NUMERICAL LINEAR ALGEBRA WITH APPLICATIONS
Numer. Linear Algebra Appl. 2005; **0**:1–16 *Prepared using* nlaauth.cls *[Version: 2002/09/18 v1.02]*

An energy-based AMG coarsening strategy

J. Brannick[1], M. Brezina[1], S. MacLachlan[1,*], T. Manteuffel[1], S. McCormick[1], and J. Ruge[1]

[1] *Department of Applied Mathematics, University of Colorado at Boulder, UCB 526, Boulder, CO, 80309-0526*

SUMMARY

Algebraic multigrid (AMG) is an iterative method that is often optimal for solving the matrix equations that arise in a wide variety of applications, including discretized partial differential equations. It automatically constructs a sequence of increasingly smaller matrix problems that hopefully enables efficient resolution of all scales present in the solution. The methodology is based on measuring how a so-called *algebraically smooth* error value at one point depends on its value at another. Such a concept of *strength of connection* is well understood for operators whose principal part is an M-matrix; however, the strength concept for more general matrices is not yet clearly understood, and this lack of knowledge limits the scope of AMG applicability. The purpose of this paper is to motivate a general definition of strength of connection, discuss its implementation, and present the results of initial numerical experiments.

KEY WORDS: multigrid,AMG

1. Introduction

Initially introduced over twenty years ago, the algebraic multigrid algorithm (AMG) [5, 23] has become a workhorse of large-scale computational simulation, due in large part to its algorithmic and parallel scalability for many important problems [1, 3, 9, 24, 25, 27]. In the past two decades, many variations on the AMG algorithm have been introduced, including modifications to the coarse-grid selection algorithms [15, 17, 20, 21], the definition of interpolation [10], or both [7, 8, 13, 26]. A summary and comparison of many of these variations within the classical AMG framework appears in [22]. Such variations arise because, while the guiding framework of algebraic multigrid ideas addresses the need for solvers that can automatically adjust to variations in the fine-scale dynamics of a discretized operator, the applicability of the original algorithm and its variants is limited by the heuristics upon which they are based.

The present work is part of a project aimed at transcending these limitations by asking not what heuristic covers a particular choice, but, instead, by looking for a measure that definitively

*Correspondence to: S. MacLachlan, Department of Applied Mathematics, University of Colorado at Boulder, UCB 526, Boulder, CO, 80309-0526, email: maclachl@colorado.edu

answers the question. Such a measure is not generally practical (or feasibly computable) for many problems where a simpler heuristic is known to be effective, but this is not our goal. Instead, we aim to address the class of problems for which efficient multigrid solution is possible, but current heuristics break down for various reasons.

Our approach is motivated by the observation that classical AMG coarsening heuristics are typically applied well outside of the regime for which they were intended. While appropriate for the diagonally dominant M-matrices arising from finite difference discretizations of elliptic PDEs, the heuristics of [5] are often applied to matrices that substantially lack these properties. Thus, we develop a new definition of strength of connection that is not dependent on such assumptions but, instead, addresses the fundamental strength present in a discrete operator. This measure of strength cannot, however, be easily computed in any setting of practical interest. We further investigate the efficient approximation of the proposed strength measure and find that, for the examples considered here, the measure can be accurately approximated for short-range connections with an acceptable amount of computation.

Many other approaches to generalizing the AMG coarsening heuristics have been proposed recently. The approach of compatible relaxation [4, 6, 7, 18, 21] uses a modified relaxation scheme to expose the character of the slow-to-converge error. Coarse-grid points are then selected where this algebraically smooth error is largest. As such errors are typically large over much of the domain, heuristics are employed to prevent slow coarsening; however, these heuristics often block or delay the selection of the dense coarse grids that are needed for some problems. Chow [14] uses samples of algebraically smooth error to determine the directions in which this error varies slowly and, thus, interpolation can be very effective. For many problems, however, such a direction may not exist, and the proposed technique does not apply. An approach that is similar to that proposed here is given by Bröker [12], where the relative sizes of the entries of the SPAI preconditioning matrix [19] are used to indicate strength of connection. In some respects, this approach is a special case of ours; however, the measure used in [12] is based on the L^2 sizes of the approximate inverse, not on the energy norm that is more compatible with the objective of computation.

Background on the classical AMG coarsening algorithm is given in Section 2. We propose our strength measure in Section 3 and give details needed to make it practical in Section 4. Samples of the selected coarse grids are given in Section 5 for a number of problems where these can be analyzed intuitively. Two-level and V-cycle convergence results using the Adaptive AMG interpolation definition of [10] are given in Section 6.

2. Classical AMG Coarse-Grid Selection

The coarse-grid selection performed in the Classical AMG algorithm [5, 23] may be viewed in terms of heuristics based on the properties of diagonally dominant M-matrices. These heuristics identify properties of the errors that pointwise relaxation on the discrete problem is slow to resolve and use these properties to define the important (or strong) connections within the linear system. The coarse-grid points are then selected using maximal independent subset heuristics to ensure a significant reduction in grid size, but also maintain accurate approximation properties.

Consider the Gauss-Seidel iteration, with error propagation operator $I - L^{-1}A$, for L the lower-triangular part of A. Fixing this to be the fine-scale relaxation used in a multigrid method

for the symmetric and diagonally dominant M-matrix, A, the goal of coarse-grid correction is to effectively reduce the error components that are not significantly reduced by Gauss-Seidel relaxation. It is shown in [5] that such an error, $\mathbf{e}$, must satisfy

$$\sum_{i,j}(-a_{ij})(e_i - e_j)^2 \ll \sum_i a_{ii}e_i^2.$$

Thus, if a_{ij} is large (relative to $\max_{k\neq i}|a_{ik}|$ or $\max_{k\neq i}|a_{jk}|$), then it must be true that $e_i \approx e_j$.

This observation leads to the definitions of strong dependence and influence central to the selection of AMG coarse-grid points. For a given degree of freedom, i, the set of points that i strongly depends upon is denoted S_i and defined as

$$S_i = \left\{ j : -a_{ij} \geq \theta \max_{k\neq i}\{-a_{ik}\} \right\}, \tag{1}$$

for some suitable choice of $0 < \theta \leq 1$. The set of points that are strongly influenced by i is denoted S_i^T, as $j \in S_i^T$ implies that $i \in S_j$.

Once strong connections are determined, a coarse grid is chosen so that all strongly connected neighbors of any fine-grid point (that is not itself a coarse-grid point) are available for direct or a path-length two indirect interpolation. That is, if two fine-grid points are strongly connected to one another, they must have a common coarse-grid neighbor, so that this strong connection may be accounted for indirectly in interpolation. This condition is usually expressed as the first AMG coarsening heuristic,

H1: For each fine-grid point i, each point $j \in S_i$ must either be a coarse-grid neighbor or strongly depend on at least one coarse-grid neighbor of i.

If used alone, this rule tends to create large coarse grids, as it is most easily satisfied by adding points to the coarse grid. In practice, a second heuristic is used to limit the size of the coarse grid,

H2: The set of coarse-grid points, C, should be a maximal subset of the original set of fine-grid points such that no coarse-grid point strongly depends on any other coarse-grid point.

Because these two goals may be contradictory, typical AMG implementations rely on enforcing **H1**, using **H2** as a guide. Recent work [17] examined the possibility of replacing **H1** with a weaker condition, requiring only a single strongly connected coarse-grid neighbor for each fine-grid point, a technique that may be useful in reducing complexities in 3D parallel implementations of AMG.

Because **H2** is a weaker condition than **H1**, the selection of coarse-grid points is usually accomplished by a two-pass algorithm that picks a set satisfying **H2**, then checks for any points where **H1** is violated, adding new coarse-grid points to compensate if this occurs. The first stage is often implemented as a coloring algorithm [11, Chap. 8], where coarse points are selected based on their number of strongly connected neighbors. Initially, all points are weighted with the number of points that strongly depend upon them (that is, the size of S_i^T). The point with the largest weight is then selected to be a coarse-grid point. Since each $j \in S_i^T$ is now strongly connected to a coarse-grid point, all such j are made fine-grid points so that

H2 is not violated. All strongly connected neighbors of these points (that is, $k \in S_j^T$ for any $j \in S_i^T$) are then made more attractive as coarse-grid points, since they reflect unresolved strong connections of fine-grid points. Thus, the weights of all such k are incremented for each $j \in S_k$ that was made a fine-grid point. The algorithm then repeats, selecting the new largest-weighted point as a coarse-grid point.

In this way, an initial coarse grid is chosen that gives a maximal independent set over all strong connections. Additional points are then added to the coarse grid, if necessary, adding the minimal number of points needed to ensure that property **H1** is enforced. Once the complete coarse grid has been selected, an interpolation operator is defined so that it is accurate for slowly varying errors along strong connections. We do not deal with the details of AMG interpolation here and, instead, refer the reader to [11, Chap. 8] for details. The results in Section 6 rely on a modification to the classical AMG interpolation procedure previously developed by the authors in [10].

3. Measuring Strength of Connection

While the definition of strong dependence is appropriate for the case of diagonally dominant M-matrices for which it was intended, it is frequently seen to break down when applied in various other cases. Diagonally dominant M-matrices typically have near null spaces that can be characterized as being locally slowly varying (or locally constant), and the AMG strength heuristic relies on this being reflected in the coefficients of the matrix, A. If either the near null space cannot be accurately characterized as locally constant or this is not reflected in the matrix coefficients, then AMG performance typically suffers.

To derive a more general measure of strength of connection, we must consider what, exactly, we are trying to quantify as strength. Two gridpoints can be considered related only through the matrix equations,

$$A\mathbf{x} = \mathbf{b} \quad \text{or} \quad A\mathbf{e} = \mathbf{r}. \tag{2}$$

The classical AMG definition of strength says that if a change in the solution (or error) at point j significantly changes the source vector (or residual) at point i in an L^2 sense, then i strongly depends on j. Here, we propose a change in both *what* is measured and *how* it is measured.

The 'forward' measurement used in Equation (1) captures one sense of connection. The corresponding 'backward' measurement, that a change in the source vector (or residual) at point j significantly changes the solution (or error) at point i, is, however, a more appropriate sense of connection for AMG. This follows because the coarse-grid correction process in AMG is needed to complement a given relaxation procedure, such as Gauss-Seidel, that make pointwise adjustments based on the residual and not the error. Thus, when we change the residual at point j in relaxation, this affects the local character of the remaining error, and it is the general size of this effect that we are interested in measuring to characterize strength.

This point of view suggests that, for the purposes of defining a strength measure, we should rewrite Equation (2) to highlight the changes in the solution or error in terms of the source or residual:

$$A^{-1}\mathbf{b} = \mathbf{x} \quad \text{or} \quad A^{-1}\mathbf{r} = \mathbf{e}.$$

Thus, instead of looking at the entries in A to determine the strong connections, we consider the entries of A^{-1} to be indicative of strength.

Appropriately quantifying this strength is also an important question. Classical approaches to AMG consider the relative sizes of the off-diagonal entries as direct measures of strength. While this pointwise approach is reasonable in the case of a diagonally dominant M-matrix, it is not robust. Consider, for example, a symmetric diagonal scaling of the matrix, A, $\hat{A} = DAD$. Even if the size of the entries in A^{-1} were a good reflection of the true strong connections, the entries of the scaled inverse, $\hat{A}^{-1} = D^{-1}A^{-1}D^{-1}$, may bear no relation to those of A^{-1} if the diagonal of D is chosen with sufficient variation. Yet, this sort of diagonal scaling does not change the essential behavior of the linear system. The slow-to-converge modes of Jacobi or Gauss-Seidel relaxation are simply scaled versions of the slow-to-converge modes of the unscaled system, and the spectrum of relaxation is not changed at all. One of our objectives is to find a measure that is insensitive to such scalings.

Additionally, the measure should be consistent with the variational setting of AMG for symmetric and positive-definite matrices. As the components of the AMG algorithm are based on minimizing the A-norm (or energy norm) of the error, this norm should affect the choice of the coarse-grid points. That is, the "sizes" of the entries in A^{-1} should be measured, in some fashion, using the A-norm. Looking for the strongest connections to a node i, we need relative measures of the strength of connection of all nodes j to i that are reflected in the coefficients $(A^{-1})_{ij}$. Thus, we define the column vector, $\mathbf{G}^{(i)}$, with entries $\left(\mathbf{G}^{(i)}\right)_j = (A^{-1})_{ij}$. The contribution of node j to the energy of this vector is then taken to be the measure of the strength of dependence of node i on node j. That is, we define the strength measure,

$$S_{ij} = \frac{\|\mathbf{G}^{(i)} - \left(\mathbf{G}^{(i)}\right)_j \mathbf{I}^{(j)}\|_A}{\|\mathbf{G}^{(i)}\|_A}, \tag{3}$$

where $\mathbf{I}^{(j)}$ is the j^{th} canonical unit vector.

Note that S_{ij} is, in fact, insensitive to diagonal scaling of A.

Theorem 1. *Let A be a symmetric positive-definite $n \times n$ matrix and D a diagonal $n \times n$ matrix. Take $\hat{A} = DAD$, and let S_{ij} and $\hat{S}_{ij}$ be the strength measures of A and $\hat{A}$, respectively, as defined in Equation* (3)*, for $1 \leq i, j \leq n$, $i \neq j$. Then*

$$S_{ij} = \hat{S}_{ij}.$$

Proof: Note first that

$$S_{ij}^2 = \frac{\|\mathbf{G}^{(i)} - \left(\mathbf{G}^{(i)}\right)_j \mathbf{I}^{(j)}\|_A^2}{\|\mathbf{G}^{(i)}\|_A^2} = 1 + \frac{\|\left(\mathbf{G}^{(i)}\right)_j \mathbf{I}^{(j)}\|_A^2}{\|\mathbf{G}^{(i)}\|_A^2},$$

because $A\mathbf{G}^{(i)} = \mathbf{I}^{(i)}$, since A (and, thus, A^{-1}) is symmetric. Thus,

$$S_{ij}^2 = 1 + \frac{A_{jj}\left(\left(\mathbf{G}^{(i)}\right)_j\right)^2}{(A^{-1})_{ii}}.$$

Similarly,

$$\hat{S}_{ij}^2 = 1 + \frac{\hat{A}_{jj}\left(\left(\hat{\mathbf{G}}^{(i)}\right)_j\right)^2}{(\hat{A}^{-1})_{ii}} = 1 + \frac{D_{jj}^2 A_{jj}\left(D_{jj}^{-1}D_{ii}^{-1}\left(\mathbf{G}^{(i)}\right)_j\right)^2}{D_{ii}^{-2}(A^{-1})_{ii}},$$

where $\left(\hat{\mathbf{G}}^{(i)}\right)_j = (\hat{A}^{-1})_{ij}$, as $\hat{A}^{-1} = D^{-1}A^{-1}D^{-1}$. So

$$\hat{S}_{ij}^2 = 1 + \frac{A_{jj}\left(\left(\mathbf{G}^{(i)}\right)_j\right)^2}{(A^{-1})_{ii}} = S_{ij}^2,$$

and the theorem follows. □

4. Practical Implementation

Clearly, the measure in Equation (3) is not a practical choice for computation. Even in its reduced form, $S_{ij}^2 = 1 + \frac{A_{jj}\left((A^{-1})_{ij}\right)^2}{(A^{-1})_{ii}}$, computing S_{ij} for all pairs of nodes, $i \neq j$, would require $O(n^2)$ operations, assuming A^{-1} is already known. Computing A^{-1} itself is even less practical. Thus, we look to approximate S_{ij} for some set of j that are 'near' i.

Efficient approximation of S_{ij} requires efficient approximation of A^{-1}. A natural choice to do this is to use the relaxation scheme of the AMG solver on the Gauss-Jordan system $AG = I$ or, column-wise, $A\mathbf{G}^{(i)} = \mathbf{I}^{(i)}$. In fact, because this approximation directly takes the smoother into account, it may be a more appropriate measure than the "exact" S_{ij}. If the relaxation scheme is somehow local (that is, writing the error propagation operator of relaxation as $I - M^{-1}A$, the matrix, M, is block diagonal with $O(1)$ block size), then a few steps of relaxation on $A\mathbf{G}^{(i)} = \mathbf{I}^{(i)}$ with a zero initial guess leads to an approximation to $\mathbf{G}^{(i)}$ that is zero everywhere except in an $O(1)$-size neighborhood of i. For a non-local relaxation procedure, such as Gauss-Seidel, restriction of the relaxation to a local neighborhood is necessary for a practical method.

Using local relaxation in this manner allows local approximations of S_{ij} to be computed in $O(1)$ time, for each node i. Relaxation on $A\mathbf{G}^{(i)} = \mathbf{I}^{(i)}$ requires computing a residual, $\mathbf{r} = \mathbf{I}^{(i)} - A\mathbf{G}^{(i)}$, then computing $M^{-1}\mathbf{r}$. Starting with a zero initial guess, no computation is needed to form the first residual, which has non-zero values only at point i. Because M has, by assumption, $O(1)$ block size, $M^{-1}\mathbf{r}$ and, thus, the new approximation also have $O(1)$ nodes with non-zero values. As we iterate this process, computing the residual for an approximation to $\mathbf{G}^{(i)}$ requires a matrix-vector product, $A\mathbf{G}^{(i)}$. If $\mathbf{G}^{(i)}$ has $O(1)$ non-zero values, then applying the matrix, A, to it results in a vector still has $O(1)$ non-zero values, under the assumption that the stencil size of A is also $O(1)$ (as is true for the matrices arising from discretization of elliptic PDEs that we consider here). The residual then has $O(1)$ non-zero values and so does the relaxed residual, $M^{-1}\mathbf{r}$, because M has $O(1)$ block size. Thus, any $O(1)$ number of

sweeps of a local relaxation routine results in an approximation to $\mathbf{G}^{(i)}$ that has $O(1)$ non-zero values. Computing S_{ij} requires computing A-norms of two vectors, but both of these have $O(1)$ non-zero values. Thus, the required matrix-vector products need only be computed at the nodes where the vectors have non-zero values, an $O(1)$ computation, followed by an $O(1)$ dot product. So, a single S_{ij} can be computed in $O(1)$ time. For each node, i, however, the computed $S_{ij} = 1$, unless the approximation to $G^{(i)}$ is non-zero at node j, which is only true for $O(1)$ nodes, j. Thus, all non-trivial values, S_{ij}, can be computed in $O(1)$ time for each node i. Note that the nodes for which $S_{ij} \neq 1$ are determined based on the stencils of A and M and, thus, reflect the propagation of errors in the relaxation process itself, with large values reflecting nodes that are strongly connected to i by a few sweeps of relaxation. Because we can compute these values in $O(1)$ time for each node i, it follows that these local strength measures can be computed for all nodes in $O(n)$ time and, so, this computation does not change the asymptotic complexity of the AMG algorithm.

The algorithm for computing S_{ij} can, thus, be summarized as follows:

- For each node i, $1 \leq i \leq n$:
  - Relax μ times on $A\mathbf{G}^{(i)} = \mathbf{I}^{(i)}$, with a zero initial guess, using a local smoother, exploiting the locality by only performing operations where the answer is expected to be non-zero
  - For each j such that the approximation to $\mathbf{G}^{(i)}$ is non-zero, compute

$$S_{ij} = \frac{\|\mathbf{G}^{(i)} - \left(\mathbf{G}^{(i)}\right)_j \mathbf{I}^{(j)}\|_A}{\|\mathbf{G}^{(i)}\|_A},$$

    again exploiting the locality of the approximation to reduce computation.

Note that if the relaxation procedure chosen is such that the relaxation matrix, $\hat{M}$, for the scaled matrix, $\hat{A} = DAD$, is given by $\hat{M} = DMD$, where M is the relaxation matrix for A, then the computed strength measures, S_{ij} and $\hat{S}_{ij}$, are equal.

Theorem 2. *Let A be a symmetric positive-definite $n \times n$ matrix, D a diagonal $n \times n$ matrix, and consider a relaxation process for A with error propogation operator $I - M^{-1}A$. Take $\hat{A} = DAD$, relaxation for $\hat{A}$ with error propogation operator $I - \hat{M}^{-1}\hat{A}$, for $\hat{M} = DMD$, and let S_{ij} and $\hat{S}_{ij}$ be the strength measures of A and $\hat{A}$, respectively, as defined in Equation (3), for $1 \leq i, j \leq n$, $i \neq j$, with $\mathbf{G}^{(i)}$ and $\hat{\mathbf{G}}^{(i)}$ each computed by μ steps of relaxation on $A\mathbf{G}^{(i)} = \mathbf{I}^{(i)}$ and $\hat{A}\hat{\mathbf{G}}^{(i)} = \mathbf{I}^{(i)}$, respectively. Then*

$$S_{ij} = \hat{S}_{ij}.$$

Proof: This can be seen in two, more general, pieces.

First, consider relaxation on $\hat{A}\hat{\mathbf{G}}^{(i)} = D\mathbf{I}^{(i)}$, with initial guess $\hat{\mathbf{G}}^{(i)} = \mathbf{0}$. The true solution is $\hat{\mathbf{G}}^{(i)} = D^{-1}\mathbf{G}^{(i)}$, where $A\mathbf{G}^{(i)} = \mathbf{I}^{(i)}$, and, so, the initial error is $D^{-1}\mathbf{G}^{(i)}$. Choosing relaxation based on $\hat{M} = DMD$, relaxation on $\hat{A}\hat{\mathbf{G}}^{(i)} = D\mathbf{I}^{(i)}$ has error propagation matrix

$$I - \hat{M}^{-1}\hat{A} = I - D^{-1}M^{-1}AD = D^{-1}(I - M^{-1}A)D,$$

and the error propagation operator for μ sweeps of relaxation is $D^{-1}(I - M^{-1}A)^{\mu}D$. Thus, the error after μ sweeps of relaxation is $D^{-1}(I - M^{-1}A)^{\mu}\mathbf{G}^{(i)}$, and the approximation after μ sweeps of relaxation on $\hat{A}\hat{\mathbf{G}}^{(i)} = D\mathbf{I}^{(i)}$ is

$$D^{-1}\mathbf{G}^{(i)} - D^{-1}(I - M^{-1}A)^{\mu}\mathbf{G}^{(i)} = D^{-1}(I - (I - M^{-1}A)^{\mu})\mathbf{G}^{(i)}.$$

Taking $D = I$, we see that the approximation after μ sweeps of relaxation on $A\mathbf{G}^{(i)} = \mathbf{I}^{(i)}$ is $(I - (I - M^{-1}A)^{\mu})\mathbf{G}^{(i)}$.

Now, for any matrix B, consider relaxation on $B\mathbf{y} = \mathbf{b}$ and on $B\tilde{\mathbf{y}} = \alpha\mathbf{b}$, with zero initial guesses. The initial error for $B\mathbf{y} = \mathbf{b}$ is the exact solution, $\mathbf{y}$, whereas the initial error for $B\tilde{\mathbf{y}} = \alpha\mathbf{b}$ is $\tilde{\mathbf{y}} = \alpha\mathbf{y}$. After μ sweeps of relaxation, the error for $B\mathbf{y} = \mathbf{b}$ will then be $(I - M^{-1}A)^{\mu}\mathbf{y}$, while the error for $B\tilde{\mathbf{y}} = \alpha\mathbf{b}$ will be $\alpha(I - M^{-1}A)^{\mu}\mathbf{y}$. Likewise, the approximation for $B\mathbf{y} = \mathbf{b}$ will be $(I - (I - M^{-1}A)^{\mu})\mathbf{y}$, while the approximation for $B\tilde{\mathbf{y}} = \alpha\mathbf{b}$ will be $\alpha(I - (I - M^{-1}A)^{\mu})\mathbf{y}$. Thus, after any number of relaxations, the two approximations will differ only by the same factor, α by which their right sides differed. Thus, the approximation to $\hat{\mathbf{G}}^{(i)}$ after μ sweeps of relaxation on $\hat{A}\hat{\mathbf{G}}^{(i)} = \mathbf{I}^{(i)}$ is $(D_{ii})D^{-1}(I - (I - M^{-1}A)^{\mu})\mathbf{G}^{(i)}$.

Now, consider the computed strength measures, S_{ij} and $\hat{S}_{ij}$.

$$\begin{aligned}
\hat{S}_{ij} &= \frac{\|(D_{ii})D^{-1}(I - (I - M^{-1}A)^{\mu})\mathbf{G}^{(i)} - ((D_{ii})D^{-1}(I - (I - M^{-1}A)^{\mu})\mathbf{G}^{(i)})_j\mathbf{I}^{(j)}\|_{\hat{A}}}{\|(D_{ii})D^{-1}(I - (I - M^{-1}A)^{\mu})\mathbf{G}^{(i)}\|_{\hat{A}}} \\
&= \frac{\|D^{-1}(I - (I - M^{-1}A)^{\mu})\mathbf{G}^{(i)} - (D^{-1}(I - (I - M^{-1}A)^{\mu})\mathbf{G}^{(i)})_j\mathbf{I}^{(j)}\|_{DAD}}{\|D^{-1}(I - (I - M^{-1}A)^{\mu})\mathbf{G}^{(i)}\|_{DAD}} \\
&= \frac{\|(I - (I - M^{-1}A)^{\mu})\mathbf{G}^{(i)} - ((I - (I - M^{-1}A)^{\mu})\mathbf{G}^{(i)})_j\mathbf{I}^{(j)}\|_{A}}{\|(I - (I - M^{-1}A)^{\mu})\mathbf{G}^{(i)}\|_{A}} \\
&= S_{ij},
\end{aligned}$$

because $(D^{-1}\mathbf{v})_j\mathbf{I}^{(j)} = (D^{-1})_{jj}(\mathbf{v})_j\mathbf{I}^{(j)} = D^{-1}((\mathbf{v})_j\mathbf{I}^{(j)})$, for any vector, $\mathbf{v}$. □

The role of relaxation may also be filled by a suitable preconditioned conjugate gradient algorithm, allowing the computation of approximations to $\mathbf{G}^{(i)}$ that are optimal in the sense of minimizing the A-norm of the error in the approximations. Using the minimization property of PCG [2], we can show that this, as well, is invariant to diagonal scaling.

Theorem 3. *Let A be a symmetric positive-definite $n \times n$ matrix and D a diagonal $n \times n$ matrix. Let M, the preconditioning matrix for A, be symmetric and positive definite. Take $\hat{A} = DAD$ and $\hat{M} = DMD$, and let S_{ij} and $\hat{S}_{ij}$ be the strength measures of A and $\hat{A}$, respectively, as defined in Equation (3) for $1 \leq i, j \leq n$, $i \neq j$, with $\mathbf{G}^{(i)}$ and $\hat{\mathbf{G}}^{(i)}$ each computed by μ steps of PCG applied to $A\mathbf{G}^{(i)} = \mathbf{I}^{(i)}$ and $\hat{A}\hat{\mathbf{G}}^{(i)} = \mathbf{I}^{(i)}$, respectively. Then*

$$S_{ij} = \hat{S}_{ij}.$$

Proof: Following the approach of the previous theorem, we first show that the PCG iteration on $\hat{A}\tilde{\mathbf{G}}^{(i)} = D\mathbf{I}^{(i)}$ produces scaled versions of the iterates from the unscaled problem, and then show that the iterates produced from iteration on $\hat{A}\hat{\mathbf{G}}^{(i)} = \mathbf{I}^{(i)}$ differ from those of $\hat{A}\tilde{\mathbf{G}}^{(i)} = D\mathbf{I}^{(i)}$ by a scalar factor.

The preconditioned conjugate gradient algorithm applied to the system, $A\mathbf{x} = \mathbf{b}$, with (symmetric and positive-definite) preconditioner M and initial guess $\mathbf{x}_0 = \mathbf{0}$, produces iterates $\mathbf{x}_\mu$ that satisfy

$$\mathbf{x}_\mu = \underset{y \in \mathcal{K}_\mu(M^{-1}\mathbf{b}, M^{-1}A)}{\operatorname{argmin}} \langle A(\mathbf{x} - \mathbf{y}), \mathbf{x} - \mathbf{y} \rangle,$$

where the Krylov space,

$$\mathcal{K}_\mu(M^{-1}\mathbf{b}, M^{-1}A) = \operatorname{span}\{M^{-1}\mathbf{b}, M^{-1}AM^{-1}\mathbf{b}, \ldots, (M^{-1}A)^{\mu-1}M^{-1}\mathbf{b}\}.$$

Let $\mathbf{G}_\mu^{(i)}$ and $\tilde{\mathbf{G}}_\mu^{(i)}$ be the iterates produced by μ steps of the PCG iteration on $A\mathbf{G}^{(i)} = \mathbf{I}^{(i)}$ and $\hat{A}\tilde{\mathbf{G}}^{(i)} = D\mathbf{I}^{(i)}$, respectively. Then

$$\begin{aligned}
\mathbf{G}_\mu^{(i)} &= \underset{y \in \mathcal{K}_\mu(M^{-1}\mathbf{I}^{(i)}, M^{-1}A)}{\operatorname{argmin}} \langle A(\mathbf{G}^{(i)} - \mathbf{y}), \mathbf{G}^{(i)} - \mathbf{y} \rangle, \\
\tilde{\mathbf{G}}_\mu^{(i)} &= \underset{y \in \mathcal{K}_\mu(\hat{M}^{-1}D\mathbf{I}^{(i)}, \hat{M}^{-1}\hat{A})}{\operatorname{argmin}} \langle \hat{A}(D^{-1}\mathbf{G}^{(i)} - \mathbf{y}), D^{-1}\mathbf{G}^{(i)} - \mathbf{y} \rangle.
\end{aligned}$$

Note that Krylov space $\mathcal{K}_\mu(\hat{M}^{-1}D\mathbf{I}^{(i)}, \hat{M}^{-1}\hat{A}) = \mathcal{K}_\mu(D^{-1}M^{-1}\mathbf{I}^{(i)}, D^{-1}M^{-1}AD)$ and, so,

$$\begin{aligned}
&\mathcal{K}_\mu(\hat{M}^{-1}D\mathbf{I}^{(i)}, \hat{M}^{-1}\hat{A}) \\
&\quad = \operatorname{span}\{D^{-1}M^{-1}\mathbf{I}^{(i)}, D^{-1}M^{-1}AM^{-1}\mathbf{I}^{(i)}, \ldots, D^{-1}(M^{-1}A)^{\mu-1}M^{-1}\mathbf{I}^{(i)}\} \\
&\quad = D^{-1}\operatorname{span}\{M^{-1}\mathbf{I}^{(i)}, M^{-1}AM^{-1}\mathbf{I}^{(i)}, \ldots, (M^{-1}A)^{\mu-1}M^{-1}\mathbf{I}^{(i)}\} \\
&\quad = D^{-1}\mathcal{K}_\mu(M^{-1}\mathbf{I}^{(i)}, M^{-1}A).
\end{aligned}$$

Thus,

$$\begin{aligned}
\tilde{\mathbf{G}}_\mu^{(i)} &= \underset{y \in D^{-1}\mathcal{K}_\mu(M^{-1}\mathbf{I}^{(i)}, M^{-1}A)}{\operatorname{argmin}} \langle \hat{A}(D^{-1}\mathbf{G}^{(i)} - \mathbf{y}), D^{-1}\mathbf{G}^{(i)} - \mathbf{y} \rangle, \\
&= \underset{Dy \in \mathcal{K}_\mu(M^{-1}\mathbf{I}^{(i)}, M^{-1}A)}{\operatorname{argmin}} \langle A(\mathbf{G}^{(i)} - D\mathbf{y}), \mathbf{G}^{(i)} - D\mathbf{y} \rangle,
\end{aligned}$$

yielding $\tilde{\mathbf{G}}_\mu^{(i)} = D^{-1}\mathbf{G}_\mu^{(i)}$.

To see that $\tilde{\mathbf{G}}_\mu^{(i)}$ and $\hat{\mathbf{G}}_\mu^{(i)}$ differ by a constant, consider the iterates produced by μ steps of PCG on $A\mathbf{x} = \mathbf{b}$ and $A(\alpha\mathbf{x}) = \alpha\mathbf{b}$,

$$\begin{aligned}
\mathbf{x}_\mu &= \underset{y \in \mathcal{K}_\mu(M^{-1}\mathbf{b}, M^{-1}A)}{\operatorname{argmin}} \langle A(\mathbf{x} - \mathbf{y}), \mathbf{x} - \mathbf{y} \rangle, \\
\tilde{\mathbf{x}}_\mu &= \underset{y \in \mathcal{K}_\mu(\alpha M^{-1}\mathbf{b}, M^{-1}A)}{\operatorname{argmin}} \langle A(\alpha\mathbf{x} - \mathbf{y}), \alpha\mathbf{x} - \mathbf{y} \rangle.
\end{aligned}$$

The scalar, $\alpha \neq 0$, does not affect the span of the Krylov space and, so,

$$\begin{aligned}
\tilde{\mathbf{x}}_\mu &= \underset{\frac{1}{\alpha}y \in \mathcal{K}_\mu(M^{-1}\mathbf{b}, M^{-1}A)}{\operatorname{argmin}} \langle A(\alpha\mathbf{x} - \mathbf{y}), \alpha\mathbf{x} - \mathbf{y} \rangle, \\
&= \underset{y \in \mathcal{K}_\mu(M^{-1}\mathbf{b}, M^{-1}A)}{\operatorname{argmin}} \langle A(\mathbf{x} - \frac{1}{\alpha}\mathbf{y}), \mathbf{x} - \frac{1}{\alpha}\mathbf{y} \rangle.
\end{aligned}$$

Thus, $\tilde{\mathbf{x}}_\mu = \alpha \mathbf{x}_\mu$.

Applying this to the iterates $\tilde{\mathbf{G}}_\mu^{(i)}$ and $\hat{\mathbf{G}}_\mu^{(i)}$, we have that the right sides for the matrix problems $\hat{A}\tilde{\mathbf{G}}^{(i)} = D\mathbf{I}^{(i)}$ and $\hat{A}\hat{\mathbf{G}}^{(i)} = \mathbf{I}^{(i)}$ differ by a scalar factor, because $D\mathbf{I}^{(i)} = d_{ii}\mathbf{I}^{(i)}$ and, thus,

$$\hat{\mathbf{G}}_\mu^{(i)} = d_{ii}\tilde{\mathbf{G}}_\mu^{(i)} = d_{ii}D^{-1}\mathbf{G}_\mu^{(i)}.$$

Computing $\hat{S}_{ij}$ as in Equation (3), we then get

$$\hat{S}_{ij} = \frac{\|d_{ii}D^{-1}\mathbf{G}_\mu^{(i)} - d_{ii}D^{-1}(\mathbf{G}_\mu^{(i)})_j\mathbf{I}^{(j)}\|_{DAD}}{\|d_{ii}D^{-1}\mathbf{G}_\mu^{(i)}\|_{DAD}} = S_{ij}.$$

□

5. Grids Chosen

Having computed the strength measures, it remains to be specified how to determine a coarse grid. Strong connections are determined by a thresholding of the strength measures, with node i said to strongly depend on nodes j in

$$S_i = \{j : S_{ij} - 1 \geq \theta \max_{k \neq i} \{S_{ik} - 1\}\},$$

where, in the tests shown here, we take $\theta = 0.25$. We know that $S_{ij} \geq 1$ in the case that $\mathbf{G}^{(i)}$ is computed exactly, so we consider only the difference between the measures and 1 to more easily distinguish between relative strength. Applying this threshold gives, for each node i, a list of nodes that strongly influence i. This list is then passed to a coloring scheme, such as that used in the first pass of the classical AMG coarse-grid selection algorithm (see [11, Chap. 8] for details). This coloring algorithm then selects a maximal independent subset of the graph of the strong connections of A as the coarse grid.

A first test of this algorithm is to see whether it chooses coarse grids that coincide with our intuition for cases where this intuition is known to produce good multigrid algorithms. As such, we consider first the grids produced for the diffusion equation,

$$-\nabla \cdot \mathcal{K}(x, y)\nabla p(x, y) = Q(x, y),$$

in two dimensions, for various diffusivities, $\mathcal{K}(x, y)$. Choosing constant $\mathcal{K}(x, y) = \mathcal{K}$, we consider both isotropic and anisotropic diffusion on the unit square. In this section, the strength measure is computed based on sweeps of both Jacobi and a Jacobi-preconditioned conjugate gradient algorithm. We consider using PCG instead of straight Jacobi because, in our experience, it is faster to resolve the local character of the inverse of A, measured in the A-norm, than weighted-Jacobi alone, but has approximately the same cost. For the examples considered here, however, good coarse grids are chosen in both situations.

We first consider the Poisson equation, $\mathcal{K} \equiv 1$. Using 2 sweeps of Jacobi or Jacobi-preconditioned conjugate gradients generates the grids shown in Figure 1, with coarse-grid points indicated in white, whereas fine-grid points that are not selected are shown in gray. In both cases, the coarsening algorithm naturally selects the fully coarsened grid, known to be

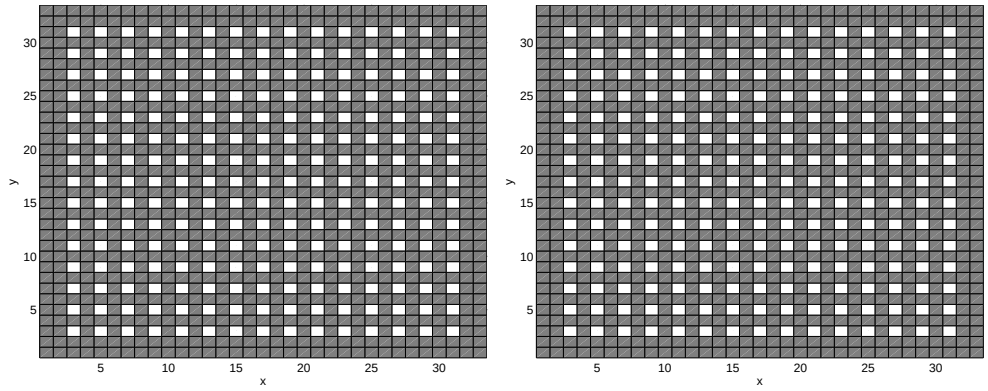

Figure 1. Coarse grids chosen for the Poisson problem with 2 iterations of Jacobi (left) and Jacobi-preconditioned CG (right)

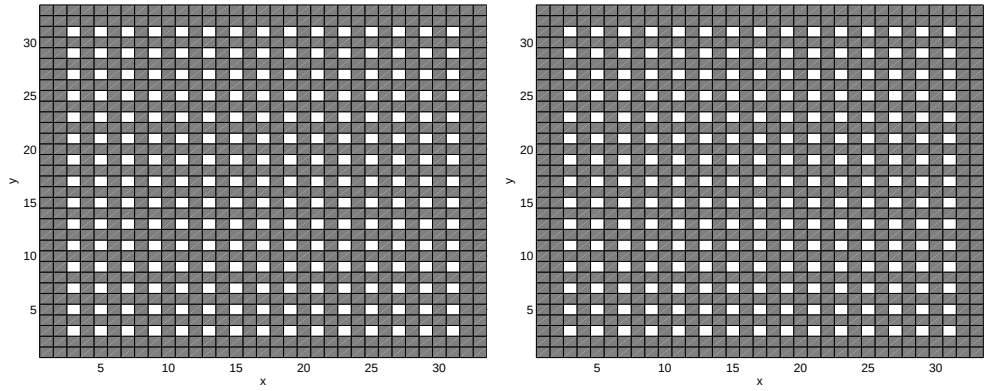

Figure 2. Coarse grids chosen for the diagonally scaled Poisson problem with 2 iterations of Jacobi (left) and Jacobi-preconditioned CG (right)

effective for this problem. Scaling the resulting discretization matrix by a diagonal matrix, D, chosen such that $d_{ii} = 10^{5r_i}$ for r_i chosen randomly from a uniform distribution on the interval $(0, 1)$, does not affect the results, as can be seen in Figure 2.

The proposed method also appears robust to anisotropies. Choosing a diffusion tensor, $\mathcal{K} \equiv Q^T \left[\begin{smallmatrix} 1 & 0 \\ 0 & \epsilon \end{smallmatrix}\right] Q$, for a rotation matrix, $Q = \begin{bmatrix} \cos(\alpha) & \sin(\alpha) \\ -\sin(\alpha) & \cos(\alpha) \end{bmatrix}$, is known to cause difficulties for classical AMG without tuning of the strong-connection threshold, θ [16, 23]. For grid-aligned anisotropy ($\alpha = 0$), the proposed coarsening algorithm chooses a semicoarsened grid, known to be effective in multigrid algorithms with the pointwise smoothing considered. Coarse grids for this problem with $\epsilon = 0.01$ are shown in Figure 3. For non-grid-aligned anisotropy, with $\alpha = \frac{\pi}{4}$, the coarse grids chosen are shown in Figure 4. The L-shaped coarsening pattern is a good grid selection for this problem, as each interior fine-grid node can interpolate from its

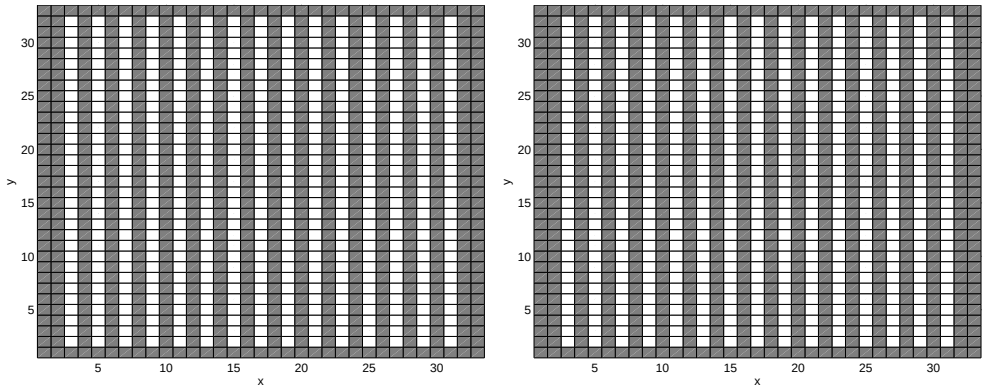

Figure 3. Coarse grids chosen for the grid-aligned anisotropic problem with 2 iterations of Jacobi (left) and Jacobi-preconditioned CG (right)

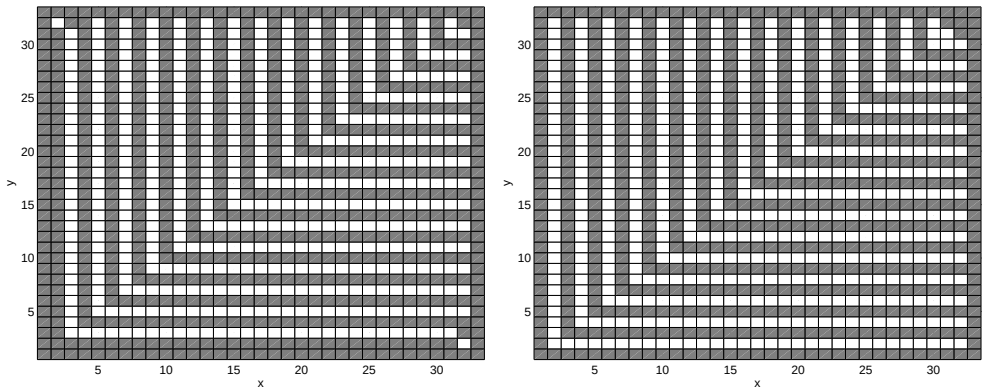

Figure 4. Coarse grids chosen for the non-grid-aligned anisotropic problem with 2 iterations of Jacobi (left) and Jacobi-preconditioned CG (right)

lower-left and upper-right neighbors along the direction of strong connection in the differential operator.

6. Numerical Results

Choosing appropriate coarse grids is only one ingredient in designing an effective AMG coarse-grid correction scheme. Equally important is the definition of the interpolation operator from the coarse to the fine grid. Here, we combine the proposed coarsening approach with the adaptive AMG interpolation algorithm of [10] to yield an effective variational AMG algorithm.

The adaptive AMG interpolation operator is determined by computing a prototype of algebraically smooth error (that which is not being effectively damped by relaxation) and

then fitting an AMG-style interpolation to this error. Given a prototypical algebraically smooth error, $\mathbf{x}$, the interpolation formula is defined by considering the small-residual equation, $A\mathbf{e} \approx \mathbf{0}$, that is known to hold for the algebraically smooth error of many matrices, A, with many pointwise relaxation schemes. Writing this out componentwise, we have

$$a_{ii}e_i \approx -\sum_{j\in C_i} a_{ij}e_j - \sum_{k\in F_i} a_{ik}e_k, \tag{4}$$

where C_i is the set of all strongly connected neighbors of i that are also coarse-grid points ($C_i = C \cap S_i$), and F_i is the set of all remaining points such that $a_{ik} \neq 0$ ($F_i = \{k : a_{ik} \neq 0, k \neq i, k \notin C_i\}$).

Connections to points $k \in F_i$ are collapsed based on the prototypical error. The quantity d^i_{kk} is defined as

$$d^i_{kk} = \frac{-\sum_{j\in C_i} a_{kj}x_j}{x_k},$$

so that

$$d^i_{kk}x_k = -\sum_{j\in C_i} a_{kj}x_j.$$

Interpolation is then determined by replacing a general algebraically smooth error, $\mathbf{e}$, at points $k \in F_i$ in Equation (4) with the approximation,

$$e_k = -\sum_{j\in C_i} \left(\frac{a_{kj}}{d^i_{kk}}\right) e_j,$$

giving the interpolation formula,

$$e_i = -\sum_{j\in C_i} \frac{1}{a_{ii}} \left(a_{ij} + \sum_{k\in F_i} a_{ik} \frac{a_{kj}x_k}{\sum_{j'\in C_i} a_{kj'}x_{j'}} \right) e_j.$$

Here, we make one small modification to this formula, because it is possible that $d^i_{kk} = 0$, giving an ill-posed definition of interpolation. If, for fine-grid node k, $\sum_{j\in C_i} a_{kj}x_j = 0$, then either k has no connection to a point in C_i, or the net result of the connections of k to C_i is zero. In either case, point k can be thought of as being especially weakly connected to i, in that it is a weak connection for i that is not connected to any of i's strong connections. Such a point should not have a significant effect on interpolation, but its connection to i must be accounted for in order to ensure the accuracy of interpolation for algebraically smooth errors. We follow classical AMG and collapse these points directly to the diagonal, weighted by the prototypical error, $\mathbf{x}$. That is, set F_i is partitioned into F^w_i and F^c_i, where points $k \in F^w_i$

grid	Laplace	Scaled Laplace	Anisotropic	Scaled Anisotropic
32×32	0.06	0.06	0.10	0.10
64×64	0.07	0.07	0.10	0.10
128×128	0.07	0.07	0.10	0.10
256×256	0.07	0.07	0.10	0.10
512×512	0.07	0.07	0.10	0.10

Table I. Asymptotic Convergence Factors of Resulting V(1,1) Cycles

satisfy $\sum_{j \in C_i} a_{kj} x_j = 0$ and $F_i^c = F_i \setminus F_i^w$. The resulting interpolation formula is then

$$e_i = -\sum_{j \in C_i} \frac{\left(a_{ij} + \sum_{k \in F_i^c} a_{ik} \frac{a_{kj} x_k}{\sum_{j' \in C_i} a_{kj'} x_{j'}} \right) e_j}{a_{ii} + \sum_{k \in F_i^w} a_{ik} \frac{x_k}{x_i}}. \tag{5}$$

Thus, we arrive at the setup algorithm:

- Given A,$\mathbf{b}$
- Relax ν_0 times on $A\mathbf{x} = \mathbf{0}$ with a random initial guess
- On each level:
  - Determine local strong connections by μ relaxations on $A\mathbf{G}^{(i)} = \mathbf{I}^{(i)}$ with a zero initial guess for each i
  - Choose coarse grid by coloring algorithm
  - Relax ν_1 times on $A\mathbf{x} = \mathbf{0}$ to improve representation of algebraically smooth error
  - Form interpolation, P, based on $\mathbf{x}$, as in Equation (5)
  - Compute $A_c = P^T A P$, inject $\mathbf{x}_c = (\mathbf{x})_c$

Here, we use Gauss-Seidel relaxation to expose the prototypical algebraically smooth error, $\mathbf{x}$, with ν_0 and ν_1 fixed to be 15. The strong connections are exposed with $\mu = 2$ local sweeps of pointwise Jacobi. Once the V-cycle components have been computed, V(1,1) AMG cycles with Gauss-Seidel relaxation are used in the solve phase. We compute the asymptotic convergence factors of these V(1,1) cycles and report them for both the Poisson problem and the grid-aligned anisotropic problem from the previous section in Table I.

These convergence factors are bounded independent of problem size, although parameters ν_0 and ν_1 were chosen so that the 512×512 grid problems showed optimal performance. In practice, the minimal values of ν_0 and ν_1 needed to achieve optimal multigrid performance increase with problem difficulty and size. Successive coarsenings maintain the structure of the fine-scale problems and, so, the complexities of the multigrid cycles remain nicely bounded,

independent of problem size, with grid complexities (defined as the number of points on all grids divided by the number of finest-grid points) of approximately $\frac{4}{3}$ for the Laplace problems and 2 for the anisotropic problems.

The non-grid-aligned anisotropic problem from the previous section provides a more difficult test for the algorithm. The first coarse grid, as shown in Figure 4, allows for an intuitively good coarse-grid correction along the diagonal axis. Two-level multigrid results confirm this expectation, with convergence factors below 0.20 for fine grids of 64×64 and 128×128 elements, for appropriate choices of ν_0 and ν_1. Multilevel results, however, degrade significantly, due to a break-down in the interplay of the adaptive AMG definition of interpolation, the exposure of prototypical algebraically smooth error, and the selection of further coarse grids. Resolution of this difficulty is the subject of current research.

One significant disadvantage of this approach for coarsening is its computational intensity. For each fine-grid node i, μ relaxations are needed to determine its strong connections, followed by the evaluation of an energy norm of the resulting vector. Even when this computation is designed to take advantage of the sparsity of the resulting vectors (they are non-zero only in a local neighborhood of i), the computation is still time consuming. It is very important that the number of relaxations, μ, be kept as small as possible and that the computation of the strength measures, S_{ij}, be done in a way to exploit the local structure of the approximations, $\mathbf{G}^{(i)}$. Finding the optimal parameters for a robust coarsening scheme that is as efficient as possible is currently under investigation.

7. Conclusions

A relaxation-based coarsening scheme for algebraic multigrid methods is proposed. The scheme uses relaxation to compute local approximations to the matrix inverse, which are then turned into strength measures using an A-norm-based measure. The resulting coarsening scheme is invariant under diagonal scalings of the matrix. When paired with the adaptive AMG interpolation scheme, the resulting algorithm produces efficient AMG V-cycles for many problems, but the interplay between the interpolation and coarse-grid selection is not as robust as is hoped. Understanding this relationship is the subject of ongoing research.

ACKNOWLEDGEMENTS

This work was sponsored by the Department of Energy under grant numbers DE-FC02-01ER25479 and DE-FG02-03ER25574, Lawrence Livermore National Laboratory under contract number B533502, Sandia National Laboratory under contract number 15268, and the National Science Foundation under VIGRE grant number DMS-9810751.

REFERENCES

1. M. Adams, M. Brezina, J. Hu, and R. Tuminaro, *Parallel multigrid smoothing: polynomial versus Gauss-Seidel*, J. Comput. Phys., 188 (2003), pp. 593–610.
2. S. F. Ashby, T. A. Manteuffel, and P. E. Saylor, *A taxonomy for conjugate gradient methods*, SIAM J. Numer. Anal., 27 (1990), pp. 1542–1568.

3. P. Bochev, C. Garasi, J. Hu, A. Robinson, and R. Tuminaro, *An improved algebraic multigrid method for solving Maxwell's equations*, Siam. J. Sci. Comput., 25 (2003), pp. 623–642.
4. A. Brandt, *General highly accurate algebraic coarsening*, Elect. Trans. Numer. Anal., 10 (2000), pp. 1–20.
5. A. Brandt, S. F. McCormick, and J. W. Ruge, *Algebraic multigrid (AMG) for sparse matrix equations*, in Sparsity and Its Applications, D. J. Evans, ed., Cambridge University Press, Cambridge, 1984.
6. J. Brannick and R. Falgout, *Compatible relaxation and coarsening in algebraic multigrid*, (2005). In preparation.
7. J. Brannick and L. Zikatanov, *Algebraic multigrid methods based on compatible relaxation and energy minimization*, in Proceedings of the 16th International Conference on Domain Decomposition Methods, 2005. Submitted.
8. M. Brezina, A. Cleary, R. Falgout, V. Henson, J. Jones, T. Manteuffel, S. McCormick, and J. Ruge, *Algebraic multigrid based on element interpolation (AMGe)*, SIAM J. Sci. Comp., 22 (2000), pp. 1570–1592.
9. M. Brezina, R. Falgout, S. MacLachlan, T. Manteuffel, S. McCormick, and J. Ruge, *Adaptive smoothed aggregation (αSA)*, SIAM J. Sci. Comp., 25 (2004), pp. 1896–1920.
10. ———, *Adaptive algebraic multigrid*, SIAM J. Sci. Comp., (2005). To Appear.
11. W. L. Briggs, V. E. Henson, and S. F. McCormick, *A Multigrid Tutorial*, SIAM Books, Philadelphia, 2000. Second edition.
12. O. Bröker, *Parallel Multigrid Methods using Sparse Approximate Inverses*, PhD thesis, Swiss Federal Institute of Technology, Zurich, Zurich, Switzerland, 2003.
13. T. Chartier, R. Falgout, V. Henson, J. Jones, T. Manteuffel, S. McCormick, J. Ruge, and P. Vassilevski, *Spectral AMGe (ρAMGe)*, SIAM J. Sci. Comp., (2003), pp. 1–26.
14. E. Chow, *An unstructured multigrid method based on geometric smoothness*, Numer. Linear Algebra Apps., 10 (2003), pp. 401–421.
15. A. Cleary, R. Falgout, V. Henson, and J. Jones, *Coarse-grid selection for parallel algebraic multigrid*, in Proc. of the Fifth International Symposium on Solving Irregularly Structured Problems in Parallel, vol. 1457 of Lecture Notes in Computer Science, Springer–Verlag, New York, 1998, pp. 104–115.
16. 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*, SIAM J. Sci. Comput., 21 (2000), pp. 1886–1908.
17. H. De Sterck, U. Yang, and J. Heys, *Reducing complexity in parallel algebraic multigrid preconditioners*, SIAM J. on Matrix Analysis and Applications, (2004). Submitted, Also available as LLNL Technical Report UCRL-JRNL-206780.
18. R. D. Falgout and P. S. Vassilevski, *On generalizing the AMG framework*, SIAM J. Numer. Anal., 42 (2004), pp. 1669–1693. Also available as LLNL technical report UCRL-JC-150807.
19. M. Grote and T. Huckle, *Parallel preconditioning with sparse approximate inverses*, SIAM J. Sci. Comput., 18 (1997), pp. 838–853.
20. V. Henson and U. Yang, *BoomerAMG: a parallel algebraic multigrid solver and preconditioner*, Applied Numerical Mathematics, 41 (2002), pp. 155–177.
21. O. Livne, *Coarsening by compatible relaxation*, Num. Lin. Alg. Appl., 11 (2004), pp. 205–227.
22. G. A. Meurant, *Numerical experiments with algebraic multilevel preconditioners*, Elect. Trans. Numer. Anal., 12 (2001), pp. 1–65.
23. J. W. Ruge and K. Stüben, *Algebraic multigrid (AMG)*, in Multigrid Methods, S. F. McCormick, ed., vol. 3 of Frontiers in Applied Mathematics, SIAM, Philadelphia, PA, 1987, pp. 73–130.
24. K. Stüben, *An introduction to algebraic multigrid*, in Multigrid, U. Trottenberg, C. Oosterlee, and A. Schüller, eds., Academic Press, San Diego, CA, 2001, pp. 413–528.
25. R. Tuminaro and C. Tong, *Parallel smoothed aggregation multigrid: Aggregation strategies on massively parallel machines*, in SuperComputing 2000 Proceedings, J. Donnelley, ed., 2000.
26. P. Vaněk, J. Mandel, and M. Brezina, *Algebraic multigrid by smoothed aggregation for second and fourth order elliptic problems*, Computing, 56 (1996), pp. 179–196.
27. U. Yang, *On the use of relaxation parameters in hybrid smoothers*, Num. Lin. Alg. Appl., 11 (2004), pp. 155–172.