Algebraic multigrid methods based on compatible relaxation and energy minimization

James Brannick[1] and Ludmil Zikatanov[2]

[1] Department of Applied Mathematics, University of Colorado, Boulder, CO, Email: James.Brannick@colorado.edu
[2] Department of Mathematics, The Pennsylvania State University, University Park, PA, Email: ludmil@psu.edu

Summary. This paper presents an adaptive algebraic multigrid method for the solution of positive definite linear systems arising from the discretizations of elliptic partial differential equations. The proposed method uses *compatible relaxation* to adaptively construct the set of coarse variables. The nonzero supports for the coarse-space basis is determined by approximation of the so-called two-level "ideal" interpolation operator. Then, an energy minimizing coarse basis is formed using an approach aimed to minimize the trace of the coarse-level operator. The presented approach maintains multigrid-like optimality, without the need for parameter tuning, for some problems where current algorithms exhibit degraded performance. Numerical experiments are presented that demonstrate the efficacy of the approach.

Key words: algebraic multigrid, compatible relaxation, trace minimization

1 Introduction

We consider solving linear systems of equations,

$$A\mathbf{u} = \mathbf{f}, \tag{1}$$

via algebraic multigrid (AMG), where $A \in \mathbb{R}^{n\times n}$ is assumed to be symmetric positive definite (SPD). Our AMG approach for solving (1) involves a stationary linear iterative smoother and a coarse-level correction. The corresponding two-grid method gives rise to an error propagation operator having the following form,

$$E_{TG} = (I - P(P^tAP)^{-1}P^tA)(I - M^{-1}A), \tag{2}$$

where $P : \mathbb{R}^{n_c} \mapsto \mathbb{R}^n$ is the interpolation operator and M is the approximate inverse of A that defines the iterative method. If A is symmetric, then this variational form of the correction step is optimal in the energy norm. A

multilevel algorithm is given by recursively solving the coarse-level problem, involving $A_c = P^t AP$, using a two-level method.

The efficiency of such an approach depends on proper interplay between the smoother and the coarse-level correction. Typically, the AMG smoother is fixed and the coarse-level correction is formed to compensate for its deficiencies. The primary task is, of course, the selection of P. It is quite common to use only the information from the current level in order to compute P and, hence, the next coarser space, because such a procedure can be implemented efficiently and at a low computational cost. A general process for constructing P is described by the following generic two-level algorithm:

- Choose a set of n_c coarse degrees of freedom;
- Choose a sparsity pattern of interpolation $P \in \mathbb{R}^{n \times n_c}$;
- Define the weights of the interpolation (i.e., the entries of P), giving rise to the next level operator as $A_c = P^t AP \in \mathbb{R}^{n_c \times n_c}$.

Standard algebraic multigrid methods use algorithms that rely on heuristics based on properties of M-matrices (e.g., strength of connection and algebraically-smooth error) in their setup to construct P. Although these traditional approaches have been shown to be extremely effective for a wide range of problems [Bra86, TOS01, RS87, VMB96], the use of heuristics based on M-matrix properties still limits their range of applicability. In fact, the components and parameters associated with these approaches are often problem dependent.

Developing more robust AMG solvers is currently a topic of intense research. General approaches for selecting the set of coarse variables are presented in [Liv04, BF05, BBM$^+$ar]. These approaches use compatible relaxation (CR) to gauge the quality of (as well as construct) the coarse variable set, an idea first introduced by Brandt [Bra00]. Recent successes in developing a more general form of interpolation include [BFM$^+$04a, BFM$^+$04b, VZ04, XZ04]. These methods are designed to allow efficient attenuation of error in a subspace characterized locally by a given set of error components, regardless of whether they are smooth or oscillatory in nature. In [BFM$^+$04a, BFM$^+$04b], the setup procedure computes these error components in a multilevel scheme based on the power method for the error propagation operator of the method itself.

The algorithm we propose for constructing P is motivated by the recently developed two-level theory introduced in [FV03] and [FVZ04]. We explore the use of this theory in developing a robust setup procedure in the setting of classical AMG. In particular, as in classical AMG, we assume that the coarse-level variables are a subset of the fine-level variables. Our coarsening algorithm constructs the coarse variable set using the CR-based algorithm introduced by Brannick and Falgout in [BF05]. The notion of strength of connection we use in determining the nonzero sparsity pattern of the columns of P is based on a sparse approximation of the so-called two-level ideal interpolation operator. Given the sparsity pattern of the columns of P, the values of the

nonzero entries of the columns of P are computed using the trace minimization algorithm proposed by Wan, Chan, and Smith [WCS00], based on the efficient implementation by Xu and Zikatanov [XZ04].

The remaining sections are organized as follows. In § 2, we introduce some notation and give a brief review of the theory motivating our approach. Next, § 3 gives a description of our CR-based AMG coarsening algorithm and § 4 describes our more general notion of strength of connection. In § 5 we describe our trace minimization form of interpolation. We give numerical results for the proposed method applied to a set of scalar PDE in § 6 and, in §7, we give some concluding remarks and future directions of our research.

2 Preliminaries and motivation

We begin by introducing notation. Since, in the presented algorithm, the coarse-level degrees of freedom are viewed as a subset of the fine-level degrees of freedom, prolongation P has the form $P = \begin{bmatrix} W \\ I \end{bmatrix}$, where I is the $n_c \times n_c$ identity and $W \in \mathbb{R}^{n_s \times n_c}$, $n_s = n - n_c$, contains the rest of the interpolation weights. In this way the coarse space $V_c \subset \mathbb{R}^n$ is defined as Range(P).

In what follows, we use several projections on the Range(P). These projections are defined for any SPD matrix X as follows:

$$\pi_X = P(P^t X P)^{-1} P^t X,$$

where, for $X = I$, we omit the subscript and write π instead of π_I. To relate the construction of interpolation to a compatible relaxation procedure, we introduce two operators: $R = [0, I]$ and S, where R has the dimensions of P^t and S has the dimensions of P. The fact that the coarse-level degrees of freedom are a subset of the fine-level degrees of freedom is reflected in the form of R. The matrix S corresponds to the complementary degrees of freedom, i.e. fine-level degrees of freedom, and can be chosen in many different ways, as long as $RS = 0$. In the approach presented here, we assume that $S = [I, 0]^t$. With R and S in hand, we define the 2×2 block splitting of any $X \in \mathbb{R}^{n \times n}$ by

$$X = \begin{bmatrix} X_{ff} & X_{fc} \\ X_{cf} & X_{cc} \end{bmatrix}, \tag{3}$$

where $X_{ff} = S^t X S$, $X_{fc} = S^t X R^t$, $X_{cf} = R X S$, and $X_{cc} = R X R^t$. We also need the Schur complement of X with respect to this splitting, defined as $\mathcal{S}(X) = X_{cc} - X_{cf} X_{ff}^{-1} X_{fc}$.

Given the smoother's M, the F-relaxation form of compatible relaxation (CR) we use in our algorithm yields an error propagation operator having the following form:

$$E_f = (I - M_{ff}^{-1} A_{ff}). \tag{4}$$

The associated symmetrized smoother is then defined as $\widetilde{M} := M^t(M^t + M - A)^{-1}M$, where $M^t + M - A$ is assumed to be SPD, a sufficient condition for convergence. To simplify the presentation here, we also assume that M is symmetric, in which case $2M - A$ being SPD is also necessary for the convergence of the smoothing iteration.

2.1 Some convergence results

The convergence result motivating our approach is a theorem proved in [FVZ04], giving the precise convergence rate of the two-grid algorithm.

Theorem 1. *Let E_{TG} be defined as in (2). Then*

$$\|E_{TG}\|_A^2 = 1 - \frac{1}{K(P)}, \qquad K(P) = \sup_{\mathbf{v}} \frac{\|(I - \pi_{\widetilde{M}})\mathbf{v}\|_{\widetilde{M}}^2}{\|\mathbf{v}\|_A^2}.$$

Assuming that the set of coarse degrees of freedom have been selected (i.e. R is defined), the remaining task is defining a P to minimize $K(P)$. Finding such a P is of course not at all straightforward, because the dependence of $K(P)$ on P given in Theorem 2 is complicated. To make this more practical we consider minimizing an upper bound of K, which is easily obtained by replacing $\pi_{\widetilde{M}}$ with π, the ℓ_2 projection on Range(P). We then obtain a measure for the quality of the coarse space defined as follows:

$$\mu(P) = \sup_{\mathbf{v}} \frac{\|(I - \pi)\mathbf{v}\|_{\widetilde{M}}^2}{\|\mathbf{v}\|_A^2}.$$

Note that $\mu(P) \geq K(P)$ for all P. Also, this measure suggests that error components consisting of eigenvectors associated with small eigenvalues (i.e., error not effectively treated by relaxation) must be well approximated by P. The following result from [FV03] gives $P_\star$ that minimizes $\mu(P)$.

Theorem 2. *Assume that R, S, and μ are defined as above. Then*

$$\mu(P_\star) = \min_P \mu(P), \quad \textit{where} \quad P_\star = \left[-A_{fc}^t A_{ff}^{-1}\ , I\right]^t.$$

Moreover, the asymptotic convergence factor of CR provides an upper bound for the above minimum as follows (see Theorem 5.1 in [FV03]).

Theorem 3. *If the number of non-zeros per row in A is bounded, then there exists a constant c, such that*

$$\mu(P_\star) \leq \frac{c}{1 - \rho_f}, \qquad \rho_f = \|E_f\|_{A_{ff}}^2.$$

A conclusion that follows immediately from this theorem is that ρ_f provides a *computable* measure of the quality of the coarse space, that is, a measure of the ability of the set of coarse variables to represent error not eliminated by relaxation.

The main ideas of our algorithm, described next, are based on observations and conclusions drawn from the above results.

3 Compatible relaxation based coarsening

In this section, we give more details on the first step of the algorithm, selecting the coarse degrees of freedom. The quality of the set of coarse-level degrees of freedom, C, depends on two conflicting criteria:

C1: *algebraically-smooth error should be approximated well by some vector interpolated from C, and*

C2: *C should have substantially fewer variables than on the fine level.*

In our adaptive AMG solver, the set of coarse variables is selected using the CR-based coarsening approach developed in [BF05]. This coarsening scheme is based on the two-level multigrid theory outlined in § 2: for a given splitting of fine-level variables Ω into C and F, F denoting the fine-level only variables, if CR is fast to converge, then there exists a P such that the resulting two-level method is uniformly convergent. The algorithm ties the selection of C to the smoother. The set of coarse variables is constructed using a multistage coarsening algorithm, where a single stage consists of: (1) running several iterations of CR (based on the current C) and (2) if CR is slow to converge, adding an independent set of fine-level variables (not effectively treated by CR) to C. Steps (1) and (2) are applied repeatedly until the convergence of CR is deemed sufficient, giving rise to a sequence of coarse variable sets:

$$\emptyset = C_0 \subseteq C_1 \subseteq ... \subseteq C_m,$$

where, for the accepted coarse set $C := C_m$, convergence of CR is below a prescribed tolerance. Hence, this algorithm constructs C so that **C1** is strictly enforced and **C2** is satisfied as much as possible. The details of this algorithm are given in [BF05].

An advantage of this approach, over the two-pass algorithm employed in classical AMG, is the use of the asymptotic convergence factor of compatible relaxation as a measure of the quality of C and, thus, the ability to adapt C when necessary. An additional advantage of this approach is that the algorithm does not rely on the notion of strength of connections to form C, instead, only the graph of matrix A and the error generated by the CR process are used to form C. This typically results in more aggressive coarsening than in traditional coarsening approaches, especially on coarser levels where *stencils* tend to grow. Additionally, this approach has been shown to work for a wide range of problems without the need for parameter tuning [BF05].

We conclude this section by proving the following proposition relating the spectral radii of E_f to the condition number of A_{ff}.

Proposition 1. *Consider compatible relaxation defined by E_f and let*

$$\rho(E_f) \leq a < 1. \tag{5}$$

Then

$$\kappa(A_{ff}) \leq \kappa(M_{ff}) \frac{1+a}{1-a}.$$

Proof. Let λ be any eigenvalue of $M_{ff}^{-1}A_{ff}$. Then $1-\lambda$ is an eigenvalue of $(I - M_{ff}^{-1}A_{ff})$. From (5) we have that

$$|1-|\lambda|| \leq |1-\lambda| \leq a, \qquad \text{implying} \qquad 1-a \leq |\lambda| \leq 1+a.$$

Thus $\kappa(M_{ff}^{-1}A_{ff}) \leq (1+a)/(1-a)$. From the assumption on the CR rate, it follows that M_{ff} is positive definite. The smallest eigenvalue of A_{ff} is then estimated as follows:

$$\begin{aligned}\lambda_{\min}(A_{ff}) &= \inf_{x\neq 0} \frac{(A_{ff}x,x)}{(x,x)} \geq \frac{\lambda_{\min}(M_{ff}^{-1/2}A_{ff}M_{ff}^{-1/2})}{\lambda_{\max}(M_{ff}^{-1})} \\ &= \frac{\lambda_{\min}(M_{ff}^{-1}A_{ff})}{\lambda_{\max}(M_{ff}^{-1})} \geq (1-a)\lambda_{\min}(M_{ff}).\end{aligned}$$

Estimating the maximum eigenvalue of A_{ff} in a similar fashion leads to the inequality

$$\lambda_{\max}(A_{ff}) \leq (1+a)\lambda_{\max}(M_{ff}). \tag{6}$$

The proof is then completed by using the last two inequalities in an obvious way.

Hence, fast-to-converge CR and M_{ff} being well conditioned imply that A_{ff} is well conditioned. Many problems that arise from PDE discretizations, M_{ff} is very well conditioned. This, together with the result from the next section, shows that fast convergence of CR indicates the existence of a sparse and local approximation to the inverse of A_{ff} and, hence, a good approximation to the two-level ideal interpolation operator. When M is ill conditioned, simple rescaling can sometimes be used to reduce the problem to the well-conditioned case. For example, replacing A by $D^{-1/2}AD^{-1/2}$ and M by$D^{-1/2}MD^{-1/2}$, where D is the diagonal of A, may produce a well conditioned M_{ff} so that the above conclusions apply.

4 Inverse of sparse matrices and supports of coarse grid basis vectors

We describe now the part of our algorithm that relate to the choice of the sparsity pattern of P. Set $\Omega = \{1,\ldots,n\}$ and assume that the coarse grid degrees of freedom are $C = \{n_s+1,\ldots,n\}$, where $n_s = n - n_c$. This leads to a 2×2 splitting of A, as given by (3). We aim to construct a covering of Ω with n_c sets $\{\Omega_i\}_{i=1}^{n_c}$, such that $\cup_{i=1}^{n_c}\Omega_i = \Omega$ contain information on the non-zero structure of the entries of P. We desribe our approach using some elementary tools from graph theory.

With matrix A_{ff}, we associate a graph, G, whose set of vertices is $\Omega \setminus C$, and set of edges is

$$\mathcal{E} = \{(i,j) \in \Omega \setminus C \quad \text{if and only if} \quad [A_{ff}]_{ij} \neq 0\}.$$

By graph distance between vertices i and j, denoted by $|i-j|_G$, we mean the length (i.e., the number of edges) of a shortest path connecting i and j in G. We assume without loss of generality that G is connected, so that the graph distance between any i and j is well defined. An important observation (see, for example, [Gib85]) related to the sparsity of A is that $(A_{ff}^k e_i, e_j) = 0$ holds for all k, i, and j such that $1 \leq k < |i-j|_G$. This in turn shows that, for any polynomial $p(x)$ of degree less than $|i-j|_G$, we have that

$$[A_{ff}^{-1}]_{ij} = (A_{ff}^{-1} e_i, e_j) = ((A_{ff}^{-1} - p(A_{ff}))e_i, e_j).$$

Taking the infimum over all such polynomials and using a standard approximation theory result for approximating $1/x$ with polynomials on the interval $[\lambda_{\min}(A_{ff}), \lambda_{\max}(A_{ff})]$, we arrive at the following inequality:

$$[A_{ff}^{-1}]_{ij} \leq c\, q^{|i-j|_G - 1}, \tag{7}$$

where $q < 1$ depends on condition number, κ, of A_{ff} and can be taken to be $\dfrac{\kappa^{1/2}-1}{\kappa^{1/2}+1}$, and c is a constant. The estimate on the decay of $[A_{ff}^{-1}]_{ij}$ given in (7) was contributed by Vassilevski [Vas04]. It is related to similar results for banded matrices due to Demko [DMS84]. This reference was also brought to our attention by Vassilevski [Vas04].

A simple and important observation from (7) is that a polynomial (or close to polynomial) approximation to the inverse A_{ff}^{-1} indicates exactly where the large entries of A_{ff}^{-1} are. Such an approximation can be constructed efficiently, since if A_{ff} is well-conditioned, the degree of the polynomial can taken to be rather small and, hence, the approximation will be sparse.

We use this observation in our algorithm to construct sets Ω_i in the following way: We first fix the cardinality of each Ω_i to be n_i (i.e. the number of nonzeros per column of P). Then, starting with initial guess $W_0 = 0 \in \mathbb{R}^{n_s \times n_c}$, we iterate towards the solution of $A_{ff}W = A_{fc}$ by ℓ steps of damped Jacobi iteration ($\ell \leq 5$):

$$W_k = W_{k-1} + \omega D_{ff}^{-1}(A_{fc} - A_{ff}W_{k-1}), \quad k = 1, \ldots, \ell. \tag{8}$$

Since this iteration behaves like a polynomial approximation to A_{ff}^{-1}, by (7), it follows that the largest entries in A_{ff}^{-1} will in fact show as large entries in W_ℓ. Thus to define Ω_i we pick the largest n_i entries in each column of W_ℓ.

There are also other methods that we are currently implementing for obtaining a polynomial approximation of A_{ff}^{-1}, such as a Conjugate Gradient approximation and also changing n_i adaptively. This is ongoing research. We point out that for the numerical results reported in 6, the approximations are based on the Jacobi iteration given in (8) with n_i fixed at the beginning.

5 On the best approximation to $P_\star$ in the trace norm

Since a covering of Ω was constructed in § 4, we proceed with the part of the algorithm for finding the interpolation weights. From the form of the iteration given in (8) for the sets $\{\Omega_i\}_{i=1}^{n_c}$, we have the following

$$\text{Each } \Omega_i \text{ contains exactly one index from } C. \tag{9}$$

To explore the relations between P obtained via trace minimization and the minimizer of $\mu(\cdot)$ introduced in § 2 consider the following affine subspaces of $\mathbb{R}^{n\times n_c}$:

$$\begin{aligned} \mathcal{X} &= \{Q \ : Q = \begin{bmatrix} W \\ I \end{bmatrix},\ W \in \mathbb{R}^{n_s\times n_c}\}, \\ \mathcal{X}_H &= \{Q \ : Q \in \mathcal{X},\ \ Q_{ji} = 0,\ \ \text{for all}\ \ j \notin \Omega_i;\ Q\mathbf{1}_c = \mathbf{e}\}. \end{aligned} \tag{10}$$

Here, $\mathbf{e}$ is an arbitrary nonzero element of $\mathbb{R}^n$ (as seen from (9) $\mathbf{e}$ is subject to the restriction that it is equal to 1 at the coarse grid degrees of freedom).

The interpolation that we use in our algorithm is then defined as the unique solution of the following constrained minimization problem:

$$P = \arg\min J(Q) := \text{trace}(Q^t A Q), \quad Q \in \mathcal{X}_H. \tag{11}$$

Various relevant properties of this minimizer can be found in the literature. Existence and uniqueness are shown in [WCS00, XZ04]. A proof that P is piecewise "harmonic" if $\mathbf{e}$ is harmonic can be found in [XZ04]. It is also well known that the i-th column of the solution to (11) is given by

$$[P]_i = I_i A_i^{-1} I_i^t M_a \mathbf{e}, \quad M_a^{-1} = \sum_{i=1}^{n_c} I_i A_i^{-1} I_i^t, \tag{12}$$

where $I_i \in \mathbb{R}^{n\times n_i}$ and $(I_i)_{kl} = \delta_{kl}$ if both k and l are in Ω_i and zero otherwise, and $A_i = I_i^t A I_i$. Associate with each Ω_i a vector space, V_i, defined as:

$$V_i = \text{span}\{e_j,\ j \in \Omega_i\}, \quad \dim V_i = n_i.$$

where e_j is the j-th standard canonical Euclidean basis vectors. Then, in (12), the matrix M_a^{-1} is the standard additive Schwarz preconditioner for A based on the splitting $\sum_{i=1}^{n_c} V_i = \mathbb{R}^n$.

We also have that, for any pair $Q_1 \in \mathcal{X}$ and $Q_2 \in \mathcal{X}$,

$$(Q_1 - Q_2)^t A P_\star = 0. \tag{13}$$

From this relation, in the extreme case, when each Ω_i contains $\{1, \ldots, n_s\}$ and $\mathbf{e} = P_\star \mathbf{1}_c$, we can easily obtain that $P_\star \in \mathcal{X}_H$, $P_\star$ minimizes $J(\cdot)$ and $J(P_\star) =$

trace($\mathcal{S}(A)$). Remember that $\mathcal{S}(A)$ is the Schur complement associated with the 2×2 splitting of A.

Since $J(Q)$ is in fact also a norm (equivalent to the usual Frobenius norm for Q), for convenience, we denote it by $|||Q|||_A^2 := J(Q)$. We have the following result:

Theorem 4. *Let P be the unique solution of (11). Then*

$$|||P_\star - P|||_A = \min_{Q \in \mathcal{X}_H} |||P_\star - Q|||_A \tag{14}$$

Proof. Let $Q \in \mathcal{X}_H$ be arbitrary. We use formula (13) and write

$$J(Q) = J(P_\star + (Q - P_\star)) = \text{trace}(\mathcal{S}(A)) + |||P_\star - Q|||_A^2 \,. \tag{15}$$

If we take the the minimum on the left side in (15) with respect to all $Q \in \mathcal{X}_H$, then we must also achieve a minimum on the right side. Hence

$$|||P_\star - P|||_A = \min_{Q \in \mathcal{X}_H} |||P_\star - Q|||_A \,,$$

which concludes the proof of the theorem.

In fact, this theorem, provides a way to estimate $|||P_\star - P|||_A$, and also to choose $\mathbf{e}$ (an error component to be represented exactly on coarser level). Since, as is well known (and can be directly computed), $J(P) = (M_a \mathbf{e}, \mathbf{e})$, from (15), we have that

$$|||P_\star - P|||_A^2 = (M_a \mathbf{e}, \mathbf{e}) - \text{trace}(\mathcal{S}(A)). \tag{16}$$

We can now take the minimum with respect to $\mathbf{e}$ on both sides of (16) and arrive at

$$|||P_\star - P|||_A^2 = \text{trace}[\mathcal{S}(M_a) - \mathcal{S}(A)], \tag{17}$$

where $\mathcal{S}(M_a)$ is the Schur complement of M_a and this equality holds for $\mathbf{e} = \begin{bmatrix} -M_{a,ff}^{-1} M_{a,fc} \mathbf{1}_c \\ \mathbf{1}_c \end{bmatrix}$. If we want to estimate the actual error of the best approximation, we need to estimate both quantities on the right side of (17). In fact, the first term, trace$[\mathcal{S}(M_a)]$, can be obtained explicitly since (9) implies that $\mathcal{S}(M_a)$ is diagonal. This can be easily seen by using the expression for M_a^{-1}, given in (12), in terms of A_i and I_i, and also the obvious relation $M_a^{-1} = \begin{bmatrix} * & * \\ * & [\mathcal{S}(M_a)]^{-1} \end{bmatrix}$. To get an accurate and computable estimate on the other quantity appearing on the right side of (16), namely, trace$(\mathcal{S}(A))$, we use the result from § 4 to get the following approximation

$$\text{trace}(\mathcal{S}(A)) \approx \text{trace}(A_{cc} - G_{cc}),$$

where, as in § 4, $G_{cc} = A_{cf} p(A_{ff}) A_{fc}$, and $p(x)$ is a polynomial approximating x^{-1} on $[\lambda_{\min}(A_{ff}), \lambda_{\max}(A_{ff})]$. Such estimates and also the relations between

optimizing the right hand side of (17), CR, and the optimal $\mathbf{e}$ (optimal for the norm $|||\cdot|||_A$), are also subject to an ongoing research. Currently in the numerical experiments we use an error component, $\mathbf{e}$, obtained during the CR iteration.

6 Numerical Results

We consider several problems of varying difficulty to demonstrate the effectiveness of our approach. Our test problems correspond to the bilinear finite element discretization of

$$-\nabla \cdot D(x,y)\nabla u(x,y) = f \quad \texttt{in} \quad \Omega = [0,1]\times[0,1] \tag{18}$$

$$u(x,y) = 0 \quad \texttt{on} \quad \partial\Omega \tag{19}$$

on a uniform rectangular grid. Our first test problem is Laplace's equation ($D \equiv 1$), a problem for which AMG works well. We consider the more difficult second problem defined by taking $D = \begin{bmatrix} 1 & 0 \\ 0 & 10^{-1} \end{bmatrix}$. In [BCF$^+$00], numerical experiments demonstrate the degraded performance classical AMG exhibits for this problem without appropriate tuning of the strength parameter (θ). This is an example of the fragility of current AMG methods. For our last test, we let $D = 10^{-8}$ in 20 percent of the elements (randomly selected) and $D = 1$ in the remaining elements. This type of rough coefficient problem becomes increasingly difficult with problem size. Classical AMG performance has been shown to degrade with increasing problem size for this problem as well [BFM$^+$04a].

To test asymptotic convergence factors, we use $\mathbf{f} = 0$ and run 40 iterations of $V(1,1)$ cycles with Gauss-Seidel relaxation. The trace minimization form of interpolation is computed using five iterations of an additive Schwarz preconditioned Conjugate Gradient solver.

The results in Table 1 demonstrate that our algorithm exhibits multigrid-like optimality for test problems one and two. Test two points to one advantage of our approach, namely, that our solver maintains optimality without any parameter tuning being necessary. Although the convergence factor of our solver grows with increasing problem size for test problem three, this is a rather difficult problem for any iterative solver, and our results are promising when compared to existing multilevel algorithms. To obtain a more complete picture of the overall effectiveness of our multigrid iteration, we examine also *operator complexity*, defined as the number of nonzero entries stored in the operators on all levels divided by the number of non-zero entries in the finest-level matrix. The operator complexity can be viewed as indicating how expensive the entire V-cycle is compared to performing only the finest-level relaxations of the V-cycle. We note that the operator complexities are acceptable for all of the test problems and remain bounded with repsect to problem size.

N	Problem 1	Problem 2	Problem 3
128^2	.085 / 5 / 1.29	.110 / 5 / 1.31	.098 / 5 / 1.79
256^2	.113 / 6 / 1.31	.124 / 6 / 1.35	.139 / 7 / 1.83
512^2	.118 / 7 / 1.33	.125 / 7 / 1.38	.197 / 9 / 1.87

Table 1. Asymptotic convergence factors / number of levels / operator complexities for test Problems 1-3.

7 Conclusions

Our current approach is only a first step towards developing a more general AMG algorithm. Using CR in constructing C and a trace minimization form of interpolation, we are able to efficiently solve problems arising from scalar PDEs. For systems of PDEs, there are other approaches that fit quite well in the framework described here. The CR algorithm can be extended in a straightforward way to include block smoothers as well as to incorporate more general algorithms for trace minimization (such as the one described in [VZ04]). Another attractive alternative is presented by using adaptive coarse space definition, namely by running simultaneous V-cycle iterations on the linear system that we want to solve and the corresponding homogeneous system (the latter with random initial guess) and using the error of the homogeneous iteration to define the constraint in the trace minimization formulation. Although expensive (part of the setup process has to be performed on every iteration), this procedure should be very robust and work in cases when there are many algebraically smooth error components that need to be approximated.

Acknowledgments

The authors would like to thank Rob Falgout and Panayot Vassilevski of the Center of Applied and Scientific Computing at the Lawrence Livermore National Laboratory (LLNL), Marian Brezina, Scott MacLachlan, Tom Manteuffel, Steve McCormick and John Ruge of the CU Boulder Applied Math Department, and Jinchao Xu of the Penn State Math Department for their insightful comments on this work. This work was sponsored by the Department of Energy under grant numbers DE-FC02-01ER25479 and DE-FC02-04ER25595, Lawrence Livermore National Laboratory under contracts number B533502 and number B551021, Sandia National Laboratories under contract number 15268, and the National Science Foundation under grant number DMS-0410318.

References

[BBM$^+$ar] J. Brannick, M. Brezina, S. MacLachlan, T. Manteuffel, S. McCormick, and J. Ruge. AMG coarsening. *Num. Lin. Alg. with App.*, to appear.

[BCF+00] 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(5):1570–1592, 2000.

[BF05] J. Brannick and R. Falgout. Compatible relaxation and coarsening in algebraic multigrid. *Num. Lin. Alg.*, 2005.

[BFM+04a] M. Brezina, R. Falgout, S. MacLachlan, T. Manteuffel, S. McCormick, and J. Ruge. Adaptive algebraic multigrid. To appear in SIAM J. Sci. Comp., 2004.

[BFM+04b] M. Brezina, R. Falgout, S. MacLachlan, T. Manteuffel, S. McCormick, and J. Ruge. Adaptive smoothed aggregation (αSA). *SIAM J. Sci. Comp.*, 25:1806–1920, 2004.

[Bra86] A. Brandt. Algebraic multigrid theory: The symmetric case. *Appl. Math. Comput.*, 19:23–56, 1986.

[Bra00] A. Brandt. General highly accurate algebraic coarsening. *ETNA*, 10:1–20, 2000. Special issue on the Ninth Copper Mountain Conference on Multilevel Methods.

[DMS84] S. Demko, W. Moss, and P. Smith. Decay rates of inverse band matrices. *Mahtematics of Computation*, 43:491–499, 1984.

[FV03] R. D. Falgout and P. S. Vassilevski. On generalizing the amg framework. Technical Report UCRL-JC-150807, Lawrence Livermore National Laboratory, 2003. To appear in SIAM J. Numer. Anal.

[FVZ04] R. D. Falgout, P. S. Vassilevski, and L. T. Zikatanov. On two–grid convergence estimates. Technical Report UCRL-JRNL-203843, Lawrence Livermore National Laboratory, April 2004. To appear in Num. Lin. Alg. Appl.

[Gib85] A. Gibbons. *Algorithmic Graph Theory*. Cambridge University Press, New York, NY, 1985.

[Liv04] O. Livne. Coarsening by compatible relaxtion. *Num. Lin. Alg. Appl.*, 11:205–227, 2004.

[RS87] J. Ruge and K. Stüben. Algebraic multigrid (AMG). In S. F. McCormick, editor, *Multigrid Methods*, volume 3 of *Frontiers in Applied Mathematics*, pages 73–130. SIAM, Philadelphia, PA, 1987.

[TOS01] Ulrich Trottenberg, Cornelis Oosterlee, and Anton Schüller. *Multigrid*. Academic Press, London, 2001.

[Vas04] P. S. Vassilevski. Exponential decay in sparse matrix inverses. Private communication, July 2004.

[VMB96] P. Vaněk, J. Mandel, and M. Brezina. Algebraic multigrid by smoothed aggregation for second and fourth order elliptic problems. *Computing*, 56(3):179–196, 1996.

[VZ04] P. S. Vassilevski and L. T. Zikatanov. Multiple vector preserving interpolation mappings in algebraic multigrid. Technical Report UCRL-JRNL-208036, Lawrence Livermore National Laboratory, November 2004. Submitted to SIAM J. Mat. Anal. Appl.

[WCS00] W. Wan, T. Chan, and B. Smith. An energy-minimizing interpolation for robust multigrid methods. *SIAM J. Sci. Comp.*, 21(4):1632–1649, 2000.

[XZ04] J. Xu and L. Zikatanov. On an energy minimizing basis for algebraic multigrid methods. *Computing and Visualization in Science*, 7:121–127, 2004.