we assume that the approximate solutions $\bar{u}_h$ converge to u more rapidly than u_h. This is expressed in terms of the *saturation assumption*

$$|\!|\!| u - \bar{u}_h |\!|\!| \leq \beta \, |\!|\!| u - u_h |\!|\!|, \tag{4.3}$$

where $\beta < 1$ independent of h. (We note that since $\mathcal{M} \subset \bar{\mathcal{M}}$, $\beta \leq 1$ is insured by the best approximation property.) In a typical situation, due to the higher degree of approximation for the space $\bar{\mathcal{M}}$, one can anticipate that $\beta = O(h^r)$, for some $r > 0$. In this case, $\beta \to 0$ as $h \to 0$, which is stronger than required by our theorems.

We seek to approximate the error $u - u_h$ in the space $\mathcal{W}$. Our first a posteriori error estimator $e_h \in \mathcal{W}$ is defined by

$$a(e_h, v) = (f, v) - a(u_h, v) \tag{4.4}$$

for all $v \in \mathcal{W}$.

To express this using matrix notation, we consider the linear system of equations corresponding to (4.1), expressed in terms of the hierarchical basis

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} \bar{U}_1 \\ \bar{U}_2 \end{bmatrix} = \begin{bmatrix} F_1 \\ F_2 \end{bmatrix}. \tag{4.5}$$

The vector $(\bar{U}_1^t, \bar{U}_2^t)$ corresponds to the function $\bar{u}_h = v + w \in \bar{\mathcal{M}}$ expanded in terms of the hierarchical basis, with $\bar{U}_1$ corresponding to $v \in \mathcal{M}$ and $\bar{U}_2$ corresponding to $w \in \mathcal{W}$. In this notation, the linear system solved to compute $u_h \in \mathcal{M}$ is given by $A_{11}U_1 = F_1$. If we combine this with the linear system for e_h corresponding to (4.4), we have

$$\begin{bmatrix} A_{11} & 0 \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} U_1 \\ E_2 \end{bmatrix} = \begin{bmatrix} F_1 \\ F_2 \end{bmatrix}, \tag{4.6}$$

where the vector E_2 corresponds to $e_h \in \mathcal{W}$.

We begin our analysis by noting the orthogonality relations

$$\begin{aligned}
a(u - u_h, v) &= 0 \qquad \text{for all } v \in \mathcal{M}, && (4.7)\\
a(u - \bar{u}_h, v) &= 0 \qquad \text{for all } v \in \bar{\mathcal{M}}, && (4.8)\\
a(\bar{u}_h - u_h, v) &= 0 \qquad \text{for all } v \in \mathcal{M}, && (4.9)\\
a(u - u_h - e_h, v) &= 0 \qquad \text{for all } v \in \mathcal{W}, && (4.10)\\
a(\bar{u}_h - u_h - e_h, v) &= 0 \qquad \text{for all } v \in \mathcal{W}. && (4.11)
\end{aligned}$$

Equations (4.7)-(4.11) are proved using various combinations of (2.4), (2.6), (4.1), and (4.4), restricted to the indicated subspaces. We can use the orthogonality relationships (4.7)-(4.9) to show

$$|||u - u_h|||^2 = |||u - \bar{u}_h|||^2 + |||\bar{u}_h - u_h|||^2. \tag{4.12}$$

Using (4.12) in conjunction with the saturation assumption (4.3) shows

$$(1 - \beta^2)|||u - u_h|||^2 \leq |||\bar{u}_h - u_h|||^2 \leq |||u - u_h|||^2, \tag{4.13}$$

demonstrating $\bar{u}_h - u_h$ to be a good approximation to the error. However, our goal is to show the easily computed function e_h also yields a good approximation of the error. This is shown next.

Theorem 1 Let $\bar{\mathcal{M}} = \mathcal{M} \oplus \mathcal{W}$ as above and assume (4.3) and Lemma 2 hold. Then

$$(1 - \beta^2)(1 - \gamma^2)\,|||u - u_h|||^2 \leq |||e_h|||^2 \leq |||u - u_h|||^2. \tag{4.14}$$

Proof. The right inequality in (4.14) is a simple consequence of (4.10) for the choice $v = e_h$. Now let $\bar{u}_h = \hat{u}_h + \hat{e}_h$, where $\hat{u}_h \in \mathcal{M}$, and $\hat{e}_h \in \mathcal{W}$. Then, using (4.9) with $v = \hat{u}_h - u_h$ and (4.11) with $v = \hat{e}_h$, we obtain

$$|||\bar{u}_h - u_h|||^2 = a(\bar{u}_h - u_h, \hat{e}_h) = a(e_h, \hat{e}_h). \tag{4.15}$$

Combining this with (4.12), we get

$$|||u - u_h|||^2 = |||u - \bar{u}_h|||^2 + a(\hat{e}_h, e_h). \tag{4.16}$$

To complete the proof, we must estimate $|||\hat{e}_h|||$ in terms of $|||e_h|||$. We apply the strengthened Cauchy inequality (3.4) to obtain

$$\begin{aligned} |||\bar{u}_h - u_h|||^2 &\geq |||\hat{u}_h - u_h|||^2 + |||\hat{e}_h|||^2 - 2\gamma\,|||\hat{u}_h - u_h|||\;|||\hat{e}_h||| \\ &\geq (1 - \gamma^2)|||\hat{e}_h|||^2. \end{aligned} \tag{4.17}$$

Combine this with (4.15) to obtain

$$(1 - \gamma^2)|||\hat{e}_h||| \leq |||e_h|||. \tag{4.18}$$

Using(4.16) and (4.18), we have

$$|||u - u_h|||^2 \leq \beta^2|||u - u_h|||^2 + \frac{1}{1 - \gamma^2}|||e_h|||^2.$$

Rearranging this inequality leads directly to the left-hand inequality in (4.14). □

We note that computing e_h in (4.4) requires the solution of a linear system involving the matrix A_{22} in (4.6). This is a rather expensive calculation, given that typically the dimension of the space $\mathcal{W}$ is much larger than that of $\mathcal{M}$. Therefore it is of great interest to explore ways in which this calculation can be made more efficient. In situations where Lemma 4 can be applied, one possibility is to replace A_{22} by its diagonal $D_{22} = \text{diag}A_{22}$. In finite element notation, let $d(\cdot, \cdot)$ be the bilinear form corresponding to D_{22}. If

$w = \sum_j w_j \phi_j \in \mathcal{W}$, and $z = \sum_j z_j \phi_j \in \mathcal{W}$, and $\{\phi_j\}$ are the basis functions used in Lemma 4, then

$$d(z, w) = \sum_j z_j w_j a(\phi_j, \phi_j).$$

We compute an approximation $\tilde{e}_h \in \mathcal{W}$ satisfying

$$d(\tilde{e}_h, v) = (f, v) - a(u_h, v). \tag{4.19}$$

In our proof of Theorem 1, we replace the orthogonality relations (4.10)-(4.11) with

$$\begin{aligned} a(u - u_h, v) &= d(\tilde{e}_h, v) \qquad \text{for all } v \in \mathcal{W}, &(4.20)\\ a(\bar{u}_h - u_h, v) &= d(\tilde{e}_h, v) \qquad \text{for all } v \in \mathcal{W}. &(4.21)\end{aligned}$$

Theorem 2 Let $d(\cdot,\cdot)$ be defined as above, and assume Theorem 1 and Lemma 4 hold. Then

$$\bar{\mu}^{-1}(1-\beta^2)(1-\gamma^2)\,|\!|\!| u - u_h |\!|\!|^2 \le |\!|\!| \tilde{e}_h |\!|\!|^2 \le \underline{\mu}^{-1} |\!|\!| u - u_h |\!|\!|^2. \tag{4.22}$$

Proof. One can follow the proof of Theorem 1 with small modifications to show (4.22). However, we will take a more direct approach. From (4.10) and (4.20), we have

$$d(\tilde{e}_h, v) = a(e_h, v)$$

for all $v \in \mathcal{W}$. Taking $v = \tilde{e}_h$ and $v = e_h$, and applying Lemma 4, we have

$$\underline{\mu} |\!|\!| \tilde{e}_h |\!|\!|^2 \le |\!|\!| e_h |\!|\!|^2 \le \bar{\mu} |\!|\!| \tilde{e}_h |\!|\!|^2.$$

Combining this with Theorem 1 proves (4.22). □

A second possibility for improving the efficiency of the computation of the a posteriori error estimate is to use a nonconforming space $\bar{\mathcal{W}}$ of *discontinuous* piecewise polynomials to approximate the error. We assume that $\mathcal{W} \subset \bar{\mathcal{W}}$, but $\bar{\mathcal{W}} \not\subset \mathcal{H}$. The advantage of this approach is that the resulting stiffness matrix $\bar{A}_{22}$ is block diagonal, with each diagonal block corresponding to a single element. Thus the error can be computed element-by-element by solving a small system for each triangle.

To analyze such an error estimator, we need to consider the effect of using nonconforming elements. First, we consider the continuous problem. Let $\mathcal{E}$ denote the set of interior edges of $\mathcal{T}$. For each edge $e \in \mathcal{E}$, we denote a fixed unit normal n_e, chosen arbitrarily from the two possibilities. For w discontinuous along e, let w_A and w_J denote the average and jump of w on e, the sign of w_J being chosen consistently with the choice of n_e. Let $v \in \mathcal{H} \cup \bar{\mathcal{W}}$ and u be the solution of (2.4). Then a straightforward calculation shows that

$$a(u, v) = (f, v) + g(u, v), \tag{4.23}$$

where

$$g(u,v) = \sum_{e\in\mathcal{E}} \int_e \{a\nabla u \cdot n_e\}_A v_J \ \mathrm{d}x, \tag{4.24}$$

and

$$a(u,v) = \sum_{t\in\mathcal{T}} a(u,v)_t.$$

The error estimator $\bar{e}_h \in \bar{\mathcal{W}}$ based on this formulation is given by

$$a(\bar{e}_h, v) = (f,v) + g(u_h, v) - a(u_h, v) \tag{4.25}$$

for all $v \in \bar{\mathcal{W}}$. Note that (4.25) consists of a collection of decoupled problems having the appearance of local Neumann problems on each element; since the space $\bar{\mathcal{W}}$ cannot contain local constants, all problems must be nonsingular and have unique solutions.

To analyze this process, we note that the orthogonality conditions (4.10)-(4.11) are now replaced by

$$\begin{aligned} a(u - u_h - \bar{e}_h, v) &= g(u - u_h, v) \qquad \text{for all } v \in \bar{\mathcal{W}}, &(4.26)\\ a(\bar{u}_h - u_h - \bar{e}_h, v) &= 0 \qquad \text{for all } v \in \mathcal{W}. &(4.27)\end{aligned}$$

Here $\bar{u}_h \in \bar{\mathcal{M}}$ is still the conforming finite element approximation defined in (4.1). The bilinear form $g(\cdot,\cdot)$ does not appear in (4.27) since $v_J = 0$ for $v \in \mathcal{W}$.

In examining the proof of Theorem 1, we note that the argument used in proving the left inequality in (4.14) remains unchanged when applied to $|||\bar{e}_h|||$. The difficulty arises only in the upper bound, where the choice $v = \bar{e}_h$ in (4.26) leads to

$$|||\bar{e}_h|||^2 \le |||u - u_h|||\, |||\bar{e}_h||| + |g(u - u_h, \bar{e}_h)|.$$

Obtaining a bound for the nonconforming term is fairly technical and lengthy, and we will only sketch the arguments here. The interested reader is referred to Bank and Weiser (1985) [14] for a more complete discussion. First note that the presence of the nonconforming term demands more (local) regularity of the solution since line integrals of $\nabla(u - u_h)\cdot n_e$ appear. Here we will make the simplifying assumption

$$\sum_{t\in\mathcal{T}} h_t^2 \|\nabla^2(u - u_h)\|_t^2 \le \alpha^2 |||u - u_h|||^2, \tag{4.28}$$

which essentially states that a standard a priori estimate for $|||u - u_h|||$ is sharp. A more complicated form of the saturation assumption could be used in place of (4.28).

Using standard trace inequalities edge-by-edge for $e \in \mathcal{E}$, we are led to the estimate

$$|g(u-u_h,\bar{e}_h)|^2 \leq C\left(\sum_{t\in\mathcal{T}}\|\sqrt{a}\nabla(u-u_h)\|_t^2+h_t^2\|\sqrt{a}\nabla^2(u-u_h)\|_t^2\right)\left(\sum_{t\in\mathcal{T}}h_t^{-2}\|\sqrt{a}\bar{e}_h\|_t^2+\|\sqrt{a}\nabla\bar{e}_h\|_t^2\right).$$

See Brenner and Scott (1994) [20] for a discussion of trace inequalities.

Now, using (4.28), and a slight generalization of Lemma 4,

$$\underline{\nu}\|w\|_t^2 \leq h_t^{-2}\|w\|_t^2 \leq \bar{\nu}\|w\|_t^2,$$

for all $w\in\mathcal{W}_t$, we obtain the bound

$$|g(u-u_h,\bar{e_h})| \leq \delta\|u-u_h\|\,\|\bar{e}_h\|.$$

Thus we have shown

Theorem 3 Let $\bar{e}_h\in\bar{\mathcal{W}}$ satisfy (4.25). Assume (4.3), (4.28), and Lemmas 2 and 4. Then

$$(1-\beta^2)(1-\gamma^2)\,\|u-u_h\|^2 \leq \|\bar{e}_h\|^2 \leq (1+\delta)^2\|u-u_h\|^2, \tag{4.29}$$

where β and γ are as in Theorem 1 and $\delta=\delta(\alpha_0,\beta_0,\delta_0,\mathcal{V}_r,\mathcal{W}_r)$.

We remark that one could make the diagonal approximation to the systems of linear equations to be solved in computing $\bar{e}_h$. One would then have an estimate modified as in Theorem 2. However, there is less advantage to be gained in the current situation because $\bar{A}_{22}$ is already block diagonal with diagonal blocks of small order. Another possibility is to use a different bilinear form $b(\cdot,\cdot)$ in place of $a(\cdot,\cdot)$ on the left-hand side of (4.25). Such a algorithm would calculate $\check{e}_h\in\bar{\mathcal{W}}$ such that

$$b(\check{e}_h,v)=(f,v)+g(u_h,v)-a(u_h,v). \tag{4.30}$$

One choice suggested by Ainsworth and Oden (1992 and 1993) [1] [2] is to let $b(\cdot,\cdot)$ correspond to the Laplace operator $-\Delta$. If there exist finite, positive constants $\underline{\mu}$ and $\bar{\mu}$ such that

$$\underline{\mu}\|w\|^2 \leq b(w,w) \leq \bar{\mu}\|w\|^2$$

in analogy to (3.14), then the analysis of such approximations may be carried out in a fashion similar to Theorem 2. Durán and Rodríguez (1992) [24] and Durán, Muschietti and Rodríguez (1991) [25] analyze the asymptotic exactness of error estimates of the type developed here, a topic we will not consider in detail.

We now develop some examples of a posteriori error estimates for the space of continuous piecewise linear polynomials. We let $\bar{\mathcal{M}}$ be the space of continuous piecewise quadratic polynomials, and $\mathcal{W}$ the space of quadratic bump functions. The basis functions, denoted $\{\psi_i\}$, will be the standard quadratic nodal basis functions associated with edge midpoints for all edges of the triangulation $\mathcal{T}$. We first consider the estimate $\tilde{e}_h$ defined in (4.19). Let

$$\tilde{e}_h = \sum_i \tilde{E}_i \psi_i.$$

Let ψ_i be associated with an interior edge e of the triangulation and have support in triangles t_1 and t_2, the two triangles sharing edge e. Then

$$\tilde{E}_i = \frac{(f, \psi_i)_{t_1} - a(u_h, \psi_i)_{t_1} + (f, \psi_i)_{t_2} - a(u_h, \psi_i)_{t_2}}{a(\psi_i, \psi_i)_{t_1} + a(\psi_i, \psi_i)_{t_2}}.$$

Here we see that the calculation of $\tilde{E}_i$ involves only local computations. Standard element-by-element assembly techniques can be used to compute all the relevant quantities.

We next consider the computation of $\bar{e}_h$ in (4.25). Let $\bar{\mathcal{W}}$ be the space of discontinuous piecewise quadratic bump functions. There are now two basis functions associated with each interior edge, one with support in each element sharing that edge, so the dimension of $\bar{\mathcal{W}}$ is approximately twice that of $\mathcal{W}$. However, at the level of a single element t, we have $\mathcal{W}_t = \bar{\mathcal{W}}_t$. Let $\{\bar{\psi}_i\}$ be the basis for $\bar{\mathcal{W}}$. Then the function $\bar{e}_h$ of (4.25) can be expressed as

$$\bar{e}_h = \sum_i \bar{E}_i \bar{\psi}_i.$$

Suppose $\bar{\psi}_i$, $\bar{\psi}_j$, and $\bar{\psi}_k$ are the three discontinuous quadratic bump functions having support in the element $t \in \mathcal{T}$. Then we must assemble and solve the 3×3 linear system

$$\begin{bmatrix} a(\bar{\psi}_i, \bar{\psi}_i)_t & a(\bar{\psi}_j, \bar{\psi}_i)_t & a(\bar{\psi}_k, \bar{\psi}_i)_t \\ a(\bar{\psi}_i, \bar{\psi}_j)_t & a(\bar{\psi}_j, \bar{\psi}_j)_t & a(\bar{\psi}_k, \bar{\psi}_j)_t \\ a(\bar{\psi}_i, \bar{\psi}_k)_t & a(\bar{\psi}_j, \bar{\psi}_k)_t & a(\bar{\psi}_k, \bar{\psi}_k)_t \end{bmatrix} \begin{bmatrix} \bar{E}_i \\ \bar{E}_j \\ \bar{E}_k \end{bmatrix} =$$

$$\begin{bmatrix} (f, \bar{\psi}_i)_t - a(u_h, \bar{\psi}_i)_t \\ (f, \bar{\psi}_j)_t - a(u_h, \bar{\psi}_j)_t \\ (f, \bar{\psi}_k)_t - a(u_h, \bar{\psi}_k)_t \end{bmatrix} + \begin{bmatrix} g(u_h, \bar{\psi}_i)_t \\ g(u_h, \bar{\psi}_j)_t \\ g(u_h, \bar{\psi}_k)_t \end{bmatrix}.$$

As in the case of $\tilde{e}_h$, only local computations are involved. All are completely standard except for the evaluation of the nonconforming terms. For example, to evaluate $g(u_h, \bar{\psi}_i)_t$, we first note that $\bar{\psi}_i$ is nonzero on only one edge of t, say edge e. Thus

$$g(u_h, \bar{\psi}_i)_t = \int_e \{a\nabla u_h \cdot n\}_A \bar{\psi}_i \; \mathrm{d}x,$$

where n is the outward normal for t. To evaluate the average, we must compute $a\nabla u_h$ for both t and the adjacent triangle sharing edge e.

5. Two-Level Iterative Methods

In this section we analyze several two-level iterations for solving (2.6) (in finite element notation) or, equivalently, (2.7) (in matrix notation). Much of our development is based on Bank and Dupont (1980) [10] and Bank, Dupont, and Yserentant (1988) [11]. See also the books of Hackbusch (1985) [30] and Bramble (1993) [17].

Let $\mathcal{M} = \mathcal{V} \oplus \mathcal{W}$, let A be the stiffness matrix computed using the hierarchical basis, and partitioned according to (2.8), and let

$$A = L + D + L^t, \tag{5.1}$$

where

$$D = \begin{bmatrix} A_{11} & 0 \\ 0 & A_{22} \end{bmatrix} \quad \text{and} \quad L = \begin{bmatrix} 0 & 0 \\ A_{21} & 0 \end{bmatrix}.$$

We consider the following iteration for solving (2.6). Let $u_0 \in \mathcal{M}$ be given. We define the sequence $u_k = v_k + w_k$, with $v_k \in \mathcal{V}$ and $w_k \in \mathcal{W}$ by

$$a(v_{k+1} - v_k, \chi) = \omega\{(f, \chi) - a(u_k, \chi)\} \tag{5.2}$$

for $\chi \in \mathcal{V}$, and

$$a(w_{k+1} - w_k, \chi) = \omega\{(f, \chi) - a(u_k, \chi)\} \tag{5.3}$$

for $\chi \in \mathcal{W}$. The iteration (5.2)-(5.3) can be written in matrix notation as

$$D(x_{k+1} - x_k) = \omega\{F - Ax_k\}, \tag{5.4}$$

where the vector $x_k \in I\!R^N$ corresponds to the finite element function $u_k \in \mathcal{M}$. Equations (5.2)-(5.4) represent a standard block Jacobi iteration for solving (2.6)-(2.7) Although we have written (5.4) as a stationary iteration, practically we expect to use D as a preconditioner in the conjugate gradient procedure. We refer the interested reader to Golub and Van Loan (1983) [27] or Golub and O'Leary (1989) [28] for a complete discussion of the preconditioned conjugate gradient algorithm. Here we analyze the generalized condition number of the preconditioned system.

Theorem 4 Let $A = L + D + L^t$ as defined above. Then for all $x \neq 0$,

$$\frac{1}{1+\gamma} \leq \frac{x^t D x}{x^t A x} \leq \frac{1}{1-\gamma}, \tag{5.5}$$

where $0 \leq \gamma < 1$ is given in Lemma 2.

Proof. It is easiest to analyze (5.5) using finite element notation. Let $u = v + w$, with $v \in \mathcal{V}$ and $w \in \mathcal{W}$, correspond to $x \in I\!R^N$. Then

$$x^t A x = \|\!|u|\!\|^2 \quad \text{and} \quad x^t D x = \|\!|v|\!\|^2 + \|\!|w|\!\|^2.$$

Now

$$\|\!|u|\!\|^2 = \|\!|v|\!\|^2 + \|\!|w|\!\|^2 + 2a(v, w).$$

Applying Lemma 2, we have

$$(1-\gamma)(|||v|||^2 + |||w|||^2) \le |||u|||^2 \le (1+\gamma)(|||v|||^2 + |||w|||^2),$$

proving (5.5). □

The generalized condition number $\mathcal{K}$ is given by

$$\mathcal{K} = \frac{1+\gamma}{1-\gamma}.$$

The optimum value for ω for the stationary iteration (5.4) is $\omega = 1$, and the rate of convergence is given by

$$\frac{\mathcal{K}-1}{\mathcal{K}+1} = \gamma.$$

See Dupont, Kendall and Rachford (1968) [23] for an analysis of the stationary method.

If conjugate gradient acceleration is used, the estimate for the rate of convergence is bounded by

$$\frac{\sqrt{\mathcal{K}}-1}{\sqrt{\mathcal{K}}+1} = \frac{\gamma}{1+\sqrt{1-\gamma^2}}.$$

We note that (5.4) requires the solution of linear systems involving the diagonal blocks A_{11} and A_{22} in each iteration. We next show that the systems involving A_{22} can be effectively solved using an inner iteration. Those involving A_{11} should either be solved directly, or solved recursively, using a multilevel iteration.

Let $\hat{A}_{22}$ be a symmetric, positive definite preconditioner for A_{22}, and suppose we approximately solve the linear system $A_{22}x = b$, using $m \ge 1$ steps of the iterative process

$$\hat{A}_{22}(x_{k+1} - x_k) = b - A_{22}x_k. \tag{5.6}$$

The iteration (5.6) should not be accelerated, but should be implemented as a stationary iteration to allow the use of conjugate gradient acceleration for the overall (outer) iteration. We assume that any fixed parameters for (5.6) have been already incorporated in the definition of $\hat{A}_{22}$. Let

$$G = I - A_{22}^{1/2}\hat{A}_{22}^{-1}A_{22}^{1/2}.$$

We assume G is symmetric with

$$\|G\|_{\ell_2} = \rho < 1. \tag{5.7}$$

Let

$$R_m = G^m(I - G^m)^{-1}. \tag{5.8}$$

The eigenvalues of R_m lie on the interval

$$0 \leq \lambda \leq \frac{\rho^m}{1-\rho^m} \tag{5.9}$$

when m is even or if all eigenvalues of G are nonnegative. In the latter case, G is sometimes called a *smoother*. If G is not a smoother and m is odd, we must use the weaker bound

$$-\frac{\rho^m}{1+\rho^m} \leq \lambda \leq \frac{\rho^m}{1-\rho^m}. \tag{5.10}$$

A induction argument shows the m-step process in (5.6) is mathematically equivalent to the solution of

$$A_{22}^{1/2}(I+R_m)A_{22}^{1/2}x_m = b + A_{22}^{1/2}R_m A_{22}^{1/2}x_0. \tag{5.11}$$

In our current situation, the initial guess $x_0 = 0$, simplifying the right hand side of (5.11). Our overall preconditioner, using m inner iterations, is thus

$$\hat{D} = D + \begin{bmatrix} 0 & 0 \\ 0 & A_{22}^{1/2}R_m A_{22}^{1/2} \end{bmatrix}. \tag{5.12}$$

Theorem 5 Let $A = L + D + L^t$ and $\hat{D}$ be defined as above. Then for all $x \neq 0$,

$$\frac{1}{(1+\gamma)(1+\rho^m)} \leq \frac{x^t\hat{D}x}{x^tAx} \leq \frac{1}{(1-\gamma)(1-\rho^m)}. \tag{5.13}$$

Proof. As in the proof of Theorem 4, we let $u = v + w \in \mathcal{M}$ correspond to $x \in I\!R^N$. Then

$$\|\!|v|\!\|^2 + (1+\rho^m)^{-1}\|\!|w|\!\|^2 \leq x^t\hat{D}x \leq \|\!|v|\!\|^2 + (1-\rho^m)^{-1}\|\!|w|\!\|^2.$$

Thus

$$\frac{1}{1+\rho^m} \leq \frac{x^t\hat{D}x}{x^tDx} \leq \frac{1}{1-\rho^m},$$

and the theorem follows from Theorem 4 and

$$\frac{x^t\hat{D}x}{x^tAx} = \left(\frac{x^t\hat{D}x}{x^tDx}\right)\left(\frac{x^tDx}{x^tAx}\right).$$

□

The generalized condition number $\mathcal{K}$ is bounded by

$$\mathcal{K} \leq \left(\frac{1+\gamma}{1-\gamma}\right)\left(\frac{1+\rho^m}{1-\rho^m}\right).$$

Here we see that the use of inner iterations has only a modest effect on the generalized condition number, provided that ρ is small or m is large.

We remark that by bounding $x^t\hat{D}x/x^tAx$ directly, instead of bounding $x^t\hat{D}x/x^tDx$ and x^tDx/x^tAx separately, one can achieve a somewhat smaller but more complicated bound for $\mathcal{K}$. If G is a smoother, then the bound on $\mathcal{K}$ can be improved to

$$\mathcal{K} \le \left(\frac{1+\gamma}{1-\gamma}\right)\left(\frac{1}{1-\rho^m}\right).$$

We now consider the symmetric block Gauss-Seidel iteration

$$\begin{aligned}(D+L)(x_{k+1/2}-x_k) &= F - Ax_k \\ (D+L^t)(x_{k+1}-x_{k+1/2}) &= F - Ax_{k+1/2}.\end{aligned} \tag{5.14}$$

In finite element notation, we may write (5.14) as

$$a(v_{k+1/2}+w_k,\chi) = (f,\chi) \tag{5.15}$$

for $\chi \in \mathcal{V}$,

$$a(v_{k+1/2}+w_{k+1},\chi) = (f,\chi) \tag{5.16}$$

for $\chi \in \mathcal{W}$, and

$$a(v_{k+1}+w_{k+1},\chi) = (f,\chi) \tag{5.17}$$

for $\chi \in \mathcal{V}$. A careful analysis of (5.15)-(5.17) will show that block Gauss-Seidel and block symmetric Gauss-Seidel are equivalent as stationary iterative methods (i.e. $v_{k+1/2} = v_k$), but this is no longer true when symmetric Gauss-Seidel is used as a preconditioner for the conjugate gradient algorithm.

Let $e_k = x - x_k$. Then from (5.14),

$$\begin{aligned}e_{k+1/2} &= \{I-(D+L)^{-1}A\}e_k, \\ e_{k+1} &= \{I-(D+L)^{-t}A\}e_{k+1/2},\end{aligned}$$

from which it follows that

$$\begin{aligned}e_{k+1} &= \{I-(D+L)^{-t}A\}\{I-(D+L)^{-1}A\}e_k \\ &= \{I-[(D+L)^{-t}+(D+L)^{-1}]A+(D+L)^{-t}A(D+L)^{-1}A\}e_k \\ &= \{I-(D+L)^{-t}(L+2D+L^t-A)(D+L)^{-1}A\}e_k \\ &= \{I-(D+L)^{-t}D(D+L)^{-1}A\}e_k \\ &= \{I-B^{-1}A\}e_k,\end{aligned} \tag{5.18}$$

where

$$B = (D+L)D^{-1}(D+L^t) = A + LD^{-1}L^t. \tag{5.19}$$

Once again, our task is to determine the generalized condition number by estimating the Rayleigh quotient.

Theorem 6 Let $A = L + D + L^t$ as defined above, and let B be given by (5.19). Then

$$1 \le \frac{x^t B x}{x^t A x} \le \frac{1}{1-\gamma^2}, \tag{5.20}$$

where $0 \le \gamma < 1$ is given in Lemma 2.

Proof. Since $LD^{-1}L^t$ is symmetric, positive semidefinite, it is clear from (5.19) that the lower bound is one. The upper bound is given by $1+\mu$ where

$$\mu = \max_{x \ne 0} \frac{x^t L D^{-1} L^t x}{x^t A x}. \tag{5.21}$$

This can be written as

$$\mu = \max_{x \ne 0} \frac{y^t D y}{x^t A x},$$

where

$$Dy = L^t x.$$

In finite element notation, this becomes

$$\mu = \max_{u \ne 0} \frac{\|\hat{v}\|^2}{\|u\|^2}, \tag{5.22}$$

where $u = v + w$, $v \in \mathcal{V}$, $w \in \mathcal{W}$ and $\hat{v} \in \mathcal{V}$ satisfies

$$a(\hat{v}, \chi) = a(w, \chi) \tag{5.23}$$

for all $\chi \in \mathcal{V}$. Written in finite element language, (5.22)-(5.23) is easy to analyze in terms of the strengthened Cauchy inequality. We take $\chi = \hat{v}$ in (5.23) to see

$$\|\hat{v}\| \le \gamma \|w\|.$$

On the other hand

$$\begin{aligned} \|u\|^2 &= \|v\|^2 + \|w\|^2 + 2a(v,w) \\ &\ge \|v\|^2 + \|w\|^2 - 2\gamma \|v\| \, \|w\| \\ &\ge (1-\gamma^2)\|w\|^2 \\ &\ge (1-\gamma^2)\gamma^{-2}\|\hat{v}\|^2. \end{aligned}$$

The theorem now follows from combining this estimate and (5.22). □

The analysis of the block symmetric Gauss-Seidel scheme with inner iterations is a little more complicated. We formally consider the iteration

$$\begin{aligned} (\hat{D} + L)(x_{k+1/2} - x_k) &= F - Ax_k, \qquad (5.24) \\ (\hat{D} + L^t)(x_{k+1} - x_{k+1/2}) &= F - Ax_{k+1/2}, \end{aligned}$$

where $\hat{D}$ is given in (5.12). A calculation similar to (5.18) shows that

$$\begin{aligned} e_{k+1} &= \{I - (\hat{D}+L)^{-t}A\}\{I - (\hat{D}+L)^{-1}A\}e_k \\ &= \{I - [(\hat{D}+L)^{-t} + (\hat{D}+L)^{-1}]A + (\hat{D}+L)^{-t}A(\hat{D}+L)^{-1}A\}e_k \\ &= \{I - (\hat{D}+L)^{-t}(L + 2\hat{D} + L^t - A)(\hat{D}+L)^{-1}A\}e_k \qquad (5.25) \\ &= \{I - (\hat{D}+L)^{-t}(2\hat{D} - D)(\hat{D}+L)^{-1}A\}e_k \\ &= \{I - \hat{B}^{-1}A\}e_k, \end{aligned}$$

where

$$\begin{aligned} \hat{B} &= (\hat{D}+L)(2\hat{D} - D)^{-1}(\hat{D}+L)^t \\ &= A + (D - \hat{D} + L)(2\hat{D} - D)^{-1}(D - \hat{D} + L)^t \qquad (5.26) \\ &= A + LD^{-1}L^t + \Delta. \end{aligned}$$

Here

$$\Delta = \begin{bmatrix} 0 & 0 \\ 0 & A_{22}^{1/2} R_m^2 (I + 2R_m)^{-1} A_{22}^{1/2} \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & A_{22}^{1/2} R_{2m} A_{22}^{1/2} \end{bmatrix},$$

and R_m is defined in (5.8).

Theorem 7 Let $A = L + D + L^t$ as defined above, and let $\hat{B}$ be given by (5.26). Then

$$1 \le \frac{x^t \hat{B} x}{x^t A x} \le \frac{1}{(1-\gamma^2)(1-\rho^{2m})}, \qquad (5.27)$$

where $0 \le \gamma < 1$ is given in Lemma 2, and ρ is given in (5.7).

Proof. Since $LD^{-1}L^t + \Delta$ is symmetric, positive semidefinite, the lower bound is one. For the upper bound, $x^t LD^{-1}L^t x / x^t A x$ was estimated in the proof of Theorem 6. Let $u = v + w \in \mathcal{M}$ correspond to $x \in I\!R^N$. Then, using (5.7)-(5.8) and Lemma 2, we have

$$\frac{x^t \Delta x}{x^t A x} \le \left(\frac{\rho^{2m}}{1-\rho^{2m}} \right) \frac{\|w\|^2}{\|u\|^2} \le \left(\frac{\rho^{2m}}{1-\rho^{2m}} \right) \left(\frac{1}{1-\gamma^2} \right).$$

Combining these estimates, we have

$$\frac{x^t \hat{B} x}{x^t A x} \le 1 + \left(\frac{\gamma^2}{1-\gamma^2} \right) + \left(\frac{\rho^{2m}}{1-\rho^{2m}} \right) \left(\frac{1}{1-\gamma^2} \right) = \frac{1}{(1-\gamma^2)(1-\rho^{2m})}.$$

□

We now consider some possibilities for the inner iterations. One obvious choice is a Jacobi method based on the diagonal matrix $D_{22} = \mathrm{diag} A_{22}$ with $\hat{A}_{22} = D_{22}/\omega$. Using Lemma 4, for the choice $\omega = 2/(\underline{\mu} + \bar{\mu})$, we have

$$\rho \le \frac{\kappa - 1}{\kappa + 1},$$

where $\kappa = \bar{\mu}/\underline{\mu}$.

A second possibility is to use a symmetric Gauss-Seidel iteration. Let $A_{22} = L_{22} + D_{22} + L_{22}^t$, where L_{22} is lower triangular. We then take

$$\hat{A}_{22} = (D_{22} + L_{22}) D_{22}^{-1} (D_{22} + L_{22})^t. \tag{5.28}$$

Lemma 5 Suppose the hypotheses of Lemma 4 hold, and let $\hat{A}_{22}$ be given by (5.28). Then there exists a finite positive constant η depending only on α_0, β_0, and δ_0, such that

$$1 \le \frac{x^t \hat{A}_{22} x}{x^t A_{22} x} \le 1 + \eta. \tag{5.29}$$

Proof. As usual, the lower bound is one, since $\hat{A}_{22} = A_{22} + L_{22} D_{22}^{-1} L_{22}^t$, and $L_{22} D_{22}^{-1} L_{22}^t$ is symmetric and positive semidefinite. Now

$$\eta = \max_x \frac{y^t D_{22} y}{x^t A_{22} x},$$

where

$$D_{22} y = L_{22}^t x.$$

In finite element notation, this is

$$\eta = \max_w \frac{\sum_j \hat{w}_j^2 \| \phi_j \|^2}{\| w \|^2},$$

where $\hat{w} \in \mathcal{W}$ corresponds to y, $w \in \mathcal{W}$ corresponds to x, and $\{\phi_j\}$ are the basis functions for $\mathcal{W}$. Since the basis functions for $\mathcal{W}$ are developed from a fixed set of functions defined on the reference element, the support of a given basis function can intersect that of only a small number of other basis functions (there are at most a fixed number of nonzeros in any row of L_{22}^t, independent of the number of elements in the mesh). Therefore we must have

$$\sum_j \hat{w}_j^2 \| \phi_j \|^2 \le C \sum_j w_j^2 \| \phi_j \|^2,$$

where $C = C(\delta_0)$. The result now follows directly from Lemma 4. □

Using Lemma 5, we can estimate

$$\rho = \| I - A_{22}^{1/2} \hat{A}_{22}^{-1} A_{22}^{1/2} \|_{\ell_2} \le \frac{\eta}{1 + \eta}.$$

Thus we see that although these inner iterations perturb the rate of convergence, they do not affect the essential feature that the rate depends only on local properties of the finite element spaces, and is independent of such things as the dimension of the space, uniformity or nonuniformity of the mesh, and regularity of the solution.

6. Multilevel Cauchy Inequalities

In this section we will develop several strengthened Cauchy inequalities of use in analyzing hierarchical basis iterations with more than two levels. These estimates are developed for the special case of continuous piecewise linear finite elements; they can be combined with the two-level analysis of Section 5 to develop multilevel algorithms for higher degree polynomial spaces. We will return to this point in Section 7. Much of the material here is based on Bank and Dupont (1979) [9], Yserentant (1986) [39], and Bank, Dupont and Yserentant (1988) [11]. See also the books of Hackbusch (1985) [30], Bramble (1993) [17], and Oswald (1994) [33].

Let $\mathcal{T}_1$ be a coarse, shape regular triangulation of Ω. We will inductively construct a sequence of uniformly refined triangulations $\mathcal{T}_j$, $2 \le j \le k$, as follows. For each triangle $t \in \mathcal{T}_{j-1}$, we will construct 4 triangles in $\mathcal{T}_j$ by pairwise connecting the midpoints of t. All triangulations will be shape regular, as every triangle $t \in \mathcal{T}_j$ will be geometrically similar to the triangle in $\mathcal{T}_0$ which contains it. We could also allow nonuniform refinements that control shape regularity, for example those of the type used in the adaptive finite element program $PLTMG$ (Bank (1994) [8]). See also Rüde (1993) [34] and Deuflhard, Leinen and Yserentant (1989) [22].

With this definition, it is easy to introduce the notion of the level of a given vertex in the triangulation $\mathcal{T}_j$. All vertices in the original triangulation $\mathcal{T}_1$ are called level-1 vertices. The new vertices created in forming $\mathcal{T}_j$ from $\mathcal{T}_{j-1}$ are called level-j vertices. Notice that all vertices in $\mathcal{T}_j$ have a level less than or equal to j. Also note that each vertex has a unique level, and this unique level is the same in all triangulations that contain it.

Let $\mathcal{M}_j$ be the space of continuous piecewise linear polynomials associated with $\mathcal{T}_j$. Functions in $\mathcal{M}_j$ will be represented using the hierarchical basis, which is easily constructed in an inductive fashion. Let $\{\phi_i\}_{i=1}^{N_1}$ denote the usual nodal basis functions for the space $\mathcal{M}_1$; this is also the hierarchical basis for $\mathcal{M}_1$. To construct the hierarchical basis for $\mathcal{M}_j$, $j > 1$, we take the union of the hierarchical basis for $\mathcal{M}_{j-1}$, $\{\phi_i\}_{i=1}^{N_{j-1}}$, with the nodal basis functions associated with the newly introduced level j vertices, $\{\phi_i\}_{i=N_{j-1}+1}^{N_j}$.

Let $\mathcal{V}_j$ be the subspace spanned by the basis functions associated with the level-j vertices, $\{\phi_i\}_{i=N_{j-1}+1}^{N_j}$, where $N_0 = 0$. Note that $\mathcal{V}_1 = \mathcal{M}_1$. Then we can write for $j > 1$,

$$\mathcal{M}_j = \mathcal{M}_{j-1} \oplus \mathcal{V}_j = \mathcal{V}_1 \oplus \mathcal{V}_2 \oplus \ldots \oplus \mathcal{V}_j.$$

Let $\mathcal{N}_j$, $1 \le j \le k-1$ be defined by

$$\mathcal{N}_j = \mathcal{V}_{j+1} \oplus \mathcal{V}_{j+2} \oplus \ldots \oplus \mathcal{V}_k$$

with $\mathcal{N}_k = \emptyset$. Then we have the decompositions

$$\mathcal{M}_k = \mathcal{M}_j \oplus \mathcal{N}_j$$

for $1 \leq j \leq k$.

Before proceeding to the Cauchy inequalities, we need a preliminary technical result.

Lemma 6 Let $t \in \mathcal{S}$, where $\mathcal{S}$ is defined as in Section 2. Let $\mathcal{T}'$ be a shape regular triangulation of t, whose elements have a minimum diameter of h. Let $\mathcal{M}'$ be the space of continuous piecewise linear polynomials associated with $\mathcal{T}'$. Then there exists a constant $c = c(\delta_0)$, independent of h, such that, for all $v \in \mathcal{M}'$,

$$\|v\|_{\infty,t} \leq c|\log h|^{1/2}\|v\|_{1,t}. \tag{6.1}$$

Proof. Here we will only sketch a proof, following ideas in Bank and Scott (1989) [12], but see Yserentant [39] (1986) for a more detailed, but also more elementary proof. We remark that estimate (6.1) is restricted to two space dimensions.

Our proof is based on an inverse inequality, and the Sobolev inequality; see Brenner and Scott (1994) [20] or Ciarlet (1980) [21] for a general discussion of these topics. Let t' be a shape regular triangle of size $h_{t'}$, and let v be a linear polynomial. The inverse inequality we require states

$$\|v\|_{\mathcal{L}^\infty(t')} \leq C_0 h_{t'}^{-2/p}\|v\|_{\mathcal{L}^p(t')}$$

for $1 \leq p \leq \infty$. Let D be a closed bounded region with a piecewise smooth boundary; then the Sobolev inequality we need states

$$\|v\|_{\mathcal{L}^p(D)} \leq C_1\sqrt{p}\|v\|_{\mathcal{H}^1(D)}$$

for all $v \in \mathcal{H}^1(D)$ and all $p < \infty$. Now let $t \in \mathcal{S}$ and $v \in \mathcal{M}'$; then

$$\begin{aligned}
\|v\|_{\mathcal{L}^\infty(t)} &= \max_{t'\in\mathcal{T}'}\|v\|_{\mathcal{L}^\infty(t')} \\
&\leq C_0 h^{-2/p}\max_{t'\in\mathcal{T}'}\|v\|_{\mathcal{L}^p(t')} \\
&\leq C_0 h^{-2/p}\|v\|_{\mathcal{L}^p(t)} \\
&\leq C_0 C_1 h^{-2/p}\sqrt{p}\|v\|_{\mathcal{H}^1(t)}
\end{aligned}$$

The proof is now completed by taking $p \approx -4\log h$. $\square$

Lemma 7 Let $\mathcal{M}_k = \mathcal{M}_j \oplus \mathcal{N}_j$ as above. Then there exist positive constants γ_j, $1 \leq j \leq k-1$ such that

$$\gamma_j \leq 1 - \frac{c}{k-j}, \tag{6.2}$$

and the strengthened Cauchy inequality

$$|a(v,w)| \le \gamma_j |||v|||\, |||w||| \tag{6.3}$$

holds for $v \in \mathcal{M}_j$ and $w \in \mathcal{N}_j$. The positive constant c in (6.2) is independent of j and k.

Proof. Our proof is based on that of Bank and Dupont (1979) [9]. Following the pattern used in proving Lemma 2, we first reduce the estimate (6.3) to an elementwise estimate for $t \in \mathcal{T}_j$. If we show

$$|a(v,w)_t| \le \gamma_{j,t} |||v|||_t |||w|||_t, \tag{6.4}$$

then

$$\gamma_j = \max_{t\in\mathcal{T}_j} \gamma_{j,t}.$$

Let $t \in \mathcal{T}_j$, and let x_i, $1 \le i \le 3$ denote the three vertices of t. We map t to a triangle $\hat{t} \in \mathcal{S}$ using the change of variable

$$\hat{x} = \frac{x - x_1}{h_t}.$$

As in the proof of Lemma 2, this verifies that $\gamma_{j,t}$ is independent of h_t. Notice that $\mathcal{M}_{j,t}$, the restriction of $\mathcal{M}_j$ to t, is just the space of linear polynomials on t and has dimension three. In the case of uniform refinement, the space $\mathcal{N}_{j,t}$ is the space of piecewise linear polynomials on a uniform grid of 4^{k-j} congruent triangles, which are zero at the three vertices of t. The (local) constant function is thus contained in $\mathcal{M}_{j,t}$, and $\mathcal{M}_{j,t} \oplus \mathcal{N}_{j,t}$ is just the space of continuous piecewise linear polynomials on t.

Let $v \in \mathcal{M}_{j,t}$ and $w \in \mathcal{N}_{j,t}$. Then

$$\begin{aligned} \gamma_{j,t} &= \max_{|||v|||_t = |||w|||_t = 1} a(v,w)_t \\ &= \max_{|||v|||_t = |||w|||_t = 1} 1 - \frac{|||v-w|||_t^2}{2} \\ &\le \max_{|||v|||_t = |||w|||_t = 1} 1 - c\|v-w\|_{1,t}^2, \end{aligned}$$

where $c = c(\alpha_0, \beta_0)$.

We now apply Lemma 6, noting that $h \approx 2^{k-j}$ for the triangulation of $\hat{t}$.

$$\gamma_{j,t} \le \max_{|||v|||_t = |||w|||_t = 1} 1 - \frac{C\|v-w\|_{\infty,t}^2}{\log 2^{k-j}},$$

where $C = C(\alpha_0, \beta_0, \delta_0)$.

Next we note that, since v is just a linear polynomial on t with $|||v|||_t = 1$, and $w(x_i) = 0$, $1 \le i \le 3$, we have a fixed constant $c' > 0$, independent of j

and k, such that

$$c' < \max_{x_i} |v(x_i)| = \max_{x_i} |v(x_i) - w(x_i)| \le \|v - w\|_{\infty,t}.$$

Thus it follows that

$$\gamma_{j,t} \le 1 - \frac{Cc'}{\log 2^{k-j}},$$

and the lemma follows. $\Box$

We next describe the result of Lemma 7 in terms of interpolation operators.

Lemma 8 Let $u = v_j + w_j \in \mathcal{M}_k$, $v_j \in \mathcal{M}_j$ and $w_j \in \mathcal{N}_j$. Define the interpolation operator $\mathcal{I}_j$, mapping $\mathcal{M}_k$ to $\mathcal{M}_j$, by $\mathcal{I}_j(u) = v_j$. Then

$$|\!|\!|\mathcal{I}_j(u)|\!|\!| \le C\sqrt{k-j}\,|\!|\!|u|\!|\!|. \tag{6.5}$$

The positive constant C is independent of j and k.

Proof. Apply Lemmas 3 and 7. See also Yserentant (1986) [39], and Bank, Dupont and Yserentant (1988) [11]. $\Box$

We finish this section with

Lemma 9 Let $\mathcal{V}_i$ and $\mathcal{V}_j$ for $1 \le i, j \le k$ be defined as above. Then there exist positive constants $\Gamma_{i,j}$ satisfying

$$\Gamma_{i,j} \le c2^{-|i-j|/2}, \tag{6.6}$$

such that

$$|a(v, w)| \le \Gamma_{i,j} |\!|\!|v|\!|\!|\,|\!|\!|w|\!|\!| \tag{6.7}$$

for all $v \in \mathcal{V}_i$ and $w \in \mathcal{V}_j$. The constant c in (6.6) is independent of i and j.

Proof. Our proof is similar to that given by Yserentant (1986) [39]. Without loss of generality, suppose $i < j$. We need to consider no triangulation finer than $\mathcal{T}_j$, since subsequent refinements do not affect either v or w. As in the other Cauchy inequalities, one first reduces the estimate to a single element $t \in \mathcal{T}_i$, that is

$$|a(v, w)_t| \le \Gamma_{i,j,t} |\!|\!|v|\!|\!|_t |\!|\!|w|\!|\!|_t. \tag{6.8}$$

We then consider the gradient terms and the lower order terms separately as in (3.6)-(3.7). For the highest order term, we must again consider the special importance of the (local) constant function, which in this case belongs to $\mathcal{V}_{i,t}$. Following the pattern in the proof of Lemma 2, we next map $t \in \mathcal{T}_i$ to an element $\hat{t} \in \mathcal{S}$ by scaling and translation, showing that the estimate must be independent of h_t. Also note that under this mapping, triangles in $\mathcal{T}_j$ become triangles with size $\hat{h} \approx 2^{i-j}$.

The central estimate is to show that

$$|a(\hat{v},\hat{w})_{1,\hat{t}}| \leq \Gamma_{i,j,1,t} \|\hat{v}\|_{1,\hat{t}} \|\hat{w}\|_{1,\hat{t}}, \tag{6.9}$$

where

$$\begin{aligned} a(\hat{v},\hat{w})_{1,\hat{t}} &= \int_{\hat{t}} \hat{a} \nabla \hat{v} \cdot \nabla \hat{w} \, \mathrm{d}\hat{x} \\ \|\hat{v}\|^2_{1,\hat{t}} &= a(\hat{v},\hat{v})_{1,\hat{t}}. \end{aligned}$$

We will also use the norms

$$\|\hat{v}\|^2_{\hat{t}} = \int_{\hat{t}} \hat{v}^2 \, \mathrm{d}\hat{x} \quad \text{and} \quad \|\hat{v}\|^2_{\partial\hat{t}} = \int_{\partial\hat{t}} \hat{v}^2 \, \mathrm{d}\hat{x}.$$

The function $\hat{v}$ is just a linear polynomial on $\hat{t}$, while $\hat{w}$ is a piecewise linear polynomial vanishing at all the vertices with level smaller than j. Such a function is necessarily very oscillatory, and for such a function the differential operator behaves very much like $\hat{h}^{-1}$ times the identity operator. In particular, we have the estimates

$$\|\hat{w}\|_{\hat{t}} \leq C\hat{h} \|\hat{w}\|_{1,\hat{t}} \leq \hat{C} 2^{i-j} \|\hat{w}\|_{1,\hat{t}} \tag{6.10}$$

and

$$\|\hat{w}\|_{\partial\hat{t}} \leq C\hat{h}^{1/2} \|\hat{w}\|_{1,\hat{t}} \leq \hat{C} 2^{(i-j)/2} \|\hat{w}\|_{1,\hat{t}}, \tag{6.11}$$

where $\hat{C} = \hat{C}(\alpha_0, \delta_0)$.

Now, using integration by parts, the fact that $\Delta v = 0$ in $\hat{t}$, and (6.10)-(6.11) we have

$$\begin{aligned} a(\hat{v},\hat{w})_{1,\hat{t}} &= \int_{\hat{t}} -\nabla \hat{a} \cdot \nabla \hat{v} \hat{w} \, \mathrm{d}\hat{x} + \int_{\partial\hat{t}} \hat{a} \nabla \hat{v} \cdot n \hat{w} \, \mathrm{d}\hat{s} \\ &\leq C\{\|\nabla \hat{v}\|_{\hat{t}} \|w\|_{\hat{t}} + \|\nabla \hat{v}\|_{\partial\hat{t}} \|\hat{w}\|_{\partial\hat{t}}\} \\ &\leq C' 2^{(i-j)/2} \|\hat{v}\|_{1,\hat{t}} \|\hat{w}\|_{1,\hat{t}}. \end{aligned}$$

The lower order term is easy to treat in this case because of (6.10). □

7. Multilevel Iterative Methods

In this section, we will analyze block Jacobi and block symmetric Gauss-Seidel iterations using the hierarchical decomposition

$$\mathcal{M}_k = \mathcal{V}_1 \oplus \mathcal{V}_2 \oplus \ldots \oplus \mathcal{V}_k$$

defined in Section 6. Much of this material comes from Bank, Dupont and Yserentant (1988) [11], but see also Bramble (1993) [17], Bramble, Pasciak, and Xu (1990) [19], Bramble, Pasciak, Wang, and Xu (1991) [18], Griebel (1994) [29], Hackbusch (1985) [30], Ong (1989) [32], Xu (1989) and (1992) [37] [38], and Yserentant (1986) and (1992) [39] [40].

As before, we let $\{\phi_i\}_{i=N_{j-1}+1}^{N_j}$ denote piecewise linear nodal basis functions for the level-j vertices in $\mathcal{T}_k$. Then the stiffness matrix A can be expressed as the symmetric, positive definite block $k \times k$ matrix

$$A = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1k} \\ A_{21} & A_{22} & & A_{2k} \\ \vdots & & \ddots & \vdots \\ A_{k1} & A_{k2} & \cdots & A_{kk} \end{bmatrix}, \tag{7.1}$$

where A_{jj} is the $(N_j - N_{j-1}) \times (N_j - N_{j-1})$ matrix of energy inner products involving just the level-j basis functions. In similar fashion to the analysis in Section 5, we set

$$A = L + D + L^t, \tag{7.2}$$

where

$$D = \begin{bmatrix} A_{11} & & & \\ & A_{22} & & \\ & & \ddots & \\ & & & A_{kk} \end{bmatrix} \quad \text{and} \quad L = \begin{bmatrix} 0 & & & \\ A_{21} & 0 & & \\ \vdots & & \ddots & \\ A_{k1} & A_{k2} & \cdots & 0 \end{bmatrix}.$$

We first consider the block Jacobi iteration. Let $u_0 \in \mathcal{M}_k$ be given. We define the sequence

$$u_i = v_{1,i} + v_{2,i} + \ldots + v_{k,i},$$

where $v_{j,i} \in \mathcal{V}_j$, $1 \le j \le k$. In finite element notation, the block Jacobi iteration is written

$$a(v_{j,i+1} - v_{j,i}, \chi) = \omega\{(f, \chi) - a(u_i, \chi)\} \tag{7.3}$$

for $\chi \in \mathcal{V}_j$, $1 \le j \le k$. The iteration (7.3) can be written in matrix notation as

$$D(x_{i+1} - x_i) = \omega\{F - Ax_i\}, \tag{7.4}$$

where the vector $x_i \in \mathbb{R}^{N_k}$ corresponds to the finite element function $u_i \in \mathcal{M}_k$.

To estimate the rate of convergence, we must bound the Rayleigh quotient

$$0 < \underline{\lambda} \le \frac{x^t D x}{x^t A x} \le \bar{\lambda} \tag{7.5}$$

for $x \ne 0$. In finite element notation, this is written

$$0 < \underline{\lambda} \le \frac{\sum_{i=1}^k |||v_i|||^2}{|||v|||^2} \le \bar{\lambda}, \tag{7.6}$$

where $v_i \in \mathcal{V}_i$ and $v = \sum_{i=1}^k v_i \ne 0$.

For any $v = v_1 + v_2 + \ldots + v_k$, we define

$$z_j = v_1 + v_2 + \ldots + v_j, \tag{7.7}$$

for $1 \le j \le k$, with $z_0 = 0$,

$$w_j = v_{j+1} + v_{j+2} + \ldots + v_k, \tag{7.8}$$

for $0 \le j \le k-1$, with $w_k = 0$. Thus we have $v = z_j + w_j$, $0 \le j \le k$. Note $z_j \in \mathcal{M}_j$, while $w_j \in \mathcal{N}_j$.

We begin our analysis with an upper bound for (7.6). First note that the angle between the spaces $\mathcal{V}_1 \oplus \mathcal{V}_2 \oplus \ldots \oplus \mathcal{V}_{j-1} = \mathcal{M}_{j-1}$ and $\mathcal{V}_j$ is just the angle between the spaces $\mathcal{V}$ and $\mathcal{W}$ of Lemma 2. Therefore the constant in the strengthened Cauchy inequality for these spaces, which we will denote by $\tilde{\gamma}$, does not depend on j. Now

$$\begin{aligned} |||z_j|||^2 &= |||z_{j-1} + v_j|||^2 \\ &= |||z_{j-1}|||^2 + |||v_j|||^2 + 2a(z_{j-1}, v_j) \\ &\ge |||z_{j-1}|||^2 + |||v_j|||^2 - 2\tilde{\gamma}|||z_{j-1}|||\,|||v_j||| \\ &\ge (1-\tilde{\gamma}^2)|||v_j|||^2. \end{aligned}$$

We now use Lemma 7 to deduce

$$\begin{aligned} |||v|||^2 &= |||z_j + w_j|||^2 \\ &= |||z_j|||^2 + |||w_j|||^2 + 2a(z_j, w_j) \\ &\ge |||z_j|||^2 + |||w_j|||^2 - 2\gamma_j|||z_j|||\,|||w_j||| \\ &\ge (1-\gamma_j^2)|||z_j|||^2 \\ &\ge (1-\gamma_j^2)(1-\tilde{\gamma}^2)|||v_j|||^2. \end{aligned}$$

Thus we have

$$\sum_{i=1}^{k} |||v_i|||^2 \le \frac{|||v|||^2}{1-\tilde{\gamma}^2}\sum_{i=1}^{k}\frac{1}{1-\gamma_i^2} \le Ck^2|||v|||^2.$$

To find a lower bound, we note that

$$|||\sum_{i=1}^{k} v_i|||^2 = \sum_{i=1}^{k}\sum_{j=1}^{k} a(v_i, v_j) \le \sum_{i=1}^{k}\sum_{j=1}^{k}\Gamma_{i,j}|||v_i|||\,|||v_j||| = E^t\Gamma E,$$

where $E_i = |||v_i|||$, and Γ is the $k \times k$ matrix introduced in Lemma 9. One can easily see that $\|\Gamma\|_{\ell^2} < C$, so that

$$|||v|||^2 = |||\sum_{i=1}^{k} v_i|||^2 \le C\sum_{i=1}^{k}|||v_i|||^2.$$

Thus we have proved

Theorem 8 Let $A = L + D + L^t$ as defined above. Then

$$C_1 \le \frac{x^t D x}{x^t A x} \le C_2 k^2, \tag{7.9}$$

where $C_i = C_i(\alpha_0, \beta_0, \delta_0)$, $i = 1, 2$.

Note that the generalized condition number $\mathcal{K} \le ck^2$ now depends on the number of levels. For the case of uniform refinement, $k = O(\log N_k)$, so this introduces a logarithmic-like term into the convergence rate. Note that $\sqrt{\mathcal{K}} \le \tilde{c}k$, so that conjugate gradient acceleration can be expected to have a more significant impact on the k-level iteration than on the two-level method.

As in the case of the two-level iteration, we may solve linear systems of the form $A_{ii}x = b$ by an inner iteration for all $i > 1$. Following the development given in Section 5, let $\hat{A}_{ii}$ be the preconditioner for A_{ii} and let $G_i = I - A_{ii}^{1/2}\hat{A}_{ii}^{-1}A_{ii}^{1/2}$. Suppose

$$\max_{i>1} \|G_i\|_{\ell_2} = \rho < 1,$$

and assume for simplicity that $m \ge 1$ inner iterations are used for all $i > 1$. Let $R_{i,m} = G_i^m(I - G_i^m)^{-1}$. Then, using reasoning similar to that of (5.12), we replace (7.4) with

$$\hat{D}(x_{i+1} - x_i) = \omega\{F - Ax_i\} \tag{7.10}$$

where

$$\hat{D} = D + D^{1/2} \begin{bmatrix} 0 & & & \\ & R_{2,m} & & \\ & & \ddots & \\ & & & R_{k,m} \end{bmatrix} D^{1/2} = D + Z.$$

Theorem 9 Let $A = L + D + L^t$ and $\hat{D}$ be defined as above. Then

$$\frac{C_1}{1+\rho^m} \le \frac{x^t \hat{D} x}{x^t A x} \le \frac{C_2 k^2}{1-\rho^m}, \tag{7.11}$$

where C_i, $i = 1, 2$ are given in Theorem 8.

Proof. Following the proof of Theorem 5, we see for all $x \ne 0$,

$$\frac{1}{1+\rho^m} \le \frac{x^t \hat{D} x}{x^t D x} \le \frac{1}{1-\rho^m}.$$

The theorem then follows easily from this estimate and Theorem 8. □

We next consider the symmetric block Gauss-Seidel iteration. In finite element notation, we may write this as

$$a(v_{j,i+1/2} - v_{j,i}, \chi) = (f, \chi) - a(z_{j-1,i+1/2} + w_{j-1,i}, \chi) \tag{7.12}$$

for $\chi \in \mathcal{V}_j$, $j = 1, 2, \ldots, k$, and

$$a(v_{j,i+1} - v_{j,i+1/2}, \chi) = (f, \chi) - a(z_{j,i+1/2} + w_{j,i+1}, \chi) \tag{7.13}$$

for $\chi \in \mathcal{V}_j$, $j = k, k-1, \ldots, 1$. Here $z_{j,i}$ and $w_{j,i}$ are defined analogously to v_j and w_j in (7.7)-(7.8). In matrix notation the iteration is written

$$\begin{aligned} (D + L)(x_{i+1/2} - x_i) &= F - Ax_i, \\ (D + L^t)(x_{i+1} - x_{i+1/2}) &= F - Ax_{i+1/2}. \end{aligned} \tag{7.14}$$

As in the two-level scheme, the preconditioner B is given by

$$B = (D + L)D^{-1}(D + L^t) = A + LD^{-1}L^t. \tag{7.15}$$

Theorem 10 Let $A = L + D + L^t$ and B be defined as above. Then

$$1 \le \frac{x^t Bx}{x^t Ax} \le 1 + \mu, \tag{7.16}$$

where

$$\mu \le C_3 k^2, \tag{7.17}$$

and $C_3 = C_3(\alpha_0, \beta_0, \delta_0)$.

Proof. The lower bound is clear since $LD^{-1}L^t$ is symmetric and positive semidefinite. For the upper bound, we estimate

$$\mu = \max_{x \neq 0} \frac{y^t Dy}{x^t Ax}$$

where

$$Dy = L^t x.$$

Let $v = v_1 + v_2 + \ldots + v_k = z_j + w_j$, with $v_i \in \mathcal{V}_i$ and $z_j \in \mathcal{M}_j$ and $w_j \in \mathcal{N}_j$ as in (7.7)-(7.8). Then in finite element notation, we have

$$\mu = \max_{v \neq 0} \frac{\sum_{i=1}^{k-1} |||\tilde{v}_i|||^2}{|||v|||^2}, \tag{7.18}$$

where

$$a(\tilde{v}_i, \chi) = a(w_i, \chi) \tag{7.19}$$

for all $\chi \in \mathcal{V}_i$.

Taking $\chi = \tilde{v}_i$ in (7.19) and applying Lemma 7, we have

$$|||\tilde{v}_i||| \le \gamma_i |||w_i|||,$$

and

$$\begin{aligned} |||v|||^2 &= |||z_i + w_i|||^2 \\ &= |||z_i|||^2 + |||w_i|||^2 + 2a(z_i, w_i) \end{aligned}$$

$$\begin{aligned}
&\geq \|z_i\|^2 + \|w_i\|^2 - 2\gamma_i \|z_i\| \, \|w_i\| \\
&\geq (1-\gamma_i^2)\|w_i\|^2 \\
&\geq (1-\gamma_i^2)\gamma_i^{-2}\|\tilde{v}_i\|^2.
\end{aligned}$$

Thus we have

$$\mu \leq \sum_{i=1}^{k-1} \frac{\gamma_i^2}{1-\gamma_i^2} < C_3 k^2.$$

□

We next analyze the effect of inner iterations on the symmetric block Gauss-Seidel iteration. Thus we replace D with $\hat{D}$ in (7.14) and obtain the iteration

$$\begin{aligned}
(\hat{D}+L)(x_{i+1/2}-x_i) &= F - Ax_i \\
(\hat{D}+L^t)(x_{i+1}-x_{i+1/2}) &= F - Ax_{i+1/2}
\end{aligned} \tag{7.20}$$

Following arguments similar to (5.26), we have

$$\begin{aligned}
\hat{B} &= (\hat{D}+L)(2\hat{D}-D)^{-1}(\hat{D}+L^t) \\
&= A + (D-\hat{D}+L)(2\hat{D}-D)^{-1}(D-\hat{D}+L^t) \\
&= A + (L-Z)(D+2Z)^{-1}(L^t-Z).
\end{aligned} \tag{7.21}$$

As usual, we need to estimate the Rayleigh quotient $x^t\hat{B}x/x^tAx$. Since $(L-Z)(D+2Z)^{-1}(L^t-Z)$ is symmetric, positive semidefinite, the lower bound is just 1. To obtain an upper bound, the essential estimate we must make is

$$\begin{aligned}
\hat{\mu} &= \max_{x\neq 0} \frac{x^t(L-Z)(D+2Z)^{-1}(L^t-Z)x}{x^tAx} \\
&= \|(D+2Z)^{-1/2}(L^t-Z)A^{-1/2}\|_{\ell^2}^2 \\
&\leq \Big(\|(D+2Z)^{-1/2}D^{1/2}\|_{\ell^2}\,\|D^{-1/2}L^tA^{-1/2}\|_{\ell^2} \\
&\qquad +\|(D+2Z)^{-1/2}ZD^{-1/2}\|_{\ell^2}\,\|D^{1/2}A^{-1/2}\|_{\ell^2}\Big)^2.
\end{aligned}$$

Now

$$\|(D+2Z)^{-1/2}D^{1/2}\|_{\ell^2} \leq \frac{1+\rho^m}{1-\rho^m}$$

and

$$\|(D+2Z)^{-1/2}ZD^{-1/2}\|_{\ell^2} \leq \frac{\rho^{2m}}{1-\rho^{2m}}.$$

The norms $\|D^{-1/2}L^tA^{-1/2}\|_{\ell^2}$ and $\|D^{1/2}A^{-1/2}\|_{\ell^2}$ are estimated using Theorems 10 and 8, respectively. Combining these estimates, we have

Theorem 11 Let $A = L + D + L^t$ and $\hat{B}$ be defined as above. Then

$$1 \leq \frac{x^t \hat{B} x}{x^t A x} \leq 1 + \hat{\mu}, \tag{7.22}$$

where

$$\hat{\mu} \leq \left(\sqrt{\frac{1+\rho^m}{1-\rho^m} C_3} + \sqrt{\frac{\rho^{2m}}{1-\rho^{2m}} C_2} \right)^2 k^2 \leq C_4 k^2, \tag{7.23}$$

and C_2 and C_3 are given in Theorems 8 and 10, respectively.

If G is a smoother, then using (5.9) we have $\|(D+2Z)^{-1/2} D^{1/2}\|_{\ell^2} \leq 1$, and the improved estimate

$$\hat{\mu} \leq \left(\sqrt{C_3} + \sqrt{\frac{\rho^{2m}}{1-\rho^{2m}} C_2} \right)^2 k^2 \leq C_4 k^2.$$

We conclude with several remarks about the two-level and k-level methods. Although the k-level method was developed for only the case of continuous piecewise linear polynomials, this is sufficient to construct efficient methods for higher-degree spaces. For example, we consider the case of continuous piecewise quadratic polynomials on a sequence of meshes $\mathcal{T}_j$, $1 \leq j \leq k$. At first glance, one might be tempted to try to develop a method in which one used piecewise quadratic spaces on all levels. Further reflection would lead one to the conclusion that such a method could potentially be very complicated, as it is not clear that there is a simple way to develop a hierarchical basis. It is also not clear that the analysis of such a method could be based on the results in this work.

On the other hand, we could begin by making the usual two-level decomposition $\mathcal{M} = \mathcal{V} \oplus \mathcal{W}$, where $\mathcal{V}$ is the space of piecewise linear polynomials on $\mathcal{T}_k$ and $\mathcal{W}$ is the space of piecewise quadratic bump functions that are zero at the vertices of $\mathcal{T}_k$. The dimension of $\mathcal{W}$ is then approximately $3N/4$ where N is the dimension of $\mathcal{M}$. For the space $\mathcal{V}$, which is just the space of piecewise linear polynomials on $\mathcal{T}_k$, we can make the hierarchical decomposition

$$\mathcal{V} = \mathcal{V}_1 \oplus \mathcal{V}_2 \oplus \ldots \oplus \mathcal{V}_k$$

as described here. Overall, we have the hierarchical decomposition

$$\mathcal{M} = \mathcal{V}_1 \oplus \mathcal{V}_2 \oplus \ldots \oplus \mathcal{V}_k \oplus \mathcal{W}.$$

Based on this decomposition, there is an obvious multilevel hierarchical basis iteration that can be developed. This iteration could be viewed as a two-level iteration, with an elaborate k-level inner iteration used to solve the linear systems associated with the space $\mathcal{V}$. Alternatively, this iteration could be viewed as a $k+1$ level iteration, in which the the first k levels are the standard ones, but level $k+1$ is special, in that the degree of approximation is increased instead of the mesh being refined. For either viewpoint, the algorithm is the same, and its analysis is straightforward using the results in Sections 3-7.

Another possibility along these lines is to make some further hierarchical decomposition of the space $\mathcal{W}$. For example, suppose now that $\mathcal{M}$ is the space of continuous piecewise quartic polynomials on $\mathcal{T}_k$. We can begin by making a decomposition $\mathcal{M} = \mathcal{V} \oplus \mathcal{W}$, where $\mathcal{V}$ is the space of continuous piecewise linear polynomials and $\mathcal{W}$ is the space of quartic polynomials that are zero at the vertices of $\mathcal{T}_k$. We make a further decomposition of $\mathcal{V}$ as in the previous example. We can also conveniently make the further decomposition $\mathcal{W} = \mathcal{W}_2 \oplus \mathcal{W}_4$, where $\mathcal{W}_2$ is the space of continuous piecewise quadratic polynomials that are zero at the vertices of $\mathcal{T}_k$. This is the same as the space $\mathcal{W}$ in our last example. The space $\mathcal{W}_4$ is now the space of continuous piecewise quartic polynomials that are zero at the vertices and edge midpoints of $\mathcal{T}_k$ (i.e. all the nodes associated with the piecewise linear and piecewise quadratic spaces). This space can be characterized in terms of a subset of the standard nodal basis functions for the piecewise quartic space, the bump functions associated with the 1/4 and 3/4 points on each edge, and the bubble functions associated with the barycentric coordinates $(1/4, 1/4, 1/2)$, $(1/4, 1/2, 1/4)$, and $(1/2, 1/4, 1/4)$ in each element. This leads to an overall decomposition

$$\mathcal{M} = \mathcal{V}_1 \oplus \mathcal{V}_2 \oplus \ldots \oplus \mathcal{V}_k \oplus \mathcal{W}_2 \oplus \mathcal{W}_4.$$

The resulting hierarchical basis iteration could then be viewed as a basic two-level iteration in which elaborate inner iterations are used for solving linear systems associated with both the $\mathcal{V}$ and $\mathcal{W}$ spaces, or as a $k+2$ level scheme in which the last two levels involve an increase in degree of approximation rather than a refinement of the mesh.

REFERENCES

Mark Ainsworth and J. Tinsley Oden. A procedure for a posteriori error estimation for h-p finite element methods. *Comp. Meth. Appl. Mech. Engrg.*, 101:73–96, 1992.

Mark Ainsworth and J. Tinsley Oden. A unified approach to a posteriori error estimation using element residual methods. *Numerische Mathematik*, 65:23–50, 1993.

A. K. Aziz and I. Babuška. Part I, survey lectures on the mathematical foundations of the finite element method. In *The Mathematical Foundations of the Finite Element Method with Applications to Partial Differential Equations*, pages 1–362. Academic Press, New York, 1972.

I. Babuška and W. Gui. Basic principles of feedback and adaptive approaches in the finite element method. *Comp. Meth. Appl. Mech. Engrg.*, 55:27–42, 1986.

I. Babuška, O. C. Zienkiewicz, J. P. de S. R. Gago, and E. R. de Arantes e Oliveira. *Accuracy Estimates and Adaptive Refinements in Finite Element Computations.* John Wiley and Sons, New York, 1986.

Ivo Babuška and Werner C. Rheinboldt. Error estimates for adaptive finite element computations. *SIAM J. Numerical Analysis*, 15:736–754, 1978.

Ivo Babuška and Werner C. Rheinboldt. A posteriori error estimates for the finite element method. *Internat. J. Numer. Methods Engrg.*, 12:1597–1615, 1978.

Randolph E. Bank. *PLTMG: A Software Package for Solving Elliptic Partial Differential Equations, Users' Guide 7.0.* Frontiers in Applied Mathematics, Vol. 15, SIAM, Philadelphia, 1994.

Randolph E. Bank and Todd F. Dupont. Notes on the k-level iteration. unpublished notes, 1979.

Randolph E. Bank and Todd F. Dupont. Analysis of a two level scheme for solving finite element equations. Technical Report CNA-159, Center for Numerical Analysis, University of Texas at Austin, 1980.

Randolph E. Bank, Todd F. Dupont, and Harry Yserentant. The hierarchical basis multigrid method. *Numerische Mathematik*, 52:427–458, 1988.

Randolph E. Bank and L. Ridgway Scott. On the conditioning of finite element equations with highly refined meshes. *SIAM J. Numerical Analysis*, 26:1383–1394, 1989.

Randolph E. Bank and R. Kent Smith. A posteriori error estimates based on hierarchical bases. *SIAM J. Numerical Analysis*, 30:921–935, 1993.

Randolph E. Bank and Alan Weiser. Some a posteriori error estimates for elliptic partial differential equations. *Mathematics of Computation*, 44:283–301, 1985.

Folkmar Bornemann and Harry Yserentant. A basic norm equivalence for the theory of multilevel methods. *Numerische Mathematik*, 64:455–476, 1993.

Dietrich Braess. The contraction number of a multigrid method for solving the Poisson equation. *Numerische Mathematik*, 37:387–404, 1981.

James H. Bramble. *Multigrid Methods.* Pitman Research Notes in Mathematical Sciences 294, Longman Sci. & Techn., Harlow, Essex, 1993.

James H. Bramble, Joseph E. Pasciak, Junping Wang, and Jinchao Xu. Convergence estimate for product iterative methods with application to domain decomposition and multigrid. *Math. Comp.*, 57:1–21, 1991.

James H. Bramble, Joseph E. Pasciak, and Jinchao Xu. Parallel multilevel preconditioners. *Math. Comp.*, 55:1–22, 1990.

Susanne C. Brenner and L. Ridgway Scott. *The Mathematical Theory of Finite Element Methods.* Springer-Verlag, Heidelberg, 1994.

Philippe G. Ciarlet. *The Finite Element Method for Elliptic Problems.* North-Holland, Amsterdam, 1980.

Peter Deuflhard, Peter Leinen, and Harry Yserentant. Concepts of an adaptive hierarchical finite element code. *IMPACT of Comput. in Sci. and Eng.*, 1:3–35, 1989.

Todd F. Dupont, Richard P. Kendall, and Henry H. Rachford. An approximate factorization procedure for self-adjoint elliptic difference equations. *SIAM J. Numerical Analysis*, 5:559–573, 1968.

Ricardo Durán, María Amelia Muschietti, and Rodolfo Rodríguez. On the asymptotic exactness of error estimators for linear triangular finite elements. *Numerische Mathematik*, 59:107–127, 1991.

Ricardo Durán and Rodolfo Rodríguez. On the asymptotic exactness of Bank-Weiser's estimator. *Numerische Mathematik*, 62:297–303, 1992.

Victor Eijkhout and Panayot Vassilevski. The role of the strengthened Cauchy-Buniakowskii-Schwarz inequality in multilevel methods. *SIAM Review*, 33:405–419, 1991.

Gene H. Golub and Charles F. Van Loan. *Matrix Computations*. Johns Hopkins University Press, 1983.

Gene H. Golub and Dianne P. O'Leary. Some history of the conjugate gradient and Lanczos algorithms: 1948-1976. *SIAM Review*, 31:50–102, 1989.

Michael Griebel. Multilevel algorithms considered as iterative methods on semidefinite systems. *SIAM J. Sci. Comput.*, 15:547–565, 1994.

Wolfgang Hackbusch. *Multigrid Methods and Applications*. Springer-Verlag, Berlin, 1985.

J. F. Maitre and F. Musy. The contraction number of a class of two level methods; an exact evaluation for some finite element subspaces and model problems. In *Multigrid Methods: Proceedings, Cologne 1981 (Lecture Notes in Mathematics 960)*, pages 535–544. Springer-Verlag, Heidelberg, 1982.

Elizabeth Ong. *Hierarchical Basis Preconditioners for Second Order Elliptic Problems in Three Dimensions*. PhD thesis, University of Washington, 1989.

Peter Oswald. *Multilevel Finite Element Approximation: Theory and Applications*. Teubner Skripten zur Numerik, B. G. Teubner, Stuttgart, 1994.

Ulrich Rüde. *Mathematical and Computational Techniques for Multilevel Adaptive Methods*. Frontiers in Applied Mathematics, Vol. 13, SIAM, Philadelphia, 1993.

Rudiger Verfürth. *A Posteriori Error Estimation and Adaptive Mesh Refinement Techniques*. Teubner Skripten zur Numerik, B. G. Teubner, Stuttgart, 1995.

Alan Weiser. *Local-Mesh, Local-Order, Adaptive Finite Element Methods with A-Posteriori Error Estimators for Elliptic Partial Differential Equations*. PhD thesis, Yale University, 1981.

Jinchao Xu. *Theory of Multilevel Methods*. PhD thesis, Cornell University, 1989.

Jinchao Xu. Iterative methods by space decomposition and subspace correction. *SIAM Review*, 34:581–613, 1992.

Harry Yserentant. On the multi-level splitting of finite element spaces. *Numerische Mathematik*, 49:379–412, 1986.

Harry Yserentant. Old and new convergence proofs for multigrid methods. In *Acta Numerica*. Cambridge University Press, 1992.

O. C. Zienkiewicz, D. W. Kelley, J. P. de S. R. Gago, and I. Babuška. Hierarchical finite element approaches, adaptive refinement, and error estimates. In *The Mathematics of Finite Elements and Applications*, pages 313–346. Academic Press, New York, 1982.