PROCEEDINGS OF SCIENCE

Progress Report on Staggered Multigrid

Richard C. Brower
Boston University, Boston, MA 02215, USA
E-mail: brower@bu.edu

M. A. Clark
NVIDIA Corporation, Santa Clara, CA 95050, USA
E-mail: mclark@nvidia.com

Alexei Strelchenko
Scientific Computing Division, Fermilab, Batavia, IL 60510-5011, USA
E-mail: astrel@fnal.gov

Evan S. Weinberg[*]
Boston University, Boston, MA 02215, USA
E-mail: weinbe2@bu.edu

Multigrid methods have become the optimal solvers for Lattice QCD, first for the Wilson-clover discrete representation [1] and more recently for the domain wall formulation [2]. However, at present, the ensembles with the largest lattices use a third discretization, staggered. At the physical pion mass, propagator inversions with the staggered operator require $\mathcal{O}(10,000)$ iterations using the Conjugate Gradient algorithm. The staggered discretization raises new challenges for a multigrid implementation. It is a first order discretization which in four-dimensions corresponds to four copies of a Dirac fermion in the continuum. Here we report on our investigation into a new geometric adaptive multigrid for staggered fermions both for the normal equations and for a first order projection based on a novel blocking scheme stabilized by a second order gauged Laplacian. Current numerical tests, applied to the two-dimensional staggered representation of the two-flavor Schwinger model, are promising. Further improvements are under way as well as generalizations and tests for four-dimensional QCD using multigrid preconditioned staggered solvers in the QUDA library [3] for multi-GPU architectures.

PoS(LATTICE2016)273

34th annual International Symposium on Lattice Field Theory
24-30 July 2016
University of Southampton, UK

[*] Speaker.

1. Introduction

A major challenge facing the lattice field theory approach to quantum chromodynamics is the divergent increase in cost as one approaches the chiral limit required for the experimental values of the up and down quark masses. The cause is well known: as the fermion mass approaches zero, the Dirac operator becomes singular, causing "critical slowing down" of the standard Krylov solvers typically used to compute propagators. This is unavoidable for all single-grid solvers. Early progress of a gauge-invariant projective multigrid (MG) for LQCD goes back to the 1990s [4, 5] with modest results for weak gauge fields near the continuum. Only with the application of truly adaptive geometric MG algorithms [1, 6] was the ill-conditioning removed completely in the limit of zero quark mass.

In 4-d, the staggered discretization introduces a single component fermion (or Grassmann) variable on each lattice site which in the continuum reassembles into a so-called taste quartet of Dirac fermions. Similar lattice fermion representations occur in field theories for $\mathcal{N} = 4$ super symmetry and materials such as graphene. As with the development of both Wilson and domain wall MG algorithms, we begin by testing our methods on the 2-d Abelian analog to 4-d QCD, the Schwinger model, taking care to formulate the methods in general terms that we anticipate will apply to QCD in 4-d.

2. Staggered Fermions

The continuum Dirac PDE is a first order anti-Hermitian operator with real mass shift,

$$\gamma_\mu(\partial_\mu - iA_\mu)\psi(x) + m\psi(x) = b(x) \,. \tag{2.1}$$

The Dirac field ($\psi^{ia}(x)$) is four component tensor coupled to the four by four Dirac Hermitian spin matrix (γ^{ij}_μ) and three component color tensor coupled to the three by three Hermitian gauge field ($A^{ab}_\mu(x)$). The staggered fermion is a remarkable discretization which closely resembles the continuum by introducing a gauge-invariant central difference,

$$D(U,m)_{x,y} = -\frac{1}{2}\sum_{\mu=1}^{d} \eta_\mu(\vec{x})\left[U(x,x+\mu)\delta_{x+\mu,y} - U(x+\mu,x)\delta_{x,y+\mu}\right] + m\delta_{x,y} \,, \tag{2.2}$$

on a hypercubic lattice and replaces the spin matrix γ_μ by staggered ±1 signs: $\eta_\mu = (-1)^{\sum_{\nu<\mu} x^\nu}$. In 4-d QCD, the gauge variables, $U(x,x+\mu) = U^\dagger(x+\mu,x)$, defined on the link between $\langle x, x+\mu\rangle$, live in the $SU(3)$ group. Unlike the Wilson and domain wall methods, the staggered discretization preserves the continuum property of anti-Hermiticity up to a linear, real mass shift. The discrete operator (2.2) has an even-odd block structure, $D(U,m) = D_{eo} + D_{oe} + m$, which is responsible for the preservation of an exact chiral symmetry at zero mass. A drawback of the staggered discretization is the appearance of 2^d zero modes in the continuum at zero mass, referred to as "doublers." In 4-d, the 16 modes reassemble into four Dirac (four component) fermions or one Kahler-Dirac fermion. In 2-d, the four modes reassemble into two Dirac (two component) fermions. The 2-d staggered fermion coupled to an Abelian gauge theory, $U(x,x+\mu) = exp[i\theta_\mu(x)]$, is the two-flavor Schwinger model in the continuum, which has many features in common with QCD: confinement, topology, gauge invariance, to name a few [7]. Thus it provides an excellent prototype for testing and development of new MG methods.

3. Multigrid for the Staggered Fermions

For both Wilson and domain wall fermions the first MG constructions relied on the Galerkin projection into a near-null space for the normal matrix. For the Wilson operator the real breakthrough was to develop a Galerkin projection of the non-Hermitian Wilson Dirac matrix itself. We follow the same two step strategy for the staggered discretization.

3.1 Normal Equation Multigrid

Conventional staggered solvers use Conjugate Gradient (CG) on the even/odd preconditioned operator, which because the normal operator is block diagonal,

$$D^\dagger(U,m)D(U,m) = \begin{bmatrix} m & -D_{oe} \\ -D_{eo} & m \end{bmatrix} \begin{bmatrix} m & D_{oe} \\ D_{eo} & m \end{bmatrix} = \begin{bmatrix} m^2 - D_{eo}D_{oe} & 0 \\ 0 & m^2 - D_{oe}D_{eo} \end{bmatrix} \tag{3.1}$$

is equivalent to solving the normal equation, $D^\dagger(U,m)D(U,m)\psi = D^\dagger(U,m)b$, restricted to the even sub-lattice. In the free field limit ($U = 1$), the normal operator is composed of 2^d decoupled free Laplace operators with a doubling of the lattice spacing. This is illustrated on the left hand side of Figure 1 for 2-d. The red lines each carry a factor of -1 with the consequence that the "corner" terms cancel exactly. In a non-free field, these corner terms form the two link Γ_5 operator that measures the topological density $F_{12}(x)$ on the plaquette [8]. The normal operator is Hermitian with a positive-definite spectrum for $m > 0$, so the conventional Galerkin adaptive MG is on a sound theoretical basis.

Following the method of adaptive smooth aggregation algebraic MG (α SA MG), we construct near-null vectors ψ_0 by relaxing the system, $D^\dagger D\psi = 0$, starting from an initial random vector ψ and relaxing to a tolerance $|D^\dagger D\psi_0|/|D^\dagger D\psi|$ of 10^{-7} using CG. After computing null vectors, we decompose them into chiral pieces defined by the projectors, $P_\pm = \frac{1}{2}(1 \pm \Gamma_{55}(x))$, on to even and odd sub-lattices. The factor, $\Gamma_{55}(x) = (-1)^{x^1+\cdots x^d}$, is the unbroken staggered chiral symmetry matrix. Subsequently the near-null vectors are decomposed into blocks and locally orthonormalized to form the prolongator, $P_{x,\hat{x}}$, and the interpolator, $R = P^\dagger$, where x is a fine coordinate and $\hat{x}$ is a coarse coordinate. Using the Galerkin prescription, the first level coarse operator is

$$(\widehat{D^\dagger D})_{\hat{x},\hat{y}} = R_{\hat{x},x}\left(D^\dagger D\right)_{x,y} P_{y,\hat{y}}. \tag{3.2}$$

In practice we have chosen 20 near-null vectors (10 even and 10 odd) and a coarsening by eight by eight blocks from the finest level.

The coarse operator becomes a preconditioner for the fine operator. The preconditioning step consists of three pieces: (1) A pre-smoother, where we construct and then relax on the current residual. (2) Restricting the smoothed residual, solving the coarse system to some tolerance (in practice 0.2), then prolongating the coarse error to the fine grid. (3) Post-smoothing on the current error, which is the combination of the error from pre-smoothing and the prolonged coarse error. Because the coarse solve is not stationary, we use GCR as a flexible outer solver. For consistency with MG on D, discussed later, we use GCR(8) as a smoother. The overall prescription for null vector generation and preconditioning with a coarse system can be employed recursively to develop a multi-level algorithm. We have tested a four-level algorithm (three coarse levels) on the normal equation, using a block size of 8^2 on the top level and 2^2 on all subsequent levels.

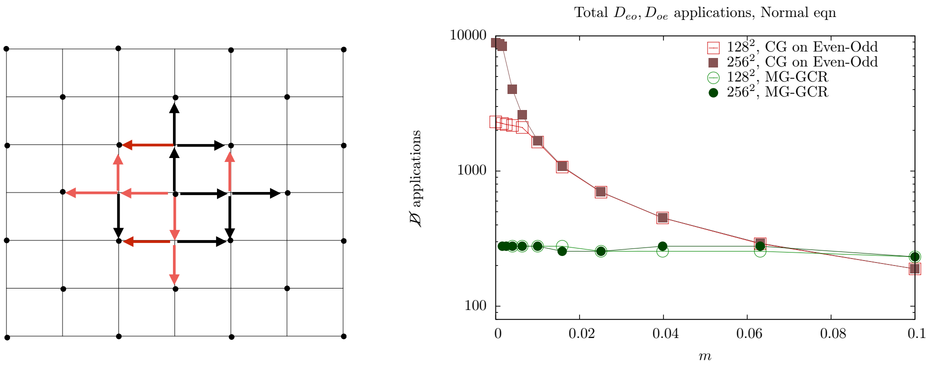

Figure 1: On the left is the stencil for the 2-d normal operator connecting even to even sites. The red arrows carry a factor of -1. On the right is a comparison of the MG convergence rate measured in units of fine $D(U,m)$ applications. Convergence is defined by solving to a tolerance of 10^{-12}.

A successful MG algorithm should eliminate critical slowing down both as a function of the fermion mass and of the volume. We demonstrate this on the right hand side of Figure 1. The red curves consider the canonical method of solving the staggered linear system, solving the even-odd preconditioned system described by the operator $m^2 - D_{eo}D_{oe}$, which diverges as $1/m$, ignoring finite volume effects. The green curves count the number of applications of the fine operator. The relative mass independence indicates the removal of critical slowing down. This algorithm works recursively: there are $\mathscr{O}(50)$ applications of the first coarse operator per outer fine iteration independent of mass and volume. The application of this algorithm to worth extending to 4-d QCD, where it is guaranteed to outperform the conventional algorithm asymptotically.

3.2 Spurious Modes in First Order Projected Multigrid

With the success of staggered MG for the normal operator, it is natural to try the same Galerkin projection on the first order operator $D(U,m)$ itself. This method produces an effective two-level algorithm with the fine operator, D, and coarse operator, $\widehat{D} = P^\dagger DP$. However we found a second coarse operator, $\widehat{\widehat{D}}$ fails to precondition $\widehat{D}$. We have traced this problem to the appearance of spurious small eigenvalues introduced into the first coarse operator.

On the left hand side of Figure 2, we compare the spectrum of the fine operator, D, and two further Galerkin projections, $\widehat{D}$ and $\widehat{\widehat{D}}$. Close to zero, the fine operator has only a few small eigenvalues. These eigenvalues are well preserved in the coarse operator, as noted by the near-perfect overlap between the two sets of eigenvalues. However, there are a plurality of "spurious" small eigenvalues which do not correspond to low modes of the original operator. These spurious eigenvalues spell disaster for a recursive algorithm: the null space generated by $\widehat{D}$ is predominantly spanned by these spurious low eigenvalues, so the still coarser operator $\widehat{\widehat{D}}$ is not a good preconditioner.

3.3 Decoupling Spurious Eigenvalues

To solve the problem of spurious low eigenvalues, we depart from the Galerkin prescription

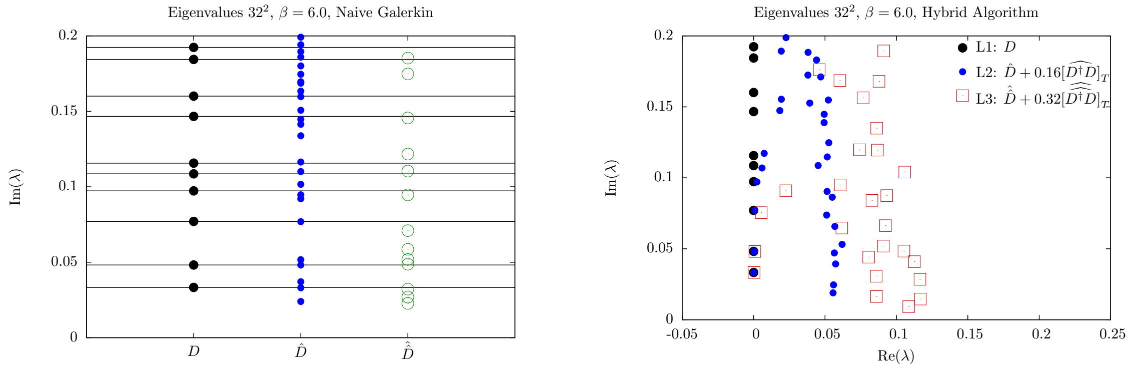

Figure 2: On the left, the imaginary part of the staggered D spectrum using a naïve Galerkin projection. $\widehat{D}$ preserves the low modes well but also introduces spurious small eigenvalues. $\widehat{\widehat{D}}$ fails to preserve the original low modes. On the right, the low spectrum of the staggered D spectrum and the stabilized coarse operators with corner term truncation. True low modes are well preserved on each level, while spurious eigenvalues are separated down the real axis.

by defining the coarse operator $\widehat{D} = P^\dagger D_T(U,m,w)P$ in terms of a stabilized fine operator,

$$D_T(U,m,w) \;=\; D(U,m) + w[D^\dagger(U,0)D(U,0)]_T = \begin{bmatrix} m - wD_{eo}D_{oe} & D_{eo} \\ D_{oe} & m - wD_{oe}D_{eo} \end{bmatrix}_T , \tag{3.3}$$

by adding a second order term with w a tunable parameter and suitable truncation (indicated by $[\cdots]_T$) described below. Using the full normal operator without truncation would guarantee that all of the fine eigenvectors are preserved but the complex eigenvalues are shifted, $\pm i\lambda + m \to \pm i\lambda + w\lambda^2 + m$. For the free field case the spectrum is exactly

$$\lambda(\vec{p};m,w) = m \pm i\sqrt{\sin^2\left(\frac{p_x}{2}\right) + \sin^2\left(\frac{p_y}{2}\right)} + w\left[\sin^2\left(\frac{p_x}{2}\right) + \sin^2\left(\frac{p_y}{2}\right)\right] . \tag{3.4}$$

Similar to the classic Wilson term, this new term destroys the remaining staggered symmetry, generating an additive shift. This is not an issue for the low modes: in practice, the real contribution $w\lambda^2$ is $\mathscr{O}(10^{-4})$ on a 32^2 configuration, 100 times smaller than the mass where finite volume effects become prevalent. This shift becomes even smaller at larger volumes. The benefit of the addition of the normal operator is that it deforms the high spectrum away from the imaginary axis on the complex plane: eigenvalues with large imaginary part are *quadratically* shifted in the positive real direction. This method may seem like a fools errand: the point of a MG algorithm for D was to simplify the stencil by avoiding the two-link "corner terms." To achieve this we apply a truncation procedure that removes the few corner terms at the edge of each block on the *fine* level which contribute to corner terms on the first *coarse* level. This keeps the coarse stencil *nearest-neighbor*. The truncation commutes with the Galerkin projections so that we may write the new coarse operator as the stabilizing Galerkin projection followed by a truncation,

$$\widehat{D(U,m)} \to \widehat{D(U,m)} + w[P^\dagger D^\dagger(U,0)D(U,0)P]_T \; . \tag{3.5}$$

For conciseness we will denote the stabilized coarse operator as $\widehat{D}_T = \widehat{D} + w[\widehat{D^\dagger D}]_T$. The majority of the two-link terms in the fine $D^\dagger D$ still contribute to the coarse operator as they are internal to each block. For example, in our 2-d method, there are four such terms at the corners on each coarse block relative to a total of $4 \times 8^2 = 256$ next-to-nearest-neighbor terms in the normal operator.

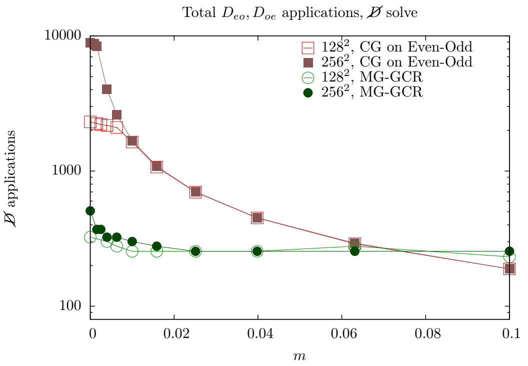

Figure 3: Convergence to a tolerance of 10^{-12} in terms of the number of fine D applications.

Deferring details on null vector generation to the next section, the effect of the truncated normal operator on the coarse spectrum can be noted in the right hand side of Figure 2. The low eigenvalues of the fine operator D (L1 in Figure 2) are well preserved by the coarse operator with the truncated normal contribution, $\widehat{D} + 0.16[\widehat{D^\dagger D}]_T$ (L2), without spurious low modes near zero. The clustered line of eigenvalues closer to $\mathrm{Re}(\lambda) = 0.05$ are the spurious eigenvalues. These do not ruin a three-level algorithm because they are shifted along the real axis.

Before discussing our results, we make a remark about a further recursive algorithm. While the normal operator contribution shifts the spurious eigenvalues along the real axis, these spurious eigenvalues can "re-collapse" towards zero in the real and imaginary plane on subsequent coarsenings. One method to avoid this issue is to rescale w on each level. This can be achieved inexpensively because chirality for staggered fermions is exactly even and odd, which is preserved in our projection, and thus we can identify which components of the coarse self-interaction and hopping terms come from projecting D as opposed to $D^\dagger D$. The effect of rescaling w can be noted in the right hand side of Figure 2 via the points from the third level, $\widehat{\widehat{D}} + 0.32[\widehat{\widehat{D^\dagger D}}]_T$ (L3).

3.4 Numerical Tests of Stabilized Staggered Multigrid

In testing our hybrid staggered MG algorithm, we largely follow our prescription for the normal operator but with a few modifications owing to not having a Hermitian positive-definite operator. We generate null vectors by relaxing with BiCGstab-ℓ [9] with $\ell = 6$ to a tolerance of 5×10^{-5}, a deviation from the normal operator. On the top level we generate null vectors using D even though we coarsen $D + 0.16[D^\dagger D]_T$. We use GCR(8) at each level for both pre- and post-smoothing; this relatively large number can be understood by noting we are not performing a canonical Galerkin projection from level to level.

Analogous to the case for the normal equation, in Figure 3 we show the number of fine-level $D(U,m)$ applications for an MG solve. In this case, we are only applying D and not the full normal operator. Again, we see a near complete independence in the number of D applications with both mass and volume. Similar to MG on the normal operator, the method works recursively, with $\mathscr{O}(50)$ applications of the first coarse operator per outer iteration. We again emphasize that this new coarse operator has the potential to be more efficient because it is a *nearest-neighbor* stencil. This anticipated boost in efficiency will come from a reduction in memory traffic, floating point operations, and in a multi-node code, network traffic.

4. Conclusions

In this proceedings we have demonstrated two promising directions of study for MG preconditioners to the staggered operator in four dimensions. The first method, in line with traditional literature on adaptive MG algorithms, is a Galerkin projection directly on the normal operator. The major drawback for an MG algorithm on the normal operator is the extended stencil. Compared to 2-d, with four corner terms or next-to-nearest-neighbor terms, in 4-d there are 24 corner terms, implying more floating-point operations and network traffic. This motivated our second, perhaps more interesting method: stabilized staggered MG, where we coarsen the staggered operator with a stabilizing normal operator contribution, then truncate the corner terms. We believe the introduction of truncated and/or stabilized modifications to the Galerkin projection represents a new method worthy of careful study for the 4-d staggered MG solver for lattice field theory. We have developed all of the basic software tools within the QUDA library [3] for GPUs to investigate our approach to staggered MG and test its efficiency on the largest available lattice lattices which at the current time are up to $144^3 \times 288$, with the prospect of doubling this linear size when Exascale hardware is in production.

Acknowledgements: We would like to thank Peter Boyle, Andreas Frommer, Matthias Rottmann, Martin Lüscher and James Osborn for useful conversations. R.C.B. and E.S.W. acknowledge support by DOE grant DE-SC0015845. Work by A.S. was done for Fermi Research Alliance, LLC under Contract No. DE-AC02-07CH11359.

References

[1] J. Brannick, R. C. Brower, M. A. Clark, J. C. Osborn, and C. Rebbi, *Adaptive Multigrid Algorithm for Lattice QCD*, *Phys.Rev.Lett.* **100** (2008) 041601, [arXiv:0707.4018].

[2] P. A. Boyle, *Hierarchically deflated conjugate gradient*, arXiv:1402.2585.

[3] http://lattice.github.com/quda, 2011.

[4] A. Hulsebos, J. Smit, and J. C. Vink, *Multigrid inversion of the staggered fermion matrix*, *Nucl. Phys. Proc. Suppl.* **20** (1991) 94–97.

[5] R. C. Brower, R. G. Edwards, C. Rebbi, and E. Vicari, *Projective multigrid for Wilson fermions*, *Nucl. Phys.* **B366** (1991) 689–705.

[6] R. Babich, J. Brannick, R. C. Brower, M. A. Clark, T. A. Manteuffel, et al., *Adaptive multigrid algorithm for the lattice Wilson-Dirac operator*, *Phys.Rev.Lett.* **105** (2010) 201602, [arXiv:1005.3043].

[7] D. H. Adams, *Theoretical foundation for the Index Theorem on the lattice with staggered fermions*, *Phys. Rev. Lett.* **104** (2010) 141602, [arXiv:0912.2850].

[8] J. Smit and J. C. Vink, *Topological Charge and Fermions in the Two-dimensional Lattice U(1) Model. 1. Staggered Fermions*, *Nucl. Phys.* **B303** (1988) 36–56.

[9] G. L. Sleijpen and D. R. Fokkema, *Bicgstab (l) for linear equations involving unsymmetric matrices with complex spectrum*, *Electronic Transactions on Numerical Analysis* **1** (1993), no. 11 2000.