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
Outline

- The Battleground: $\mathcal{O}$ 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 $D$ 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 $\mathcal{O}$ inversions.

**Multigrid** methods can completely eliminate critical slowing down.
MG for the Wilson-Clover operator has a rich history:

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

**Figure**: QUDA-MG update, February 2016
Beyond Wilson-Clover

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

- Hierarchically deflated conjugate gradient: [arXiv 1402.2585: P.A. Boyle]
Beyond Wilson-Clover

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

- Hierarchically deflated conjugate gradient: [arXiv 1402.2585: P.A. Boyle]
- Hierarchically deflated conjugate residual: [Lattice 2016 Poster: A. Yamaguchi, P.A. Boyle.]
Beyond Wilson-Clover

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

- 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!
Given the staggered $\mathcal{D}$ operator:

$$D_{xy} = \sum_{\mu} \eta_{\mu}(x) \left[ U_{\mu}^\dagger(x) \delta_{y+\mu,x} - U_{\mu}(x) \delta_{x+\mu,y} \right] + m \delta_{x,y}$$

$$= iA + mI \quad \leftarrow \text{Anti-Hermitian + Hermitian piece}$$
Given the staggered $\mathcal{D}$ operator:

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

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

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

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

Normal, solve with CG! 

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

\[
\left( m^2 - D_{eo}D_{oe} \right) \psi_e = mb_e - D_{eo}b_o; \quad \psi_o = \frac{1}{m} (b_o - D_{oe}\psi_e)
\]
The E/O preconditioned operator is Hermitian positive definite: Rich theory of multigrid exists.

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

**Issues:**

- Two link sparsity pattern is computationally inefficient.
- Further issue: Even and odd decouple in chiral limit.
The E/O preconditioned operator is Hermitian positive definite:
Rich theory of multigrid exists.

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

**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 $\mathcal{D}$. 
Algorithm setup for 2-level V-Cycle:

1. Relax on homogeneous problem $\mathcal{D}\vec{x}_i = 0$ 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 $\left(1 - PP^+\right) \vec{x}_i = 0$
5. Define coarse grid operator $\mathcal{D}_c = P^+\mathcal{D}P$
Chiral Blocks

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

\[
\vec{x}_i \rightarrow (1 + \gamma_5)\vec{x}_i \quad (1 - \gamma_5)\vec{x}_i
\]
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:

\[
\vec{x}_i = \begin{pmatrix} \vec{x}^e_i \\ \vec{x}^o_i \end{pmatrix} \rightarrow \begin{pmatrix} \vec{x}^e_i \\ 0 \\ \vec{x}^o_i \end{pmatrix}
\]

Unique to staggered: \( \gamma^{stag}_5 \equiv \epsilon(x) = (-1)^\sum_\mu x_\mu \) (1 \( \pm \gamma_5 \)) is just even/odd!!
Similar to Wilson-Clover, we do not block in chirality.

For every null vector $\vec{x}_i$, split into two vectors:

$$\vec{x}_i = \begin{pmatrix} x^e_i \\ x^o_i \end{pmatrix} \rightarrow \begin{pmatrix} x^e_i \\ 0 \end{pmatrix} \begin{pmatrix} 0 \\ x^o_i \end{pmatrix}$$

Unique to staggered: $\gamma^\text{stag}_5 \equiv \epsilon(x) = (-1) \sum_\mu x_\mu$

$$(1 \pm \gamma_5) \text{ is just even/odd!!}$$

Leads to nice sparisty pattern for coarse operator.
• 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
  ▶ Vortices (2 dimensional "instantons")
  ▶ Topology
  ▶ Two flavor theory has a "pion"-like state: \[ M_{\text{gap}}(m) = A_{\text{gap}} m^{2/3} g^{1/3} \]

• Comparatively very inexpensive to look at large volumes: \( 128^2 \)
Remark: Our prescription isn’t final—this just fixes our tests!
Remark: Our prescription isn’t final—this just fixes our tests!

- Null vector generation:
  - Use BiCGStab
  - Generate random vector $\vec{x}_0$.
  - Solve residual equation $\mathcal{D}\vec{e} = \vec{r}(\equiv -A\vec{x}_0)$ to tolerance $5 \times 10^{-5}$.
  - Generate $N_{\text{vec}}$ null vectors.
Remark: Our prescription isn’t final—this just fixes our tests!

- **Null vector generation:**
  - Use BiCGStab
  - Generate random vector $\vec{x}_0$.
  - Solve residual equation $D\vec{\epsilon} = \vec{r}(\equiv -A\vec{x}_0)$ to tolerance $5 \times 10^{-5}$.
  - Generate $N_{vec}$ null vectors.

- **MG Solve:**
  - One-level V cycle
  - Outer solver: GCR(24) to tolerance $5 \times 10^{-7}$
  - Block size: $4 \times 4$
  - 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

\[
\begin{array}{cccccc}
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & 1 & 0 & 1 & 0 & \cdot \\
\cdot & 0 & 0 & 0 & 0 & \cdot \\
\cdot & 1 & 0 & 1 & 0 & \cdot \\
\cdot & 0 & 0 & 1 & 0 & \cdot \\
\cdot & 0 & 0 & 0 & 0 & \cdot \\
\end{array}
\]

\[
\begin{array}{cccccc}
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & 0 & 1 & 0 & 1 & \cdot \\
\cdot & 0 & 0 & 0 & 0 & \cdot \\
\cdot & 0 & 1 & 0 & 1 & \cdot \\
\cdot & 0 & 0 & 0 & 0 & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\end{array}
\]

\[
\begin{array}{cccccc}
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & 0 & 0 & 0 & 0 & \cdot \\
\cdot & 1 & 0 & 1 & 0 & \cdot \\
\cdot & 0 & 0 & 0 & 0 & \cdot \\
\cdot & 1 & 0 & 1 & 0 & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\end{array}
\]

\[
\begin{array}{cccccc}
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & 0 & 0 & 0 & 0 & \cdot \\
\cdot & 0 & 1 & 0 & 1 & \cdot \\
\cdot & 0 & 0 & 0 & 0 & \cdot \\
\cdot & 0 & 1 & 0 & 1 & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\end{array}
\]

\[
\begin{array}{cccccc}
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & 0 & 0 & 0 & 0 & \cdot \\
\cdot & 1 & 0 & 1 & 0 & \cdot \\
\cdot & 0 & 0 & 0 & 0 & \cdot \\
\cdot & 1 & 0 & 1 & 0 & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\end{array}
\]

\[
\begin{array}{cccccc}
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & 0 & 1 & 0 & 1 & \cdot \\
\cdot & 0 & 0 & 0 & 0 & \cdot \\
\cdot & 0 & 1 & 0 & 1 & \cdot \\
\cdot & 0 & 0 & 0 & 0 & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\end{array}
\]

\[
\begin{array}{cccccc}
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & 0 & 1 & 0 & 1 & \cdot \\
\cdot & 0 & 0 & 0 & 0 & \cdot \\
\cdot & 0 & 1 & 0 & 1 & \cdot \\
\cdot & 0 & 0 & 0 & 0 & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\end{array}
\]
Free Staggered Operator

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

\[
\begin{array}{cccccc}
\ldots & 1 & 0 & 1 & 0 & \ldots \\
\ldots & 0 & 0 & 0 & 0 & \ldots \\
\ldots & 1 & 0 & 1 & 0 & \ldots \\
\ldots & 0 & 0 & 0 & 0 & \ldots \\
\ldots & 1 & 0 & 1 & 0 & \ldots \\
\ldots & 0 & 0 & 0 & 0 & \ldots \\
\ldots & 1 & 0 & 1 & 0 & \ldots \\
\ldots & 0 & 0 & 0 & 0 & \ldots \\
\ldots & 1 & 0 & 1 & 0 & \ldots \\
\ldots & 0 & 0 & 0 & 0 & \ldots \\
\ldots & 1 & 0 & 1 & 0 & \ldots \\
\ldots & 0 & 0 & 0 & 0 & \ldots \\
\end{array}
\]

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.
MG removes critical slowing down!
(For large mass, there’s no critical slowing down to remove.)
Quenched Schwinger model: Non-Compact 2D $U(1)$ gauge action.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$l_\sigma$</td>
<td>3.30</td>
<td>4.35</td>
<td>Gauge correlation length via Wilson loop: $W \approx e^{-A/l_\sigma^2}$</td>
</tr>
<tr>
<td>$\mu^{-1}$</td>
<td>$\approx 5.5$</td>
<td>$\approx 6.3$</td>
<td>Pseudoscalar meson correlation length for $m = 0.01$</td>
</tr>
</tbody>
</table>
Quenched Schwinger model: Non-Compact 2D $U(1)$ gauge action.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$l_\sigma$</td>
<td>3.30</td>
<td>4.35</td>
<td>Gauge correlation length via Wilson loop: $W \approx e^{-A/l_\sigma^2}$</td>
</tr>
<tr>
<td>$\mu^{-1}$</td>
<td>$\approx 5.5$</td>
<td>$\approx 6.3$</td>
<td>Pseudoscalar meson correlation length for $m = 0.01$</td>
</tr>
</tbody>
</table>

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

Fine $\mathcal{D}$: Interacting, $\beta = 10.0$

For small $m$, critical slowing down is suppressed.
Residual per Iteration: Interacting, $\beta = 10.0$

![Graph showing relative residual per iteration for BiCGStab and GCR-MG methods with target residual $5 \times 10^{-7}$ and $128^2$, $\beta = 10.0$, $N_{vec} = 4$, $m = 10^{-3}$.

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, $\beta = 10.0$

![Graph showing comparison of different $N_{vec}$ for $128^2$, $\beta = 10.0$.]

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

Diminishing returns with more $N_{vec}$: not a problem!
Critical slowing down moves to coarse operator.
Still coarser levels will cure this.

Figure: Inner Iterations for $128^2, \beta = 10.0, N_{vec} = 4$
Fine $\mathcal{D}$: Interacting, $\beta = 6.0$

Figure: Fine $\mathcal{D}$ for $128^2$, $\beta = 6.0$, $N_{vec} = 4$

MG still works for coarser $\beta$. 
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](http://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!