Update on a Staggered Multigrid Algorithm (with hope in 4D dimensions)

Evan Weinberg

Boston University, Boston, MA

weinbe2@bu.edu

in collaboration with Richard Brower, Kate Clark, Alexei Strelchenko

June 19, 2017

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.
- Modern analysis requires 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.
- Modern analysis requires 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]

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]

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

- Hierarchically deflated conjugate residual: [Lattice 2016 Poster: A. Yamaguchi, P.A. Boyle.] update meee

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]

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

- Hierarchically deflated conjugate residual: [Lattice 2016 Poster: A. Yamaguchi, P.A. Boyle.] update meee

And for Twisted Mass:

- On $\not{D}$: [Phys. Rev. D 94, 114509: C. Alexandrou, S. Bacchio, J. Finkenrath, A. Frommer, K. Kahl, and M. Rottmann]
- Implementation in QUDA

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]

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

- Hierarchically deflated conjugate residual: [Lattice 2016 Poster: A. Yamaguchi, P.A. Boyle.] update meee

And for Twisted Mass:

- On $\not{D}$: [Phys. Rev. D 94, 114509: C. Alexandrou, S. Bacchio, J. Finkenrath, A. Frommer, K. Kahl, and M. Rottmann]
- Implementation in QUDA

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:

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

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

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

Staggered State of the Art

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

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

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

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

Issues:

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

Testing Environment for the Interacting Case

- 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: 256^2
- Constant physics and box size: double L, half g (quadruple β), half m.

Algebraic Multigrid in the Interacting Case

Standard story...

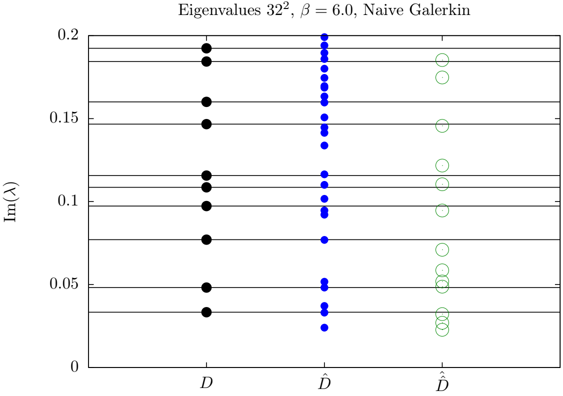
- Generate null vectors (I use BiCGstab-8 to a tolerance of 5×10^{-5} *or* maximum of 500 mat-vec).
  - For this talk: 8 null vectors.
- Block orthonormalize, Galerkin projection.
- Preserve chirality $(1 \pm \gamma_5) \rightarrow$ even/odd for staggered.
  - For this talk: 16 degrees of freedom per coarse site.

Galerkin Projection: Failure

Naïvely using what worked for Wilson fails for staggered.

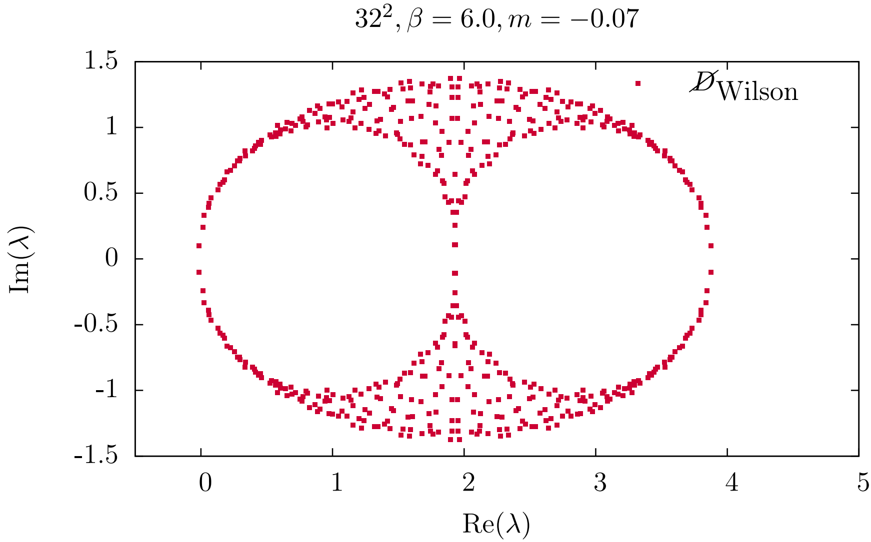
Galerkin Projection: Failure

Naïvely using what worked for Wilson fails for staggered.

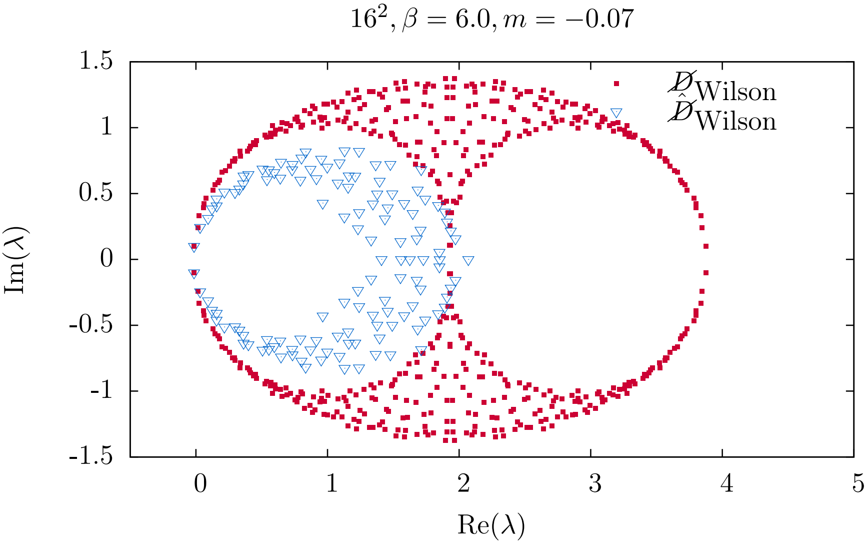


Spurious low modes.

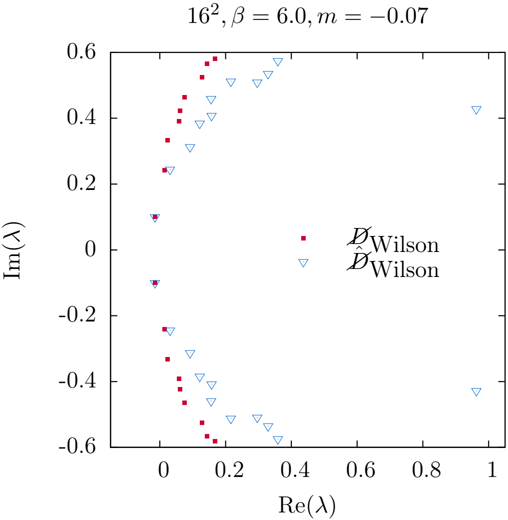
Why (?) Wilson Works



Why (?) Wilson Works



Why (?) Wilson Works



Kahler-Dirac Thinking

Staggered fermions distribute d fermions over 2^d sites.
Think of this as one "super-site" (or Kahler-Dirac block)

[arXiv:0509026: S. Durr]

$$\mathcal{S} = b^4 \sum_{X,\mu} \bar{q}(X) \left[\nabla_\mu \left(\gamma_\mu \otimes 1 \right) - \frac{b}{2} \triangle_\mu \left(\gamma_5 \otimes \tau_\mu \tau_5 \right) + m \left(1 \otimes 1 \right) \right] q(X)$$

$$\equiv b^4 \sum_{X,\mu} \bar{q}(X) \left[\not{D} + m \right] q(X)$$

$$\left(\nabla_\mu q \right)(X) = \frac{q(X + b\hat{\mu}) - q(X - b\hat{\mu})}{2b}$$

$$\left(\triangle_\mu q \right)(X) = \frac{q(X + b\hat{\mu}) - 2q(X) + q(X - b\hat{\mu})}{b^2}$$

The free staggered operator can be transformed into this form via a *unitary transformation*.

Super-site preconditioning

Unitary: spectrum is unchanged.
What if we solve the Kahler-Dirac block exactly?

$$B(m) = \begin{pmatrix} m & 0 & -\frac{1}{2} & -\frac{1}{2} \\ 0 & m & -\frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & m & 0 \\ \frac{1}{2} & -\frac{1}{2} & 0 & m \end{pmatrix}$$

Super-site preconditioning

Unitary: spectrum is unchanged.
What if we solve the Kahler-Dirac block exactly?

$$B(m) = \begin{pmatrix} m & 0 & -\frac{1}{2} & -\frac{1}{2} \\ 0 & m & -\frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & m & 0 \\ \frac{1}{2} & -\frac{1}{2} & 0 & m \end{pmatrix}$$

- Wilson-Clover: preconditioning with the inverse clover term.
- Domain Wall: 4D preconditioned operator.

Super-site preconditioning

Unitary: spectrum is unchanged.
What if we solve the Kahler-Dirac block exactly?

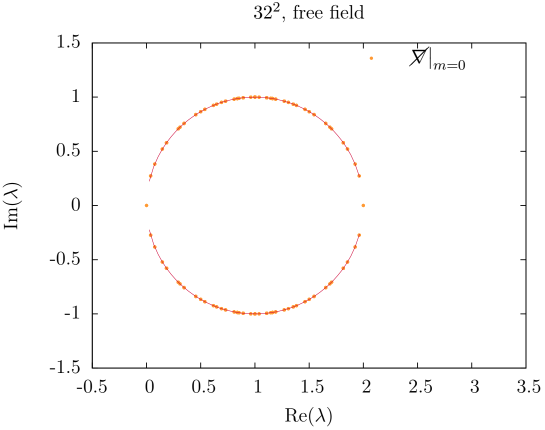
$$B(m) = \begin{pmatrix} m & 0 & -\frac{1}{2} & -\frac{1}{2} \\ 0 & m & -\frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & m & 0 \\ \frac{1}{2} & -\frac{1}{2} & 0 & m \end{pmatrix}$$

- Wilson-Clover: preconditioning with the inverse clover term.
- Domain Wall: 4D preconditioned operator.

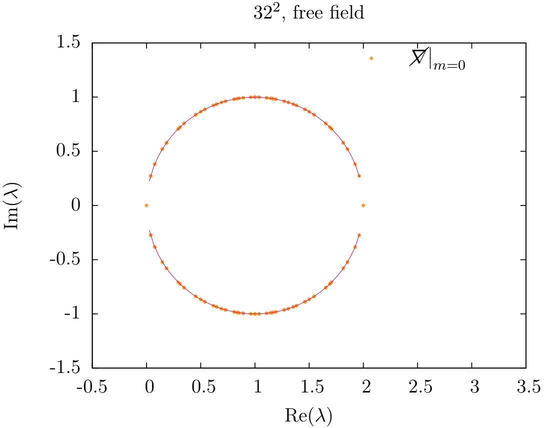
$$\not{\nabla}|_m = \left[\not{D} + m\right] B(m)^{-1}$$

$$= \left[\nabla_\mu \left(\gamma_\mu \otimes 1\right) - \frac{b}{2} \triangle_\mu \left(\gamma_5 \otimes \tau_\mu \tau_5\right) + m \left(1 \otimes 1\right)\right] \begin{pmatrix} m & 0 & -\frac{1}{2} & -\frac{1}{2} \\ 0 & m & -\frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & m & 0 \\ \frac{1}{2} & -\frac{1}{2} & 0 & m \end{pmatrix}^{-1}$$

Preconditioned Op

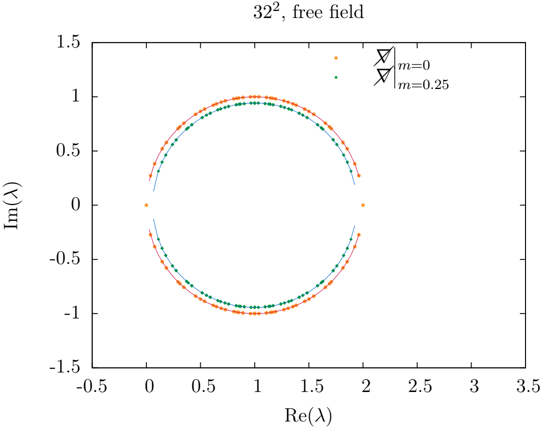


Preconditioned Op



$1 + U$, where U is a *unitary operator*. **This is an Overlap operator.**

Preconditioned Op



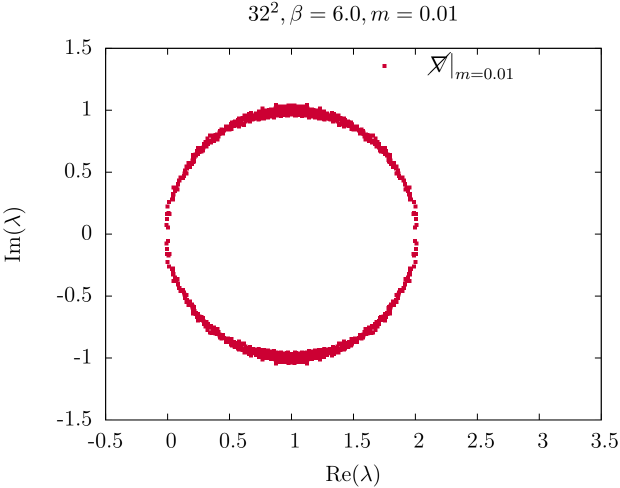
Massive case: $1 + \frac{1}{\sqrt{1+2m^2}} U$

Interacting Case

For the interacting case, we can still perform the unitary transformation then solve the Kahler-Dirac block exactly. It's now just gauge field $\rightarrow$ coordinate dependent.

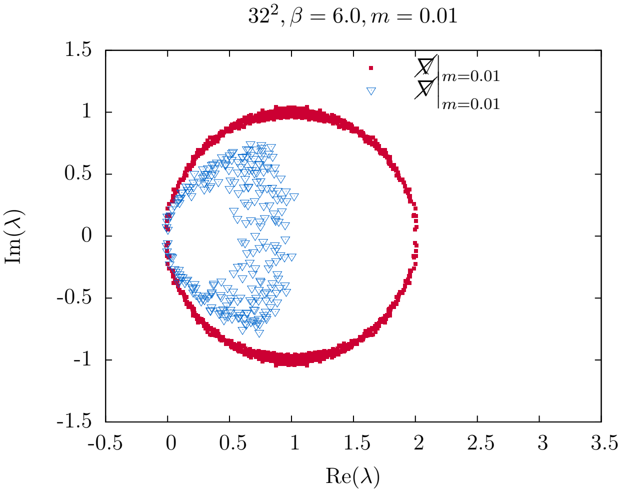
$$\begin{pmatrix} m & 0 & -\frac{1}{2}U_x(2\vec{n}) & -\frac{1}{2}U_y(2\vec{n}) \\ 0 & m & -\frac{1}{2}U_y^\dagger(2\vec{n}+\hat{x}) & \frac{1}{2}U_x^\dagger(2\vec{n}+\hat{y}) \\ \frac{1}{2}U_x^\dagger(2\vec{n}) & \frac{1}{2}U_y(2\vec{n}+\hat{x}) & m & 0 \\ \frac{1}{2}U_y^\dagger(2\vec{n}) & -\frac{1}{2}U_x(2\vec{n}+\hat{y}) & 0 & m \end{pmatrix}$$

Interacting Preconditioned Op



The circle "fuzzes".

Interacting Preconditioned Op



The circle "fuzzes"... but the spectrum collapses well!

Our Multigrid Prescription

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

Our Multigrid Prescription

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

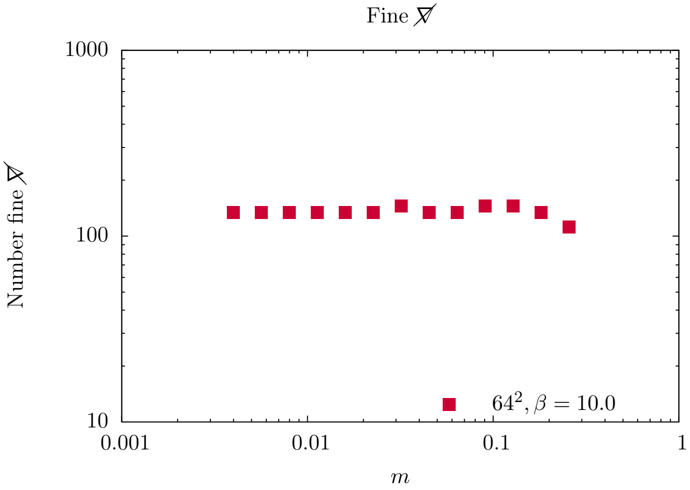
- Null vector generation:
  - Use BiCGStab-8
  - Generate random vector $\vec{x_0}$.
  - Solve residual equation $\not{D}\vec{e} = \vec{r}(\equiv -A\vec{x_0})$ to tolerance 5×10^{-5}, or to a maximum of 500 $\not{D}$.
  - Generate 8 null vectors $\rightarrow$ 16 internal dof after chiral projecting

Our Multigrid Prescription

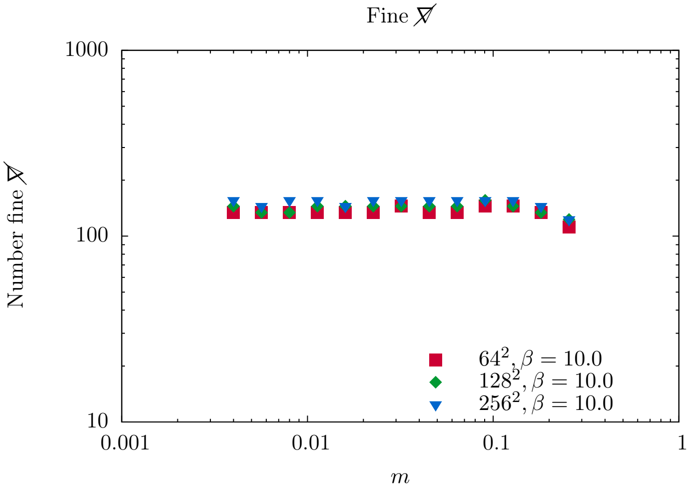
Remark: Our presecription isn't final—this just fixes our tests!

- Null vector generation:
  - Use BiCGStab-8
  - 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}, or to a maximum of 500 $\not{D}$.
  - Generate 8 null vectors $\rightarrow$ 16 internal dof after chiral projecting
- MG Solve:
  - Three level K-cycle
  - Outer solver: GCR(32) to tolerance 10^{-10}
  - Block size: 4×4 (e.g., $128^2 \rightarrow 32^2 \rightarrow 8^2$)
  - Pre- and Post-Smoother: 2 iterations of GCR
  - Inner solver: GCR(32) to tolerance 0.2

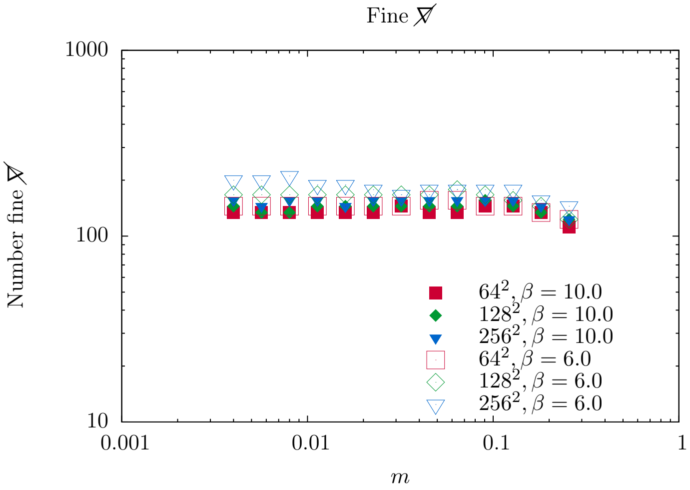
Outer $\not\nabla$



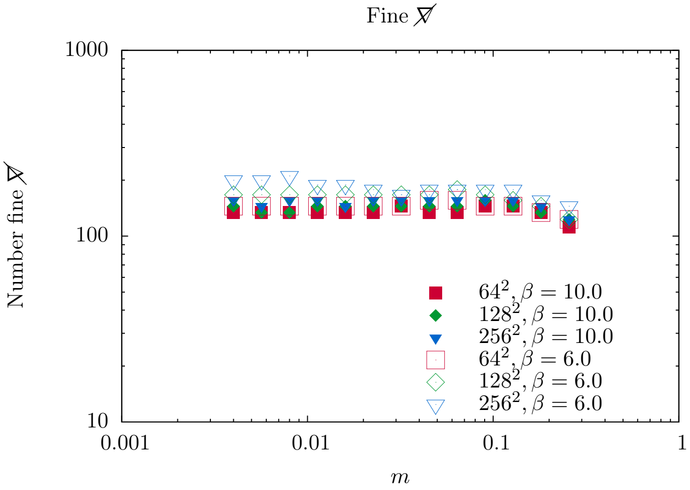
Outer $\not\nabla$



Outer $\not\nabla$

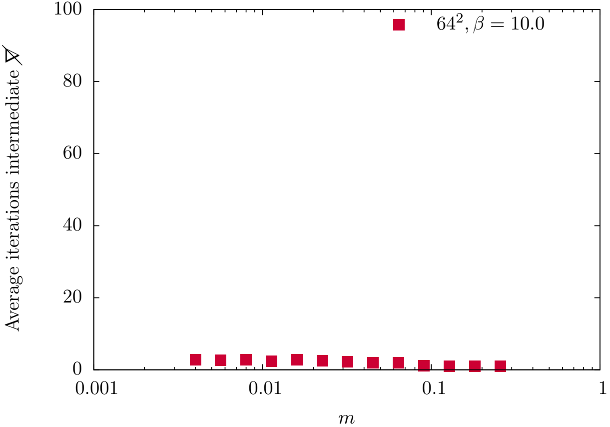


Outer $\not\nabla$

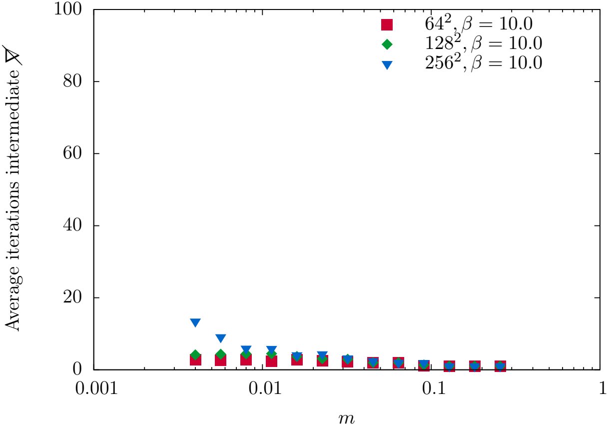


Relatively constant independent of mass, volume, β...

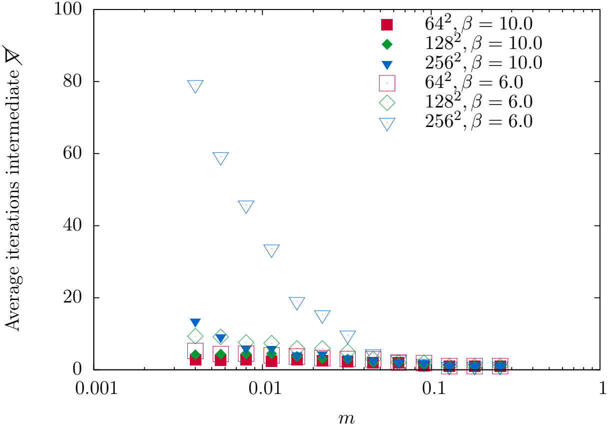
Recursive test: average iterations, intermediate $\not\nabla$



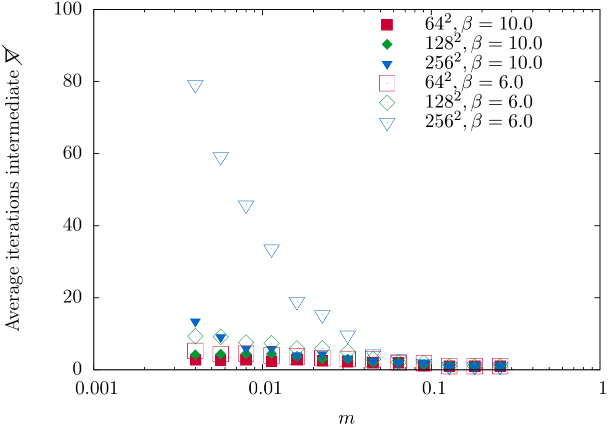
Recursive test: average iterations, intermediate $\not\nabla$



Recursive test: average iterations, intermediate $\not\nabla$

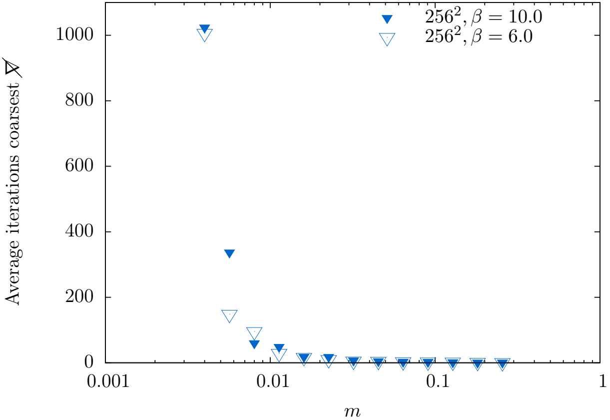


Recursive test: average iterations, intermediate $\not\nabla$



For larger volumes, smaller β, the coarsest level does *not* precondition the intermediate level well.

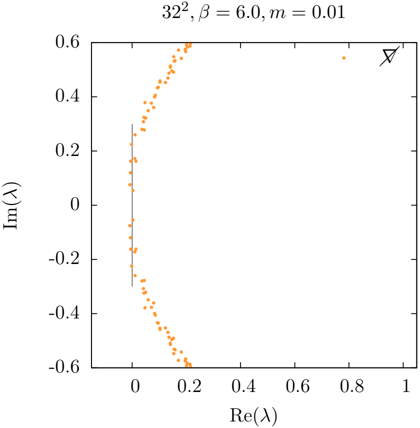
Average iterations, coarsest $\not\nabla$



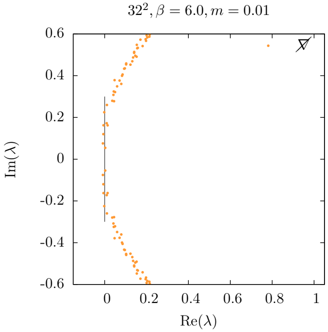
The coarsest $\not\nabla$ is becoming very ill-conditioned... which is damaging the intermediate solve.

The crux of the issue...

The recursive algorithm is falling apart at larger volume, smaller mass, moreso at smaller β. A likely culprit? The "fuzzed" circle.



Discussion



- Larger volume, the spectrum becomes more dense.
- Smaller mass, the "fuzzed" circle expands.
- Smaller β, the circle becomes more fuzzed.

These all lead to more "exceptional eigenvalues".

Constant physics

What happens with finer lattices at constant physics?
(double L, quadruple β, half m)

- $64^2, m = 0.01, \beta = 3.0$:
- $128^2, m = 0.005, \beta = 12.0$:
- $256^2, m = 0.0025, \beta = 48.0$:

Constant physics

What happens with finer lattices at constant physics?
(double L, quadruple β, half m)

- $64^2, m = 0.01, \beta = 3.0$: 200 Fine $\not\nabla$, ~20 intermediate iterations, ~16 coarse iterations.
- $128^2, m = 0.005, \beta = 12.0$:
- $256^2, m = 0.0025, \beta = 48.0$:

Constant physics

What happens with finer lattices at constant physics?
(double L, quadruple β, half m)

- $64^2, m = 0.01, \beta = 3.0$: 200 Fine $\not\nabla$, ˜20 intermediate iterations, ˜16 coarse iterations.
- $128^2, m = 0.005, \beta = 12.0$: 134 Fine $\not\nabla$, ˜5 intermediate iterations, 6̃6 coarse iterations.
- $256^2, m = 0.0025, \beta = 48.0$:

Constant physics

What happens with finer lattices at constant physics?
(double L, quadruple β, half m)

- $64^2, m = 0.01, \beta = 3.0$: 200 Fine $\not\nabla$, ˜20 intermediate iterations, ˜16 coarse iterations.
- $128^2, m = 0.005, \beta = 12.0$: 134 Fine $\not\nabla$, ˜5 intermediate iterations, 6̃6 coarse iterations.
- $256^2, m = 0.0025, \beta = 48.0$: 123 Fine $\not\nabla$, ˜2 intermediate iterations, ˜602 coarse iterations.

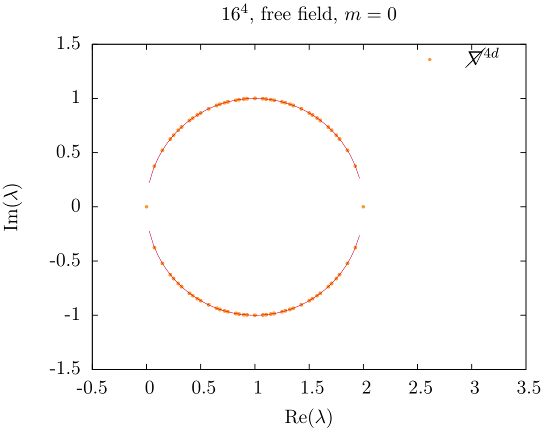
Constant physics

What happens with finer lattices at constant physics?
(double L, quadruple β, half m)

- $64^2, m = 0.01, \beta = 3.0$: 200 Fine $\not\nabla$, ˜20 intermediate iterations, ˜16 coarse iterations.
- $128^2, m = 0.005, \beta = 12.0$: 134 Fine $\not\nabla$, ˜5 intermediate iterations, ˜66 coarse iterations.
- $256^2, m = 0.0025, \beta = 48.0$: 123 Fine $\not\nabla$, ˜2 intermediate iterations, ˜602 coarse iterations.

Even though the coarsest level is becoming ill-conditioned, it becomes a better and better preconditioner.

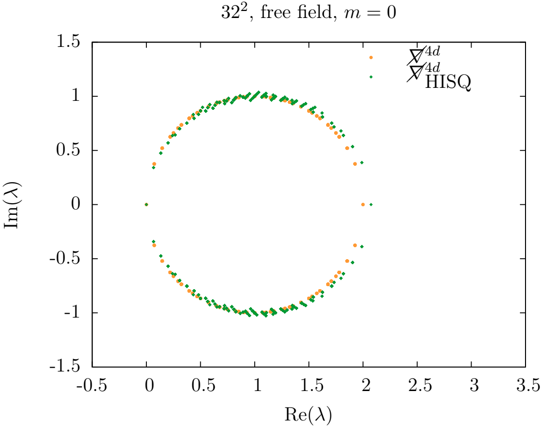
Looking forward: 4D



$$1 - \frac{1}{d}\left(\sum_\mu \cos(p_\mu)\right) \pm i\sqrt{1 - \frac{1}{d^2}\left(\sum_\mu \cos(p_\mu)\right)^2}$$

Circle of radius 1

Looking forward: 4D



HISQ term: no longer a circle, but circle-like, stretched to the right.

Overview and Future Work

Overview

- Showed a *recursive, mostly* working MG algorithm for the two-flavor Schwinger model.
  - Currently breaks down in the very coarse, large volume case.

Overview and Future Work

Overview

- Showed a *recursive, mostly* working MG algorithm for the two-flavor Schwinger model.
  - Currently breaks down in the very coarse, large volume case.

Future Work

- Extend to the *interacting* 4 dimensional staggered operator
- Continue progress on an implementation in QUDA
- Experiment with alternative methods to generate null vectors
- Try to cure this coarse preconditioning issue? Apparently twisted mass is struggling too.

Overview and Future Work

Overview

- Showed a *recursive, mostly* working MG algorithm for the two-flavor Schwinger model.
  - Currently breaks down in the very coarse, large volume case.

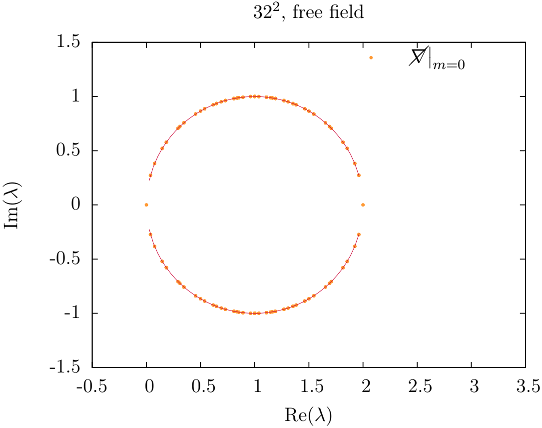
Future Work

- Extend to the *interacting* 4 dimensional staggered operator
- Continue progress on an implementation in QUDA
- Experiment with alternative methods to generate null vectors
- Try to cure this coarse preconditioning issue? Apparently twisted mass is struggling too.

Thank you!

BACKUP

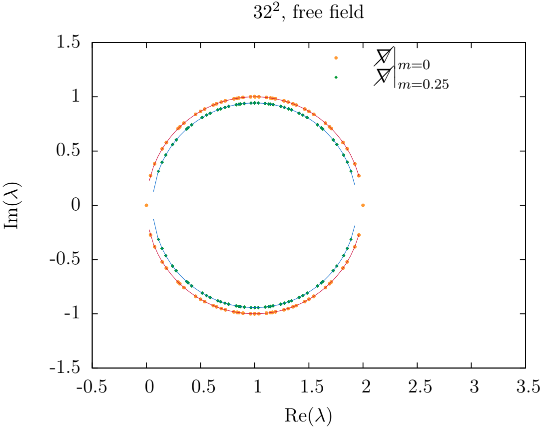
Preconditioned Op



$$1 - \frac{1}{2}\left(\cos(p_x) + \cos(p_y)\right) \pm i\sqrt{1 - \frac{1}{4}\left(\cos(p_x) + \cos(p_y)\right)^2}$$

Circle of radius 1

Preconditioned Op



$$1 - \frac{1}{1+2m^2}\left(\frac{1}{2}\left(\cos(p_x)+\cos(p_y)\right) \pm i\sqrt{1+2m^2-\frac{1}{4}\left(\cos(p_x)+\cos(p_y)\right)^2}\right)$$

Circle of radius $1/\sqrt{1+2m^2}$