New Multigrid Solver Advances in TOPS

R D Falgout[1], J Brannick[2], M Brezina[2], T Manteuffel[2] and S McCormick[2]

[1] Center for Applied Scientific Computing, Lawrence Livermore National Laboratory, P.O. Box 808, L-561, Livermore, CA 94551

[2] Department of Applied Mathematics, Campus Box 526, University of Colorado at Boulder, Boulder, CO 80309-0526

E-mail: rfalgout@llnl.gov

Abstract. In this paper, we highlight new multigrid solver advances in the Terascale Optimal PDE Simulations (TOPS) project in the Scientific Discovery Through Advanced Computing (SciDAC) program. We discuss two new algebraic multigrid (AMG) developments in TOPS: the adaptive smoothed aggregation method (αSA) and a coarse-grid selection algorithm based on compatible relaxation (CR). The αSA method is showing promising results in initial studies for Quantum Chromodynamics (QCD) applications. The CR method has the potential to greatly improve the applicability of AMG.

1. Introduction

At the core of many scientific simulation codes—especially those being developed as part of the SciDAC program—is the need to solve huge linear systems on thousands of processors. Multigrid methods are so-called *scalable* or *optimal* methods because they can solve a linear system with N unknowns with only $O(N)$ work. This property makes it possible to solve ever larger problems on proportionally larger parallel machines in constant time. Classical iterative methods like conjugate gradients are not scalable.

The algebraic multigrid (AMG) method was originally developed to solve general matrix equations using multigrid principles [1]. The fact that it used only information in the underlying matrix made it attractive as a potential black box solver. Instead, however, a wide variety of AMG algorithms have been developed that target different problem classes and have different robustness and efficiency properties.

In this paper, we consider solving the linear system

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

where A is a real symmetric positive definite (SPD) matrix. There are two main components to a multigrid method: the smoother (or relaxation method) and the coarse-grid correction step. The coarse-grid correction step involves operators that transfer information between the fine and coarse "grids", denoted more generally by the space $\Re^n$ and the lower-dimensional (coarse) vector space $\Re^{n_c}$. Let $P : \Re^{n_c} \to \Re^n$ be the *interpolation* (or *prolongation*) operator, and let the *restriction* operator be given by P^T. We use the Galerkin coarse-grid operator P^TAP. In all multigrid methods, the smoother and coarse-grid correction steps must complement each other

well. In AMG, the smoother is typically fixed, hence all of the work goes into constructing an appropriate P.

The goal of our research is to develop new AMG algorithms and techniques that broaden its range of applicability even further. In this paper, we will briefly describe two new developments. In Section 2, we will discuss the adaptive smoothed aggregation method (αSA) and show some encouraging numerical results in the area of Quantum Chromodynamics (QCD) simulations. In Section 3, we will discuss a new coarse-grid selection algorithm for AMG that is based on the idea of compatible relaxation (CR).

2. Adaptive Algebraic Multigrid

In order for a multigrid method to be optimal, the coarse-grid correction must eliminate those error components—the so-called *algebraically smooth* error components—that are not damped well by the relaxation process. The accuracy by which each of these components is approximated on the coarse grid is determined by its energy. As a consequence, the near null-space (kernel) components of the linear system must be represented almost exactly. Hence, nearly all multigrid methods today assume some a-priori information about the near null-space of the operator. However, for the Quantum Chromodynamics (QCD) application that we will discuss briefly at the end of this section, a-priori knowledge of the near kernel of the operator is not possible.

To address this problem, we have recently developed a class of AMG methods called *adaptive AMG* (αAMG). These methods employ the idea of "using the method to improve the method." Recent algorithms that utilize this basic idea are Wagner and Wittum's adaptive filtering [2], Brandt's bootstrap algebraic multigrid (bootstrap AMG) [3], adaptive smoothed aggregation [4], and a Ruge-Stüben-based adaptive AMG algorithm described in [5]. These methods exhibit the optimal convergence properties of multigrid, but do not require a-priori knowledge of the near null-space. Instead, they automatically "discover" these problematic components and make adjustments for them (i.e., they adapt).

One well-known method that can be viewed as an adaptive multigrid method is the preconditioned conjugate gradient method (PCG) (see [6] for details). The fact that PCG is generally not an optimal method illustrates the following crucial point about αAMG:

> slow-to-converge error components should not be viewed globally, but instead should be viewed only as *representatives* of locally smooth error.

In the αSA algorithm [4], this is achieved by "chopping up" near-kernel components to form the columns of a tentative interpolation operator. These column vectors can be thought of as a locally-supported basis that spans a space containing the near-kernel components, *plus much more*. The interpolation operator is then formed by smoothing the tentative interpolation operator. This is another important step in the αSA (and SA) algorithm because it lowers the overall energy of the space spanned by the range of interpolation, thereby improving its effectiveness for eliminating algebraically smooth error.

As alluded to above, QCD is an ideal application for using an adaptive method such as αSA. We have so far consider the simplified 2D Hermitian Dirac-Wilson formulation (see [7] for details). There are several difficulties solving these systems. The system becomes extremely ill-conditioned as the so-called quark mass ρ approaches certain critical values ρ_{crit}. In addition, the near null-space components of the operator are not known a-priori, and they are oscillatory as illustrated in Figure 1.

Table 1 shows results for αSA PCG and diagonally-scaled PCG. Eight near-kernel components were computed in the adaptive setup and used to define the transfer operators. The convergence factors for diagonally-scaled CG in the table illustrates the worsening of the condition number as $m_q := (\rho - \rho_{crit})$ approaches zero. On the other hand, the convergence factors for αSA remains uniformly bounded away from one for all parameter values. This is the main result in [7], and the first such result to date.

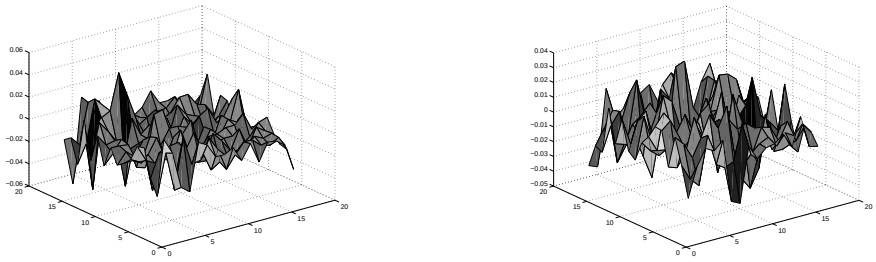

Figure 1. Real (left) and complex (right) algebraically smooth error for the Dirac system.

β / m_q	.01	.05	.1	.3
2	.33 /.99	.31 /.96	.31 /.94	.31 /.85
3	.42 /.98	.42 /.97	.40 /.93	.31 /.86
5	.31 /.99	.31 /.96	.29 /.92	.28 /.83

Table 1. Average convergence factors for αSA PCG and diagonally-scaled PCG applied to the Dirac system (size 16384) for various choices of ρ and β.

3. Compatible Relaxation and Coarse-Grid Selection

One important component of all AMG algorithms is the selection of coarse "grids" or *coarse variables*. Most coarsening algorithms are based on a particular notion of *strength of connection* defined in terms of the relative size of off-diagonal coefficients in the matrix. It is easy to show (by way of a simple example) that the size of off-diagonals is not a robust indicator of strength of connection, and as a result, these coarsening algorithms are not always robust. In this section, we discuss coarsening algorithms based on *compatible relaxation* (CR) as an alternative to using strength-of-connection. We will start first with the theory behind CR.

We will use the notation in [8]. Again, consider solving the linear system in Equation (1). Define the smoother error propagator by

$$I - M^{-1}A, \tag{2}$$

and assume that the smoother is convergent (in energy norm) so that $(M + M^T - A)$ is SPD. Denote the symmetrized smoother operator by $\widetilde{M} = M(M^T + M - A)^{-1}M^T$, i.e., $\widetilde{M}$ is the "M" in Equation (2) for the symmetric smoother $(I - M^{-T}A)(I - M^{-1}A)$.

Let $P : \Re^{n_c} \to \Re^n$ be the interpolation operator as before, and let $R : \Re^n \to \Re^{n_c}$ be some restriction operator such that $RP = I_c$, the identity on $\Re^{n_c}$. Note that R is not the multigrid restriction operator. We can think of R as defining the *coarse-grid variables*, i.e., $\mathbf{u}_c = R\mathbf{u}$. Note that PR is a projection onto range(P). Consider now the two-grid multigrid operator,

$$E_{TG} = (I - M^{-1}A)(I - P(P^TAP)^{-1}P^TA). \tag{3}$$

The next two theorems summarize the main convergence results in [8].

Theorem 3.1

$$\|E_{TG}\|_A^2 \leq 1 - \frac{1}{K}; \qquad K = \sup_{\mathbf{e}} \frac{\|(I - PR)\mathbf{e}\|_{\widetilde{M}}^2}{\|\mathbf{e}\|_A^2}. \tag{4}$$

Theorem 3.2

$$K \leq \eta K_\star; \qquad \eta = \|PR\|_A; \qquad K_\star = \inf_P \sup_{\mathbf{e}} \frac{\|(I - PR)\mathbf{e}\|_{\widetilde{M}}^2}{\|\mathbf{e}\|_A^2}. \tag{5}$$

Theorem 3.1 gives conditions that P must satisfy in order to achieve a fast uniformly convergent multigrid method. The idea is to use this theory to guide us in the development of AMG algorithms, and in particular, the selection of coarse grids. To do this, we use Theorem 3.2, which bounds the K in the first theory by two new constants, η and $K_\star$. The significance of this theorem is that it separates the construction of P into its natural two components: coarse-grid selection and definition of P's coefficients. The constant $K_\star$ is the K in the first theorem for the "best" P possible. Hence, $K_\star$ measures the quality of the coarse grid in some sense, because if it is small, we know there exists an interpolation operator that gives good AMG convergence. Once we have a coarse grid, the expression for η gives us guidance on how to define the coefficients of P in a way that is independent of the relaxation process.

To insure that $K_\star$ is bounded in practice, we can use CR. In fact, in [8], we proved that fast convergence of CR implies a small $K_\star$ (a good coarse grid). Based on this work, we have developed an algorithm for selecting coarse grids. The algorithm is as follows, where ρ_{cr} is the convergence factor of CR and θ is the stopping criteria of the CR-based coarsening algorithm:

$$\text{Initialize } \mathcal{U} = \Omega;\ \mathcal{C} = \emptyset;\ \mathbf{e} = \mathbf{1} \tag{6a}$$
$$\text{Do } \nu \text{ CR sweeps} \tag{6b}$$
$$\text{While } \rho_{cr} > \theta \tag{6c}$$
$$\gamma_i = |\mathbf{e}_i| / \|\mathbf{e}_i\|_\infty \tag{6d}$$
$$\mathcal{U} = \{i : \gamma_i > 1 - \rho_{cr}\} \tag{6e}$$
$$\mathcal{C} = \mathcal{C} \cup \{\text{independent set of } \mathcal{U}\} \tag{6f}$$
$$\text{Do } \nu \text{ CR sweeps} \tag{6g}$$

To date, we have considered only the case where the coarse grid is chosen as a subset of the fine grid variables. This is the classic AMG approach. In this setting, CR is simply F-relaxation. That is, if we decompose the fine grid variables into the coarse-grid points, C-pts, and everything else, F-pts, then CR is just relaxation on the homogeneous F-pt equations, $A_{ff}\mathbf{e} = 0$. The independent set algorithm is based on the nonzeros of A, and does not use any form of strength of connection.

As a simple first example, consider the following 2D diffusion equation

$$-au_{xx} - bu_{yy} + cu = f \tag{7}$$

on a square domain with coefficients that vary in four quadrants as depicted in Figure 2. We discretize on a uniform mesh with bilinear quadrilateral finite elements (this produces a 9-point stencil). The resulting discrete linear system is anisotropic and strongly coupled in the x-direction in the top left quadrant, anisotropic and strongly coupled in the y-direction in the top right quadrant, isotropic in the bottom right quadrant, and diagonally dominant in the bottom left quadrant. The coarse grid that our algorithm chooses is also given in the figure. We can see that the algorithm does exactly the right thing in each of the quadrants.

As a second example, again consider the diffusion problem in Equation (7), but now with $a \gg b$ uniformly throughout the domain. Hence, the problem is everywhere anisotropic and strongly coupled in the x-direction. In Figure 3, we show the coarse grids generated by our algorithm, where we use a pointwise Gauss-Seidel CR method on the left, and a line Jacobi CR method on the right. Here again, the algorithm picks the best complementary grid for the relaxation method being used. Furthermore, there are currently no coarse-grid selection algorithms based on strength of connection that are able to choose the grid on the right. This is one nice feature of our algorithm; the fact that it naturally complements the smoother.

This work was performed under the auspices of the U. S. Department of Energy by University of California Lawrence Livermore National Laboratory under contract W-7405-Eng-48.

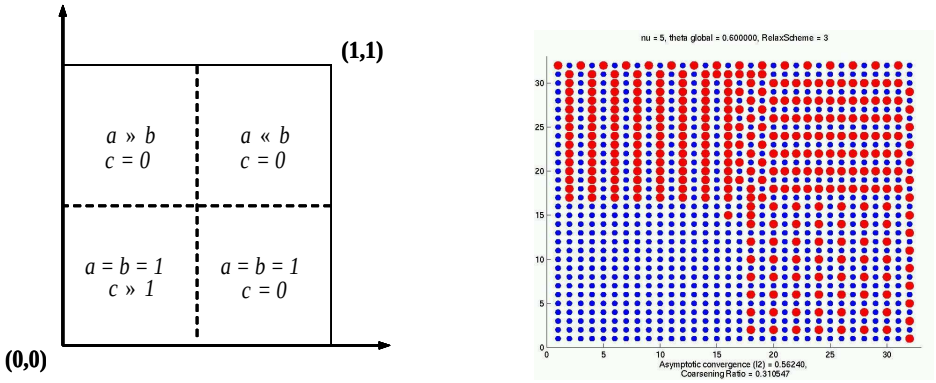

Figure 2. Coarse grid (right) for the diffusion equation Equation (7) with varying coefficient data in different quadrants of the domain (left).

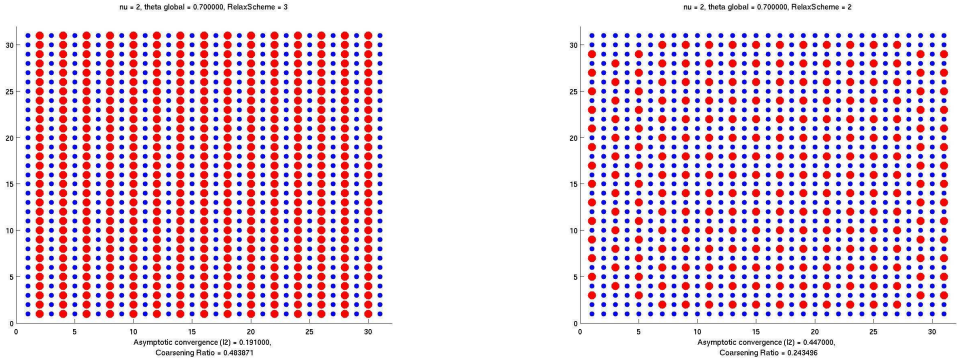

Figure 3. Coarse grids for the diffusion equation Equation (7) with uniformly anisotropic coefficient data. Two different smoothers were used in the CR algorithm: pointwise Gauss-Seidel (left), line Jacobi (right).

References

[1] 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.

[2] C. Wagner and G. Wittum. Adaptive filtering. *Numerische Mathematik*, 78(2):305–328, 1997.

[3] A. Brandt. General highly accurate algebraic coarsening. *Electronic Transactions on Numerical Analysis*, 10:1–20, 2000.

[4] M. Brezina, R. Falgout, S. MacLachlan, T. Manteuffel, S. McCormick, and J. Ruge. Adaptive smoothed aggregation (αSA). *SIAM J. Sci. Comput.*, 25(6):1896–1920, 2004. Also available as LLNL technical report UCRL-JC-151149.

[5] M. Brezina, R. Falgout, S. MacLachlan, T. Manteuffel, S. McCormick, and J. Ruge. Adaptive algebraic multigrid. *SIAM J. Sci. Comput.*, 2004. To appear. Also available as LLNL technical report UCRL-JRNL-203501.

[6] Robert D. Falgout. A note on the relationship between adaptive AMG and PCG. Technical Report UCRL-TR-205838, Lawrence Livermore National Laboratory, August 2004.

[7] J. Brannick, M. Brezina, O. Livne, I. Livshits, S. MacLachlan, T. Manteuffel, S. McCormick, J. Ruge, and L. Zikatanov. Adaptive smoothed aggregation in lattice qcd. Lecture Notes Comp. Sci. Eng. Springer–Verlag, 2005. to appear.

[8] R. D. Falgout and P. S. Vassilevski. On generalizing the AMG framework. *SIAM J. Numer. Anal.*, 42(4):1669–1693, 2004. Also available as LLNL technical report UCRL-JC-150807.