ALGEBRAIC MULTIGRID BASED ON ELEMENT INTERPOLATION (AMGe)

M. BREZINA*, A. J. CLEARY†, R. D. FALGOUT†, V. E. HENSON†, J. E. JONES†,
T. A. MANTEUFFEL*, S. F. MCCORMICK*, , AND J. W. RUGE*

Abstract. *We introduce AMGe, an algebraic multigrid method for solving the discrete equations that arise in Ritz-type finite element methods for partial differential equations. Assuming access to the element stiffness matrices, AMGe is based on the use of two local measures, which are derived from global measures that appear in existing multigrid theory. These new measures are used to determine local representations of algebraically "smooth" error components that provide the basis for constructing effective interpolation and, hence, the coarsening process for AMG. Here, we focus on the interpolation process; choice of the coarse "grids" based on these measures is the subject of current research. We develop a theoretical foundation for AMGe and present numerical results that demonstrate the efficacy of the method.*

1. Introduction. Computer simulations play an increasingly important role in scientific investigations. Indeed, as experimentation becomes more expensive, impracticable, or even proscribed, scientists are turning more and more to numerical simulation. Modern simulation packages are extremely complex, with components spanning many disciplines (e.g., hydrodynamics, radiation, transport, structures, thermal, chemistry, and electromagnetics). Also, the problems are frequently posed in multi-material regimes, with contact surfaces, interpenetrability constraints, and intricate geometries. As a result, codes are being developed to solve complex multi-physics problems on highly resolved, unstructured grids. Such large-grid simulations require the efficient union of massively parallel computing with scalable numerical algorithms such as multigrid (see e.g., [2]).

An especially effective method for many of the problems that arise in these applications is algebraic multigrid (AMG [6, 4, 5, 7, 21, 18, 20, 19]). AMG is a method for solving matrix equations that is based on multigrid concepts, but constructs the coarsening process in an algebraic way that requires no explicit knowledge of the geometry. It examines the matrix entries to determine a sequence of smaller matrix problems that serve as coarse-level equations. AMG also determines associated inter-level transfer operators (restriction and prolongation), then solves the original matrix equation in a multigrid-like process based on these automatically-constructed components. AMG has been shown to be well-suited for solving unstructured grid problems, and to work well over a wide variety of applications (see, e.g., [11]). It has been applied successfully to M-matrix problems where the so-called strength of connection is easily measured (this measure is used to determine which variables are strongly representative of the errors left by relaxation, so that they can be used to construct the coarse levels). It also applies well to scalar problems that depart substantially from M-matrix discretizations. However, for problems where strength of connection is not easily measured, AMG is not effective without certain problem-specific modifications or careful parameter tuning. For such cases, there is no systematic AMG approach that has proven effective in any kind of general context. There are still other problems (e.g., thin-body elasticity on unstructured grids) for which AMG and other iterative methods in general have failed to achieve full optimality (i.e., convergence factors bounded uniformly in the size of the problem). The goal of our research is to develop a more robust AMG for solving these difficult problems.

This paper introduces an algebraic multigrid method for solving partial differential equations discretized by Ritz-type finite element methods. As a departure from standard AMG, where only the operator matrix is required, this approach assumes access to element stiffness matrices. We thus refer to it as AMGe (AMG henceforth refers to the standard scheme). This new approach is based on the use of either of two measures (derived from global measures used in existing theory) to determine algebraically "smooth" error and to construct effective interpolation. AMGe uses a minimization principle based on the *element interpolation*

*Applied Math Department, Campus Box 526, University of Colorado at Boulder, Boulder, CO 80309-0526. *email*:{mbrezina, tmanteuf, stevem}@colorado.edu. This work was sponsored in part by the National Science Foundation under grant number DMS-9706866 and the Department of Energy under grant number DE-FG03-93ER25165.

†Center for Applied Scientific Computing, Lawrence Livermore National Laboratory, P.O. Box 808, L-561, Livermore, CA 94551. *email*:{cleary,rfalgout,vhenson,jjones}@llnl.gov

scheme first introduced in [17]. Other multigrid methods, using minimization principles for constructing energetically stable inter-grid transfer operators, have recently appeared in [24, 25, 13].

Some notation and the key ideas behind AMG are summarized in the next section. (Nevertheless, we assume that the reader is familiar with AMG methods and terminology. For more detail, see [11] and [19].) In particular, we discuss the notion of *strength of dependence* and its role in defining the basic AMG components. In section 3, we define a heuristic based on two *global measures* and establish a corresponding two-level convergence result. We "localize" these measures in section 4, and describe how they can be used to compute the interpolation operator for AMGe. We also discuss the relationship between the local and global measures in subsection 4.3. Section 5 contains numerical results supporting the theory and demonstrating the efficacy of the approach. Concluding remarks are made in section 6.

2. Preliminaries. We begin this section by describing notation. Capital Roman letters (A, B, P, R) denote matrices and bold lower case Roman and Greek letters denote vectors $(\mathbf{u}, \mathbf{v}, \boldsymbol{\varepsilon})$. The ith component of the vector $\mathbf{q}$ is denoted by q_i. Other lower case letters denote scalars, while capital calligraphic letters denote sets and spaces $(\mathcal{C}, \mathcal{F}, \mathcal{S})$, with the singular exception that $\mathcal{A}$ is used to denote finite element stiffness matrices. We define the A-inner product by $\langle \cdot,\ \cdot \rangle_A := \langle A\cdot,\ \cdot \rangle$, where $\langle \cdot,\ \cdot \rangle$ is the standard Euclidean inner product, and the A-norm (also called the energy norm) by $\|\cdot\|_A := \langle \cdot,\ \cdot \rangle_A^{1/2}$.

Assume that we are given an $n \times n$ symmetric positive definite matrix A expressed as the sum of a given set of finite element stiffness matrices: $A = \sum_{\alpha \in \mathcal{T}} \mathcal{A}_\alpha$, where $\mathcal{T}$ is the set of finite elements used to discretize the problem and each $\mathcal{A}_\alpha$ is symmetric positive semi-definite. We do not assume access to a spatial grid or the ability to create new finite element stiffness matrices.

We seek the solution $\mathbf{u} \in \mathbb{R}^n$ to the linear system

$$A\mathbf{u} = \mathbf{f}, \tag{2.1}$$

for a given $\mathbf{f} \in \mathbb{R}^n$. Standard iterative schemes, like Gauss-Seidel and Krylov space methods, tend to converge slowly for large-scale problems of this type that arise from partial differential equations. The difficulty is that smooth error components are typically attenuated very slowly by these simple processes, because they are based on local properties (i.e., local connections in A). Multigrid methods attempt to correct this limitation by representing the smooth errors on increasingly coarser, and, therefore, more global levels.

To describe how system (2.1) could be solved by a multilevel method, let P be an $n \times n_c$ *interpolation* or *prolongation* matrix that transfers level n_c corrections to level n, with $n_c < n$. P could be determined *geometrically* by, say, linear interpolation (cf. [8]) when $\mathbb{R}^{n_c}$ and $\mathbb{R}^n$ represent grids whose nodal positions are accessible. P could instead be determined *algebraically* by so-called operator interpolation (cf. [1]), which is based on the entries of A. In any event, we choose P^T as the *restriction* matrix that transfers level n residuals to level n_c. The two-grid method for solving (2.1) is then defined as follows:

$$\text{Relax } \nu_1 \text{ times on } A\mathbf{u} = \mathbf{f}. \tag{2.2a}$$

$$\text{Correct } \mathbf{u} \leftarrow \mathbf{u} + P(P^TAP)^{-1}P^T(\mathbf{f} - A\mathbf{u}). \tag{2.2b}$$

$$\text{Relax } \nu_2 \text{ times on } A\mathbf{u} = \mathbf{f}. \tag{2.2c}$$

Note the use of P^TAP in correction step (2.2b). This so-called *Galerkin* coarse-grid operator, together with the use of P^T as the restriction operator, amounts to a *variational* form of multigrid. When A is symmetric, the correction step minimizes the energy norm of the fine-grid error over all possible corrections from the range of P (cf. [8]). To solve (2.1) in practice, one would use a multilevel method that recursively applies algorithm (2.2) to solve the linear system involving P^TAP in correction step (2.2b).

Further examination of (2.2) reveals that relaxation and coarse-grid correction must be chosen to complement each other: error not reduced by one must be reduced by the other. In this paper, we fix the choice of relaxation, then determine interpolation. The relaxation we choose is a simple pointwise method, like Richardson, damped Jacobi, or Gauss-Seidel, that satisfies the following heuristic:

H1: *Error in the direction of an eigenvector associated with a large eigenvalue is rapidly reduced by relaxation, while error in the direction of an eigenvector associated with a small eigenvalue is reduced by a factor that may approach* 1 *as the eigenvalue approaches* 0.

Error that is not rapidly reduced by relaxation is called *algebraically smooth*. The actual character of algebraically smooth error depends on the operator and the type of relaxation, but it loosely means that the residual is small when compared to the error itself (we will be more precise about this shortly). This does not mean that the error is smooth in any geometric sense. Thus, error at a point may be very different from the errors at neighboring points, yet it might be difficult to reduce the error by relaxation. Such is the case for anisotropic problems, where algebraically smooth error that point Gauss-Seidel relaxation cannot effectively reduce can be geometrically oscillatory in the direction of small coefficients of the differential equation. In any case, the interpolation matrix, P, must be defined so that algebraically smooth error is effectively eliminated in step (2.2b) and the coarse-grid equations, which involve P^TAP, are amenable to solution.

2.1. AMG. To define the multigrid components in AMG, we use the following heuristic (cf. [6, 4, 19]) based on special properties of M-matrices:

H2: *Smooth error varies slowest in the direction of strong dependence.*

Here, we say that *unknown i strongly depends on unknown j* if

$$-a_{i,j} \geq \theta \max_{k \neq i}\{-a_{i,k}\}, \quad \text{for some fixed } \theta \in (0,1). \tag{2.3}$$

Thus, strong dependence is characterized by matrix coefficients that are large in the sense of (2.3). A typical choice for parameter θ is 0.25.

Although AMG was developed with M-matrices in mind, in practice it is not limited to this class of problems. However, the standard method does rely on **H2**, and our sense of strong dependence may not be suitable for many important classes of problems. For example, one simple problem with which standard AMG has difficulty is the Poisson equation on a rectangular grid, discretized with bilinear quadrilateral elements, where the fine-grid elements are stretched to a 10 : 1 aspect ratio. This yields the coefficient stencil

$$\begin{bmatrix} -1 & -3.9 & -1 \\ 1.9 & 8 & 1.9 \\ -1 & -3.9 & -1 \end{bmatrix}. \tag{2.4}$$

In (2.4), it is not readily apparent from the size of the off-diagonal entries that the direction of strongest dependence is vertical. Since **H2** is used to define all of the AMG components, and it requires a clear understanding of strong dependence, AMG can exhibit degraded performance (see Table 5.2). For this simple case, slow convergence of AMG can be ameliorated by simply tuning its parameters (e.g., setting $\theta = 0.5$) or by more elaborate algorithmic "fixes" (e.g., *iterative weight interpolation* [11] or geometric/algebraic interpolation methods [12, 10, 9]). Another approach is to replace **H2** by a heuristic that leads to a more robust AMG algorithm. Exploring this possibility, as we begin to do in the next section, is the primary aim of this paper.

3. Global Measures and Convergence Bounds. This paper takes a slightly different approach, using a heuristic based not on M-matrices, but on the eigenvectors of A. In a two-grid scheme, coarse-grid correction will completely eliminate error in Range$\,(P)$, the range of the interpolation operator. To complement the action of relaxation, which satisfies **H1**, the interpolation matrix must satisfy the following heuristic:

H3: *Interpolation must be able to approximate an eigenvector with error bound proportional to the size of the associated eigenvalue.*

To make **H3** more rigorous, define $Q : \mathbb{R}^n \rightarrow \mathbb{R}^n$ to be a convenient projection onto Range$\,(P)$, that is,

$$Q = PR, \tag{3.1}$$

for some restriction operator $R : \mathbb{R}^n \rightarrow \mathbb{R}^{n_c}$ such that $RP = I_c$, the identity on $\mathbb{R}^{n_c}$. The specific form for Q (and, hence, R) will not become important until section 4. For any vector $\mathbf{e} \in$ Range$\,(P)$, we have $Q\mathbf{e} = \mathbf{e}$. Thus, $I - Q$ can be used to measure the defect of interpolation. With this in mind, we now define two measures of how well **H3** is satisfied:

$$M_1(Q, \mathbf{e}) := \frac{\langle (I-Q)\mathbf{e},\ (I-Q)\mathbf{e} \rangle}{\langle A\mathbf{e},\ \mathbf{e} \rangle}, \tag{3.2}$$

$$M_2(Q, \mathbf{e}) := \frac{\langle A(I-Q)\mathbf{e},\ (I-Q)\mathbf{e} \rangle}{\langle A\mathbf{e},\ A\mathbf{e} \rangle}. \tag{3.3}$$

Measure M_2 was used in the early multigrid theory [16, 14, 15] to establish optimal convergence of the V-cycle algorithm under full regularity assumptions on the associated partial differential equation. Measure M_1 was introduced in [5] and used more recently to establish convergence, independent of the coarse-grid size, of a two-level method for linear elasticity [23]. It is also an essential ingredient of the regularity-free multilevel theory found in [3]. We develop the relevant two-grid theory here for both measures so that we can tailor the results to our needs.

It has not been our practice to use diagonal conditioning of A in standard AMG. Such a scaling generally changes the nature of smooth errors. Since current schemes at some point rely on a premise of how smooth error behaves (e.g., that it is locally constant), then diagonal scaling can make it more difficult for AMG to handle. However, no such premise of smoothness is made anywhere in AMGe. Thus, in the remainder of this paper, we are free to assume for convenience that matrix A has been scaled so that its diagonal is the identity. For a general symmetric positive-definite matrix with diagonal $D \neq I$, this can be assured by a diagonal scaling that replaces A by $D^{-1/2}AD^{-1/2}$. Note that this transformation must be considered in the representation of A as a sum of local stiffness matrices, but this is just a straightforward rescaling of the variables. This scaling does, however, bear on the practicality of our results because we analyze AMG based on Richardson iteration, which is not generally a good smoother for matrices that have widely varying diagonal entries. Thus, if diagonal scaling is not used, then in general it would be wise to use a relaxation scheme like damped Jacobi and adjust measures M_1 and M_2 accordingly.

Our theory assumes that either M_1 or M_2 is bounded uniformly in $\mathbf{e} \in \mathbb{R}^n \backslash \{\mathbf{0}\}$. To see how this assumption relates to **H3**, suppose that $\mathbf{e}$ is an eigenvector of A corresponding to a small eigenvalue. Then, for M_1 or M_2 to be bounded, since the denominators of the two measures are small, the numerators must also be small. Thus, Q must accurately interpolate eigenvectors belonging to small eigenvalues. On the other hand, if $\mathbf{e}$ is an eigenvector of A corresponding to a large eigenvalue, then the denominators of the two measures are large, so the numerators may be large. Thus, Q need not accurately interpolate eigenvectors belonging to large eigenvalues.

We now prove convergence results based on M_1 or M_2 for two-level algorithm (2.2).

Lemma 3.1. *Let Q be any projection onto* Range(P). *Assume that either of the following two approximation properties are satisfied for some constant K:*

$$M_1(Q, \mathbf{e}) \leq K \quad \forall \mathbf{e} \in \mathbb{R}^n \backslash \{\mathbf{0}\}, \tag{3.4}$$

$$M_2(Q, \mathbf{e}) \leq K \quad \forall \mathbf{e} \in \mathbb{R}^n \backslash \{\mathbf{0}\}. \tag{3.5}$$

If $\mathbf{e} \neq 0$ is A-orthogonal to Range(P), *then*

$$\frac{1}{K} \leq \frac{\|A\mathbf{e}\|^2}{\langle A\mathbf{e},\ \mathbf{e}\rangle} \leq \|A\|\,. \tag{3.6}$$

Proof. The upper bound in (3.6) follows easily from the definition of the matrix norm. To prove the lower bound, note that Range$(Q) =$ Range(P). Hence, if $\mathbf{e}$ is A-orthogonal to Range(P), then

$$\langle A\mathbf{e},\ Q\mathbf{v}\rangle = 0 \ \ \forall \mathbf{v} \in \mathbb{R}^n. \tag{3.7}$$

First, assume that (3.4) holds. From (3.7) and the Cauchy-Schwarz inequality, we have

$$\begin{aligned} \langle A\mathbf{e},\ \mathbf{e}\rangle &= \langle A\mathbf{e},\ (I-Q)\mathbf{e}\rangle \\ &\leq \|A\mathbf{e}\|\,\|(I-Q)\mathbf{e}\| \\ &\leq \|A\mathbf{e}\|\,\langle A\mathbf{e},\ \mathbf{e}\rangle^{1/2} K^{1/2}. \end{aligned}$$

The lower bound in (3.6) now follows by dividing through by $\langle A\mathbf{e},\ \mathbf{e}\rangle K^{1/2}$ and squaring the result.

Now, assume that (3.5) holds. From (3.7) and the Cauchy-Schwarz inequality, we have

$$\begin{aligned} \langle A\mathbf{e},\ \mathbf{e}\rangle &\leq \langle A\mathbf{e},\ \mathbf{e}\rangle + \langle AQ\mathbf{e},\ Q\mathbf{e}\rangle \\ &= \langle A\mathbf{e},\ \mathbf{e}\rangle - \langle A\mathbf{e},\ Q\mathbf{e}\rangle - \langle AQ\mathbf{e},\ \mathbf{e}\rangle + \langle AQ\mathbf{e},\ Q\mathbf{e}\rangle \\ &= \langle A(I-Q)\mathbf{e},\ (I-Q)\mathbf{e}\rangle \\ &\leq \|A\mathbf{e}\|^2 K. \end{aligned}$$

The lower bound in (3.6) now follows by dividing through by $\langle A\mathbf{e},\ \mathbf{e}\rangle K$. ∎

Denote the A-orthogonal projection onto the Range (P) by S. Thus,

$$S := P(P^T AP)^{-1}P^T A. \tag{3.8}$$

The error propagation matrix for the coarse-grid correction step (2.2b) is $I - S$. A Richardson iteration with step-size parameter $s = \omega / \|A\|$, $\omega \in (0, 2)$, has the error propagation matrix $G = I - sA$. If we choose $(\nu_1, \nu_2) = (0, 1)$ in (2.2), then the associated error propagation matrix for this simple two-grid scheme is $G(I - S)$. The following theorem analyzes its convergence by bounding its error propagation matrix in the A-norm. Convergence results for other values of (ν_1, ν_2) then follow naturally [15].

Analogous multilevel results can be found in [16, 14, 15] for approximation property (3.5), and in [3, 22] for (3.4) under the additional assumption of *energetic stability of interpolation*, a sufficient condition for which is that $\|P(P^T P)^{-1}P^T\|_A$ be bounded uniformly on all levels.

THEOREM 3.2. *Assume that either approximation property (3.4) or (3.5) is satisfied for some constant K. Then*

$$\|G(I - S)\|_A \le \left(1 - \frac{\omega(2-\omega)}{K\,\|A\|}\right)^{1/2}. \tag{3.9}$$

Proof. First note that (3.6) implies $K \ge 1/\|A\| \ge \omega(2-\omega)/\|A\|$, so that (3.9) makes sense. We have

$$\begin{aligned} \langle AG\mathbf{e},\ G\mathbf{e}\rangle &= \langle A\mathbf{e},\ \mathbf{e}\rangle - 2s\,\langle A\mathbf{e},\ A\mathbf{e}\rangle + s^2\left\langle A^2\mathbf{e},\ A\mathbf{e}\right\rangle \\ &\le \langle A\mathbf{e},\ \mathbf{e}\rangle - \frac{\omega(2-\omega)}{\|A\|}\,\langle A\mathbf{e},\ A\mathbf{e}\rangle . \end{aligned}$$

Replacing $\mathbf{e}$ with $(I - S)\mathbf{e}$ and applying the result in Lemma 3.1 yields

$$\begin{aligned} \|G(I-S)\mathbf{e}\|_A^2 &\le \langle A(I-S)\mathbf{e},\ (I-S)\mathbf{e}\rangle - \frac{\omega(2-\omega)}{\|A\|}\,\|A(I-S)\mathbf{e}\|^2 \\ &\le \left(1 - \frac{\omega(2-\omega)}{K\,\|A\|}\right)\|\mathbf{e}\|_A^2 . \end{aligned}$$

∎

Notice that the bound on the convergence factor approaches 1 as K becomes large. Conversely, smaller K yields a smaller bound on the convergence factor. Our aim is to determine P so that, for some appropriate Q, either (3.4) or (3.5) is satisfied for a reasonably small K.

We also remark that the above results can be generalized to apply when (2.1) is a consistent system with symmetric positive semi-definite matrix A. Measures M_1 and M_2 must be restricted to $\mathbf{e} \notin \mathrm{Null}(A)$. A finite bound K in (3.4) or (3.5) then implies that interpolation is exact for $\mathbf{e} \in \mathrm{Null}(A)$, which in turn implies that the correction step involves a consistent system. A zero initial guess and relaxation using a polynomial method like Richardson iteration ensures that the approximate solution remains orthogonal to Null(A).

4. Interpolation Using Local Measures. Quantities M_1 and M_2 are global measures of the quality of interpolation. Our intent is to use these measures to determine an effective strategy for constructing interpolation in AMG, but it is not practical to do this globally. In this section, we discuss an approach for localizing these measures for linear systems (2.1) that arise from finite element discretizations.

Recall that A is given as the sum of finite element stiffness matrices: $A = \sum_{\alpha\in\mathcal{T}} \mathcal{A}_\alpha$. Now, we do not assume access to an underlying spatial grid. However, we can construct an artificial grid based on the graph associated with A, with vertices $\mathcal{G} := \{1, 2, \ldots, n\}$ and edges $\mathcal{E} := \{(i, j)\ :\ a_{ij} \ne 0\ \text{ for } i \ne j\}$. Grid point (vertex) $i \in \mathcal{G}$ is associated with unknown u_i.

We first define the point set of an element:

$$\mathcal{M}_\alpha := \left\{ j\ :\ \boldsymbol{\varepsilon}_j^T \mathcal{A}_\alpha \boldsymbol{\varepsilon}_j \ne 0 \right\}, \tag{4.1}$$

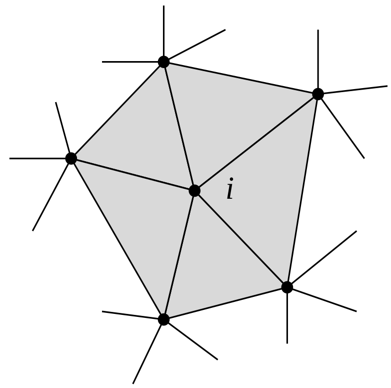

FIG. 4.1. *Local neighborhoods.*

where $\boldsymbol{\varepsilon}_j$ is the canonical basis vector associated with unknown j. Next, define the neighborhood of grid point i as the set of elements and set of points

$$\mathcal{T}_i := \{\alpha \in \mathcal{T} : \boldsymbol{\varepsilon}_i^T \mathcal{A}_\alpha \boldsymbol{\varepsilon}_i \neq 0\}, \tag{4.2}$$

$$\mathcal{N}_i := \cup_{\alpha \in \mathcal{T}_i} \mathcal{M}_\alpha, \tag{4.3}$$

respectively (see Figure 4.1). Define the local matrices on neighborhood i by

$$A_i = \sum_{\alpha \in \mathcal{T}_i} \mathcal{A}_\alpha. \tag{4.4}$$

We also assume that a coarse grid has been selected, that is, the points in $\mathcal{G}$ have been partitioned into coarse-grid points $\mathcal{C}$ and fine-grid points $\mathcal{F}$ such that $\mathcal{C} \cup \mathcal{F} = \mathcal{G}$ and $\mathcal{C} \cap \mathcal{F} = \emptyset$. We now seek the $n \times n_c$ interpolation matrix P, where $n_c = |\mathcal{C}|$, that interpolates from the coarse-grid points $\mathcal{C}$ to the entire grid $\mathcal{G}$.

Two conflicting goals drive the construction of P. The first is to minimize the bound on measure M_1 or M_2, while the second is to control the sparsity of the coarse-grid system involving P^TAP. Focusing on the second goal first, assume that the coarse-grid points interpolate to themselves exactly, that is, P restricted to $\mathcal{C}$ is the identity, while fine-grid points interpolate only from coarse-grid points in their neighborhood, that is, from $\mathcal{C}_i := \mathcal{N}_i \cap \mathcal{C}$.

To make the construction more clear, suppose that the rows and columns of A have been arranged so that the fine-grid points come first, followed by the coarse-grid points. We may then write A in block form as follows:

$$A = \begin{bmatrix} A_{ff} & A_{fc} \\ A_{cf} & A_{cc} \end{bmatrix}. \tag{4.5}$$

In this context, the interpolation matrix has the block form

$$P = \begin{bmatrix} P_{fc} \\ I_c \end{bmatrix}. \tag{4.6}$$

Alternatively, we may define the projection

$$Q = \begin{bmatrix} 0 & P_{fc} \\ 0 & I_c \end{bmatrix}, \tag{4.7}$$

which implies the choice of $R = [0, I_c]$ as the restriction in (3.1).

In what follows, we develop a strategy for constructing the rows of P_{fc}, that is, the rows of Q corresponding to each point $i \in \mathcal{F}$, which we denote

$$\mathbf{q}_i^T := \boldsymbol{\varepsilon}_i^T Q. \tag{4.8}$$

Restricting interpolation to a neighborhood of coarse-grid points is equivalent to choosing

$$\mathbf{q}_i \in \mathcal{Z}_i := \{\mathbf{v} \in \mathbb{R}^n \ : \ v_j = 0 \text{ for } j \notin \mathcal{C}_i\}. \tag{4.9}$$

We now localize measures M_1 and M_2 by defining

$$M_{i,1}(Q,\mathbf{e}) := \frac{\left\langle \varepsilon_i \varepsilon_i^T (I-Q)\mathbf{e},\ \varepsilon_i \varepsilon_i^T (I-Q)\mathbf{e} \right\rangle}{\langle A_i \mathbf{e},\ \mathbf{e}\rangle}, \tag{4.10}$$

$$M_{i,2}(Q,\mathbf{e}) := \frac{\left\langle A_i \varepsilon_i \varepsilon_i^T (I-Q)\mathbf{e},\ \varepsilon_i \varepsilon_i^T (I-Q)\mathbf{e} \right\rangle}{\langle A_i \mathbf{e},\ A_i \mathbf{e}\rangle}, \tag{4.11}$$

for any $\mathbf{e} \notin \mathrm{Null}(A_i)$. Notice for $i \in \mathcal{C}$ that $M_{i,1} = M_{i,2} = 0$, while for $i \in \mathcal{F}$ the above measures only depend on the ith row of Q, which is to be chosen in $\mathcal{Z}_i$. To emphasize this dependence, when the meaning is clear we write

$$M_{i,1}(\mathbf{q_i},\mathbf{e}) = \frac{\left\langle (\varepsilon_i - \mathbf{q_i})^T \mathbf{e},\ (\varepsilon_i - \mathbf{q_i})^T \mathbf{e} \right\rangle}{\langle A_i \mathbf{e},\ \mathbf{e}\rangle}, \tag{4.12}$$

$$M_{i,2}(\mathbf{q_i},\mathbf{e}) = \frac{\left\langle (\varepsilon_i - \mathbf{q_i})^T \mathbf{e},\ (\varepsilon_i - \mathbf{q_i})^T \mathbf{e} \right\rangle}{\langle A_i \mathbf{e},\ A_i \mathbf{e}\rangle}, \tag{4.13}$$

for $\mathbf{q_i} \in \mathcal{Z}_i$ and $\mathbf{e} \notin \mathrm{Null}(A_i)$. (Recall that A has unit diagonal.)

Heuristic **H3**, as applied to these local measures, now relates interpolation accuracy to local eigenvectors of A_i. This makes it practical to use $M_{i,1}$ and $M_{i,2}$ to compute interpolation. Since we wish to make these local measures small, interpolation is defined so that the $\mathbf{q}_i$ in (4.8) is the $\arg\min$ (that is, the argument that attains the minimum) of one of the following min-max problems:

$$K_{i,p} := \min_{\mathbf{q_i} \in \mathcal{Z}_i} \max_{\mathbf{e} \notin \mathrm{Null}(A_i)} M_{i,p}(\mathbf{q_i},\mathbf{e}), \tag{4.14}$$

for $p = 1$ or 2. Note that if there exists a $\mathbf{q_i} \in \mathcal{Z}_i$ that yields $K_{i,p} < \infty$, then $\mathbf{q_i}$ satisfies the constraint

$$(\varepsilon_i - \mathbf{q_i})^T \mathbf{e} = 0 \quad \forall \mathbf{e} \in \mathrm{Null}(A_i).$$

Thus, min-max problem (4.14) can be restated as the constrained min-max problem

$$K_{i,p} = \min_{\mathbf{q_i} \in \mathcal{Z}_i} \max_{\mathbf{e} \perp \mathrm{Null}(A_i)} M_{i,p}(\mathbf{q_i},\mathbf{e}), \quad \text{subject to} \quad (\varepsilon_i - \mathbf{q_i})^T \mathbf{e} = 0 \quad \forall \mathbf{e} \in \mathrm{Null}(A_i), \tag{4.15}$$

for $p = 1$ or 2. The next two subsections focus on solving these min-max problems. In Section 4.3, we relate the local measures to the global measures.

4.1. Computing Interpolation by Fitting Eigenvectors. One way to compute the $\mathbf{q}_i$ in (4.14) or (4.15) is to "fit" the eigenvectors of A_i, as quantified in the following theorem.

THEOREM 4.1. *Suppose we have computed the eigen-decomposition*

$$A_i V_i = V_i \Lambda_i, \quad V_i^T V_i = I. \tag{4.16}$$

The columns of V_i are the orthonormalized eigenvectors of A_i, and Λ_i is the diagonal matrix formed from the corresponding eigenvalues. Assume that this eigen-decomposition is ordered to distinguish between zero eigenvalues and positive eigenvalues (that form the diagonal matrix Λ_{i+}):

$$V_i = \begin{bmatrix} V_{i0} & V_{i+} \end{bmatrix}, \quad \Lambda_i = \begin{bmatrix} 0 & 0 \\ 0 & \Lambda_{i+} \end{bmatrix}. \tag{4.17}$$

Then min-max problem (4.15) is equivalent to the following constrained least-squares problem:

$$\min_{\mathbf{q}_i} \left\| \Lambda_{i+}^{-p/2} V_{i+}^T (\varepsilon_i - \mathbf{q}_i) \right\|^2, \quad \textit{subject to} \quad V_{i0}^T (\varepsilon_i - \mathbf{q}_i) = 0, \tag{4.18}$$

for $p = 1$ or 2.

Proof. Note that the null-space constraint in (4.15) is equivalent to that in (4.18). Assume first that $\mathbf{q}_i$ satisfies (4.15) with $p = 1$. Since $\mathbf{e} \perp \mathrm{Null}(A_i)$, we can write $\mathbf{e} = V_{i+}\Lambda_{i+}^{-1/2}\mathbf{w}$, which yields

$$\begin{aligned}\min_{\mathbf{q}_i \in \mathcal{Z}_i} \max_{\mathbf{e} \perp \mathrm{Null}(A_i)} M_{i,1}(\mathbf{q}_i, \mathbf{e}) &= \min_{\mathbf{q}_i \in \mathcal{Z}_i} \max_{\mathbf{w}} \frac{\left\|(\boldsymbol{\varepsilon}_i - \mathbf{q}_i)^T V_{i+}\Lambda_{i+}^{-1/2}\mathbf{w}\right\|^2}{\|\mathbf{w}\|^2} \\ &= \min_{\mathbf{q}_i \in \mathcal{Z}_i} \left\|\Lambda_{i+}^{-1/2} V_{i+}^T (\boldsymbol{\varepsilon}_i - \mathbf{q}_i)\right\|^2 .\end{aligned}$$

Assume now that $\mathbf{q}_i$ satisfies (4.15) with $p = 2$. Writing $\mathbf{e} = V_{i+}\Lambda_{i+}^{-1}\mathbf{w}$, we then have

$$\begin{aligned}\min_{\mathbf{q}_i \in \mathcal{Z}_i} \max_{\mathbf{e} \perp \mathrm{Null}(A_i)} M_{i,2}(\mathbf{q}_i, \mathbf{e}) &= \min_{\mathbf{q}_i \in \mathcal{Z}_i} \max_{\mathbf{w}} \frac{\left\|(\boldsymbol{\varepsilon}_i - \mathbf{q}_i)^T V_{i+}\Lambda_{i+}^{-1}\mathbf{w}\right\|^2}{\|\mathbf{w}\|^2} \\ &= \min_{\mathbf{q}_i \in \mathcal{Z}_i} \left\|\Lambda_{i+}^{-1} V_{i+}^T (\boldsymbol{\varepsilon}_i - \mathbf{q}_i)\right\|^2 .\end{aligned}$$

∎

Computing the interpolation weights $\mathbf{q}_i$ using (4.18) requires eigen-decomposition (4.16), which is not the most efficient method. We introduce a simpler approach in the next subsection. However, we include this notion of fitting eigenvectors because it is useful for understanding the basic principles involved in selecting interpolation.

4.2. A More Practical Algorithm for Computing Interpolation. Here describe a practical algorithm for determining when (4.14) or (4.15) has a (unique) solution for $i \in \mathcal{F}$, and for computing Q when a solution does exist. One important consequence of this characterization is its update property: whenever the solution with the current interpolatory set does not exist, we can add points to $\mathcal{C}_i$ and test again for solvability without redoing all of the computation.

Assume first that grid point $i \in \mathcal{F}$ has a neighborhood, as depicted in Figure 4.1, consisting of n_i points in set $\mathcal{N}_i$, with n_f fine-grid points and n_c coarse-grid points in C_i. Next, order the unknowns and equations of matrix A_i so that unknown i is first, followed by the other fine-grid points, with the coarse-grid points last. The neighborhood matrix and its square can then be written as

$$A_i = \begin{bmatrix} A_{ff}^{(1)} & A_{fc}^{(1)} \\ A_{cf}^{(1)} & A_{cc}^{(1)} \end{bmatrix} \text{ and } A_i^2 = \begin{bmatrix} A_{ff}^{(2)} & A_{fc}^{(2)} \\ A_{cf}^{(2)} & A_{cc}^{(2)} \end{bmatrix},$$

respectively, and $\boldsymbol{\varepsilon}_i$ becomes $\boldsymbol{\varepsilon}_1$.

In the remainder of this subsection, we drop the subscript i whenever the meaning is clear. Set $\mathcal{Z}_i$ restricted to the neighborhood becomes

$$\mathcal{Z} := \{\mathbf{e} \in \mathbb{R}^{n_i} \; : \; e_j = 0 \; \forall j \notin C_i\}.$$

We can then interpret (4.15) with $p = 1$ or 2 as the problem of determining a vector $\mathbf{q} \in \mathcal{Z}$ that minimizes $\max_{\mathbf{e} \notin \mathrm{Null}(A_i)} M_{i,p}(\mathbf{q}, \mathbf{e})$, subject to the constraint

$$(\boldsymbol{\varepsilon}_1 - \mathbf{q})^T \mathbf{e} = 0 \qquad \forall\, \mathbf{e} \in \mathrm{Null}(A_i) = \mathrm{Null}(A_i^2).$$

That is, we require

$$\boldsymbol{\varepsilon}_1 - \mathbf{q} \in \mathrm{Range}\,(A_i) = \mathrm{Range}\left(A_i^2\right). \tag{4.19}$$

Our first concern is the existence of such a vector $\mathbf{q}$. For this, we let $\hat{\boldsymbol{\varepsilon}}_1 \in \mathbb{R}^{n_f}$ denote the first canonical basis vector of length n_f.

Lemma 4.2. *There exists $\mathbf{q} \in \mathcal{Z}$ such that $\boldsymbol{\varepsilon}_1 - \mathbf{q} \in \mathrm{Range}\,(A_i^p)$ if and only if*

$$\hat{\boldsymbol{\varepsilon}}_1 \in \mathrm{Range}\left(A_{ff}^{(p)}\right),$$

with $p = 1$ *or* 2.

Proof. Assume that $\hat{\boldsymbol{\varepsilon}}_1 \in \text{Range}\left(A_{ff}^{(p)}\right)$ so that

$$\hat{\boldsymbol{\varepsilon}}_1 = A_{ff}^{(p)} \hat{\boldsymbol{\delta}}_1$$

for some $\hat{\boldsymbol{\delta}}_1 \in \mathbb{R}^{n_f}$. Then

$$A_i^p \boldsymbol{\delta} := \begin{bmatrix} A_{ff}^{(p)} & A_{fc}^{(p)} \\ A_{cf}^{(p)} & A_{cc}^{(p)} \end{bmatrix} \begin{pmatrix} \hat{\boldsymbol{\delta}}_1 \\ 0 \end{pmatrix} = \boldsymbol{\varepsilon}_1 - \mathbf{q} \ \in \text{Range}\left(A_i^p\right),$$

and $\mathbf{q} \in \mathcal{Z}$.

Conversely, suppose there exists $\mathbf{q} \in \mathcal{Z}$ such that $\boldsymbol{\varepsilon}_1 - \mathbf{q} \in \text{Range}\left(A_i^p\right)$, that is, there exists $\boldsymbol{\delta}$ such that

$$\boldsymbol{\varepsilon}_1 - \mathbf{q} = A_i^p \boldsymbol{\delta}.$$

This, in turn, implies that

$$\hat{\boldsymbol{\varepsilon}}_1 = \left[A_{ff}^{(p)},\ A_{fc}^{(p)}\right] \boldsymbol{\delta} \ \in \text{Range}\left(\left[A_{ff}^{(p)},\ A_{fc}^{(p)}\right]\right).$$

The proof will be completed by demonstrating that

$$\text{Range}\left(\left[A_{ff}^{(p)},\ A_{fc}^{(p)}\right]\right) = \text{Range}\left(A_{ff}^{(p)}\right).$$

This is certainly true if $A_{ff}^{(p)}$ is nonsingular. Assume otherwise, and let $\hat{\boldsymbol{\delta}}$ be a nonzero vector in $\text{Null}(A_{ff}^{(p)})$. Then

$$\left\langle \begin{bmatrix} A_{ff}^{(p)} & A_{fc}^{(p)} \\ A_{cf}^{(p)} & A_{cc}^{(p)} \end{bmatrix} \begin{pmatrix} \hat{\boldsymbol{\delta}} \\ 0 \end{pmatrix},\ \begin{pmatrix} \hat{\boldsymbol{\delta}} \\ 0 \end{pmatrix} \right\rangle = 0.$$

Since A_i^p is symmetric positive semi-definite, then 0 is an extreme value of $\langle A_i^p \mathbf{e},\ \mathbf{e}\rangle$, which implies that the vector $(\hat{\boldsymbol{\delta}},\ 0)^T$ is an eigenvector of A_i^p with eigenvalue 0. In other words, $(\hat{\boldsymbol{\delta}},\ 0)^T \in \text{Null}(A_i^p)$, which implies that

$$\text{Null}(A_{ff}^{(p)}) = \text{Null}(A_{cf}^{(p)}),$$

which, in turn, implies that

$$\text{Range}\left(A_{ff}^{(p)}\right) = \text{Range}\left(\left(A_{cf}^{(p)}\right)^T\right) = \text{Range}\left(A_{fc}^{(p)}\right),$$

and the lemma is proved. ∎

Rewriting (4.19), we want $\boldsymbol{\delta} \in \mathbb{R}^{n_i}$ such that

$$\boldsymbol{\varepsilon}_1 - \mathbf{q} = A_i^p \boldsymbol{\delta}$$

for some $\mathbf{q} \in \mathcal{Z}$. By the proof of Lemma 4.2, the set of all such $\boldsymbol{\delta}$ is

$$Y^{(p)} := \left\{ \boldsymbol{\delta} \in \mathbb{R}^{n_i} \ : \ \left[A_{ff}^{(p)},\ A_{fc}^{(p)}\right] \boldsymbol{\delta} = \hat{\boldsymbol{\varepsilon}}_1 \right\}.$$

If $Y^{(p)}$ is empty, then the constraint in (4.15) cannot be satisfied and $K_{i,p} = \infty$. In this case, more points must be added to C_i for (4.15) to have a solution. If $Y^{(p)}$ is not empty, then any $\boldsymbol{\delta} \in Y^{(p)}$ can be written as $\boldsymbol{\delta} = \boldsymbol{\delta}^* + \boldsymbol{\gamma}$, where $\boldsymbol{\delta}^*$ is a particular element of $Y^{(p)}$ and $\boldsymbol{\gamma} \in \text{Null}([A_{ff}^{(p)},\ A_{fc}^{(p)}])$. From the proof of Lemma 4.2, we may choose $\boldsymbol{\delta}^* = (\hat{\boldsymbol{\delta}}_1,\ 0)^T$, where $A_{ff}^{(p)} \hat{\boldsymbol{\delta}}_1 = \hat{\boldsymbol{\varepsilon}}_1$. We now show that

$$\boldsymbol{\varepsilon}_1 - \mathbf{q}^* = A_i^p \boldsymbol{\delta}^*$$

yields the unique solution to (4.14) or (4.15).

THEOREM 4.3. *If* $\hat{\varepsilon}_1 \notin \mathrm{Range}\left(A_{ff}^{(p)}\right)$, *then* $K_{i,p} = \infty$. *If* $\hat{\varepsilon}_1 = A_{ff}^{(p)}\hat{\boldsymbol{\delta}}_1$, *then the unique solution of (4.14) is given by*

$$\mathbf{q}^* = \begin{pmatrix} 0 \\ -A_{cf}^{(p)}\hat{\boldsymbol{\delta}}_1 \end{pmatrix} \in \mathcal{Z}, \tag{4.20}$$

and $K_{i,p} =< \hat{\varepsilon}_1, \hat{\boldsymbol{\delta}}_1 >$, *for* $p = 1$ *or* 2.

Proof. The first statement follows from Lemma 4.2. To prove the second, let $\boldsymbol{\delta}^* = (\hat{\boldsymbol{\delta}}_1,\ 0)^T$. Using the substitution

$$\boldsymbol{\varepsilon}_1 - \mathbf{q} = A^{(p)}\boldsymbol{\delta}$$

with $\boldsymbol{\delta} \in Y^{(p)}$, then (4.14) can be written as

$$\begin{aligned} \min_{\mathbf{q}\in\mathcal{Z}} \max_{\mathbf{e}\notin \mathrm{Null}(A_i^p)} \frac{\left\langle(\boldsymbol{\varepsilon}_1-\mathbf{q})^T\mathbf{e},\ (\boldsymbol{\varepsilon}_1-\mathbf{q})^T\mathbf{e}\right\rangle}{\langle A_i^p\mathbf{e},\ \mathbf{e}\rangle} &= \min_{\boldsymbol{\delta}\in Y^{(p)}} \langle A_i^p\boldsymbol{\delta}, \boldsymbol{\delta}\rangle \\ &= \min_{\boldsymbol{\gamma}\in \mathrm{Null}([A_{ff}^{(p)},\ A_{fc}^{(p)}])} \langle A_i^p(\boldsymbol{\delta}^*+\boldsymbol{\gamma}),\ (\boldsymbol{\delta}^*+\boldsymbol{\gamma})\rangle\,. \end{aligned} \tag{4.21}$$

Any solution of (4.21) is characterized by $\boldsymbol{\gamma}^* \in \mathrm{Null}([A_{ff}^{(p)},\ A_{fc}^{(p)}])$ such that

$$\langle A_i^p(\boldsymbol{\delta}^*+\boldsymbol{\gamma}^*),\ \boldsymbol{\gamma}\rangle = 0 \qquad \forall\, \boldsymbol{\gamma} \in \mathrm{Null}([A_{ff}^{(p)},\ A_{fc}^{(p)}]); \tag{4.22}$$

that is,

$$A_i^p(\boldsymbol{\delta}^*+\boldsymbol{\gamma}^*) \in \mathrm{Range}\left(\begin{bmatrix} A_{ff}^{(p)} \\ A_{cf}^{(p)} \end{bmatrix}\right). \tag{4.23}$$

But $\boldsymbol{\gamma}^* = 0$ satisfies (4.22) by construction of $\boldsymbol{\delta}^*$, which proves that (4.20) solves (4.14).

To prove uniqueness, suppose there are two such solutions to (4.22), say, $\boldsymbol{\delta}^*$ and $\boldsymbol{\beta}^*$. Then

$$A_i^p(\boldsymbol{\delta}^*-\boldsymbol{\beta}^*) = \begin{bmatrix} A_{ff}^{(p)} \\ A_{cf}^{(p)} \end{bmatrix}\hat{\mathbf{w}}$$

for some $\hat{\mathbf{w}} \in \mathbb{R}^{n_f}$. Since both $\boldsymbol{\delta}^*$ and $\boldsymbol{\beta}^*$ are in $Y^{(p)}$, we have $\hat{\mathbf{w}} \in \mathrm{Null}(A_{ff}^{(p)})$. From Lemma 4.2, we have $\mathrm{Null}(A_{ff}^{(p)}) = \mathrm{Null}(A_{cf}^{(p)})$, which implies that $A_i^p(\boldsymbol{\delta}^*-\boldsymbol{\beta}^*) = 0$ and that $\mathbf{q}^*$ is unique.

Finally, substituting $\boldsymbol{\delta}^*$ into (4.21) yields

$$K_{i,p} = \langle A_i^p\boldsymbol{\delta}^*,\ \boldsymbol{\delta}^*\rangle = \left\langle \hat{\varepsilon}_1,\ \hat{\boldsymbol{\delta}}_1 \right\rangle,$$

which completes the proof. ∎

A practical algorithm for determining Q is as follows:
For $p = 1$, set

$$A_{ff}^{(1)} = A_{ff}, \quad A_{cf}^{(1)} = A_{cf}.$$

For $p = 2$, set

$$A_{ff}^{(2)} = A_{ff}^2 + A_{fc}A_{cf}, \quad A_{cf}^{(2)} = A_{cf}A_{ff} + A_{cc}A_{cf}.$$

Perform a QR factorization on $A_{ff}^{(p)}$ using Householder reflections and column pivoting to detect rank deficiency. If

$$A_{ff}^{(p)}\hat{\boldsymbol{\delta}}_1 = \hat{\boldsymbol{\varepsilon}}_1$$

has a solution, then set

$$\mathbf{q}^* = \begin{pmatrix} 0 \\ -A_{cf}^{(p)}\hat{\boldsymbol{\delta}}_1 \end{pmatrix}$$

and $K_{i,p} = \left\langle \hat{\boldsymbol{\varepsilon}}_1,\ \hat{\boldsymbol{\delta}}_1 \right\rangle$; otherwise, set $K_{i,p} = \infty$.

4.3. Local-Global Measure. This subsection shows that if $M_{i,1}$ or $M_{i,2}$ is bounded for every $i \in \mathcal{F}$, then the global measure M_1 is also bounded.

THEOREM 4.4. *Let $p = 1$ or 2 and assume that the local approximation property*

$$M_{i,p}(Q, \mathbf{e}) \le K_{i,p} \quad \forall \mathbf{e} \in \mathbb{R}^n \tag{4.24}$$

holds for some $K_{i,p}$ and all $i \in \mathcal{F}$. Then global approximation property (3.4) is also satisfied with

$$K = \max_{\alpha \in \mathcal{T}} \sum_{i \in \mathcal{M}_\alpha \cap \mathcal{F}} K_{i,p} \|A_i\|^{p-1}. \tag{4.25}$$

Proof. We have

$$\begin{aligned} \langle (I-Q)\mathbf{e},\ (I-Q)\mathbf{e} \rangle &= \sum_{i \in \mathcal{F}} \left\langle \boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i^T (I-Q)\mathbf{e},\ \boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i^T (I-Q)\mathbf{e} \right\rangle \\ &\le \sum_{i \in \mathcal{F}} K_{i,p} \langle A_i^p \mathbf{e},\ \mathbf{e} \rangle \\ &\le \sum_{i \in \mathcal{F}} K_{i,p} \|A_i\|^{p-1} \langle A_i \mathbf{e},\ \mathbf{e} \rangle \\ &= \sum_{\alpha \in \mathcal{T}} \langle A_\alpha \mathbf{e},\ \mathbf{e} \rangle \sum_{i \in \mathcal{M}_\alpha \cap \mathcal{F}} K_{i,p} \|A_i\|^{p-1} \\ &\le K \sum_{\alpha \in \mathcal{T}} \langle A_\alpha \mathbf{e},\ \mathbf{e} \rangle \\ &= K \langle A\mathbf{e},\ \mathbf{e} \rangle . \end{aligned}$$

∎

Straightforward application of the above techniques can be used to bound M_2 in terms of $M_{i,2}$. However, the resulting bounds on M_2 can be much larger than the maximum value of $M_{i,2}$. While this may not be sharp, it is simple to construct an example where M_2 is much larger than the largest $M_{i,2}$ and, hence, much larger than M_1. In this case, using M_2 to estimate convergence could lead to the erroneous conclusion that the resulting two-level method is slow to converge.

The local measure bounds, $K_{i,p}$, can be used as a diagnostic tool: Theorem 4.4 shows that they contribute to the bound K, used to establish convergence in Theorem 3.2. While neither measure provides a sharp bound when the algorithm exhibits a small convergence factor, they can provide a warning: if $K_{i,p}$ is large for some i, it may be profitable to reexamine the choice of the coarse grid, perhaps adding more grid points to $\mathcal{C}$.

As an alternative to increasing the size of $\mathcal{C}$, we could respond to large values of $K_{i,p}$ locally by increasing the size of the neighborhood. Define the set $\mathcal{N}_i^{(k)}$ of *kth removed neighbors* recursively by letting $\mathcal{N}_i^{(1)} := \mathcal{N}_i$ and

$$\mathcal{N}_i^{(\ell+1)} := \cup_{j \in \mathcal{N}_i^{(\ell)}} \mathcal{N}_j. \tag{4.26}$$

Then interpolation could be allowed from the set $\mathcal{C}_i := \mathcal{N}_i^{(k)} \cap \mathcal{C}$, which are the coarse-grid points connected to point i by a path of length k in the graph of A. While this would yield more accurate interpolation, the complexity of $P^T AP$ would certainly increase.

5. Numerical Results. We apply the element interpolation methods numerically to two illustrative examples: a Poisson equation discretized on stretched quadrilaterals and a plane-stress cantilever beam. We compare our numerical results to the *bounds* predicted by our theory, and demonstrate the improved robustness of the new methods over AMG.

For each problem, we first present results for AMG to show that our usual approach breaks down. For standard AMG, which we now refer to as AMG1, parameter θ defining the cutoff for strong connections is set to 0.25. The interpolation formula is that found in [19] (see (5.10) and (6.4) there). For the stretched grid problem, two variants of AMG are presented that restore convergence (although these fixes will be shown to be ineffective in the elasticity problem). The first, called AMG2, uses a more restrictive definition of strong connections by taking $\theta = 0.50$. The second variant (AMG3) returns to $\theta = 0.25$ but uses iterative weight interpolation [11].

For comparison, we also include a method presented by Chang *et al.* [12, 10, 9], since they report results for both anisotropic Poisson problems and 2D elasticity. Chang *et al.* describe two basic methods, each with a user-specified parameter θ_1. The one we include (referred to in [9] as Method II, with $\theta_1 = 0$, and called the CWF method here) appears to be the most robust of their methods overall. There are two main differences between our standard AMG algorithm and theirs. The first is that their strong connections are determined by absolute value, while we consider only connections of the "right" sign (i.e., the opposite sign from that of the diagonal entry) to be strong. (As with standard AMG, they take $\theta = 0.25$.) The second difference is the use of a modified interpolation formula. As with standard AMG, interpolation to a point i is derived by writing the corresponding residual equation in terms of the error, making some approximations for those terms defined at points not used in interpolation, and solving for e_i to obtain the weights. Unlike standard AMG, these approximations use weighted averages based on absolute values of the matrix entries, and incorporate some geometric ideas based on the assumption that the size of matrix entries diminishes with the distance between grid points. A seemingly minor modification is that, for weak (small) connections j with no connection to set the interpolation points, the approximation is made that $e_j = e_i$ if $sign(a_{ij}) = -sign(a_{ii})$ or $e_j = -e_i$ if $sign(a_{ij}) = sign(a_{ii})$. Variants of their methods include modifications to restriction and the definition of the coarse grid operator. However, the algorithm we test reduces to the standard Galerkin formulation. That is, restriction is defined as the transpose of interpolation, and the coarse grid operator is taken as the product of restriction, the fine grid matrix, and interpolation. The coarsening method they describe is simply the Ruge-Stüben two-pass coarsening scheme [19] using the modified definition of strong connections. For testing their method, we use our AMG coarsening code with strong connections determined by absolute value.

The only difference here between AMGe and the AMG variants described above is that we use the element interpolation method in AMGe to construct the interpolation operators. Thus, the coarse grids are selected in the same way that they are in AMG. The possibility of using the AMGe measures to determine coarsening is a topic of current research. Three different definitions are considered for interpolation: AMG, local measure 1 (AMGe1), and local measure 2 (AMGe2).

In the multilevel algorithm, we construct "coarse element stiffness matrices" $A_{c,\alpha}$ as follows:

$$A_{c,\alpha} = P^T A_\alpha P. \tag{5.1}$$

To reduce computational complexity and storage costs, we combine coarse elements that operate on the same points by summing them. That is, we define

$$\mathcal{M}_{c,\alpha} := \left\{ j \ : \ \varepsilon_j^T A_{c,\alpha} \varepsilon_j \neq 0 \right\} \tag{5.2}$$

and, when $\mathcal{M}_{c,\alpha} = \mathcal{M}_{c,\beta}$, we combine $A_{c,\alpha}$ and $A_{c,\beta}$ to form a single coarse-element stiffness matrix.

To conform to the theory, the linear systems for the AMGe tests are scaled so that the diagonal is the identity. That is, we actually solve $\hat{A}\hat{\mathbf{u}} = \hat{\mathbf{f}}$, where $\hat{A} = D^{-1/2} A D^{-1/2}$, $\hat{\mathbf{u}} = D^{1/2}\mathbf{u}$, and $\hat{\mathbf{f}} = D^{-1/2}\mathbf{f}$. Our initial experiments use $V(0,1)$ cycles based on damped Jacobi with step-size $s = 1/2$. In the examples below, $\|A\|$ is between 2.5 and 3.0 so that $\frac{1}{\|A\|} \le s \le \frac{2}{\|A\|}$. For AMG, we use the original unscaled matrix A.

Equation (3.9) in Theorem 3.2 yields a bound on the convergence factor given by

$$\rho \le 1 - \frac{4 - \|A\|}{4K}, \tag{5.3}$$

Size	Two-Level				Multilevel			
	AMG1	AMG2	AMG3	CWF	AMG1	AMG2	AMG3	CWF
64×64	0.80	0.12	0.09	0.91	0.81	0.14	0.09	0.91
128×128	0.80	0.12	0.10	0.92	0.81	0.14	0.10	0.91

TABLE 5.1
2-level and V-cycle asymptotic convergence factors for the stretched quadrilateral problem.

where K is the bound on either M_1 or M_2. As we will see, this bound is very pessimistic. Replacing K from (4.25) by $K_p = \max_i K_{i,p}$ yields a somewhat more realistic but still pessimistic estimate for the convergence factor. These estimates are included in the numerical results below.

5.1. Stretched Quadrilateral. Consider the stretched quadrilateral problem introduced in Section 2, which consists of a Poisson equation on a rectangular grid discretized with $n_x \times n_y$ bilinear quadrilateral elements. The fine-grid elements have a $10 : 1$ aspect ratio, yielding the stencil in (2.4). The boundary conditions are Dirichlet, which are eliminated from the matrix during discretization.

In Table 5.1, we present results of tests with AMG and its variants for the stretched-grid Poisson problem. Both 2-level and V-cycle results are given. In all cases, (1,1) V-cycles are used, with C/F-ordered Gauss-Seidel relaxation. For AMG1, convergence is much worse than the 2-level factors of 0.06 and V-cycle factors of 0.10 that we would get on unstretched grids. AMG2 and AMG3 both improve convergence greatly, nearly restoring the results that would be obtained in the uniform case.

The AMG variants each produced a semi-coarsened grid for the first coarsening. Away from boundaries, AMG1 used a 6-point interpolation stencil of the form

$$P_{\mathrm{AMG}} = \begin{bmatrix} 0.084 & 0.332 & 0.084 \\ & * & \\ 0.084 & 0.332 & 0.084 \end{bmatrix} \tag{5.4}$$

Both AMG2 and AMG3 produced 2-point interpolation away from boundaries, so that weights of 0.5 were obtained for the north and south points. Here, the difference between these two methods lies in the interpolation stencils near boundaries.

In geometric multigrid, this anisotropic situation is often treated by semi-coarsening, that is, by choosing coarse-grid points along each vertical line. Interpolation is then performed only in the y direction. The typical interpolation weights used in geometric semi-coarsening do not involve corner points, so smaller weights intuitively make more sense here. In fact, it can be shown that, for this problem, smaller corner weights (up to a point) generally produce better 2-level results. Since performance is very sensitive to these weights, small changes in the algorithm can affect convergence greatly. For example, in the code AMG1R5 that is widely available to the public, in computing the interpolation operator, weak connections are treated in a manner similar to strong connections. For this problem, this modification gives corner weights of 0.0052 and results in 2-level and V-cycle factors of 0.33 and 0.54, respectively.

Poor performance of the CWF method is primarily due to the grid chosen. Using their modified definition of strong connections, all off-diagonals in stencil (2.4) are considered strong. This results in standard coarsening (coarsening by a factor of 2 in each grid direction), not semi-coarsening. It is well known that such a coarse grid cannot be used effectively for this problem without additional modifications to the algorithm, such as line relaxation. It is interesting to note that, as the grid becomes "unstretched", the connections to the left and right become smaller, and proper semi-coarsening results. If semi-coarsening were used in conjunction with the CWF interpolation, 2-level results of 0.33 would be obtained. This is still far from the results reported for anisotropic problems in [9]. In that paper, however, simple 5-point finite difference stencils are used, so that M-matrices result, and connections to the east and west actually do become small as the grid is stretched. AMG and the variants used here would also have no problem with such discretizations. It should be noted that the coarse grids we obtain may not be those that would be obtained with the algorithm as they have implemented it, and in fact appear to be coarser in most of their test problems, given the larger grid complexities they report. Nevertheless, the grids chosen satisfy the criteria they present.

The point here is not that particular fixes exist for this specific problem. There is no shortage of variations of the AMG algorithm: each works well for some specific cases, but can break down for others.

Size	Two-Level AMG1	Two-Level AMGe1	Two-Level AMGe2	Multilevel AMG	Multilevel AMGe1	Multilevel AMGe2
64×64	0.82	0.27	0.27	0.84	0.32	0.27
128×128	0.82	0.28	0.28	0.84	0.31	0.28
Bound	0.97	0.90	–	–	–	–
Estimate		0.81	0.87	–	–	–

TABLE 5.2

Asymptotic convergence factors, bound predicted by theory, and 'improvement' of observed over predicted for the stretched quadrilateral problem.

Our goal is instead to obtain robust AMG methods that are much more difficult to break. This goal is the motive for the development of AMGe.

In the AMGe tests, the AMG coarsening algorithm used for all three methods again produces semi-coarsened grids, and the interpolation formulas for AMGe1 and AMGe2 are as follows:

$$P_{\mathrm{AMGe1}} = \begin{bmatrix} 0.007 & 0.486 & 0.007 \\ & * & \\ 0.007 & 0.486 & 0.007 \end{bmatrix}, \tag{5.5}$$

$$P_{\mathrm{AMGe2}} = \begin{bmatrix} 0.003 & 0.494 & 0.003 \\ & * & \\ 0.003 & 0.494 & 0.003 \end{bmatrix}. \tag{5.6}$$

The stencils at boundaries are similar. Note that interpolation for these algorithms also involve corner points, since these are considered strong connections, but the associated weights for AMGe1 and AMGe2 are much smaller than for AMG. The large element aspect ratio effectively decouples each vertical line of grid points from the others.

The experimental results for AMG1, AMGe1, and AMGe2 are presented in Table 5.2. Two grid sizes, 64×64 and 128×128, are used. For better comparison with the theory, two changes were made to the solution method. First, we used (1,0) V-cycles, as opposed to the (1,1) V-cycles reported in the previous table. In addition, relaxation has been changed from C/F Gauss-Seidel to a Richardson iteration with a relaxation parameter $\omega = 0.5$. For each grid, we show asymptotic convergence factors for AMG1, AMGe1, and AMGe2. Factors are shown for both two-level and multilevel cases. For the two-level case, we show the bound on the convergence factor corresponding to using (4.25) in Theorem 3.2 for M_1. This is computed using $\|A\| = 2.97$ and $K_1 = 2.68$. As expected, the bound is very pessimistic. We also show the convergence factor (labeled "estimate") that would result from substituting $K_1 = \max_i K_{i,1} = 1.34$ and $K_2 = \max_i K_{i,2} = 2.0$ for (4.25) in Theorem 3.2. This provides a somewhat improved but still very pessimistic value for the convergence factor. This behavior is typical of most multigrid theory, where results often substantially exceed theoretical estimates.

The key observation to be made from the data in Table 5.2 is that both AMGe1 and AMGe2 produce substantial improvement over AMG for stretched quadrilaterals. For this problem, AMG2 and AMG3 described above would produce results similar to AMGe1 and AMGe2. Such techniques, however, tend to be somewhat *ad hoc*, and are not based on theoretical considerations. As such, we cannot determine in advance whether such treatments will be useful for a given problem. By contrast, we expect AMGe1 and AMGe2 to perform well in more general problems involving high aspect ratios, so they should find wide applicability for problems based on unstructured grids having thin domains or regions.

5.2. Plane-Stress Cantilever Beam. Consider the 2D linear elasticity equations

$$u_{xx} + \frac{1-\nu}{2}u_{yy} + \frac{1+\nu}{2}v_{xy} = f_1,$$
$$v_{yy} + \frac{1-\nu}{2}v_{xx} + \frac{1+\nu}{2}u_{xy} = f_2,$$

where u and v are displacements in the x and y directions, respectively. We take $\nu = 4/7$ for the tests. The problem, depicted in Figure 5.1, has free boundaries, except on the left where $u = v = 0$. We discretize with

d	Two-Level				Multilevel			
	AMG1	AMG2	AMG3	CWF	AMG1	AMG2	AMG3	CWF
1	0.33	0.31	0.31	0.73	0.58	0.62	0.58	0.96
3/4	0.33	0.31	0.29	0.79	0.72	0.65	0.64	0.95
1/2	0.34	0.32	0.29	0.89	0.80	0.76	0.81	0.96
1/4	0.58	0.57	0.51	0.98	0.97	0.95	0.98	0.98
1/8	0.90	0.90	0.87	0.99	0.99	0.98	0.99	0.99
1/16	0.98	0.98	0.98	0.99	0.98	0.98	0.99	0.99
1/32	0.98	0.99	0.98	0.99	0.99	0.99	0.99	0.98
1/64	0.99	0.99	0.99	0.99	0.98	0.98	0.98	0.98

TABLE 5.3
2-Level and V-cycle asymptotic convergence factors for four methods for the plane-stress problem.

bilinear finite elements on a uniform rectangular mesh with spacing h in both directions (square elements).

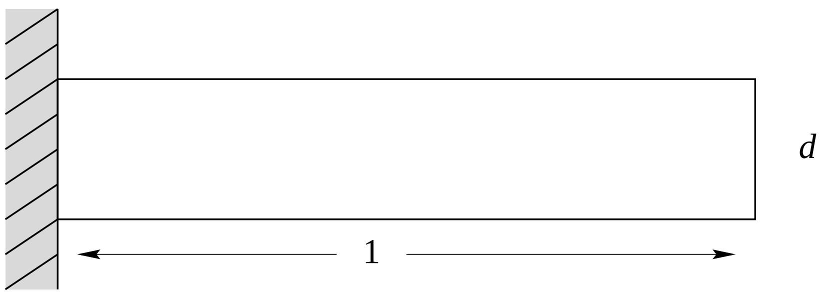

FIG. 5.1. *Plane-stress cantilever beam problem.*

Again, before testing the element-based methods, we present results for the AMG1, AMG2, AMG3, and CWF methods of the previous section. AMG1, AMG2, and AMG3 are applied in a separate fashion (the so-called "unknown approach" [19]) in which connections between u and v are completely ignored in the determination of strong connections and computation of interpolation weights. Such an approach has been shown to produce good results for the elasticity problem on the unit square when full Dirichlet boundary conditions are used, although it degrades with the number of free sides allowed (cf. [11]). As in [9], the CWF method does not differentiate between the two unknowns. (It should be noted that, in comparisons with the AMG1R5 code in that paper, AMG was also applied in a "scalar" fashion. Not unexpectedly, it did not perform well since AMG1R5 was designed for scalar problems and there is really no local relationship between pointwise values of errors in u and errors in v, even when these errors are smooth.)

Results are presented in Table 5.3. Several different thicknesses are used for the beam, ranging from a square cross section, $d = 1$, to a very thin beam, $d = 1/64$. As before, (1,1) V-cycles were used with C/F Gauss-Seidel relaxation. The factors shown were obtained after 100 cycles. (With such poor convergence factors, it generally takes many iterations to reach the asymptotic state.) Note that there is little difference between convergence factors for AMG1, AMG2, and AMG3, although AMG3 with iterative weight interpolation shows a slight advantage. For this problem, the $u - u$ and $v - v$ stencils each resemble the anisotropic stencils for the Poisson problem (with the first stretched in x and the second in y), but with grids that are less stretched. Thus, the "fixes" do not improve much on the basic algorithm, as they did for the stretched-grid Poisson problem. Note that 2-level factors are nearly constant until d is reduced to below 1/2, at which convergence degrades sharply. V-cycle factors, while somewhat acceptable for large d, degrade even more quickly than 2-level factors as the domain becomes thinner. With all three AMG variants, on the finest grid, u is semi-coarsened in x while v is semi-coarsened in y for all d used. Thus, for all $d > 1/64$, interpolation computed on the finest level does not change. This indicates strongly that the problem is not that interpolation accuracy for smooth components is affected, but that such smooth components must be interpolated *more* accurately as d is reduced. In fact, it can be shown that the energy norm relative to the Euclidean norm of the smoothest components decreases to zero with d, which by standard variational

d	Two-Level GMG	AMGe1	AMGe2	Multilevel GMG	AMGe1	AMGe2
1	0.45	0.49	0.48	0.49	0.65	0.85
1/4	0.46	0.48	0.47	0.76	0.68	0.90
1/8	0.46	0.47	0.45	0.78	0.64	0.87
1/16	0.46	0.49	0.45	0.77	0.58	0.77
1/32	0.53	0.45	0.50	0.75	0.51	0.56
1/64	0.67	0.39	0.28	0.67	0.39	0.28
Bound	–	0.97	–	–	–	–
Estimate	–	0.87	0.97	–	–	–

TABLE 5.4

Asymptotic convergence factors, bound predicted by theory, and 'improvement' of observed over predicted for the plane-stress problem with $h = 1/64$.

multigrid theory (see Section 3) suggests that interpolation must become increasingly more accurate.

For this problem, the interpolation scheme of Chang *et al.* gives very slow convergence. Even 2-level factors are large for $d = 1$, indicating that the local interpolation accuracy is not sufficient. Since quite good results for the 2D elasticity problem were presented in [9], some explanation for the behavior of the method here is warranted. As noted earlier, the CWF method does not differentiate between u and v, and, for example, u at a point can interpolate from a combination of u and v values at surrounding points. With such interpolation, however, some minimal conditions must be met before reasonable convergence could be expected. With discretizations of scalar differential operators, away from Dirichlet boundaries, the matrix generally has zero row sums, so that the constant is in the local null space of the operator. Since the smoothest grid functions are locally constant, these must be interpolated exactly, which in turn requires interpolation weights at each point to sum to 1. Similarly, in the elasticity problem, away from fixed boundaries the $u-u$, $u-v$, $v-u$, and $v-v$ connections at each point also sum to zero, so that separate independent constants in u and v are in the local null space of the operator. For these components to be interpolated exactly, the $u-u$ and $v-v$ interpolation weights both must sum to 1, while the $u-v$ and $v-u$ weights both must sum to zero. Loss of any of these conditions can severely degrade convergence. Now, one thing to note is that the tests in [9] use finite differences and full Dirichlet boundary conditions. In the finite difference case, the $u-v$ and $v-u$ connections are smaller relative to the maximum connection, and are considered weak, thus keeping interpolation completely separate for u and v, at least on the finest grid. In the finite element case here, connections from u to v and v to u are considered strong. By itself, this is not bad. In fact, when full Dirichlet conditions are used with the finite element discretization, the CWF method gives a 2-level convergence factor of 0.33 for the $d = 1$ problem. Along free boundaries, however, it was found that different geometric assumptions applied to positive and negative cross-connections resulted in a loss of 1-row sum for the $u-u$ interpolation weights and zero-row sum for the $u-v$ and $v-u$ weights, resulting in the uniformly poor convergence found.

In the AMGe tests, we use the geometric coarsening strategy of doubling the element size in both directions until there is only one element in the y direction, then doubling the element size in the x direction only. For the multilevel results, we coarsen until $h_x = 2h_y$. Such a grid is not admissible with the other AMG algorithms presented, so a direct comparison of the effect of using the different interpolation formulas on the same grid is not practical here. (Actually, a modification of iterative weight interpolation could be used, and would yield results comparable to those previously obtained.) However, it is instructive to include results of AMG where linear interpolation is actually used on these uniform grids. This method then becomes a geometric multigrid algorithm, which we call GMG.

Experimental results from three methods are shown in Table 5.4, using (0,1) V-cycles based on Jacobi sweeps with relaxation parameter $\omega = 0.5$. This should be kept in mind when comparing results to the AMG tests of Table 5.3, since (1,1) V-cycles were used there. The theoretical bounds and estimates suggest extremely slow convergence for AMGe1 and AMGe2 (when applied to AMG, they do not indicate that AMG will converge at all). In fact, however, both AMGe1 and AMGe2 achieve substantial improvement, especially for the two-level algorithm, where they greatly exceed predictions. The bound is based on $\|A\| = 2.50$ and $K_1 = 12.25$, while the predictions are based on $\max_i K_{i,1} = 2.84$ and $\max_i K_{i,2} = 8.31$.

Note that the GMG method, which is basically the best that the AMG methods using separate interpolation can aspire to, also shows the same type of degradation as the AMG methods, although less severe. That is, 2-level results are stable until d becomes small, then worsen. V-cycle results degrade faster, although they stabilize due to the artificial limitation on the coarsest grid used. This limitation was not present in the AMG tests. In fact, coarser grids could have been used here, which would result in much worse convergence factors if we were to continue with the same interpolation and relaxation schemes. For further coarsening in the x direction, smoothing of pointwise relaxation would suffer as the grid aspect ratios worsen, so "group" relaxation (here equivalent to y-line relaxation) would be needed to maintain efficiency. We did not consider this group relaxation option or its implications on our theory because the focus here is on the interpolation process.

Two observations are significant: the two-level performance of AMGe1 and AMGe2 is generally independent of the beam thickness until $d = h_x$, where even greater improvement occurs; and the multilevel performance of AMGe1 and AMGe2 improves steadily as the beam becomes thinner. This is in direct contrast to the AMG and GMG results obtained, demonstrating the effectiveness of the new interpolation methods.

While this paper concentrates on the effect of the new interpolation method, it should be kept in mind that there are other techniques that may be applied to enhance performance of the algorithm. For instance, the multilevel experiments shown here focused on Jacobi relaxation and a (0,1) V-cycle. The relaxation method and its parameters can be chosen differently. For example, the multilevel AMGe1 case with $d = 1/4$ shows a convergence factor of 0.65 in Table 5.4. A Jacobi (1,1) V-cycle improves this factor to 0.58, while a (1,1) F-cycle (see [14]) attains a convergence factor of 0.31. Nearly identical results, 0.65 for $V(0,1)$, 0.56 for $V(1,1)$, and 0.33 for $F(1,1)$, are obtained if the Jacobi relaxation is replaced by nodal Gauss-Seidel with symmetric CF relaxation, which sweeps over the C points followed by the F points on the downward leg of the V-cycle, and over the F points followed by the C points on the upward leg. Another possibility is the use of a single multigrid $V(1,1)$ cycle as a preconditioner for a conjugate gradient iteration. Applied to the plane-stress problem using the nodal relaxation described above, this yields convergence factors ranging from 0.16 to 0.26 per CG iteration.

For both sets of experiments, AMGe interpolation achieves significant improvement over conventional AMG performance. We believe that further improvement is possible using more sophisticated coarse-grid selection. We observe that local measures $M_{i,1}$ and $M_{i,2}$ carry a great deal of information about the nature of the underlying problem and its discretization, and we should be able to exploit this information to determine more effective coarse grids.

6. Conclusions.

For any multigrid method to work, errors that remain after relaxation must be well approximated by the range of interpolation. Since algebraic multigrid does not rely on geometric information, its fundamental challenge is to construct coarse grids and interpolation operators that approximate these errors. The core of this challenge is to determine errors that cannot be effectively reduced by local processing.

Two local measures were introduced here to quantify how well the coarsening processes determine algebraically smooth error, and they were used to construct new interpolation operators. Experimental data for two representative test problems confirm that these operators produce an AMGe algorithm whose convergence rates for these cases are substantially better than standard AMG.

Current research focuses on using these measures in AMGe also to assess the ability of coarse-grid points to represent the necessary error components, that is, to determine *which* points are best suited to be on the coarse grid. Combined with the improved interpolation operator, this may lead to very efficient AMGe algorithms for a much wider range of problems than is currently available.

REFERENCES

[1] R. E. Alcouffe, A. Brandt, J. E. Dendy, and J. W. Painter, *The multi–grid methods for the diffusion equation with strongly discontinuous coefficients*, SIAM J. Sci. Stat. Comput., 2 (1981), pp. 430–454.

[2] C. Baldwin, P. N. Brown, R. D. Falgout, J. Jones, and F. Graziani, *Iterative linear solvers in a 2d radiation-hydrodynamics code: methods and performance*. Submitted to Journal of Computational Physics, 1998.

[3] J. H. Bramble, J. E. Pasciak, J. Wang, and J. Xu, *Convergence estimates for multigrid algorithms without regularity assumptions*, Math. Comp., 57 (1991), pp. 23–45.

[4] A. Brandt, *Algebraic multigrid theory: The symmetric case*, in Preliminary Proceedings for the International Multigrid Conference, Copper Mountain, Colorado, April 1983.

[5] A. Brandt, *Algebraic multigrid theory: The symmetric case*, Appl. Math. Comput., 19 (1986), pp. 23–56.
[6] A. Brandt, S. F. McCormick, and J. W. Ruge, *Algebraic multigrid (AMG) for automatic multigrid solutions with application to geodetic computations*. Report, Inst. for Computational Studies, Fort Collins, Colo., October 1982.
[7] ———, *Algebraic multigrid (AMG) for sparse matrix equations*, in Sparsity and Its Applications, D. J. Evans, ed., Cambridge University Press, Cambridge, 1984.
[8] W. L. Briggs, V. E. Henson, and S. F. McCormick, *A Multigrid Tutorial, 2nd Edition*, SIAM Books, Philadelphia, in press.
[9] Q. Chang, Y. S. Wong, and H. Fu, *On the algebraic multigrid method*, J. of Comp. Phys., 125 (1996), pp. 279–292.
[10] Q. Chang, Y. S. Wong, and Z. Li, *New interpolation formulas using geometric assumptions in the algebraic multigrid method*, Appl. Math. and Comp., 50 (1992), pp. 223–254.
[11] A. J. Cleary, R. D. Falgout, V. E. Henson, J. E. Jones, T. A. Manteuffel, S. F. McCormick, G. N. Miranda, and J. W. Ruge, *Robustness and scalability of algebraic multigrid*. To appear in SIAM Journal on Scientific Computing special issue on the Fifth Copper Mountain Conference on Iterative Methods, 1998.
[12] W. Z. Huang, *Convergence of algebriac multigrid methods for symmetric positive definite matrices with weak diagonal dominance*, Appl. Math. and Comp., 46 (1991), pp. 145–164.
[13] J. Mandel, M. Brezina, and P. Vaněk, *Energy optimization of algebraic multigrid bases*. Submitted.
[14] S. F. McCormick, *Multigrid methods for variational problems: further results*, SIAM J. Numer. Anal., 21 (1984), pp. 255–263.
[15] ———, *Multigrid methods for variational problems: general theory for the V–cycle*, SIAM J. Numer. Anal., 22 (1985), pp. 634–643.
[16] S. F. McCormick and J. W. Ruge, *Multigrid methods for variational problems*, SIAM J. Numer. Anal., 19 (1982), pp. 924–929.
[17] J. Ruge, *Element interpolation for algebraic multigrid (AMG)*. Presentation at the 4th Copper Mountain Conference on Multigrid Methods, Copper Mountain, CO, 1989.
[18] J. W. Ruge and K. Stüben, *Efficient solution of finite difference and finite element equations by algebraic multigrid (AMG)*, in Multigrid Methods for Integral and Differential Equations, D. J. Paddon and H. Holstein, eds., The Institute of Mathematics and its Applications Conference Series, Clarendon Press, Oxford, 1985, pp. 169–212.
[19] ———, *Algebraic multigrid (AMG)*, in Multigrid Methods, S. F. McCormick, ed., vol. 3 of Frontiers in Applied Mathematics, SIAM, Philadelphia, PA, 1987, pp. 73–130.
[20] K. Stüben, *Algebraic multigrid (AMG): experiences and comparisons*, Appl. Math. Comput., 13 (1983), pp. 419–452.
[21] K. Stüben, U. Trottenberg, and K. Witsch, *Software development based on multigrid techniques*, in Proc. IFIP–Conference on PDE Software, Modules, Interfaces and Systems, B. Enquist and T. Smedsaas, eds., Sweden, 1983, Söderköping.
[22] P. Vaněk, M. Brezina, and J. Mandel, *Convergence analysis of algebraic multigrid based on smoothed aggregation*. Submitted.
[23] P. Vaněk, M. Brezina, and R. Tezaur, *Two-grid method for linear elasticity on unstructured meshes*, SIAM J. Sci. Comput., (1998). To appear.
[24] W. L. Wan, *An energy-minimizing interpolation for multigrid methods*, tech. rep., Department of Mathematics, UCLA, April 1997. UCLA CAM Report 97-18.
[25] W. L. Wan, T. F. Chan, and B. Smith, *An energy-minimizing interpolation for robust multigrid methods*, tech. rep., Department of Mathematics, UCLA, February 1998. UCLA CAM Report 98-6.