Adaptive Smoothed Aggregation in Lattice QCD

James Brannick[1], Marian Brezina[1], David Keyes[3], Oren Livne[5], Irene Livshits[2], Scott MacLachlan[1], Tom Manteuffel[1], Steve McCormick[1], John Ruge[1], and Ludmil Zikatanov[4]

[1] Department of Applied Mathematics, University of Colorado at Boulder, Email: brannick@newton.colorado.edu, mbrezina@math.cudenver.edu, maclachl@colorado.edu, tmanteuf@colorado.edu, stevem@colorado.edu, jruge@colorado.edu
[2] Department of Mathematical Sciences, Ball State University, Email: ilivshits@bsu.edu
[3] Department of Applied Physics and Applied Mathematics, Columbia University, Email: david.keyes@columbia.edu
[4] Department of Mathematics, The Pennsylvania State University, Email: ludmil@psu.edu
[5] SCI Institute, University of Utah, Email: livne@sci.utah.edu

Summary. The linear systems arising in lattice QCD pose significant challenges for traditional iterative solvers. For physically interesting values of the so-called quark mass, these systems are nearly singular, indicating the need for efficient preconditioners. However, multilevel preconditioners cannot easily be constructed because the Dirac operator associated with these systems has multiple near-kernel components that need to be approximated. Moreover, these components are generally both oscillatory and not known a priori. This paper presents the application of adaptive smoothed aggregation, αSA [2], to this Dirac system. Heuristic arguments and numerical results are provided to demonstrate that this recently developed extension of the smoothed aggregation methodology can be used to overcome the challenges posed by the Dirac system.

1 Introduction

In the field of particle physics, the so-called "Standard" model accounts for the interactions between the elementary particles that make up matter. The Standard model is completely described by two theories: the Electroweak theory for weak interactions, and the widely accepted Quantum Chromodynamics (QCD) theory for strong interactions. The interactions between these constituents of matter are well understood for Electroweak theory, where they can be analyzed analytically using perturbation theory. For strong interactions between fermions (quarks), the coupling forces are so strong that a perturbation theory analysis becomes increasingly complex and ultimately breaks down. In the early 1970's, Wilson proposed simulating these strong

interactions numerically using Lattice Gauge Theory (LGT), effectively discretizing QCD [4]. LGT is now the primary means for modeling such strong interactions. However, a major obstacle remains: current LGT simulations require enormous computations that become prohibitively expensive for physically interesting choices of parameters (e.g., quark mass and temperature of the physical system), even on today's supercomputers. Hence, the understanding of strong interactions is still very limited.

The majority of the computations in these numerical simulations is dedicated to solving the linear Dirac systems

$$H(\mathcal{U}_i)\mathbf{f} = \mathbf{b},$$

where H, a complex-valued matrix, is the fermionic propagator (Dirac Operator), $\mathcal{U}_i$ is a discrete configuration of the gauge field, and $\mathbf{f}$ denotes the fermionic field (located at sites on a four-dimensional space-time lattice). This is due to the fact that, in current state-of-the-art simulations, the solvers used are limited to Krylov methods with preconditioners that are suboptimal for interesting choices of the physical parameters [1, 3]. In addition, the system must be solved for various configurations, $\mathcal{U}_i$, and multiple vectors $\mathbf{b}$ for each given configuration.

Developing an appropriate preconditioner for these systems has been a topic of intense research for many years. In the 1990's, various multigrid approaches were explored [1, 3]. More recently, in [6], the use of an alternating Schwarz preconditioner was studied. While these efforts have lead to significant improvements, none of the approaches completely eliminate the so-called "critical slowing down" that results from the fact that the Dirac operator becomes increasingly ill conditioned as the masses of the quarks approach physically interesting values.

An additional difficulty associated with the Dirac operator is the oscillatory local character of its near-kernel (the so-called "algebraically smooth error" not reduced by relaxation). Standard multigrid preconditioners commonly rely in their design on geomtric smoothness of the near-kernel and, thus are not immediately suitable for the Dirac system.

This paper considers a simplified 2D Schwinger model exhibiting similar challenges to those of the four-dimensional problem of interest. We explore the use of an adaptive algebraic multigrid preconditioner for solving the corresponding system of equations. The method is a recently developed extension of smoothed aggregation [8], referred to as αSA [2], in which good convergence properties are achieved in situations where explicit knowledge of the near-kernel components may not be available. Preliminary results with our 2D model problems suggest that this approach yields typical multigrid-like performance with convergence factors bounded uniformly below one for realistic choices of quark mass and gauge-field randomness.

The remaining sections are organized as follows. In §2, we present the 2D Hermitian Dirac-Wilson Operator. In §3, we discuss the properties of the 2D operator. Numerical results demonstrating the effectiveness of our approach are given in §4. In §5, we give some concluding remarks.

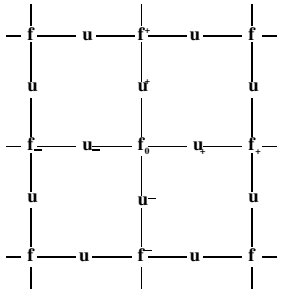

Fig. 1. Local numbering of the unknowns $\mathbf{f}$ and gauge field coefficients, $\mathbf{u}$, used in the definition of the discrete equations for H.

2 2D Hermitian Dirac-Wilson formulation

Following [7], we describe here the 2D Dirac-Wilson operator, H, and the attendant system of linear equations,

$$H(\mathbf{u})\mathbf{f} = \mathbf{b}. \tag{1}$$

Here the gauge variables, $\mathbf{u}$, are unitary, complex, and scalar valued; thus, values of $\mathbf{u}$ are given as $e^{i\theta}$, where phase θ is chosen randomly. The distribution of θ depends on the "temperature" of the physical system, prescribed by parameter β. For $\beta \to \infty$, corresponding to a cold temperature of the physical system, the distribution is *smooth* ($\lim_{\beta\to\infty} \mathbf{u} \equiv 1$). For realistic values of β, between 2 and 6, the system temperature is said to be hot, and the gauge field is random.

The domain of interest is a 2D periodic $N \times N$ uniform lattice (grid), where the lattice sites (gridpoints) are distance $h = 1$ apart. Larger systems are, thus, obtained by enlarging N, the length of the domain in both coordinate directions and not by moving lattice points closer together. The fermionic degrees of freedom are defined at the gridpoints on the lattice, while the gauge variables are defined on the lattice edges, as shown in Figure 1. At every gridpoint, x, the unknown function is a vector of length 2:

$$\mathbf{f}(x) = \begin{pmatrix} \mathbf{f}(x, s=1) \\ \mathbf{f}(x, s=2) \end{pmatrix} = \begin{pmatrix} \mathbf{v}(x) \\ \mathbf{w}(x) \end{pmatrix},$$

where $s = 1, 2$ are spin indices, with spin corresponding to angular momentum.

Let $\hat{m}$ be the unit vector in coordinate direction $m = 1, 2$ and $\mathbf{u}(x, m)$ be the gauge variable located at the link associated with gridpoints x and $x + \hat{m}$. Then the action of H is defined in terms of the covariant difference operators,

$$(\nabla_m^+ \mathbf{f})(x, s) = \mathbf{u}(x, m)\mathbf{f}(x + \hat{m}, s) - \mathbf{f}(x, s) \tag{2}$$

$$(\nabla_m^- \mathbf{f})(x, s) = \mathbf{f}(x, s) - \mathbf{u}(x - \hat{m}, m)^* \mathbf{f}(x - \hat{m}, s), \tag{3}$$

and the Pauli matrices,

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_3 = -i\sigma_1\sigma_2 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

The action of the Pauli matrices are denoted by $\sigma_m \mathbf{f}$. Explicitly,

$$(\sigma_m \mathbf{f})(x,s) = \sum_{s'} (\sigma_m)_{s,s'} \mathbf{f}(x,s').$$

With these definitions, H is defined implicitly by its action on $\mathbf{f}(x,s)$ at lattice site x:

$$(H\mathbf{f})(x,s) = \sigma_3 \Big[\sum_{m=1}^{2} \frac{1}{2h} \sigma_m (\nabla_m^+ + \nabla_m^-) \mathbf{f}(x,s) + \frac{1}{2h} \sum_{m=1}^{2} (-\nabla_m^+ + \nabla_m^-) \mathbf{f}(x,s) + \rho \mathbf{f}(x,s) \Big], \tag{4}$$

where ρ is the quark mass and $h = 1$. Here, h is included to emphasize the comparason to familiar matrices from PDEs, and will be used later in §3 to scale the matrix. Note that, due to the σ_3 term in (4), H is Hermitian and indefinite.

The corresponding system of linear equations can be written in the following 2×2 block form:

$$\begin{pmatrix} -\rho I - A & B \\ B^* & \rho I + A \end{pmatrix} \begin{pmatrix} \mathbf{v} \\ \mathbf{w} \end{pmatrix} = \begin{pmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{pmatrix}.$$

Operators $A, B \in C^{n \times n}$, $n = N \times N$, are defined by their actions on the components of $\mathbf{f}$, corresponding to spin indices $s = 1, 2$. For example, using the local numbering in Figure 1, the actions of A and B on $\mathbf{v}$ are given as follows:

$$(A\mathbf{v})_0 = \frac{1}{2h}(\overline{\mathbf{u}}_- \mathbf{v}_- + \mathbf{u}_+ \mathbf{v}_+ + \overline{\mathbf{u}}^- \mathbf{v}^- + \mathbf{u}^+ \mathbf{v}^+) - \frac{2}{h}\mathbf{v}_0, \tag{5}$$

$$(B\mathbf{v})_0 = \frac{1}{2h}(\mathbf{u}_+ \mathbf{v}_+ - \overline{\mathbf{u}}_- \mathbf{v}_- - i\mathbf{u}^+ \mathbf{v}^+ + i\overline{\mathbf{u}}^- \mathbf{v}^-). \tag{6}$$

The right-hand side, $\mathbf{b}$, is called the *fermionic source vector* and is equal to one at a given gridpoint x and is zero elsewhere.

3 Spectral properties of the Dirac system

Here, we analyze the properties of the 2D Hermitian Dirac-Wilson operator, H, for different values of ρ and β. In particular, we consider the dependence of the conditioning of H on ρ, recalling that, as the quark mass approaches its physical value, H becomes nearly singular.

We begin with an analysis of H in the absence of an external field (i.e., $\mathbf{u} \equiv 1$), referred to as the "free" case. For $\mathbf{u} \equiv 1$, the covariant difference operators defined in

(2) and (3) reduce to standard first-order forward and backward difference operators. It is easy to see that, for $u \equiv 1$, operators A and B in (5) and (6) have zero-row-sum. Hence, H is singular for $\rho = 0$, and H becomes nearly singular for $0 < \rho \ll 1$, the physical value of ρ for $\mathbf{u} \equiv 1$. Recall that operator H is indefinite. Considering a (preconditioned) conjugate gradient algorithm, the equivalent system of normal equations,

$$H^*H\mathbf{f} = H^*\mathbf{b}, \tag{7}$$

can be solved instead. Given an eigenpair, $(\lambda, \mathbf{x})$, of H, we have $H^*H\mathbf{x} = \lambda^2\mathbf{x}$, and, so, the normal form also becomes increasingly ill conditioned as $\rho \to 0$, since the maximum eigenvalue remains $O(1)$ for all ρ. Clearly, CG is inefficient as a stand-alone solver for this system. However, classical multigrid provides a suitable preconditioner in this "free" case, as can be easily seen by relating H^2 to a decoupled 2×2 system of partial differential equations (PDEs) as follows.

Consider the equivalent system, where the problem is rescaled such that $h \to 0$ on a fixed domain. For $\mathbf{u} \equiv 1$, $A = \frac{h}{2}\Delta_h$ and $B^*B = BB^* = -\Delta_h$, where Δ_h is the five-point discrete Laplacian obtained using second-order central differences. Thus, H^*H has the following 2×2 block diagonal form:

$$H^*H = \begin{pmatrix} -\rho I - A & B \\ B^* & \rho I + A \end{pmatrix}^2 = \begin{pmatrix} (\rho I + \frac{h}{2}\Delta_h)^2 - \Delta_h & 0 \\ 0 & (\rho I + \frac{h}{2}\Delta_h)^2 - \Delta_h \end{pmatrix}.$$

Denoting the diagonal blocks by C, for $\rho = 0$ we have $C = -\Delta_h(I - \frac{h^2}{4}\triangle_h)$. Since $\sigma(-\Delta_h) \subseteq [0, \frac{8}{h^2}]$, then $\sigma(I - \frac{h^2}{4}\Delta_h) \subseteq [1, 3]$, and we have that C is spectrally equivalent to $-\Delta_h$. The theory of equivalent preconditioning [5] then suggests that preconditioning C with a standard multigrid method for the Laplacian would be efficient. In practice, if we apply AMG preconditioned CG, we observe good solver performance for physical values of the quark mass (i.e., $0 < \rho \ll 1$) as well.

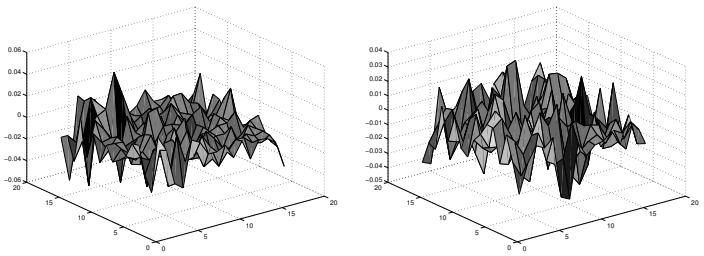

Fig. 2. Real (left) and complex (right) algebraically smooth error of H^2 for $m_q = .05, \beta = 3$ and $N = 16$. This error was computed using 200 Gauss-Seidel iterations on $H^2\mathbf{f} = \mathbf{0}$ with a random initial guess for $\mathbf{f}$.

We now focus on the more challenging case with the presence of the gauge field, when H^2 is no longer a block diagonal system. Unfortunately, in the presence of an

external field, where $\mathbf{u}$ is random (e.g., $\beta \in [2,6]$), H^2 does not appear to be related to a system of PDEs. Similar to the free case, H is indefinite and H^2 becomes ill conditioned as ρ approaches its so-called critical value. More precisely, take $\rho = 0$ for H defined in (4) and let $\hat{H} = \sigma_3 H$. $\hat{H}$ is then non-Hermitian and has eigenvalues with positive *real part*. The critical value of the quark mass, ρ_{cr}, is then defined by $\rho_{cr} = \min_i |\Re(\lambda_i(\hat{H}))|$. For physical values of ρ, the *mass gap*, $m_q := \rho - \rho_{cr}$, tends to zero, and H^2 becomes near-singular. This is the primary reason that all existing local algorithms grow in computational complexity for the Dirac system, a phenomenon referred to as *critical slowing down*.

An additional difficulty, not encountered in the free case, is that the local character of the near-kernel components is oscillatory. This is demonstrated in Figure 2, where plots of the real and imaginary parts of a near-kernel component, computed using 200 iterations of Gauss-Seidel on the homogeneous problem, are given. Our experiments indicate that this oscillatory local character of near-kernel components is dependent on the distribution of the gauge field, which is itself randomly specified.

In order to successfully solve the Dirac system for the random case, it is therefore imperative that our iterative solver be able to efficiently attenuate such error components. Standard geometric and algebraic multigrid methods typically construct coarse-level corrections based on the assumption that the error not effectively reduced by the multigrid relaxation procedure is locally constant or, in general, smooth in the geometric sense, and, thus, would not be immediately suitable as a solver for the random case.

Smoothed aggregation multigrid (SA) [8] was designed to allow efficient attenuation of error in a subspace characterized locally by a given set of error components, regardless of whether these are smooth or oscillatory in nature. The Dirac system poses an additional difficulty for the iterative solver, in that we do not have a priori knowledge of these near-kernel components. For this reason, we use the recently developed adaptive version of the smoothed aggregation multigrid method (αSA, [2]), which allows its setup procedure to identify the requisite error components and modify the method to ensure they can be efficiently eliminated. The αSA setup procedure is a multilevel scheme based on the power method for the error propagation operator of the method itself. In the interest of brevity, we refer for details of the method to [2].

4 Numerical results

In this section, we present numerical results for various solvers applied to the 2D Dirac system defined in §2. Instead of (1), we solve the equivalent normal system of equations, (7), reformulated as a 2×2 block *real* system,

$$\begin{bmatrix} X & -Y \\ Y & X \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} = \begin{bmatrix} \mathbf{c}_1 \\ \mathbf{c}_2 \end{bmatrix}, \tag{8}$$

where X, Y are real-valued matrices and satisfy $H^*H = X + iY$, $\mathbf{f} = \mathbf{x} + i\mathbf{y}$, and $H^*\mathbf{b} = \mathbf{c}_1 + i\mathbf{c}_2$.

Results of our numerical experiments are given in Table 1. The experiments were carried out on the problem with $\mathbf{b} = \mathbf{0}$ and random initial guesses for $[\mathbf{x}, \mathbf{y}]^T$. For the αSA preconditioner, we used $V(2,2)$-cycles with an SOR-type relaxation. We note that, due to the more aggressive coarsening used in smoothed aggregation multigrid, V(2,2) cycles are commonly used even when solving the Poisson equation. Eight near-kernel components were computed in the adaptive setup and used to define the transfer operators. For our αSA preconditioned CG results, one iteration of the resulting $V(2,2)$-cycle was used as a preconditioner in the conjugate gradient iteration. The Krylov solver used for comparison in Table 1 was a standard diagonally preconditioned conjugate gradient iteration.

In current state-of-the-art QCD simulations, typical problem sizes are $N = 32$ and 64. As already mentioned, the asymptotic convergence factors of existing solvers, applied to the Dirac system, quickly tend towards one for critical choices of mass and temperature. This is demonstrated in Table 1 for diagonally preconditioned CG. Further, the numerical results in Table 1 suggest that when using eight computed near-kernel components in defining interpolation, the asymptotic convergence factor of αSA remains uniformly bounded away from one for $\beta \in [2.0, 6.0]$ and $\rho \to \rho_{cr}$. This is the main result of this paper, and the first such result to date.

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. Even though we use eight near-kernel components to define prolongation, the operator complexity in our experiments stayed bounded by 3.0. These low values result from the fact that the problem size is aggressively reduced in forming the smoothed aggregation coarse problems.

One drawback of using the adaptive version of SA is the nontrivial cost associated with identifying the error components on which we base the transfer operators in SA. Of course, the hope is that this cost is optimal in that it is proportional to the number of degrees of freedom in the problem. Our experiments suggest that this is so. In addition, a QCD simulation requires solving the system of equations for many right-hand sides. The setup is performed only once, with the resulting method used to solve the system with many right-hand sides. We demonstrate that, even with the cost of the adaptive setup, the resulting method is more efficient than a diagonally preconditioned conjugate gradient algorithm when solving with a small number ($O(1)$) of right-hand sides. Indeed, for the experiments given in Table 1, only four right-hand sides need be solved to justify the cost of the adaptive setup for the smallest value of mass gap ($m_q = .01$), i.e., the most ill conditioned system. For $\beta = 3$ and $m_q = .01$, the CPU time required for the adaptive setup was 13.7 seconds and the CPU time needed to reduce the relative residual by a factor of 10^5 for a single right-hand side, using αSA preconditioned CG, was 0.8 seconds. Solving the same system using diagonally preconditioned CG required 4.7 sec CPU time.

Problem size 4096

β / m_q	.01	.05	.1	.3
2	.25 / .98	.27 /.96	.24 /.91	.22 /.83
3	.29 /.98	.27 /.94	.26 /.92	.27 /.84
5	.28 /.96	.29 /.95	.26 /.92	.25 /.81

Problem size 16384

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 preconditioned CG and diagonally preconditioned CG applied to (8) for two problem sizes and various choices of ρ and β.

5 Conclusions

Our experiments demonstrate that the αSA algorithm provides an efficient solver for the 2D lattice QCD problems considered. It remains to be verified whether the favorable convergence properties carry over to the full 4D case. Note that the cost of each iteration critically depends on the ratio of the number of degrees of freedom on the fine level to that on the coarse level. The coarsening for our 2D problem was very aggressive, leading to acceptable operator complexities even with eight adaptively computed near-kernel components.

Acknowledgments

The authors thank Rich Brower of the Boston University Physics Department for his many insightful comments on this work and for supplying the gauge variable data used in the numerical tests. This work was sponsored by the Department of Energy under grant numbers DE-FC02-01ER25479, DE-FC02-04ER25595 and DE-FG03-99ER25217, Lawrence Livermore National Laboratory under contract number B533502, Sandia National Laboratories under contract number 15268, and the National Science Foundation under grant number DMS-0410318.

References

1. A. Brandt. Multigrid methods in lattice field computations. *Nuclear Phys. B Proc. Suppl.*, 26:137–180, 1992.
2. M. Brezina, R. Falgout, S. MacLachlan, T. Manteuffel, S. McCormick, and J. Ruge. Adaptive smoothed aggregation (αSA). *SIAM J. Sci. Comp.*, 25(6):1896–1920, 2004.
3. R. Brower, R. Edwards, C. Rebbi, and E. Vicari. Projective multigrid for Wilson fermions. *Nucl. Phys.*, B366:689–709, 1991.
4. M. Creutz. *Quarks, Gluons and Lattices*. Cambridge Univ. Press, 1982.
5. V. Faber, T. Manteuffel, and S. Parter. On the theory of equivalent operators and application to the numerical solution of uniformly elliptic partial differential operators. *Advances in Applied Mathematics*, 11:109–163, 1989.
6. M. Lüscher. Lattice QCD and the Schwarz alternating procedure. Tech. Report, CERN, Theory Division, 2004.
7. Claudio Rebbi. Program hwilson2d. Code description. Boston University, October, 2003.
8. Petr Vaněk, Marian Brezina, and Jan Mandel. Convergence of algebraic multigrid based on smoothed aggregation. *Numerische Mathematik*, 88:559–579, 2001.