Adaptive Multigrid Algorithm for Lattice QCD

J. Brannick,[1] R. C. Brower,[2,3] M. A. Clark,[2] J. C. Osborn,[2] and C. Rebbi[2,3]

[1]*Department of Mathematics, The Pennsylvania State University, 230 McAllister Building, University Park, Pennsylvania 16802, USA*
[2]*Center for Computational Sciences, Boston University, 3 Cummington Street, Boston, Massachusetts 02215, USA*
[3]*Department of Physics, Boston University, 590 Commonwealth Avenue, Boston, Massachusetts 02215, USA*

We present a new multigrid solver that is suitable for the Dirac operator in the presence of disordered gauge fields. The key behind the success of the algorithm is an adaptive projection onto the coarse grids that preserves the near null space. The resulting algorithm has weak dependence on the gauge coupling and exhibits very little critical slowing down in the chiral limit. Results are presented for the Wilson-Dirac operator of the 2D U(1) Schwinger model.

The most demanding computational task in lattice QCD simulations consists of the calculation of quark propagators, which are needed both for generating gauge field configurations with the appropriate measure and for the evaluation of most observables. The calculation of a quark propagator, which in the course of a simulation must be carried out innumerous times with varying sources and gauge backgrounds, consists in turn of solving a very large system of linear equations,

$$D(U)\psi = \chi, \tag{1}$$

where ψ is the quark propagator, χ is the source term, and $D(U)$ is the discretized Dirac operator, with elements dependent on the gauge field background U.

In the language of applied mathematics, Eq. (1) is a discretized elliptic partial differential equation. For definiteness,

$$D_{x,y} = -\frac{1}{2} \sum_{\mu=1}^{d} [(1 - \gamma_\mu) U_x^\mu \delta_{x+\hat{\mu},y}$$
$$+ (1 + \gamma_\mu) U_{x-\hat{\mu}}^{\mu\dagger} \delta_{x-\hat{\mu},y}] + (2d + m)\delta_{x,y}$$

is the discretized Dirac operator describing a fermion in d dimensions with mass m in the Wilson discretization of the Dirac equation. In the full 4-dimensional QCD problem the matrices γ_μ are the 4×4 Dirac spin matrices and U is the SU(3) gauge field. The Wilson discretization is not the only one available, but it enters as a crucial ingredient in the chirality preserving "overlap" and "domain wall" discretizations [1–3]. Moreover, many of the problems encountered in solving the Wilson-Dirac equation extend to other formulations, such as the "staggered" fermion discretization. For these reasons, in this Letter we will concentrate on the Wilson discretization.

For any realistic QCD simulation the size of the matrix in Eq. (1) is too large for using a direct solver. Iterative Krylov-space methods, made possible by the sparsity of the matrix, must be used for calculating the propagators, and very efficient solvers have been developed. Yet, as the

system being considered grows in size [for forefront simulation on a 64^4 lattice, $D(U)$ is a $200M \times 200M$ complex matrix] and the quark mass in lattice units is brought toward zero, the condition number of the matrix increases rapidly and so does the computational cost of the solution.

In the field of applied mathematics it has been known for some time that in such circumstances the separation of physical length scales can be a very effective paradigm for improving the effectiveness of numerical algorithms. This paradigm has proven correct whether for evolving Monte Carlo processes, modeling chemical reactions, or molecular dynamics. This is especially true when it comes to solving systems of the form $Ax = b$, where A is the sparse matrix that arises from the discretization of continuum differential equations, b is a source vector, and x is the desired solution vector. For such systems the multigrid (MG) approach, where discretizations on successively coarser grids are used to accelerate the solution finding process, has proven to be the method to beat.

One exception to the above statement is in solving the Dirac operator in lattice QCD: here the nature of the underlying gauge field in the Dirac operator has proven to be especially resistant to various MG approaches. Previous attempts at MG solvers have relied on renormalization group arguments to define the coarse grids without realizing why the MG approach succeeds, and this has invariably led to failure as the physically interesting regime is approached [4,5]. In this Letter we demonstrate a MG algorithm for the Dirac operator normal equations, i.e., the positive definite operator given by

$$A = D^\dagger D,$$

that is shown to work in all regimes and vastly reduces the notorious critical slowing of the solver as the renormalized fermion mass is brought to zero. We do so in the context of a 2-dimensional system with U(1) gauge field (Schwinger model). This system captures many of the physical properties (confinement, chiral symmetry breaking, existence

of nontrivial topological sectors) of the more complex 4-dimensional QCD.

The original formulation of MG is best viewed with the example of the free Dirac operator. Multigrid solvers are based on the observation that relaxation methods (e.g., Jacobi, Gauss-Seidel, etc.) are only effective at reducing local error components leaving slow to converge, low wave-number components in the error. For the free Dirac operator these slow modes will be smooth and can be accurately represented on a coarser grid using simple linear averaging. However, on the coarse grid these low wave-number error modes have shorter range and so relaxation is effective at removing them. This process can recurse to coarser grids until the degrees of freedom have thinned enough to solve the system exactly. We then transfer this coarsest-level solution back to increasingly finer levels, where at each level we relax on the corrected residual to remove any high wave-number error components that are introduced. This process is known as a V cycle [6] and can be used as a solver in its own right, or more effectively as a preconditioner for a Krylov method.

To help facilitate our discussion we introduce the notation where the integer l represents the degree of coarseness: $l = 1$ represents the finest grid (i.e., where our actual problem is defined) and $l = N$ is the coarsest grid in an N-level MG algorithm. The operator used to promote a coarse grid vector on grid $l + 1$ to the adjacent fine grid l is known as the prolongator $P^{(l,l+1)}$, and it is convenient to take $P^{(l+1,l)} = P^{(l,l+1)\dagger}$ as the auxiliary restriction operator. [This guarantees Hermiticity of the coarse grid operator in Eq. (3).] Typically the Galerkin definition is used to define the coarse grid operator [6],

$$A^{(l+1)} = P^{(l,l+1)\dagger} A^{(l)} P^{(l,l+1)}. \qquad (2)$$

That this is the best definition for Hermitian positive definite A can be found by minimizing the error of the coarse grid corrected solution vector in the A norm.

Each level $l < N$ of an N-level V cycle acting on source $b^{(l)}$, returning trial solution vector $x^{(l)}$, is given by

(1) Relax on the input vector, $x_0^{(l)} = R^{(l)\dagger} b^{(l)}$, where $R^{(l)\dagger}$ is a suitable relaxation procedure. (The relaxation operator need not be Hermitian for the entire V cycle to be Hermitian: the postrelaxation operator need only be Hermitian conjugate to the prerelaxation operator.)

(2) The $l + 1$ source is the resulting residual restricted to this grid, $b^{(l+1)} = P^{(l,l+1)\dagger}(b^{(l)} - A^{(l)} x_0^{(l)})$.

(3) Apply the $l + 1$ V cycle on the coarse source vector, $e^{(l+1)} = V^{(l+1)} b^{(l+1)}$ ($V^{(N)}$ is an exact solve).

(4) Correct the current solution with coarse grid correction, $x_1^{(l)} = x_0^{(l)} + P^{(l,l+1)} e^{(l+1)}$.

(5) Relax on resulting residual, $x_2^{(l)} = R^{(l)}(b^{(l)} - A x_1^{(l)})$.

Written explicitly in terms of operators the lth level of the V cycle thus takes the following form:

$$V^{(l)} = R^{(l)} + R^{(l)\dagger} + R^{(l)} A^{(l)} R^{(l)\dagger}$$
$$+ [(1 - R^{(l)} A^{(l)}) P^{(l,l+1)} V^{(l+1)} P^{(l,l+1)\dagger} (1 - A^{(l)} R^{(l)\dagger})]. \qquad (3)$$

In this form the Hermiticity of the V cycle is obvious. The cost of applying the V cycle becomes apparent from this explicit form: on each level we must apply the operator $A^{(l)}$ a total of $2\nu + 2$ times for each l, where ν is the number of steps within the relaxation operator.

The problem in the early application of the above procedure to the interacting theory is that, in the presence of a nontrivial gauge field, the eigenvectors responsible for slow convergence are no longer low wave-number modes with smooth variation over the lattice. They are instead modes that exhibit localized lumps, typically extending over several lattice spacings. In such circumstances, trying to use smooth components of the fermion field, defined through a suitable gauge fixing or by some gauge covariant procedure, for the definition of the prolongator is bound to produce only a limited advantage. This is the method that was followed, e.g., in Refs. [4,5], where some acceleration was obtained but critical slowing down was not fully removed.

A breakthrough in the application of multiscale methods to more complex problems, such as the one at hand, has occurred with the discovery of adaptive MG techniques [7,8]. In the adaptive algorithm one lets the MG process itself define the appropriate prolongator by an iterative procedure which we now concisely describe.

In the first pass, one uses relaxation alone to solve the homogenous problem $Ax = 0$ with a randomly chosen initial vector e (for the homogeneous problem e is exactly the error vector). After a certain number, ν, of relaxation steps, the relaxation procedure, which we symbolically represent by

$$e \rightarrow e' = (I - \omega A)^\nu e \equiv (I - \omega D^\dagger D)^\nu e, \qquad (4)$$

produces an e' that essentially belongs to the space spanned by the near-zero modes of A, so e' is used to define a first approximation to the prolongator. One blocks the variables of the original lattice into subsets, which we denote by S_j. From e' we construct the vectors e'_j, which are identical to e' within S_j and 0 outside S_j, and the vectors of unit norm $v_{1j} = e'_j/|e'_j|$. The extra "1" index in v_{1j} has been introduced for a discussion that follows. The prolongator $P^{(1,2)} \equiv P_{ij}^{(1,2)}$ which maps a vector $\psi_j^{(2)}$ in the coarse lattice, indexed by j, to the original lattice, where i denotes collectively the site, spin, and possible internal symmetry indices, is then defined by

$$P_{i,j}^{(1,2)} = v_{1j,i}, \qquad (5)$$

where we have made explicit the fine lattice indices of i.

There are variations on how to block the fine lattice, i.e., how to define the sets S_j. In the so-called "algebraic

adaptive MG'' one partitions the fine lattice into subsets on the basis of the magnitude of the matrix elements of A. Since such matrix elements in lattice gauge theories are typically of uniform magnitude, differing rather in phase or, in a broader sense, in orientation within the space of gauge transformations, we chose instead to partition the lattice geometrically into fixed blocks of neighboring sites. Maintaining a regular lattice on coarse levels allows more efficient parallel code with exact load balancing.

A refinement of the technique consists of applying one further relaxation step [Eq. (4) with $\nu = 1$] to the vectors v_{1j} when defining the prolongator. Thus the final coarse grid operator takes the form

$$A^{(l+1)} = P^{(l,l+1)\dagger}(1 - \omega A^{(l)})^\dagger A^{(l)}(1 - \omega A^{(l)})P^{(l,l+1)}.$$

Here $\omega \sim \lambda_{\max}^{-1}$ to minimize the condition number of the resulting coarse grid operator, while retaining positivity ($\lambda_{\max}$ being the maximum eigenvalue of A). This additional step, known as "smoothed aggregation" drastically improves convergence properties at the expense of an increase in coarse operator complexity [7].

As with all MG methods, the procedure is recursive: to define the prolongator for subsequent coarsening, the vector $e^{(l+1)} = [(1 - \omega A^{(l)})P^{(l,l+1)}]^\dagger e'^{(l)}$ is the initial vector for relaxing upon the system $A^{(l+1)}x^{(l+1)} = 0$. From this we produce $v_{1j'}^{(l+1)}$ from which the prolongator $P^{(l+1,l+2)}$ is defined.

We come now to the crux of the adaptive MG method. We use the prolongator defined above to implement a standard V cycle and apply it (in place of relaxation) to the homogeneous problem, again with a randomly chosen error vector. There are two possibilities. Either the V cycle effectively reduces the error or some large error, e'', survives the cycle. In the first case, of course, one need not proceed: the MG procedure works as is. In the second case, we define another set of vectors v_{2j} by restricting e'' to the subsets S_j, making the new vectors orthogonal to the vectors v_{1j} and normalizing them to 1. The smoothed aggregation procedure is now applied to the set $v_{sj} \equiv (v_{1j}, v_{2j})$. A new prolongator is defined by projecting over these vectors

$$P_{i,sj}^{(1,2)} = v_{sj,i},$$

where the index s, now taking values $1, 2$, can be considered as an intrinsic index over the coarse lattice. This process is repeated at all levels of the V cycle.

The procedure described in the above paragraph is repeated as necessary, until the application of a V cycle reduces a random initial error to 0 without critical slowing down. The method works if critical slowing down is eliminated with a few iterations of the adaptive procedure. If this occurs with M vector sets, then the coarse lattice will carry M degrees of freedom per site.

In testing this algorithm for lattice QCD we generated quenched U(1) gauge field configurations on a 128×128 lattice with the standard Wilson gauge field action

$$S = \sum_{x, \nu < \mu} \beta \operatorname{Re} U_x^{\mu\nu} \equiv \sum_{x, \nu < \mu} \beta \operatorname{Re} U_x^\mu U_{x+\hat{\mu}}^\nu U_{x+\hat{\nu}}^{\mu\dagger} U_x^{\nu\dagger}$$

and periodic boundary conditions at $\beta = 6$ and $\beta = 10$ at a wide range of mass parameters. These values of β define gauge field correlation lengths $l_\sigma = 3.30$ and $l_\sigma = 4.35$, respectively, via the area law for the Wilson loop: $W \sim \exp[-A/l_\sigma^2]$. On these lattices, a fermion mass gap $\hat{m} = m - m_{\text{crit}} = 0.01$ corresponds to the pseudoscalar meson correlation lengths $\mu^{-1} = 6.4$ and $\mu^{-1} = 12.7$, respectively. (All quantities are expressed in lattice units.) In the 2-dimensional U(1) gauge theory, one can identify a gauge invariant topological charge Q, which in the continuum limit is proportional to the quantized magnetic flux flowing through the system. A gauge field with nonzero Q results in exactly real eigenvalues in the Dirac operator and, hence, as the mass gap is brought towards zero the condition number becomes infinite. Thus, it is important to test both trivial ($Q = 0$) and nontrivial ($Q \neq 0$) gauge fields.

We blocked the lattice into 4×4 blocks and implemented the adaptive MG procedure described above. (We used a 4×4 blocking over 2×2 to help regain sparsity of the coarse grid operator lost through smoothed aggregation.) We used a degree 2 polynomial smoother for our relaxation procedure, where the coefficients were chosen by running two iterations of an underrelaxed minimum residual solver ($\omega = 0.8$) and subsequently held fixed (hence, we have $R = R^\dagger$). The coarsening procedure was repeated twice maintaining $M = 8$ vectors in all coarsenings, down to an 8×8 lattice, over which the equations were solved exactly. For each gauge field we performed the setup procedure for the MG preconditioner for the lightest mass parameter only, and reused these null space vectors for the heavier masses. We used this constructed V cycle as
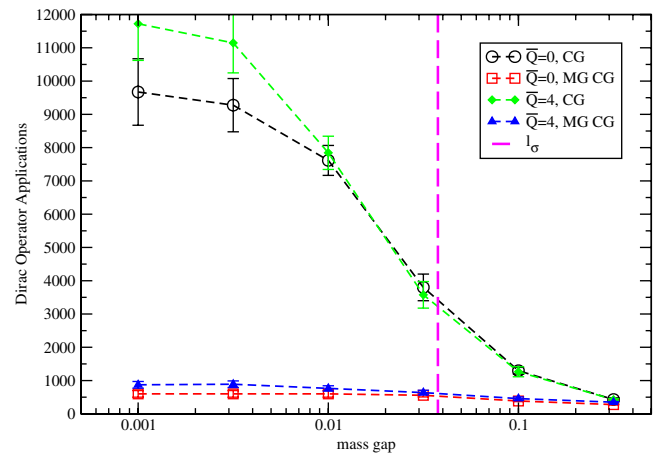

FIG. 1 (color online).   Number of Dirac operator applications of standard CG versus MG CG techniques as a function of the fermion mass gap at $\beta = 6$ with topological number $Q = 0$ and $Q = 4$ (average over 10 gauge fields, relative solver residual tolerance $|r| = 10^{-14}$).
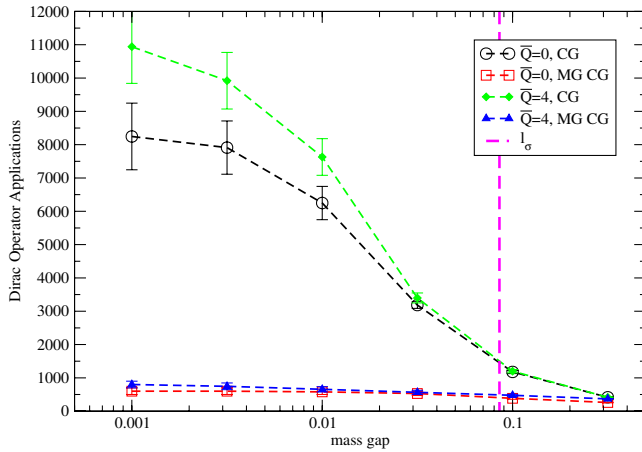
FIG. 2 (color online).   Number of Dirac operator applications of standard CG versus MG CG as a function of the fermion mass gap at $\beta = 10$ with topological number $Q = 0$ and $Q = 4$ (average over 10 gauge fields, relative solver residual tolerance $|r| = 10^{-14}$).

a preconditioner for the conjugate gradient (CG) technique where the operator defined in Eq. (4) is applied at each iteration to the CG direction vector.

If one compares the number of CG iterations needed to achieve convergence with or without MG preconditioning, the gain obtained with the MG method is dramatic: for example, with $\beta = 6$, $\hat{m} = 0.01$, and $Q = 0$, it takes 3808 iterations to achieve convergence, in the sense above, with a straightforward application of the CG technique, whereas it takes only 26 iterations using the MG CG technique. However, this comparison does not take into account the fact that many more operations per iteration must be performed when applying the MG preconditioner. To achieve a more balanced comparison, in Figs. 1 and 2 we plot the total number of applications of D and $D^\dagger$ done on the fine lattice. This reflects better the actual cost of the calculations (at each iteration of the MG CG method there are 6 applications of $D^\dagger D$: 1 application in the outer CG, and 2 precoarsening and 2 postcoarsening smoothing applications and 1 further application required to form the residual). The advantage from using the MG CG method is very dramatic: critical slowing down, if not totally eliminated, is very substantially reduced, even in the limit $\mu^{-1} \gg l_\sigma$. We note that there is a small increase in the cost of the MG CG algorithm when moving to nontrivial topology. We have not included the cost arising from the coarse lattices since this is expected to be a small overhead, but even given a doubling in cost, the MG CG algorithm is competitive.

From the point of view of computational complexity, one should also take into account the cost of setting up the MG preconditioner, i.e., of constructing the prolongator P. This cost is however heavily amortized, to the point of being negligible, if, as is often the case, one must apply the solver to systems with multiple given vectors [for example,

solving for all color and spin components of a quark propagator or, in the calculation of disconnected diagrams where $O(1000)$ inverses are required to estimate the trace of the inverse Dirac operator].

Our results, albeit for now limited to a 2-dimensional example, provide a clear indication that adaptive MG can be made to work with the lattice Dirac operator. What appears to be at the root of its success is that, although the modes responsible for slow convergence of the Dirac solver on a fine lattice are not low wave-number excitations, like in the free case, their span can be well approximated by a set of vectors of limited dimensionality on the blocks that define the coarse lattice. Earlier attempts [4,5] tried to find the approximating subspaces on the basis of smoothness, failing to eliminate critical slowing down when the pseudoscalar length exceeded the disorder length of the gauge field: $\mu^{-1} > l_\sigma$. Adaptive MG finds the coarse subspaces through the iterative application of the method itself. It is crucial that the approximation to the space of slow modes can be achieved with a small number of vectors on the individual blocks, otherwise the application of the method would not be cost effective: this appears to be the case in the examples we studied and, if the results hold true in general, adaptive MG has the potential to substantially speed up lattice QCD calculations as the increase of available computational power leads one to consider ever larger lattices. The observation that the space of slow modes may be of limited span is at the root of a method recently proposed by Lüscher in Ref. [9], although the technique there is quite different from the one we follow. The application of the method to 4-dimensional systems is in progress.

[1] D. B. Kaplan, Phys. Lett. B **288**, 342 (1992).
[2] Y. Shamir, Nucl. Phys. **B406**, 90 (1993).
[3] H. Neuberger, Phys. Lett. B **417**, 141 (1998).
[4] R. C. Brower, R. G. Edwards, C. Rebbi, and E. Vicari, Nucl. Phys. **B366**, 689 (1991).
[5] P. G. Lauwers and T. Wittlich, Int. J. Mod. Phys. C **4**, 609 (1993).
[6] A. Brandt, Math. Comput. **31**, 333 (1977).
[7] M. Brezina, R. Falgout, S. MacLachlan, T. Manteuffek, S. McCormick, and J. Ruge, SIAM J. Sci. Comput. **25**, 1896 (2004).
[8] J. Brannick, M. Brezina, D. Keyes, O. Livne, I. Livshits, S. MacLachlan, T. Manteuffel, S. McCormick, J. Ruge, and L. Zikatanov, in *Domain Decomposition Methods in Science and Engineering XVI*, Lecture Notes in Computational Science and Engineering Vol. 55, edited by O. B. Widlund and D. E. Keyes (Springer, New York, 2006), pp. 499–506.
[9] M. Lüscher, J. High Energy Phys. 07 (2007) 081.