Electronic Transactions on Numerical Analysis.
Volume 37, pp. 276-295, 2010.

ISSN 1068-9613.

ETNA
Kent State University
http://etna.math.kent.edu

ADAPTIVE REDUCTION-BASED MULTIGRID FOR NEARLY SINGULAR AND HIGHLY DISORDERED PHYSICAL SYSTEMS*

J. BRANNICK†, A. FROMMER‡, K. KAHL‡, S. MACLACHLAN§, AND L. ZIKATANOV†

Abstract. Classical multigrid solution of linear systems with matrices that have highly variable entries and are nearly singular is made difficult by the compounding difficulties introduced by these two model features. Efficient multigrid solution of nearly singular matrices is known to be possible, provided the so-called Brandt-McCormick (or eigenvector approximation) criterion is satisfied, which requires building interpolation to fit the near-null-space modes with high accuracy. When these modes are known, traditional multigrid approaches may be very effective. In this paper, we consider the case of matrices describing highly disordered systems, such as those that arise in lattice quantum chromodynamics (QCD), where the near-null modes cannot be easily expressed in closed form. We develop a variational adaptive reduction-based algebraic multigrid preconditioner for such systems and present a two-level convergence theory for the approach for Hermitian and positive-definite systems. The proposed method is applied to a two-dimensional model known as the Gauge Laplacian, a common test problem for development of solvers in quantum dynamics applications, showing promising numerical results. The proposed reduction-based setup uses compatible relaxation coarsening together with a sparse approximation to the so-called ideal interpolation operator to recursively construct the coarse spaces.

Key words. algebraic multigrid, lattice QCD, gauge Laplacian, iterative methods, adaptive multigrid

AMS subject classifications. 65N55, 65N22, 65F10

1. Introduction.

Multigrid methods employ two complementary processes: *smoothing* and *coarse-grid correction*. In the classical setting, for scalar elliptic problems, the smoother (or relaxation method) is a simple iterative method, such as Gauss-Seidel, that is effective at reducing high-frequency error. The remaining low-frequency error is then accurately represented and efficiently eliminated on coarser grids via the coarse-grid correction step. To achieve their optimality, AMG methods employ a fixed smoother and generally exploit the character of the error of the relaxation method. Such error is referred to as algebraically smooth and, for most AMG relaxation schemes, is characterized by the near-nullspace (near-kernel) of the discrete operator: the span of all vectors x such that $Ax \approx 0$. For simpler problems, such as scalar elliptic partial differential equations, these methods are often optimal, since the near-nullspace components are known and AMG can be designed to resolve these types of components using a hierarchy of coarse-scale problems.

In recent years, significant effort has been focused on improving the range of applicability of black-box multigrid techniques. While there are many approaches to achieving robust linear solvers for wide classes of matrices, adaptive multigrid methods [10, 11, 12] offer many advantages because of the efficiency they inherit from the algebraic multigrid approaches on which they are based [31, 33, 32]. The key idea behind adaptive multigrid algorithms is to experimentally use the multigrid relaxation process itself to expose those error components that

*Received October 2, 2009. Accepted for publication June 28, 2010. Published online September 7, 2010. Recommended by R. Lehoucq.

†Department of Mathematics, Pennsylvania State University, University Park, PA 16802, USA ({brannick,ludmil}@psu.edu). This work was supported by the National Science Foundation grants OCI-0749202 and DMS-810982.

‡Fachbereich Mathematik und Naturwissenschaften, Bergische Universität Wuppertal, D-42097 Wuppertal, Germany ({frommer,kkahl}@math.uni-wuppertal.de). This work was supported by the Deutsche Forschungsgemeinschaft through the Collaborative Research Centre SFB-TR 55 "Hadron physics from Lattice QCD".

§Department of Mathematics, Tufts University, 503 Boston Avenue, Medford, MA 02155, USA (scott.maclachlan@tufts.edu). This work was supported in part by the European Community's Sixth Framework Programme, through a Marie Curie International Incoming Fellowship, MIF1-CT-2006-021927, and by the National Science Foundation grants OCI-0749317 and DMS-0811022.

must be accurately accounted for in the coarse-grid correction process, the so-called "algebraically smooth" errors that relaxation is slow to reduce. In its simplest form, this amounts to simply iterating many times with a fixed stationary iterative (relaxation) method on the homogeneous problem, $Ax = 0$, with a random initial guess. The dominant error left after many relaxation sweeps must, by definition, reflect the algebraically smooth errors of the problem. These errors can then be built into the coarse-grid correction process in the usual way. In practice, exposing these errors by simple relaxation alone is very inefficient and, so, the process is accelerated by a multilevel relaxation process that exposes the local and global characteristics of these slow-to-converge errors simultaneously.

These approaches have been shown to be successful for a wide range of problems [10, 11, 12]. An important new class of problems that can be effectively treated by these techniques arises in numerical models of quantum dynamics, e.g., quantum electrodynamics (QED) and quantum chromodynamics (QCD) [7, 8]. The caveat here is the large setup costs required by these "classical" adaptive solvers – the setup costs reported in [7] are roughly equivalent to that of solving the original system with 3 or 4 different right-hand sides using diagonally-preconditioned CG. Of course, certain QCD calculations require solves with thousands of right-hand sides and, so, these costs can be amortized. For other calculations, e.g., evolution of the gauge fields in a Monte Carlo process, each system needs to be solved with only a few right-hand sides, and these methods are not yet competitive.

The linear systems arising in numerical models of QED and QCD applications are very challenging for traditional multigrid methods, due mainly to two properties of the system matrix. The first is the extremely disordered nature of the matrix elements; each non-zero off-diagonal entry of the matrix is chosen at random from a specific distribution function, with little correlation between neighboring coefficients. The second property is that the system matrix is shifted so that it is very ill-conditioned (with smallest eigenvalue close to zero), much more so than one arising from a typical discretization of an elliptic PDE. As a result, classical multigrid assumptions are not satisfied by the discrete operator and, thus, such algorithms offer very little in terms of improved convergence over relaxation alone.

Individually, these difficulties can be treated by the classical approach to the adaptive setup procedure or by preconditioning CG with classical AMG, respectively. Indeed, for homogeneous but nearly singular elliptic problems, for example the shifted-Laplace operator, while AMG fails as a standalone solver, using AMG as a preconditioner for CG gives a very efficient and scalable solution technique [22]. Moreover, for heterogeneous problems that aren't nearly singular, adaptive AMG works very well, giving a good stand-alone solver [22, 10, 11, 6, 12, 9]. The systems encountered in numerical models of quantum field theories, however, exhibit both these difficulties; the resulting systems are very heterogeneous, so that adaptivity is needed, and are also nearly singular, in which case care must be taken to ensure that the computed prototypes give a suitable local representation of the algebraically smooth error (as they are assumed to in the "classical" adaptive process). Here, we consider an adaptive reduction-based multigrid algorithm as a preconditioner to CG for such systems. As a first step, we explore the applicability of these methods to a simplified Gauge Laplace system.

Multigrid methods for nearly singular problems have been considered before [2, 19, 18, 5, 29, 13], but not in this context. Whereas for classical elliptic operators such as Poisson's equation, accurate local fitting of the slowest to converge mode of relaxation (or the lowest-energy mode of the system matrix) is sufficient to ensure effective reduction of all error modes by the multigrid process, this is not the case for the extremely disordered systems that arise in quantum dynamical systems. For these operators, the smallest eigenvalues are typically well separated from the remainder of the spectrum and, moreover, the eigenvectors associated with these eigenvalues do not provide for a good local representation of the algebraically smooth error over the entire domain. Attempting to solve these systems by directly applying the adaptive multigrid methodology presents a difficult challenge, as classical multigrid wisdom [26, 3] requires that modes in the near-null space of the matrix be represented in the range of interpolation with accuracy inversely proportional to their energy norms and, furthermore, interpolation is typically based on the single lowest eigenmode of the system matrix. We demonstrate, however, that with careful design of the adaptive process, optimal performance of a multigrid-preconditioned Krylov iteration can be recovered for such systems. In addition, we explore various issues that must be considered in algorithmic development of adaptive methods for such systems. We also prove the two-level convergence of the method for Hermitian and positive-definite (HPD) systems and extend the theory of reduction-based AMG to allow for smoothing on all variables (using, for example, Jacobi or Gauss-Seidel smoothers) instead of only $\mathcal{F}$-smoothing.

The remainder of this paper is organized as follows. First, in Section 2, we introduce an important model problem for the operators that appear in quantum electro- and chromodynamics, the Gauge Laplacian system. In addition, we discuss some of the properties of this operator. In Section 3, we present an adaptive reduction-based algorithm and related two-level theory for general HPD systems. In addition, we explore several practical issues that arise in designing an adaptive AMG algorithm for disordered nearly singular problems such as the Gauge Laplacian in Section 4. Following this, in Section 5, we present numerical results of our modified adaptive reduction-based AMG (“αAMGr”) method for a variety of configurations of the Gauge Laplacian.

2. The Gauge Laplacian. The aim of this paper is to develop adaptive multigrid methods appropriate for the highly disordered nearly singular systems that arise in numerical simulations of quantum dynamics. We consider a simplified two-dimensional model problem called the “Gauge Laplacian”, as was done previously in [15, 20, 24]. The inverse of the Gauge Laplacian operator is the simplest form of a propagator satisfying a gauge theory [14] (a necessary and fundamental property for physical relevance of the calculation) and, thus, provides a good initial test problem in the development of AMG schemes for quantum dynamics applications.

Consider a uniform $N \times N$ periodic (toroidal) quadrilateral lattice, with $n = N^2$ node points $\{(k, \ell) \mid k, \ell = 1, \ldots, N\}$. Such a lattice has $n_e = 2n$ edges, which can be numbered individually from 1 through n_e or be associated in pairs with the lattice nodes, connecting a node (k, ℓ) to its “eastern” and “northern” neighbors, $(k + 1, \ell)$ and $(k, \ell + 1)$, respectively, where all numbers are understood to be mod N. On such a lattice, we are given values on each edge in the form of a “$U(1)$ gauge field”, $\mathcal{U} = \{u_j := e^{i\theta_j} \mid j = 1, \ldots, n_e\}$, where the values θ_j are prescribed based on some known distribution, discussed momentarily, and the “gauge links”, u_j, live on the edges of the lattice. Our interest is in the solution of systems of the form

$$A(\mathcal{U})\varphi = \psi,$$

where $A(\mathcal{U}) \in \mathbb{C}^{n\times n}$ and φ, ψ denote vectors from $\mathbb{C}^n$. The symbol x will stand for a lattice site, i.e., a point (k, ℓ) of the grid, and the operations $x \pm \mu$ for $\mu = 1, 2$ yield the neighboring lattice sites, i.e., $x \pm 1 = (k \pm 1, \ell)$ and $x \pm 2 = (k, \ell \pm 1)$; again all numbers are understood to be mod N.

The gauge links on the edges (one link, u_j, per edge j) act as coupling coefficients. Explicitly, the two-dimensional Gauge Laplace matrix $A = A(\mathcal{U})$ expresses a periodic nearest-neighbor coupling which, for a pair of lattice sites x, y with corresponding matrix entry A_{xy},

ETNA
Kent State University
http://etna.math.kent.edu

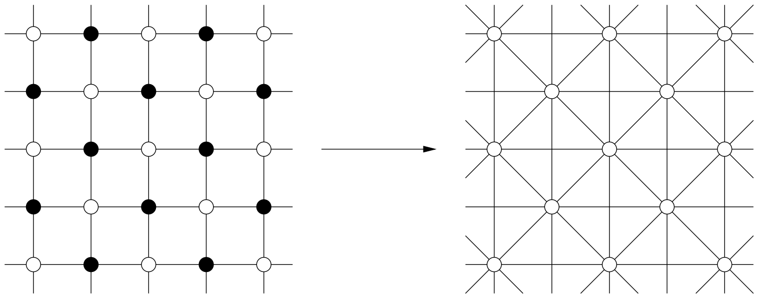

FIG. 2.1. *Connectivity of the Gauge Laplace operator before and after the odd-even reduction.*

can be described using the Kronecker δ as

$$A_{xy} = -\frac{1}{h^2}\sum_{\mu=1}^{2}\left(u_x^\mu\,\delta_{x+\mu,y} + \left(u_{x-\mu}^\mu\right)^\dagger\,\delta_{x-\mu,y}\right) + \left(\frac{4}{h^2}+m\right)\delta_{x,y}. \tag{2.1}$$

Here, u_x^μ is the gauge link defined on the edge connecting lattice sites x and $x+\mu$ and $(u_{x-\mu}^\mu)^\dagger$ is the complex conjugate of the gauge link defined on the edge connecting lattice sites $(x-\mu)$ and x. As is usually the case when considering PDEs on periodic grids, $h = 1/N$; the parameter m can be interpreted physically as a mass. It is common to explicitly scale A to have unit diagonal, yielding $A = I - \kappa D$, where $\kappa = \frac{h^2}{4+h^2m}$. In the related physics literature, the parameter κ is known as the "hopping" parameter and matrix D is known as the hopping matrix. We will work with the scaled matrix A from now on.

To be physically relevant, the gauge links u_j associated with a gauge field $\mathcal{U}$ are random variables from a given Boltzmann distribution that depends on a temperature parameter, β [14]. The case of $\beta = \infty$ is known as the so-called "cold" configuration and gives $u_j = 1$ for all j. For $\beta = 0$, the configurations are "hot", in which case the phases θ_j in $u_j := e^{i\theta_j}$ are uniformly distributed in $[0, 2\pi)$. Physically relevant configurations arise for $\beta \in (0, \infty)$.

The nearest-neighbor coupling that is inherent in the Gauge Laplacian suggests a further reduction of the problem using an odd-even (or red-black) reduction. Splitting the lattice sites into two sets, O and E, by

$$O := \{(k,\ell) : k+\ell \text{ odd}\}, \quad E := \{(k,\ell) : k+\ell \text{ even}\}$$

and ordering the variables such that all odd sites appear before the even ones, the matrix A exhibits the 2×2 block form

$$A = \begin{bmatrix} I & A^{(oe)} \\ A^{(eo)} & I \end{bmatrix}.$$

The Schur complement of A resulting from this "odd-even" reduced system is then given by $A^{(ee)} = I - A^{(eo)}A^{(oe)}$. Figure 2.1 illustrates the odd-even reduction: the 5-point stencil in the original system becomes a 9-point stencil in the odd-even reduced system.

The odd-even splitting is motivated by the fact that a solution $\tilde{\varphi}$ of the odd-even reduced system $A^{(ee)}\tilde{\varphi} = \tilde{\psi}$ can be easily interpolated exactly to the solution φ of the original system by

$$\varphi = \begin{bmatrix} A^{(oe)} \\ I \end{bmatrix}\tilde{\varphi}. \tag{2.2}$$

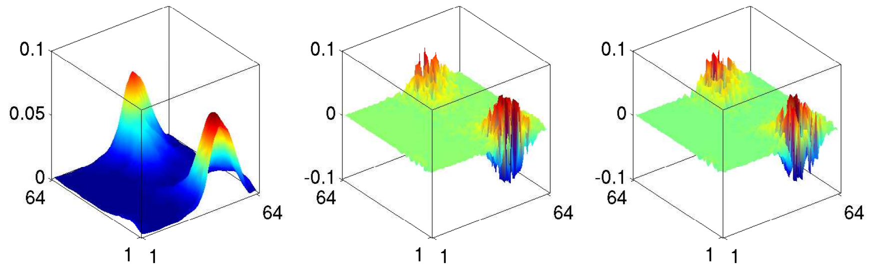

FIG. 2.2. *Modulus, real and imaginary part of the eigenmode to the smallest eigenvalue for $\beta = 5$ on a 64×64 grid, no odd-even reduction.*

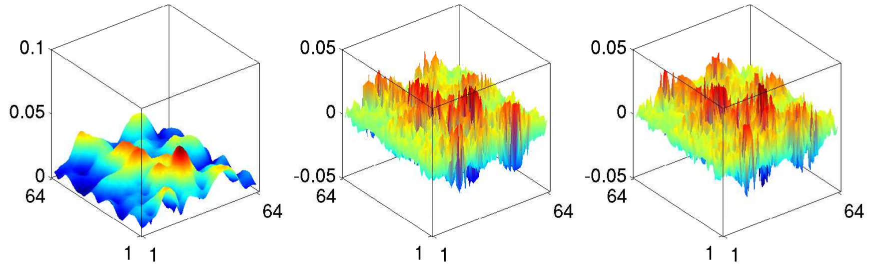

FIG. 2.3. *Modulus, real and imaginary part of an algebraically smooth error after* 50 *Gauss-Seidel iterations for $\beta = 5$ on a 64×64 grid, no odd-even reduction.*

Interpreting this splitting in our reduction-based algebraic multigrid framework, the exact Schur complement will turn out to be a suitable choice for the first coarse-grid operator using the odd-even splitting. This leads to a significant reduction in problem size with almost no additional computational cost to retrieve the solution of the original system. For this reason, we often assume this odd-even reduction has already been performed as a first coarsening step and work directly on the odd-even reduced system. Hereafter, we state explicitly when such a reduction is not used.

2.1. Spectral properties of the Gauge Laplacian. From (2.1), it follows that the Gauge Laplacian is Hermitian. In our tests, we vary the hopping parameter κ to generate matrices with varying condition number. For each gauge field configuration, $\mathcal{U}$, there exists a constant κ_{cr} for which the Gauge Laplacian with $\kappa = \kappa_{cr}$ is singular whereas, for $\kappa < \kappa_{cr}$, it is positive definite. In the following, we assume that κ is chosen to be close to κ_{cr} but smaller than κ_{cr}, so that the Gauge Laplacian operator is positive definite.

An important feature to consider when developing solvers for the Gauge Laplace system is the character of the algebraically smooth error of the system matrix, also called the near-kernel. In Figure 2.2, the modulus, real, and imaginary parts of the eigenmode with smallest eigenvalue is shown for $\beta = 5$ on a 64×64 grid, and the error after 50 Gauss-Seidel iterations applied to this same system with zero right-hand side and random initial guess is shown in

ETNA
Kent State University
http://etna.math.kent.edu

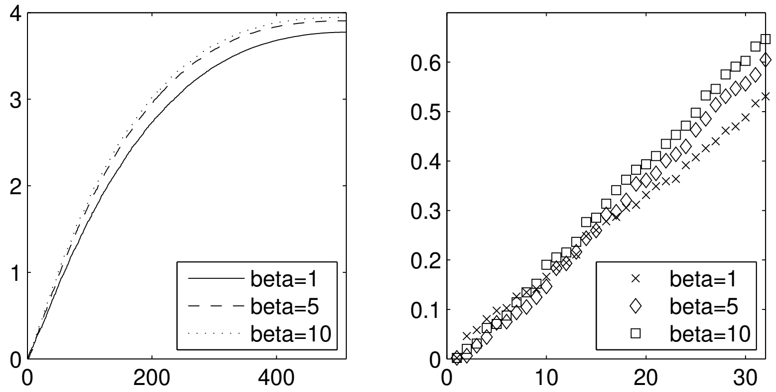

FIG. 2.4. *Eigenvalues of odd-even reduced Gauge Laplacians for* $\beta = 1, 5, 10$ *on a* 32×32 *grid. Depicted on the left hand side are the full spectra; on the right-hand side, a close-up of the smallest* 32 *eigenvalues for each temperature is shown.*

Figure 2.3. Here, we see that the algebraically smooth error varies locally, with random behavior induced by the gauge field configuration. As these plots illustrate, the support of the eigenmodes is local, further adding to the difficulty of defining an effective MG interpolation operator for the Gauge Laplace system. Our reduction-based AMG interpolation is defined adaptively to fit a given relaxed vector (or some linear combination of eigenmodes), such as the computed vector shown in Figure 2.3.

Another important aspect to consider is the spectrum of the system matrix. As depicted in Figure 2.4, the spectrum of the odd-even reduced system tends to be clustered around its upper bound and only a few eigenvalues turn out to be small, with the smallest eigenvalue well separated from the second-smallest. Note that from (2.2), it is easy to see that the eigenvalues, λ, of A come in pairs, $\lambda, 2-\lambda$, and that

$$\mathrm{spec}(A^{(ee)}) = \{\lambda(2-\lambda) : \lambda \in \mathrm{spec}(A)\}.$$

When shifting the spectrum by changing the hopping parameter, κ, the relative difference between the two smallest eigenvalues actually increases. We will see later that this property of the Gauge Laplacian (in its odd-even reduced form) has a major influence on the adaptive setup process.

3. Theoretical considerations: two-level convergence estimates for our "modified" AMGr solver. Consider a decomposition of $\mathbb{C}^n$ into two subspaces, V_c and V_f, given by a splitting of these n variables into two sets, the coarse ($\mathcal{C}$) and fine ($\mathcal{F}$) variables. This decomposition induces the following block two-by-two representation of the Hermitian and positive-definite $n \times n$ matrix A,

$$A = \begin{bmatrix} A_{ff} & A_{fc} \\ A_{cf} & A_{cc} \end{bmatrix}. \tag{3.1}$$

As explained in [23], from a variational point of view, the operator

$$P_* = \begin{bmatrix} -A_{ff}^{-1}A_{fc} \\ I \end{bmatrix}$$

is the "ideal" interpolation operator in the sense that a $V(1,0)$ cycle with Galerkin coarse-grid operator and exact relaxation on $\mathcal{F}$ leads to a direct solver. As A_{ff}^{-1} is generally dense, however, a sparse approximation to A_{ff}^{-1} is needed to define a practical interpolation operator and, thereby, a variational multigrid algorithm. In the multigrid literature, the various multilevel iterations whose design is based on such an approximation are typically referred to as reduction-based AMG methods (AMGr), following [30], because of their close relation to total-reduction approaches.

Before describing our choice for the approximation of A_{ff}^{-1} and the resulting algorithm, we briefly recall the ideas and ingredients of the adaptive AMGr method from [23]. In particular, to expose the main differences between our "modified" AMGr solver and the AMGr method introduced in [23], we first recall the main assumptions of the latter method.

In the following, the notation $A \leq B$ between Hermitian matrices A and B is meant to be with respect to the cone of positive-semidefinite matrices, i.e., $A \leq B$ if and only if $\varphi^H A\varphi \leq \varphi^H B\varphi$ for all $\varphi \in \mathbb{C}^n$, where φ^H denotes the Hermitian transpose of the vector φ.

The main assumption in [23] is that there exists an easy-to-invert approximation D to A_{ff} that can be used in both the definition of interpolation and the $\mathcal{F}$-relaxation. Interpolation is then given by

$$P_D = \begin{bmatrix} -D^{-1}A_{fc} \\ I \end{bmatrix}$$

and the relaxation operator as

$$M = \omega \begin{bmatrix} D^{-1} & 0 \\ 0 & 0 \end{bmatrix}. \tag{3.2}$$

Sufficient conditions on D that guarantee convergence of a two-level method with error-propagation operator given by

$$E = (I - P(P^H AP)^{-1}P^H A)(I - MA), \tag{3.3}$$

i.e., with one step of pre-smoothing, have been given in [23] as

$$D \leq A_{ff} \leq (1 + \epsilon)D, \tag{3.4}$$

for any fixed $\epsilon > 0$ and

$$0 \leq A_D = \begin{bmatrix} D & A_{fc} \\ A_{cf} & A_{cc} \end{bmatrix}. \tag{3.5}$$

This result also holds with an arbitrary number of pre- or post-relaxation steps. The spectral equivalence relation (3.4) can be viewed as a smoothing property of D with respect to the set of fine variables $\mathcal{F}$. Compatible relaxation [4, 21, 6, 17] or the method of greedy partitioning [25] generate splittings where the set of fine variables, $\mathcal{F}$, yields an A_{ff} block that is well approximated by a known matrix, D. In this light, relation (3.4) states that D defines a convergent smoother on the $\mathcal{F}$-variables. Relation (3.5), on the other hand, can be interpreted as a requirement on the interpolation operator and, hence, the coarse-grid operator.

Under assumptions (3.4) and (3.5), and assuming $\omega = \frac{2}{2+\epsilon}$ in (3.2), the following estimate on the convergence of the two-grid method was proved in [23],

$$\|E\|_A^2 \leq \frac{\epsilon}{1+\epsilon}\left(1 + \left(\frac{\sqrt{\epsilon}}{2+\epsilon}\right)^2\right) < 1. \tag{3.6}$$

In [23, 25], approaches for finding D were focused on satisfying (3.4) only. For the problems considered here, however, enforcing (3.5) appears to be of equal importance. In [23], D

is adaptively defined to match the action of A_{ff}^{-1} on a specific vector u. In our case, this vector and the resulting D can be complex valued; however, assumption (3.5) cannot be fulfilled for such D.

Thus, we now look to generalize these conditions, by using one approximation, D_R, for relaxation and another, D_P, for defining interpolation: $P_{D_P} = \begin{bmatrix} -D_P^{-1}A_{fc} \\ I \end{bmatrix}$. As we show below, the following requirements on D_R and D_P also imply the convergence of the two-level method with a bound on the error-propagation matrix similar to that in (3.6),

$$\lambda D_R \le A_{ff} \le \Lambda D_R \;, \tag{3.7}$$

$$\theta P_*^H A P_* \le P_{D_P}^H A P_{D_P} \le \Theta P_*^H A P_* \;, \tag{3.8}$$

for some positive constants, λ, Λ, θ, and Θ. The proof of this convergence result uses the convergence estimate from [17] for the two-grid operator, E_{tg}, with one pre- and one post-smoothing step,

$$\|E_{tg}\|_A = 1 - \frac{1}{K},$$

where the constant K can be bounded as

$$K \le \frac{1}{1-\gamma^2} \sup_w \frac{w^H \widetilde{M}_s w}{w^H A_s w}$$

for constant γ and matrices $\widetilde{M}_s$ and A_s defined below. This estimate adds further insight into the two-grid convergence of AMGr methods and also leads to a proof of convergence of AMGr-based methods with full-grid smoothers, i.e., for the case where, as opposed to (3.2), the block row of M corresponding to the $\mathcal{C}$ variables is non-zero. For the sake of consistency, we now adopt the notation in [17].

Reduction-based AMGr methods use only smoothing in the space of fine degrees of freedom, $\mathcal{F}$, and, as such, can be interpreted as multiplicative hierarchical basis methods based on the space decomposition

$$V = SV_s + PV_c \tag{3.9}$$

with associated interpolation operators $P : V_c \to V$ and $S : V_s \to V$. Note that writing $V = \mathbb{C}^n$, so that $V_c = \mathbb{C}^{n_c}$ and $V_s = \mathbb{C}^{n_s}$, gives $P \in \mathbb{C}^{n \times n_c}$ and $S \in \mathbb{C}^{n \times n_f}$. In general, we assume that $[S, P]$ is a square invertible matrix, so that $n_f + n_c = n$. This is obviously fulfilled in the case where $P = \begin{bmatrix} W \\ I \end{bmatrix}$, for $W \in \mathbb{C}^{n_s \times n_c}$ and $S = \begin{bmatrix} I \\ 0 \end{bmatrix}$. In this case, the splitting in (3.9) is direct.

In the following, we impose certain restrictions on these subspaces to define a two-grid hierarchical basis method. First, define the coarse-grid matrix A_c and its hierarchical complement A_s as

$$A_c = P^H A P, \;\; A_s = S^H A S.$$

Additionally, for a given smoother, $M_s : V_s \to V_s$ for A_s (on V_s), define its symmetrized version,

$$\widetilde{M}_s = M_s^H \left(M_s^H + M_s - A_s\right)^{-1} M_s,$$

ETNA
Kent State University
http://etna.math.kent.edu

and introduce the following variational definition of the Schur complement, S_A, of A, induced by the above space decomposition,

$$v^H S_A v = \inf_w (Sw + Pv)^H A (Sw + Pv) .$$

Note that this definition of S_A is equivalent to the usual definition, $S_A = A_{cc} - A_{cf} A_{ff}^{-1} A_{fc}$; cf. [1, Thm. 3.8].

The strengthened Cauchy-Bunyakowski-Schwarz (CBS) inequality [1, Eq. (9.5)], which provides a bound on the abstract angle between two subspaces, can also be used to bound the spectral equivalence between A_c and S_A, as needed in (3.8). Let $\gamma \in [0, 1)$ be the smallest constant such that

$$\langle Sw, Pu\rangle_A^2 \le \gamma^2 \|Sw\|_A^2 \|Pu\|_A^2 \; \forall w \in V_s, \forall u \in V_c.$$

Due to [17, Lemma 2.1], it follows that

$$\left(1 - \gamma^2\right) A_c \le S_A \le A_c$$

or, equivalently,

$$S_A \le A_c \le \frac{1}{1-\gamma^2} S_A \; . \tag{3.10}$$

The two-grid hierarchical basis method (a symmetric two-level AMGr method) is defined by the error propagation operator

$$E = I - B^{-1}A = \left(I - SM_s^{-H} S^H A\right)\left(I - PA_C^{-1} P^H A\right)\left(I - SM_s^{-1} S^H A\right).$$

Note that we assume here that the cycle uses both pre- and post-smoothing.

THEOREM 3.1. *Let M_s be Hermitian and positive-definite, and let γ be the smallest constant such that relation* (3.10) *holds. Assume that there exist positive constants, $0 < c_1 \le c_2 < 2$, such that*

$$c_1 M_s \le A_s \le c_2 M_s. \tag{3.11}$$

Then M_s is a convergent smoother for A_s. Furthermore, the two-grid multiplicative hierarchical basis method defined by E with smoother M_s satisfies

$$\|E\|_A \le 1 - \frac{1-\gamma^2}{\alpha}, \; \textit{where } \alpha = \max\left(\frac{1}{c_1\left(2-c_1\right)}, \frac{1}{c_2\left(2-c_2\right)}\right) \; .$$

Proof. First note that $c_1 M_s \le A_s \le c_2 M_s$ implies $c_1 I \le M_s^{-1/2} A_s M_s^{-1/2} \le c_2 I$ which gives $(1-c_2)I \le I - M_s^{-1/2} A_s M_s^{-1/2} \le (1-c_1)I$. It follows that

$$\rho(I - M_s^{-1}A_s) = \rho(I - M_s^{-1/2} A_s M_s^{-1/2}) \le \max\{|1-c_1|, |c_2 - 1|\} < 1,$$

and, so, M_s defines a convergent smoother for A_s. Due to [17, Theorem 4.2], we also have that

$$A \le B \le KA, \; \text{where } K \le \frac{1}{1-\gamma^2} \sup_w \frac{w^H \widetilde{M_s} w}{w^H A_s w} \; . \tag{3.12}$$

Thus,

$$\|E\|_A \le 1 - \frac{1}{K} \; . \tag{3.13}$$

ETNA
Kent State University
http://etna.math.kent.edu

Note that $\sup_w \frac{w^H \widetilde{M}_s w}{w^H A_s w}$ can equivalently be defined as the smallest β for which $\widetilde{M}_s \leq \beta A_s$. Now,

$$\begin{aligned}\widetilde{M}_s \leq \beta A_s &\Leftrightarrow M_s \left(2M_s - A_s\right)^{-1} M_s \leq \beta A_s \\ &\Leftrightarrow A_s^{-\frac{1}{2}} M_s A_s^{-\frac{1}{2}} \left(2A_s^{-\frac{1}{2}} M_s A_s^{-\frac{1}{2}} - I\right)^{-1} A_s^{-\frac{1}{2}} M_s A_s^{-\frac{1}{2}} \leq \beta I.\end{aligned}$$

From (3.11), $\frac{1}{c_2} I \leq A_s^{-\frac{1}{2}} M_s A_s^{-\frac{1}{2}} \leq \frac{1}{c_1} I$ and, thus,

$$\beta \in \left\{ \frac{t^2}{2t-1} : t \in \left[\frac{1}{c_2}, \frac{1}{c_1}\right] \right\} . \tag{3.14}$$

Note that $\frac{1}{c_2} > \frac{1}{2}$ by the hypothesis. Taking the maximum of the set in (3.14), we see that

$$\beta \leq \alpha = \max \left(\frac{1}{c_1 \left(2 - c_1\right)}, \frac{1}{c_2 \left(2 - c_2\right)} \right) .$$

Thus, $\sup_w \frac{w^H \widetilde{M}_s w}{w^H A_s w} \leq \alpha$, and

$$K \leq \frac{\alpha}{\left(1 - \gamma^2\right)}. \tag{3.15}$$

Combining this with (3.13), we have

$$\|E\|_A \leq 1 - \frac{1-\gamma^2}{\alpha}. \quad \square$$

COROLLARY 3.2. *Let the Hermitian and positive-definite matrix D_s be given and the assumptions of Theorem 3.1 be satisfied. Further, assume that there are positive constants, λ and Λ, such that $\lambda D_s \leq A_s \leq \Lambda D_s$. Define the smoothing operator M_s as*

$$M_s = \frac{1}{\sigma} D_s, \; for \; \sigma = \frac{2}{\Lambda + \lambda}. \tag{3.16}$$

Then,

$$\|E\|_A \leq 1 - \frac{4\lambda\Lambda}{\left(\Lambda + \lambda\right)^2} \left(1 - \gamma^2\right) .$$

Proof. With (3.16), we have $c_1 = \frac{2\lambda}{\Lambda+\lambda}$ and $c_2 = \frac{2\Lambda}{\Lambda+\lambda}$ so that

$$\frac{1}{\alpha} = c_1 \left(2 - c_1\right) = c_1 c_2 = \frac{4\lambda\Lambda}{\left(\Lambda + \lambda\right)^2}. \quad \square$$

The requirements in (3.7) and (3.8) are tailored to reflect the smoothing property of D_R and the quality of interpolation defined by D_P compared to the ideal interpolation operator P_*. For a given choice of D_s, the bound in (3.7) can be directly turned into a bound on $\|E\|_A$, as in (3.16). The relationship between (3.8) and (3.10) arises through the strengthened Cauchy-Bunyakowski-Schwarz inequality for the coarse subspace spanned by P and the corresponding fine subspace, measuring the abstract angle between them. A more thorough analysis of this relation can be found in [17, 28].

Following [17], we can also derive an estimate for the convergence of a two-grid method that uses full smoothing, i.e., smoothing on both $\mathcal{F}$ and $\mathcal{C}$, rather than smoothing on only $\mathcal{F}$. Hence, this result also applies to full-grid Jacobi or Gauss-Seidel smoothing, which are of interest in a final implementation as they tend to yield far superior smoothing properties than $\mathcal{F}$-smoothing alone.

To analyze this case, consider a two-grid method given by its error propagator,

$$E_{tg} = I - B_{tg}^{-1} A.$$

Generalizing the above approach, we can interpret the two-grid method with full smoothing in the same framework as was used for the analysis of the $\mathcal{F}$-smoothing case. Instead of using a smoother, M_s, on V_s with $\|I - M_s^{-1} A_s\|_{A_s} \leq 1$, we consider a smoother $M \in \mathbb{C}^{n\times n}$ with $\|I - M^{-1}A\|_A \leq 1$ and analogously to (3.10) its symmetrized versions $\widetilde{M} = M^H \left(M^H + M - A\right)^{-1} M$ and $\bar{M} = M \left(M^H + M - A\right)^{-1} M^H$. Note that this is equivalent to assuming that $S = I \in \mathbb{C}^{n\times n}$. The two-grid preconditioner B_{tg}^{-1} may then be written as

$$B_{tg}^{-1} = \begin{bmatrix} I & P \end{bmatrix} \hat{B}_{tg}^{-1} \begin{bmatrix} I \\ P^H \end{bmatrix}$$

with

$$\hat{B}_{tg}^{-1} = \begin{bmatrix} I & -M^{-H}AP \\ 0 & I \end{bmatrix} \begin{bmatrix} \bar{M}^{-1} & 0 \\ 0 & A_c^{-1} \end{bmatrix} \begin{bmatrix} I & 0 \\ -P^H A M^{-1} & I \end{bmatrix}.$$

Assuming that $\widetilde{M}$ gives a convergent smoother for A with

$$A \leq \widetilde{M} \leq \kappa A$$

and that P is chosen so that

$$A_c \leq \nu S_A,$$

then [17, Theorem 5.1] gives the bound

$$\|E_{tg}\|_A \leq 1 - \frac{1}{\nu\kappa}.$$

If ν and κ are independent of n, then this bound is independent of the problem size. For the $\mathcal{F}$-smoothing case, we can look at (3.16) as a refinement of the condition on $\widetilde{M}$, while the conditions on A_c are equivalent in the two cases.

4. Implementation details and practical issues. In this section, we give a detailed discussion of the practical issues that must be addressed when designing an effective adaptive AMGr setup algorithm for the disordered and nearly singular systems arising in lattice gauge theories. Particular attention is paid to the two-level setup algorithm, noting that the multilevel algorithm follows immediately from recursion. While many other approaches are possible, we consider only a variational construction and, thus, limit our discussion to the construction of interpolation, P. In Section 5, we will consider a fixed choice of relaxation (Gauss-Seidel on the full matrix, A).

4.1. Compatible-relaxation-based coarse-grid selection. The first main task in the AMG setup algorithm is the partitioning of the grid into appropriate sets of coarse and fine variables. Given a certain localized structure of the linear operator A, as occurs in most discretizations of PDEs, a coarse-grid variable $u_k^{(c)}$ is generally defined through a weighted linear combination of fine-grid variables and "nearby" coarse-grid variables [4],

$$u_k^{(c)} = \sum_i \mu_{ki} u_i.$$

In our approach, however, we will take the more-standard approach and assume that the coarse-grid variables are simply a subset of the fine-grid variables.

We consider various compatible relaxation (CR) based approaches for partitioning the fine degrees of freedom (dofs) into a coarse set, $\mathcal{C}$, and its complementary set, $\mathcal{F}$, the fine-level-only dofs. In its simplest form, compatible relaxation is a relaxation scheme that is confined to the fine-grid variables keeping the coarse-grid variables fixed. As shown in [16], the convergence rate of compatible relaxation is directly related to the convergence of the algebraic multigrid method that incorporates the same coarse grid and can, thus, be viewed as a quality measure of the coarse set. As such, compatible relaxation can be used to develop a practical adaptive approach for coarse-grid selection [4, 21, 6].

In our context, fast convergence of CR can be used to show that A_{ff}, the restriction of A to the set $\mathcal{F}$, is well conditioned with respect to M_{ff}, the restriction of the relaxation matrix to $\mathcal{F}$, an essential condition for bounding (3.11) with reasonable constants, c_1 and c_2. Indeed, if we consider compatible relaxation with error-propagation operator defined by $E_f = I - M_{ff}^{-1} A_{ff}$ and let

$$\rho(E_f) \le a < 1 \tag{4.1}$$

and λ be any eigenvalue of $M_{ff}^{-1} A_{ff}$, then $1 - \lambda$ is an eigenvalue of $I - M_{ff}^{-1} A_{ff}$. From (4.1), we have that

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

Thus, $\kappa(M_{ff}^{-1} A_{ff}) \le (1+a)/(1-a)$. It also follows from (4.1) that M_{ff} is positive definite. The smallest eigenvalue of A_{ff} is, then, estimated as

$$\begin{aligned} \lambda_{\min}(A_{ff}) = \inf_{\varphi \ne 0} \frac{\varphi^H A_{ff} \varphi}{\varphi^H \varphi} &\ge \frac{\lambda_{\min}(M_{ff}^{-1/2} A_{ff} M_{ff}^{-1/2})}{\lambda_{\max}(M_{ff}^{-1})} \\ &= \frac{\lambda_{\min}(M_{ff}^{-1} A_{ff})}{\lambda_{\max}(M_{ff}^{-1})} \ge (1-a)\lambda_{\min}(M_{ff}). \end{aligned}$$

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

$$\lambda_{\max}(A_{ff}) \le (1+a)\lambda_{\max}(M_{ff}). \tag{4.2}$$

It then follows that

$$\kappa(A_{ff}) \le \kappa(M_{ff}) \frac{1+a}{1-a} \text{ and, similarly, } (1-a)M_{ff} \le A_{ff} \le (1+a)M_{ff}.$$

For the Jacobi relaxation scheme, this bound can be interpreted as the spectral equivalence of A_{ff} and its diagonal, D_{ff}. This, then, can lead to a proof that fast convergence of CR implies a well-conditioned A_{ff} for certain problems that arise from PDE discretizations.

Promising adaptive coarse-grid selection techniques have been developed based on the idea of compatible relaxation [21, 6, 9]. In what follows, we discuss an implementation of a coarsening algorithm that uses compatible relaxation. For completeness, we include pseudocode for this approach in the appendix as Algorithm 1. In the compatible relaxation framework, the quality of a given $\mathcal{C}/\mathcal{F}$-splitting is measured by the convergence rate of relaxation on the $\mathcal{F}$-variables with the $\mathcal{C}$-variables fixed. Aiming at a specified convergence rate a, CR approaches successively add variables to the set of coarse variables until the target convergence factor is achieved. Starting with an empty set of $\mathcal{C}$-variables and an initial error e_0, compatible relaxation exposes variables for which relaxation convergence, when measured by the pointwise change in a known error, e.g., iterating on the homogeneous problem, is slower than a chosen threshold. Because each variable influences the convergence of variables in its neighborhood when using a local relaxation scheme (such as Richardson, Jacobi, or Gauss-Seidel), it may not be necessary to add all variables that are slow to converge in CR to the set $\mathcal{C}$. Instead, in each cycle a maximally independent set of slow-to-converge variables is added to $\mathcal{C}$. That is, in each iteration of the process we add a disconnected subset of the remaining fine-grid variables to $\mathcal{C}$, where connectivity is defined by the graph canonically associated to A. This process is repeated until the convergence of CR is deemed to be fast enough. A detailed description of the algorithm considered here can be found in [6].

Results such as Theorem 3.1 and its corollary can also be useful for development of coarsening techniques that ensure fast convergence of compatible relaxation without having to run CR iterations to test convergence, a main cost of most CR-based coarsening procedures. In [25], one such algorithm, a greedy strategy using a measure of diagonal dominance of A_{ff} for coarse-grid selection, was introduced. We briefly review this approach now, referring the reader to Algorithm 2 in the appendix for a more detailed description.

The goal of coarsening is to partition the fine-level variables into disjoint sets, $\mathcal{F}$ and $\mathcal{C}$, such that $\mathcal{F} \cup \mathcal{C}$ is the entire fine grid. For a given partition, the following function measures the diagonal dominance of row i of the resulting matrix A_{ff},

$$\theta_i = \frac{|a_{ii}|}{\sum_{j\in\mathcal{F}} |a_{ij}|}.$$

Classical diagonal dominance corresponds to $\theta_i \geq \frac{1}{2}$ for all i in this definition. Given a threshold, θ, the greedy strategy from [25] tries to find the largest subset, $\mathcal{F}$, of the variables, such that $\theta_i \geq \theta$ for all rows i of A_{ff}.

To do this partitioning, a third set of "undecided" variables, $\mathcal{U}$, is introduced to represent variables that have not been assigned to $\mathcal{C}$ or $\mathcal{F}$. A dynamic measure,

$$\hat{\theta}_i = \frac{|a_{ii}|}{\sum_{j\in\mathcal{F}\cup\mathcal{U}} |a_{ij}|}, \tag{4.3}$$

tracks the diagonal dominance of each row in $\mathcal{U}$. If $\hat{\theta}_i \geq \theta$ for any $i \in \mathcal{U}$, variable i is automatically added to the set of fine variables $\mathcal{F}$. If there are no such variables, the least diagonally dominant variable, $i \in \mathcal{U}$, is added to the set of coarse variables, $\mathcal{C}$, and the dynamic measure is updated for all variables in the neighborhood of i (in the grid-interpretation of the matrix A). This procedure is repeated until a splitting of the problem domain into $\mathcal{F}$ and $\mathcal{C}$ is achieved, i.e., until $\mathcal{U}$ is empty.

As in [25, Theorem 4], the A_{ff} block after the greedy coarse-grid selection satisfies

$$\left(2-\frac{1}{\theta}\right) D_{ff} \leq A_{ff} \leq \frac{1}{\theta} D_{ff}\ , \tag{4.4}$$

TABLE 4.1
Greedy coarsening and compatible-relaxation-based coarsening for several odd-even reduced Gauge Laplacians on a 64×64 grid, with all systems shifted so that the smallest eigenvalue $\lambda_{\min}$ is $\lambda_{\min} = 1.0 \times 10^{-4}$.

Greedy Algorithm Performance						CR Algorithm Performance			
System	θ	$\frac{\Lambda_{th}}{\lambda_{th}}$	$\frac{\Lambda_{obs}}{\lambda_{obs}}$	$\frac{\lvert\mathcal{C}\rvert}{\lvert\Omega\rvert}$	γ_{CR}	System	a	$\frac{\lvert\mathcal{C}\rvert}{\lvert\Omega\rvert}$	γ_{CR}
$\beta = 1$	.55	10	6.539	.299	.648	$\beta = 1$	0.7	.304	.655
$\beta = 1$	.60	5	4.212	.418	.513	$\beta = 1$	0.65	.367	.606
$\beta = 1$	.65	3.33	2.938	.499	.446	$\beta = 1$	0.6	.379	.568
$\beta = 5$	.55	10	6.910	.268	.691	$\beta = 5$	0.7	.266	.675
$\beta = 5$	.60	5	4.041	.419	.554	$\beta = 5$	0.65	.419	.631
$\beta = 5$	.65	3.33	3.127	.478	.538	$\beta = 5$	0.6	.440	.573
$\beta = 10$	.55	10	6.740	.267	.694	$\beta = 10$	0.7	.301	.682
$\beta = 10$	.60	5	4.113	.421	.570	$\beta = 10$	0.65	.398	.643
$\beta = 10$	.65	3.33	3.060	.477	.533	$\beta = 10$	0.6	.427	.580

where $D_{ff} = \text{diag}(A_{ff})$. Clearly, the spectral equivalence of D_{ff} and A_{ff} gets better as θ increases. Note that (4.4) is equivalent to (3.7) with $D_R = D_{ff}$ and $\lambda = 2 - \frac{1}{\theta}, \Lambda = \frac{1}{\theta}$. In Table 4.1, the performance of the greedy coarse-grid selection with respect to the theoretical bounds from (4.4) and the observed best possible bounds, which can be computed from the respective generalized eigenvalue problem, are provided. In addition, we estimate the spectral equivalence bounds between A_{ff} and M_{ff} and also report the convergence rate, γ_{CR}, of CR for these partitions. We also report on the CR convergence rates for compatible relaxation run on these same problems, for values of a that produce similar coarsening ratios, $\frac{|\mathcal{C}|}{|\Omega|}$.

In general, the best possible equivalence bounds for the greedy strategy are a lot better than what the theory predicts. In particular, for $\theta = 0.55$ we obtain a relatively aggressive coarsening along with good spectral equivalence between D_{ff} and A_{ff}. Because of the similarity between the performance of the greedy and compatible-relaxation coarsening algorithms, we only give results using compatible relaxation in Section 5.

4.2. Adaptivity in the modified AMGr framework. As proposed in [23], we use an adaptive scheme to define D_P and, hence, the interpolation operator, P_{D_P}. As fast-to-converge Jacobi-CR implies that A_{ff} can be accurately approximated by a diagonal matrix, we take D_P to be diagonal. Under this assumption, we choose D_P so that D_P^{-1} matches the action of A_{ff}^{-1} on a given vector $u = \begin{bmatrix} u_f \\ u_c \end{bmatrix}$ that corresponds to the near-kernel; i.e., we require

$$-D_P^{-1} A_{fc} u_c = u_f = -A_{ff}^{-1} A_{fc} u_c \tag{4.5}$$

for a given u_c. The key issue to consider when attempting to design an efficient adaptive AMGr solver in this setting is then reduced to development of an efficient scheme for computing the prototype, u, used to define D_P. The classical adaptive methods [10, 11, 12] use repeated application of the given relaxation scheme (or the resulting solver) to compute (or improve) the prototype. In general, the two main drawbacks of this approach are that, first, there is no theoretically founded stopping criterion available for such an approach that guarantees its optimality; and, second, such a classical adaptive process requires (roughly) $O(\log(K))$ setup iterations, where K is the condition number of the matrix, to compute a sufficiently accurate approximation of the prototype [22]. For the Gauge Laplacian, the smallest eigenmode is often not a good local representative of the algebraically smooth error,

ETNA
Kent State University
http://etna.math.kent.edu

TABLE 5.1
Odd-even reduced 5-pt discretization of the Laplace operator with Dirichlet boundary conditions shifted so that the smallest eigenvalue is $\lambda_{\min}$. V$(2,2)$-cycle asymptotic convergence rates with Gauss-Seidel smoother, using Gauss-Seidel relaxation applied to a positive random initial guess in the adaptive setup phase.

$n_{rel} \setminus \lambda_{\min}$	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}
5	.06	.02	.04	.37	.85	.98
25	.07	.02	.05	.05	.38	.86
50	.07	.02	.05	.05	.17	.66
100	.07	.02	.06	.06	.06	.16
500	.07	.02	.06	.06	.06	.06
exact	.07	.02	.06	.06	.06	.06

which further compounds the difficulty of developing an adaptive scheme for this system. Our numerical experience suggests that developing the solver using a setup scheme for the problem shifted to have only a mild smallest eigenvalue, or perhaps a large smallest eigenvalue, and, then, using the resulting multigrid solver for the unshifted system provides a much more effective preconditioner than does directly applying the setup to the problem with full shift, which typically has much larger condition number. This seems to be mainly due to the fact that, as we shift the hopping parameter towards its critical value, the relative gap between the smallest few eigenvalues and the remaining ones increases. As this relative gap becomes larger, the adaptive process becomes increasingly dominated by these few modes.

5. Numerical results. For our numerical tests, we consider Gauge Laplacians of varying size, mass, and temperature to test the AMGr-style method. As a benchmark for later tests of our method applied to the Gauge Laplace system, we first consider the $\beta = \infty$ case with Dirichlet boundary conditions, which gives the standard 5-point discrete shifted-Laplacian operator,

$$\mathcal{L} = -\Delta - (2\pi^2 - m)I, \tag{5.1}$$

obtained using a central-difference discretization. Here, the lowest eigenmode is known and has global support; specifically, this lowest mode is the restriction of $\sin(\pi x)\sin(\pi y)$ to the grid points, and the lowest eigenvalue can be determined by the choice of shift, m. This problem was a first test case in the development of the original adaptive AMG setup process [22]. To illustrate the performance of the original adaptive process for such problems, we consider this problem with the shift chosen so that the system becomes increasingly ill-conditioned for fixed problem sizes. As the numerical results provided in Table 5.1 illustrate, such an adaptive setup procedure produces an effective solver for this model problem provided that a sufficient amount of work is done to expose the lowest mode of the system matrix, i.e., a sufficient amount of work is done to ensure that the weak approximation property [26, 3] is satisfied by P, built using this computed vector for the given shift.

Next, we report the results of this original adaptive setup applied to a highly disordered system. The numerical results in Table 5.2 correspond to this scheme applied to a Gauge Laplacian with randomly configured gauge field. Here, we take $\beta = 5$ and $N = 64$ and again vary the minimal eigenvalue and number of relaxations used to approximate the lowest eigenmode of the fine-level system. As the numerical results in Table 5.2 demonstrate, in contrast to the $\beta = \infty$ case, here increasing the number of relaxations used in the adaptive process eventually leads to degradation in performance of the resulting solver based on this single mode. Further, we see that this degradation is more severe in cases where the minimal eigenvalue is $O(10^{-3})$ or $O(10^{-4})$. This is consistent in all tests, except for the last column

ETNA
Kent State University
http://etna.math.kent.edu

TABLE 5.2

Odd-even reduced Gauge Laplace operator with periodic boundary conditions shifted to a fixed smallest eigenvalue. V$(2,2)$-cycle asymptotic convergence rates with Gauss-Seidel smoother, using Gauss-Seidel applied to a complex-valued random initial guess in the adaptive setup phase. In parentheses, we report the iteration count for preconditioned CG to reduce the initial residual by a relative factor of 10^8. For the line labeled "LC", a linear combination of the eigenvectors associated with the ten smallest eigenvalues of the system matrix, weighted by the reciprocal of their eigenvalues as the vector to be fit in the adaptive setup phase. The line labeled CG contains iteration counts of the Conjugate Gradient method applied to this system as a stand-alone solver; again the (relative) residual is reduced to 10^{-8} in these tests.

$n_{rel} \setminus \lambda_{\min}$	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}
5	.4 (9)	.79 (15)	.97 (19)	.99 (21)	.99 (23)	.99 (25)
25	.32 (9)	.53 (11)	.83 (14)	.98 (15)	.99 (17)	.99 (18)
50	.31 (8)	.55 (11)	.72 (12)	.95 (14)	.99 (15)	.99 (17)
100	.28 (8)	.52 (10)	.65 (13)	.9 (14)	.99 (16)	.99 (17)
300	.32 (8)	.48 (10)	.53 (10)	.54 (10)	.61 (11)	.89 (13)
500	.33 (8)	.5 (10)	.6 (11)	.6 (11)	.60 (11)	.62 (11)
exact	.31 (8)	.53 (10)	.61 (12)	.61 (11)	.62 (12)	.62 (12)
LC	.35 (8)	.43 (9)	.67 (11)	.67 (12)	.62 (11)	.62 (12)
CG	(44)	(75)	(107)	(231)	(343)	(435)

where the minimal eigenvalue is shifted to be $O(10^{-6})$. In this case, using the exact lowest mode does provide the best overall solver. This is to be expected as the weak approximation property implies that P must be able to reproduce this mode very accurately. Because of the local support of the smoothest eigenvalues for this problem, we see that using the minimal eigenvector is, in general, a suboptimal choice for the vector in the adaptive setup scheme. While each of these modes is supported locally, their support does not, in general, overlap exactly. In such cases, a linear combination of these modes may give a better approximation to the slow-to-converge modes of the system matrix.

To test this approach, we consider an "artificial" adaptive process that uses a linear combination of the eigenvectors associated with the k smallest eigenvalues of the system matrix, weighted by the reciprocal of their eigenvalues, as the vector to be fit in the adaptive setup phase. We choose $k = 10$ as this gives good performance in our numerical tests. Results for this approach are shown in Table 5.2 in the line labeled "LC". Here, we see that the performance of the stand alone MG solver based on this approach is not, in general, better than that of the solver based on P defined using a prototype computed using relaxation. As the lowest modes can be local, using relaxation (or a linear combination of the ten smallest eigenmodes computed exactly) does not produce an AMGr-style P that satisfies the weak approximation property [26, 3], which requires accuracy in the computed prototype proportional to its Rayleigh Quotient. However, both methods produce a P that leads to an effective variational MG preconditioner.

The results in Table 5.3 are for various problem sizes and choices of β. Here, P is defined using the prototype computed by using relaxation and also by taking a linear combination of the ten lowest modes. As before, we see that both solvers perform well as a preconditioner for CG. Overall, our proposed AMGr-style method, based on a single prototype, is not expected to produce an optimal stand-alone solver for these systems. Our numerical results suggest that the approach does, however, have potential for dramatically improving CG performance for cases where the more expensive multiple-vector type adaptive methods (e.g., α SA) are not applicable. An example of such setting was mentioned earlier, where only system solves for $O(1)$ right hand sides are needed for a given gauge field configuration.

ETNA
Kent State University
http://etna.math.kent.edu

TABLE 5.3

Odd-even reduced Gauge Laplacians of various sizes and temperatures β, shifted so that the smallest eigenvalue is $\frac{1}{N^2}$. AMGr 2-level V$(2,2)$ preconditioner with Gauss-Seidel smoothing for CG using both a linear combination of the smallest ten eigenmodes, scaled by their associated inverse RQs to define P (shown first) as well as using relaxation to define the prototype in the definition of interpolation (shown second).

$\beta \setminus N$	32	64	128	256
1	11 / 12	10 / 14	15 / 15	11 / 14
5	12 / 15	11/ 15	15 / 15	14 / 16
10	7 / 11	13 / 15	17 / 16	19 / 17

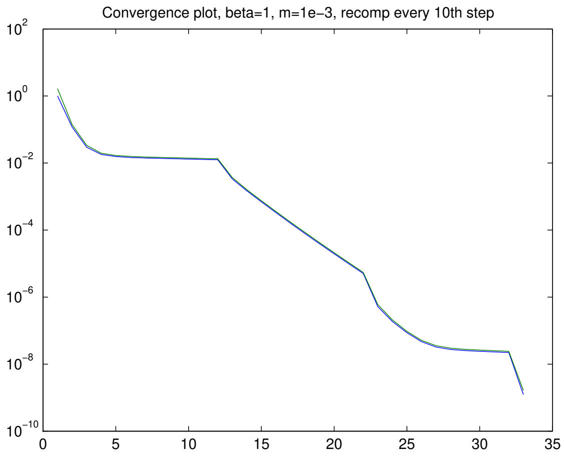

FIG. 5.1. *Performance of nonlinear adaptive solver to applied to the Gauge Laplacian with $\beta = 1$ and $N = 64$.*

5.1. Non-linear adaptive cycling schemes. A possible (practical) variant of the stationary adaptive setup schemes for the Gauge Laplace systems is given by a non-linear iteration in which we consider integrating the adaptive setup and solve phases into a single, non-linear, solution process. The most basic implementation of non-linear adaptive cycling schemes is to run the solver for the homogeneous and the inhomogeneous systems simultaneously and use the homogeneous system to improve the solver while solving the inhomogeneous problem. If we start with a fixed AMG method and apply a small number of steps of the method to both the homogeneous and inhomogeneous systems, we can adaptively tune our approach. If the convergence of the solver measured on the homogeneous system is fast enough, we continue to use this method for the non-homogeneous system of interest. If, on the other hand, the convergence factor of the method on the homogeneous system is larger than a certain threshold, we incorporate the current error computed for the homogeneous system as a new near-kernel prototype in an additional reduction-based AMG setup process to define a new method and, then, continue the iteration using this method. Heuristically, this method is motivated by the fact that each of these successive AMG methods removes certain components of the near-kernel, but fails to remove others. Incorporating the evolving error into a new AMG method yields an effective iteration for treating this error in the inhomogeneous system. To prohibit previously treated error from reappearing in the solution, we can cycle through a set of methods created in this non-linear adaptive process. In Figure 5.1, we provide a plot of the residual versus number of nonlinear adaptive AMG iterations applied to the Gauge Laplacian with $\beta = 1$ and $N = 64$. We note that the number of nonlinear iterations needed to reduce the

ETNA
Kent State University
http://etna.math.kent.edu

residual by 10^8 is again significantly less for this adaptive scheme than it is for the CG solver. Further, we mention that the iteration counts reported here are for the nonlinear solver applied as a stand alone solver, as opposed to a preconditioner to CG. Combining our nonlinear scheme with a *flexible* CG solver [27] will improve the performance of this method. Finally, we mention that our reported results are representative of the performance of this solver for varying problem sizes, shifts and configurations of the gauge field.

6. Concluding remarks. In this paper, we analyze and develop an adaptive reduction-based AMG algorithm for highly disordered nearly singular systems encountered in gauge theories discretized on a lattice. We provide two-level convergence theory for AMGr-type methods for HPD matrices. Using this theory, we develop practical measures and tools for constructing an effective MG method for such systems. Further, we explore variants of this adaptive AMGr process for a simplified two-dimensional Gauge Laplacian system and show that these approaches can provide effective preconditioners in this setting. The reduction in iteration counts of our solver over CG, coupled with the low grid and operator complexities of this MG method that results from our chosen form of interpolation are, thus, expected to significantly improve time to solution for this Gauge Laplace system. Further, as the problem size increases, this improvement is expected to become even more dramatic.

Appendix A. Coarsening algorithms. Algorithm 1 describes the implementation of the CR coarsening strategy. Herein, E_i measures how slowly the $\mathcal{F}$-variable i converges. Algorithm 2 describes our implementation of the greedy coarsening strategy. We use $Adj(j) = \{i \neq j | a_{ij} \neq 0\}$ to denote the graph neighborhood of variable j in the graph associated to A.

Algorithm 1: Coarse-grid selection using compatible relaxation.

Input : A, e_0
Output: $\mathcal{F}, \mathcal{C}$
$\mathcal{F} = \{1, 2, \dots, n\}, \mathcal{C} = \emptyset, m = 0$;
Do k CR sweeps on $Ae = 0$, measure convergence rate $\mu_{CR} := \|e^{(k)}\| / \|e^{(k-1)}\|$;
while $\mu_{CR} > a$ *AND* $m < m_{max}$ **do**
    $E = \frac{1}{\max_i(e_i)} e$;
    $\mathcal{U} = \{i,\ |E_i| > 1 - a\}$;
    Compute maximal independent set $\mathcal{C}^*$ of $\mathcal{U}$;
    Update $\mathcal{C} = \mathcal{C} \cup \mathcal{C}^*$, $\mathcal{F} = \mathcal{F} \setminus \mathcal{C}^*$;
    $m = m + 1$;
    Do k CR sweeps on $Ae = 0$, measure convergence rate μ_{CR};

REFERENCES

[1] O. AXELSSON, *Iterative Solution Methods*, Cambridge University Press, Cambridge, 1994.
[2] A. BRANDT, *Multi–level adaptive techniques (MLAT) for singular–perturbation problems*, in Numerical Analysis of Singular Perturbation Problems, P. W. Hemker and J. J. H. Miller, eds., Academic Press, New York, 1979, pp. 53–142.
[3] A. BRANDT, *Algebraic multigrid theory: The symmetric case*, Appl. Math. Comput., 19 (1986), pp. 23–56.
[4] A. BRANDT, *General highly accurate algebraic coarsening*, Electron. Trans. Numer. Anal., 10 (2000), pp. 1–20.
http://etna.math.kent.edu/vol.10.2000/pp1-20.dir.
[5] A. BRANDT AND S. TA'ASAN, *Multigrid methods for nearly singular and slightly indefinite problems*, in Multigrid Methods II, W. Hackbusch and U. Trottenberg, eds., Lecture Notes in Math., Springer, Berlin, 1986, pp. 99–121.

Algorithm 2: Greedy coarsening.

```
Input  : A
Output: F, C
U = {1, 2, ..., n};
F = C = ∅;
Compute θ̂_i as in (4.3) for all i ∈ U;
for i = 1 to n do
    if θ̂_i ≥ θ then
        F = F ∪ {i}    U = U \ {i};
while U ≠ ∅ do
    Find j = argmin_{i∈U}{θ_i};
    C = C ∪ {j}    U = U \ {j};
    for i ∈ U ∩ Adj(j) do
        Update θ̂_i;
        if θ̂_i ≥ θ then
            F = F ∪ {i}    U = U \ {i};
```

[6] J. BRANNICK, *Adaptive algebraic multigrid coarsening strategies*, PhD thesis, Department of Applied Mathematics, University of Colorado at Boulder, 2005.
[7] J. BRANNICK, M. BREZINA, D. KEYES, O. LIVNE, I. LIVSHITS, S. MACLACHLAN, T. MANTEUFFEL, S. MCCORMICK, J. RUGE, AND L. ZIKATANOV, *Adaptive smoothed aggregation in lattice QCD*, in Domain Decomposition Methods in Science and Engineering XVI, O. B. Widlund and D. E. Keyes, eds., Lect. Notes Comput. Sci. Eng., Vol. 55, Springer, Berlin, 2007, pp. 505–512.
[8] J. BRANNICK, R. BROWER, M. CLARK, J. OSBORN, AND C. REBBI, *Adaptive multigrid algorithm for lattice QCD*, Phys. Rev. Lett., 100 (2008).
[9] J. BRANNICK AND L. ZIKATANOV, *Algebraic multigrid methods based on compatible relaxation and energy minimization*, in Domain Decomposition Methods in Science and Engineering XVI, O. B. Widlund and D. E. Keyes, eds., Lect. Notes Comput. Sci. Eng., Vol. 55, Springer, Berlin, 2007, pp. 15–26.
[10] M. BREZINA, R. FALGOUT, S. MACLACHLAN, T. MANTEUFFEL, S. MCCORMICK, AND J. RUGE, *Adaptive smoothed aggregation (αSA)*, SIAM J. Sci. Comput., 25 (2004), pp. 1896–1920.
[11] ———, *Adaptive smoothed aggregation (αSA) multigrid*, SIAM Rev., 47 (2005), pp. 317–346.
[12] ———, *Adaptive algebraic multigrid*, SIAM J. Sci. Comput., 27 (2006), pp. 1261–1286.
[13] Z. CAI, J. MANDEL, AND S. MCCORMICK, *Multigrid methods for nearly singular linear equations and eigenvalue problems*, SIAM J. Numer. Anal., 34 (1997), pp. 178–200.
[14] M. CREUTZ, *Quarks, Gluons and Lattices*, Cambridge Monographs on Mathematical Physics, Cambridge University Press, Cambridge, 1983.
[15] R. EDWARDS, *Numerical simulations in lattice gauge theories and statistical mechanics*, PhD thesis, Department of Physics, New York University, 1989.
[16] R. D. FALGOUT AND P. S. VASSILEVSKI, *On generalizing the AMG framework*, SIAM J. Numer. Anal., 42 (2004), pp. 1669–1693.
[17] R. D. FALGOUT, P. S. VASSILEVSKI, AND L. T. ZIKATANOV, *On two-grid convergence estimates*, Numer. Linear Algebra Appl., 12 (2005), pp. 471–494.
[18] W. HACKBUSCH, *Multigrid convergence for a singular perturbation problem*, Linear Algebra Appl., 58 (1984), pp. 125–145.
[19] P. W. HEMKER, *Numerical aspects of singular perturbation problems*, in Asymptotic Analysis II, F. Verhulst, ed., Lecture Notes in Math., Vol. 985, Springer, Berlin, 1982, pp. 267–287.
[20] K. KAHL, *An algebraic multilevel approach for a model problem for disordered physical systems*, Master's thesis, Bergische Universität Wuppertal, Fachbereich Mathematik und Naturwissenschaften, 2006.
[21] O. LIVNE, *Coarsening by compatible relaxation*, Numer. Linear Algebra Appl., 11 (2004), pp. 205–227.
[22] S. MACLACHLAN, *Improving robustness in multiscale methods*, PhD thesis, Department of Applied Mathematics, University of Colorado at Boulder, 2004.
[23] S. MACLACHLAN, T. MANTEUFFEL, AND S. MCCORMICK, *Adaptive reduction-based AMG*, Numer. Linear Algebra Appl., 13 (2006), pp. 599–620.

[24] S. MACLACHLAN AND C. OOSTERLEE, *Algebraic multigrid solvers for complex-valued matrices*, SIAM J. Sci. Comput., 30 (2008), pp. 1548–1571.

[25] S. MACLACHLAN AND Y. SAAD, *A greedy strategy for coarse-grid selection*, SIAM J. Sci. Comput., 29 (2007), pp. 1826–1853.

[26] S. F. MCCORMICK AND J. W. RUGE, *Multigrid methods for variational problems*, SIAM J. Numer. Anal., 19 (1982), pp. 924–929.

[27] Y. NOTAY, *Flexible conjugate gradients*, SIAM J. Sci. Comput., 22 (2000), pp. 1444–1460.

[28] ———, *Algebraic multigrid and algebraic multilevel methods: A theoretical comparison*, Numer. Linear Algebra Appl., 12 (2005), pp. 419–451.

[29] A. REUSKEN, *Multigrid with matrix dependent transfer operators for a singular perturbation problem*, Computing, 50 (1993), pp. 199–211.

[30] M. RIES, U. TROTTENBERG, AND G. WINTER, *A note on MGR methods*, Linear Algebra Appl., 49 (1983), pp. 1–26.

[31] J. W. RUGE AND K. STÜBEN, *Algebraic multigrid (AMG)*, in Multigrid Methods, S. F. McCormick, ed., Frontiers in Applied Mathematics, Vol. 3, SIAM, Philadelphia, PA, 1987, pp. 73–130.

[32] P. VANĚK, M. BREZINA, AND J. MANDEL, *Convergence of algebraic multigrid based on smoothed aggregation*, Numer. Math., 88 (2001), pp. 559–579.

[33] 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.