Progress Report on Staggered Multigrid

Evan Weinberg

Boston University, Boston, MA

weinbe2@bu.edu

in collaboration with Richard Brower, Kate Clark, Alexei Strelchenko

July 28, 2016

Lattice 2016

LATTICE 2016

34th International Symposium on Lattice Field Theory, 24-30 July 2016

Outline

- The Battleground: $\not{D}$ inversions and critical slowing down
- The Target: The staggered operator
- Advanced Skills: Adaptive algebraic MG
- Training Room: Solving the free case
- First Round: Two dimensional two flavor Schwinger Model
- World Tour: Conclusions and future work

Motivation

- Push to exascale enables increasingly accurate lattice calculations.
- Physical pion mass, finer lattices: critical slowing down.
  - MILC: $144^3 \times 288$, physical pion mass, single precision multi-mass solve to rel. resid. 10^{-6}: about 25,000 iterations.
- New measurements (disconnected diagrams) require more $\not{D}$ inversions.

Motivation

- Push to exascale enables increasingly accurate lattice calculations.
- Physical pion mass, finer lattices: critical slowing down.
  - MILC: $144^3 \times 288$, physical pion mass, single precision multi-mass solve to rel. resid. 10^{-6}: about 25,000 iterations.
- New measurements (disconnected diagrams) require more $\not{D}$ inversions.

Multigrid methods can completely eliminate critical slowing down.

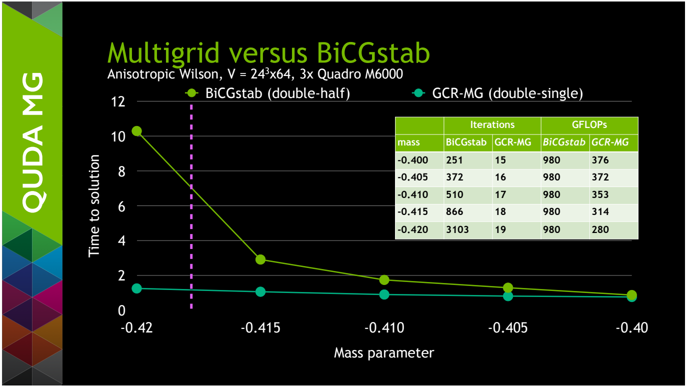
History

MG for the Wilson-Clover operator has a rich history:

[Phys.Rev.Lett. 105 (2010): R. Babich, J. Brannick, R.C. Brower, M.A. Clark, T.A. Manteuffel, S.F. McCormick, J.C. Osborn, C. Rebbi]

MG for the Wilson-Clover operator has a rich history:

[Phys.Rev.Lett. 105 (2010): R. Babich, J. Brannick, R.C. Brower, M.A. Clark, T.A. Manteuffel, S.F. McCormick, J.C. Osborn, C. Rebbi]



	Iterations		GFLOPs	
mass	BiCGstab	GCR-MG	*BiCGstab*	*GCR-MG*
-0.400	251	15	980	376
-0.405	372	16	980	372
-0.410	510	17	980	353
-0.415	866	18	980	314
-0.420	3103	19	980	280

Figure : QUDA-MG update, February 2016

There has been progress on multi-level Domain wall algorithms as well:

- Multigrid on the normal operator: [PoS LATTICE2011 (2011): S. Cohen, R.C. Brower, M.A. Clark, J.C. Osborn]
- Hierarchically deflated conjugate gradient: [arXiv 1402.2585: P.A. Boyle]

There has been progress on multi-level Domain wall algorithms as well:

- Multigrid on the normal operator: [PoS LATTICE2011 (2011): S. Cohen, R.C. Brower, M.A. Clark, J.C. Osborn]
- Hierarchically deflated conjugate gradient: [arXiv 1402.2585: P.A. Boyle]
- Hierarchically deflated conjugate residual: [Lattice 2016 Poster: A. Yamaguchi, P.A. Boyle.]

There has been progress on multi-level Domain wall algorithms as well:

- Multigrid on the normal operator: [PoS LATTICE2011 (2011): S. Cohen, R.C. Brower, M.A. Clark, J.C. Osborn]
- Hierarchically deflated conjugate gradient: [arXiv 1402.2585: P.A. Boyle]
- Hierarchically deflated conjugate residual: [Lattice 2016 Poster: A. Yamaguchi, P.A. Boyle.]

Multigrid is needed for Staggered fermions!

Staggered State of the Art

Given the staggered $\not{D}$ operator:

$$D_{xy} = \sum_{\mu} \eta_\mu(x) \left[U_\mu^\dagger(x)\delta_{y+\mu,x} - U_\mu(x)\delta_{x+\mu,y} \right] + m\delta_{x,y}$$
$$= iA + m\mathbf{I} \quad \leftarrow \text{Anti-Hermitian + Hermitian piece}$$

Staggered State of the Art

Given the staggered $\not{D}$ operator:

$$D_{xy} = \sum_{\mu} \eta_\mu(x) \left[U_\mu^\dagger(x) \delta_{y+\mu,x} - U_\mu(x) \delta_{x+\mu,y} \right] + m\delta_{x,y}$$
$$= iA + m\mathbf{I} \quad \leftarrow \text{Anti-Hermitian + Hermitian piece}$$

Perform an even-odd decomposition; $D_{eo} = -D_{oe}^\dagger$

$$D\psi = b \rightarrow \qquad \rightarrow \begin{bmatrix} m & D_{eo} \\ D_{oe} & m \end{bmatrix} \begin{bmatrix} \psi_e \\ \psi_o \end{bmatrix} = \begin{bmatrix} b_e \\ b_o \end{bmatrix}$$

$$\underbrace{\left(m^2 - D_{eo} D_{oe}\right)}_{\text{Normal, solve with CG!}} \psi_e = m b_e - D_{eo} b_o; \qquad \psi_o = \underbrace{\frac{1}{m}\left(b_o - D_{oe}\psi_e\right)}_{\text{Reconstruct odd}}$$

Efficiency Moving Forward

The E/O preconditioned operator is Hermitian positive definite:
Rich theory of multigrid exists.

$$\left(m^2 - D_{eo}D_{oe}\right)\psi_e = mb_e - D_{eo}b_o; \qquad \psi_o = \frac{1}{m}\left(b_o - D_{oe}\psi_e\right)$$

Efficiency Moving Forward

The E/O preconditioned operator is Hermitian positive definite:
Rich theory of multigrid exists.

$$\left(m^2 - D_{eo}D_{oe}\right)\psi_e = mb_e - D_{eo}b_o; \qquad \psi_o = \frac{1}{m}\left(b_o - D_{oe}\psi_e\right)$$

Issues:

- Two link sparsity pattern is computationally inefficient.
- Further issue: Even and odd decouple in chiral limit.

Efficiency Moving Forward

The E/O preconditioned operator is Hermitian positive definite:
Rich theory of multigrid exists.

$$\left(m^2 - D_{eo}D_{oe}\right)\psi_e = mb_e - D_{eo}b_o; \qquad \psi_o = \frac{1}{m}\left(b_o - D_{oe}\psi_e\right)$$

Issues:

- Two link sparsity pattern is computationally inefficient.
- Further issue: Even and odd decouple in chiral limit.

Follow lead of Wilson-Clover MG:
Directly precondition $\not{D}$.

Adaptive Algebraic Multigrid

Credit where credit is due: [QCDNA 2008]

Description shamelessly stolen from Kate Clark's 2008 QCDNA slides!

- Algorithm setup for 2-level V-Cycle:
  1. Relax on homogeneous problem $\not{D}\vec{x}_i = 0$ for N_{vec} random $\vec{x}_0$. (Each vector = extra degree of freedom per coarse site.)
  2. Cut vectors $\vec{x}_i$ into geometrically regular subsets to be aggregated (blocked)
  3. Block orthonormalize vectors $\vec{x}_i$
  4. This defines the prolongator such that $\left(1 - PP^\dagger\right)\vec{x}_i = 0$
  5. Define coarse grid operator $\not{D}_c = P^\dagger \not{D} P$

Chiral Blocks

Similar to Wilson-Clover, we do not block in chirality.
For every null vector $\vec{x}_i$, split into two vectors:

$$\vec{x}_i \rightarrow (1+\gamma_5)\vec{x}_i \quad (1-\gamma_5)\vec{x}_i$$

Chiral Blocks

Similar to Wilson-Clover, we do not block in chirality.
For every null vector $\vec{x}_i$, split into two vectors:

$$\vec{x}_i \rightarrow (1+\gamma_5)\vec{x}_i \quad (1-\gamma_5)\vec{x}_i$$

Unique to staggered: $\gamma_5^{stag} \equiv \epsilon(x) = (-1)^{\sum_\mu x_\mu}$
$(1 \pm \gamma_5)$ is just even/odd!!

Chiral Blocks

Similar to Wilson-Clover, we do not block in chirality.
For every null vector $\vec{x}_i$, split into two vectors:

$$\vec{x}_i = \begin{pmatrix} \vec{x}_i^e \\ \vec{x}_i^o \end{pmatrix} \rightarrow \begin{pmatrix} \vec{x}_i^e \\ \vec{0} \end{pmatrix} \quad \begin{pmatrix} \vec{0} \\ \vec{x}_i^o \end{pmatrix}$$

Unique to staggered: $\gamma_5^{stag} \equiv \epsilon(x) = (-1)^{\sum_\mu x_\mu}$
$(1 \pm \gamma_5)$ is just even/odd!!

Chiral Blocks

Similar to Wilson-Clover, we do not block in chirality.
For every null vector $\vec{x}_i$, split into two vectors:

$$\vec{x}_i = \begin{pmatrix} \vec{x}_i^e \\ \vec{x}_i^o \end{pmatrix} \rightarrow \begin{pmatrix} \vec{x}_i^e \\ \vec{0} \end{pmatrix} \quad \begin{pmatrix} \vec{0} \\ \vec{x}_i^o \end{pmatrix}$$

Unique to staggered: $\gamma_5^{stag} \equiv \epsilon(x) = (-1)^{\sum_\mu x_\mu}$
$(1 \pm \gamma_5)$ is just even/odd!!

Leads to nice sparisty pattern for coarse operator.

Testing Environment

- Use a simpler model *with similar physics* for initial testing:
- Two-flavor Schwinger model in two dimensions:

$$\mathcal{L} = \frac{1}{2}F^2 - i\sum_{f=1,2}\bar{\psi}_f\gamma^\mu\left(\partial_\mu - igA_\mu\right)\psi_f + m\sum_{f=1,2}\bar{\psi}_f\psi_f$$

  - Confinement
  - Chiral symmetry breaking
  - Vorticies (2 dimensional "instantons")
  - Topology
  - Two flavor theory has a "pion"-like state: $M_{gap}(m) = A_{gap}m^{2/3}g^{1/3}$ [Phys.Rev. D55 (1997)]
- Comparatively very inexpensive to look at large volumes: 128^2

Our (Current) Prescription

Remark: Our presecription isn't final—this just fixes our tests!

Our (Current) Prescription

Remark: Our presecription isn't final—this just fixes our tests!

- Null vector generation:
  - Use BiCGStab
  - Generate random vector $\vec{x_0}$.
  - Solve residual equation $\not{D}\vec{e} = \vec{r}(\equiv -A\vec{x_0})$ to tolerance 5×10^{-5}.
  - Generate N_{vec} null vectors.

Our (Current) Prescription

Remark: Our presecription isn't final—this just fixes our tests!

- Null vector generation:
  - Use BiCGStab
  - Generate random vector $\vec{x_0}$.
  - Solve residual equation $\not{D}\vec{e} = \vec{r}(\equiv -A\vec{x_0})$ to tolerance 5×10^{-5}.
  - Generate N_{vec} null vectors.
- MG Solve:
  - One-level V cycle
  - Outer solver: GCR(24) to tolerance 5×10^{-7}
  - Block size: 4×4
  - Pre- and Post-Smoother: 6 iterations of GCR
  - Inner solver: GCR(64) to tolerance 10^{-3}

Free Staggered Operator

- Free Staggered operator obeys a "shift-by-two" translational invariance: 2^D zero modes

⋱	⋮	⋮	⋮	⋮	⋰
…	1	0	1	0	…
…	0	0	0	0	…
…	1	0	1	0	…
…	0	0	0	0	…
⋰	⋮	⋮	⋮	⋮	⋱

⋱	⋮	⋮	⋮	⋮	⋰
…	0	1	0	1	…
…	0	0	0	0	…
…	0	1	0	1	…
…	0	0	0	0	…
⋰	⋮	⋮	⋮	⋮	⋱

⋱	⋮	⋮	⋮	⋮	⋰
…	0	0	0	0	…
…	1	0	1	0	…
…	0	0	0	0	…
…	1	0	1	0	…
⋰	⋮	⋮	⋮	⋮	⋱

⋱	⋮	⋮	⋮	⋮	⋰
…	0	0	0	0	…
…	0	1	0	1	…
…	0	0	0	0	…
…	0	1	0	1	…
⋰	⋮	⋮	⋮	⋮	⋱

Free Staggered Operator

- Free Staggered operator obeys a "shift-by-two" translational invariance: 2^D zero modes

⋱	⋮	⋮	⋮	⋮	⋰
…	1	0	1	0	…
…	0	0	0	0	…
…	1	0	1	0	…
…	0	0	0	0	…
⋰	⋮	⋮	⋮	⋮	⋱

⋱	⋮	⋮	⋮	⋮	⋰
…	0	1	0	1	…
…	0	0	0	0	…
…	0	1	0	1	…
…	0	0	0	0	…
⋰	⋮	⋮	⋮	⋮	⋱

⋱	⋮	⋮	⋮	⋮	⋰
…	0	0	0	0	…
…	1	0	1	0	…
…	0	0	0	0	…
…	1	0	1	0	…
⋰	⋮	⋮	⋮	⋮	⋱

⋱	⋮	⋮	⋮	⋮	⋰
…	0	0	0	0	…
…	0	1	0	1	…
…	0	0	0	0	…
…	0	1	0	1	…
⋰	⋮	⋮	⋮	⋮	⋱

MG test: Use these as null vectors (up to normalization).
Each vector becomes "internal degree of freedom".

Outer Iterations: Free Case

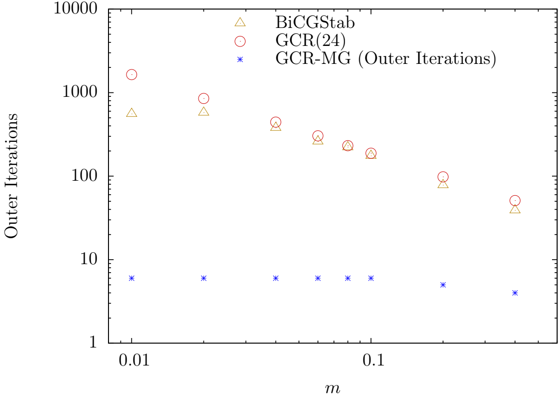


Figure : Outer Iterations for 128^2, free field

Demonstrates working (free) algorithm,
but not a “fair” comparison.

Fine $\not{D}$ Operators: Free Case

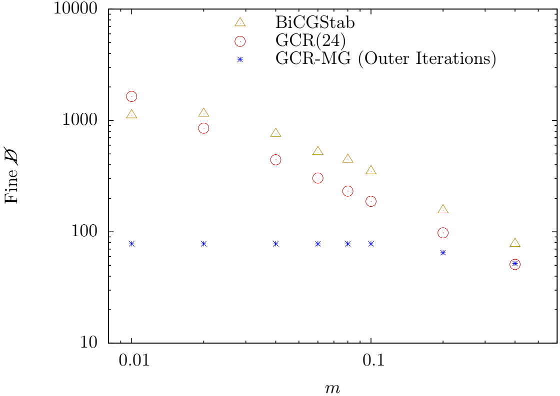


Figure : Fine $\not{D}$ for 128^2, free field

MG removes critical slowing down!

(For large mass, there's no critical slowing down to remove.)

Interacting Case

Quenched Schwinger model:
Non-Compact 2D $U(1)$ gauge action.

β	6.0	10.0	[Phys.Rev.Lett. 100 (2008): J. Brannick, R.C. Brower, M.A. Clark, J.C. Osborn, C. Rebbi]
l_σ	3.30	4.35	Gauge correlation length via Wilson loop: $W \approx e^{-A/l_\sigma^2}$
μ^{-1}	$\approx$ 5.5	$\approx$ 6.3	Pseudoscalar meson correlation length for $m = 0.01$

Interacting Case

Quenched Schwinger model:
Non-Compact 2D $U(1)$ gauge action.

β	6.0	10.0	[Phys.Rev.Lett. 100 (2008): J. Brannick, R.C. Brower, M.A. Clark, J.C. Osborn, C. Rebbi]
l_σ	3.30	4.35	Gauge correlation length via Wilson loop: $W \approx e^{-A/l_\sigma^2}$
μ^{-1}	≈ 5.5	≈ 6.3	Pseudoscalar meson correlation length for $m = 0.01$

As a representative ensemble, we will look in depth at $128^2, \beta = 10.0$.

Fine $\not{D}$: Interacting, $\beta = 10.0$

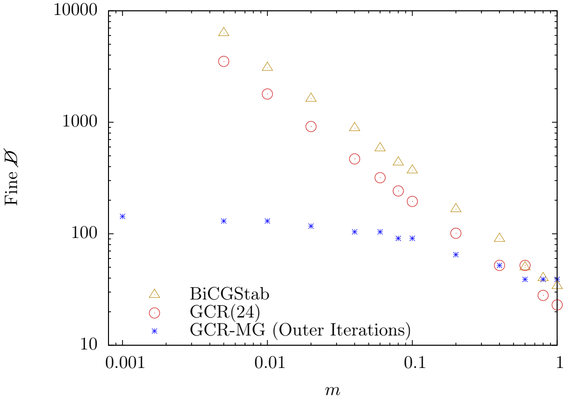


Figure : Fine $\not{D}$ for $128^2, \beta = 10.0, N_{vec} = 4$

For small m, critical slowing down is suppressed.

Residual per Iteration: Interacting, $\beta = 10.0$

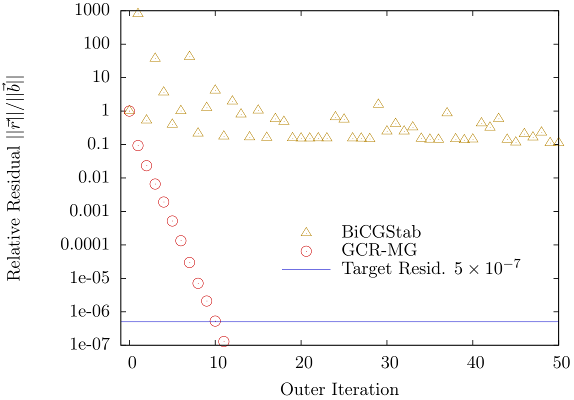


Figure : Relative Residual for $128^2, \beta = 10.0, N_{vec} = 4, m = 10^{-3}$

MG can stabilize otherwise unstable solves.

Number of Null Vectors: Interacting, $\beta = 10.0$

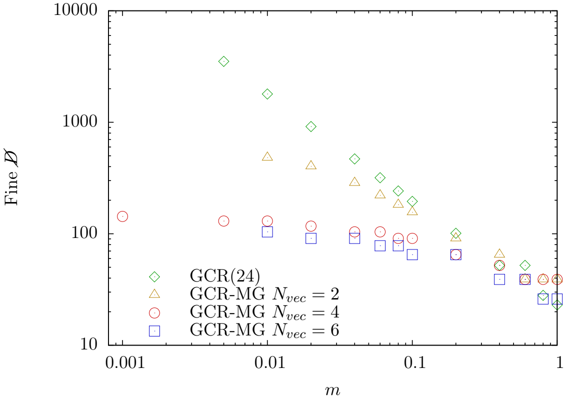


Figure : Comparison of different N_{vec} for $128^2, \beta = 10.0$

Diminishing returns with more N_{vec}: not a problem!

Coarse Solve: Interacting, $\beta = 10.0$

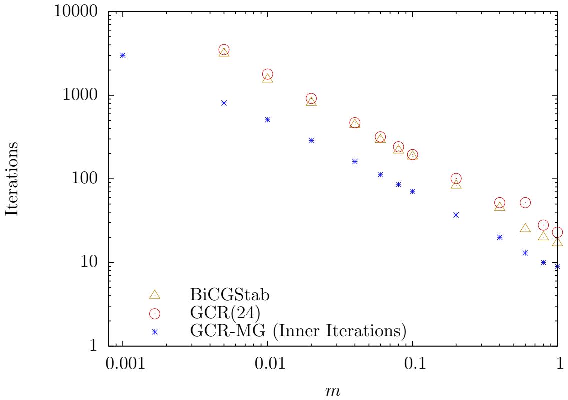


Figure : Inner Iterations for $128^2, \beta = 10.0, N_{vec} = 4$

Critical slowing down moves to coarse operator.

Still coarser levels will cure this.

Fine $\not{D}$: Interacting, $\beta = 6.0$

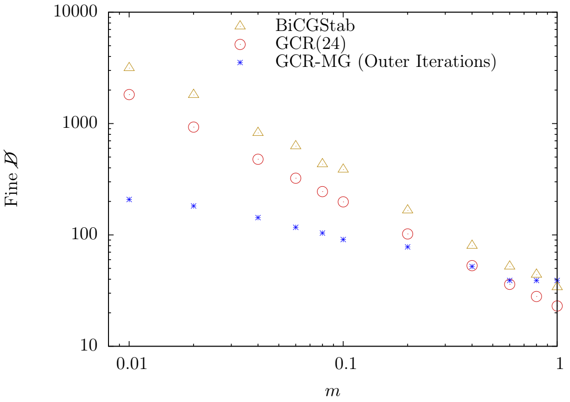


Figure : Fine $\not{D}$ for $128^2, \beta = 6.0, N_{vec} = 4$

MG still works for coarser β.

Overview and Future Work

Overview

- Showed a working MG algorithm for two-flavor Schwinger model

Overview and Future Work

Overview

- Showed a working MG algorithm for two-flavor Schwinger model

Future Work

- Extend to full 4 dimensional staggered operator: naïve, Asqtad, HISQ
- Continue progress on an implementation in QUDA
  - Extends existing infrastructure for Wilson-Clover
  - Implementation nearly there, further room for tuning and optimization.
  - `github.com/lattice/quda/tree/feature/staggered-multigrid`
- Experiment with alternative methods to generate null vectors
  - Can we use the normal equation to generate null vectors?
  - If we can, can we use approximate low vectors from EigCG instead?
    - This lets us do "real" solves while generating null vectors!
  - If we can't, can we use EigBiCG?

Overview and Future Work

Overview

- Showed a working MG algorithm for two-flavor Schwinger model

Future Work

- Extend to full 4 dimensional staggered operator: naïve, Asqtad, HISQ
- Continue progress on an implementation in QUDA
  - Extends existing infrastructure for Wilson-Clover
  - Implementation nearly there, further room for tuning and optimization.
  - `github.com/lattice/quda/tree/feature/staggered-multigrid`
- Experiment with alternative methods to generate null vectors
  - Can we use the normal equation to generate null vectors?
  - If we can, can we use approximate low vectors from EigCG instead?
    - This lets us do "real" solves while generating null vectors!
  - If we can't, can we use EigBiCG?

Thank you!