The Removal of Critical Slowing Down

Lattice 2008

College of William and Mary

Michael Clark

Boston University

James Brannick, Rich Brower, Tom Manteuffel, Steve McCormick,
James Osborn, Claudio Rebbi

Inverting the Dirac Operator

- Required to invert the Dirac operator for QCD calculation
  - Determinant: $O(1)$ inversions, continuously changing U
  - Observables: $O(10) - O(1000)$ inversions, constant U
- Direct solve infeasible
- Krylov solvers typically used
  - Using CG on $D^\dagger D x = D^\dagger b$
  - e.g., BiCGstab on $Dx = b$
- Condition number $\propto 1/m$ $(1/m^2)$
  - Critical slowing down in solver

Inverting the Dirac Operator: Recent Progress

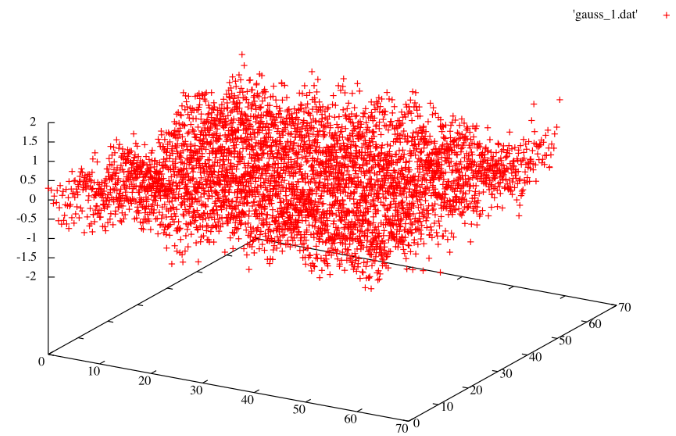
- Eigenvector deflation strategies (Morgan/Wilcox, Orginos/Stathopoulos)
  - Deflate Dirac operator by its low e-vectors
  - With sufficient e-vectors, critical slowing down gone
  - $O(N^2)$ scaling in number of e-vectors required
  - Large overhead finding e-vectors $\Rightarrow$ multiple rhs problems

- Local mode deflation (Lüscher)
  - Deflate the Dirac operator only locally over blocks
  - E-vectors locally co-linear ("Local coherence")
  - $O(N)/O(N \log N)$ cost
  - Small setup overhead $\Rightarrow$ suitable for all problems

- Adaptive multigrid (Brannick/Brower/Clark/Osborn/Rebbi)
  - Preconditioner for Dirac normal equations in $2d$
  - Critical slowing down gone
  - How about $4d$?

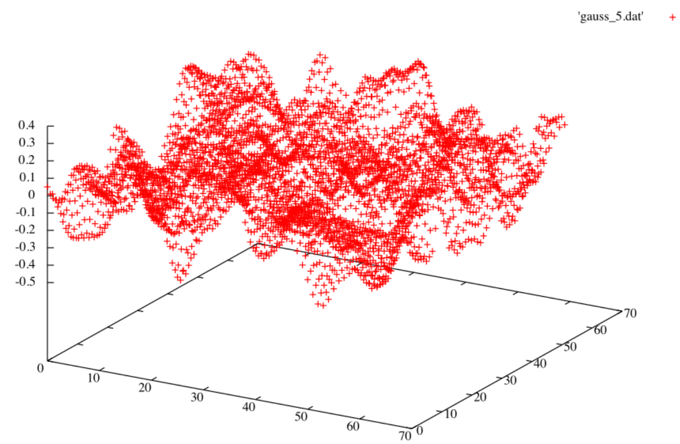
Intro to Multigrid: Failure of stationary solvers

- Iterative solvers (relaxation) effective on high frequency error
- Minimal effect on low frequency error
- Example:
  - Free Laplace operator in $2d$
  - $Ax = 0$, x_0 random
  - Gauss-Seidel relaxation
  - Plot error $= x$



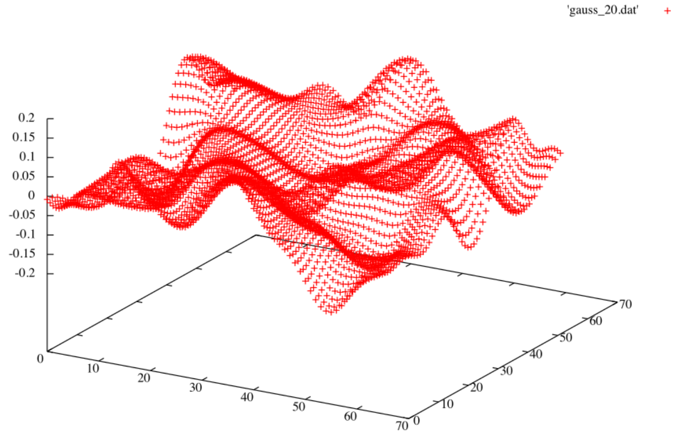
Intro to Multigrid: Failure of stationary solvers

- Iterative solvers (relaxation) effective on high frequency error
- Minimal effect on low frequency error
- Example:
  - Free Laplace operator in $2d$
  - $Ax = 0$, x_0 random
  - Gauss-Seidel relaxation
  - Plot error $= x$



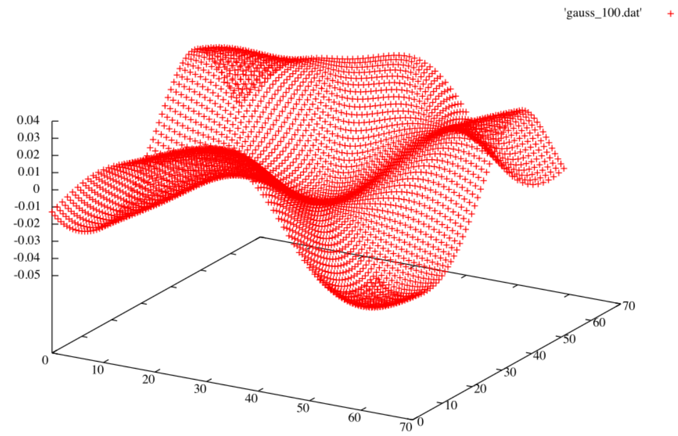
Intro to Multigrid: Failure of stationary solvers

- Iterative solvers (relaxation) effective on high frequency error
- Minimal effect on low frequency error
- Example:
  - Free Laplace operator in $2d$
  - $Ax = 0$, x_0 random
  - Gauss-Seidel relaxation
  - Plot error $= x$



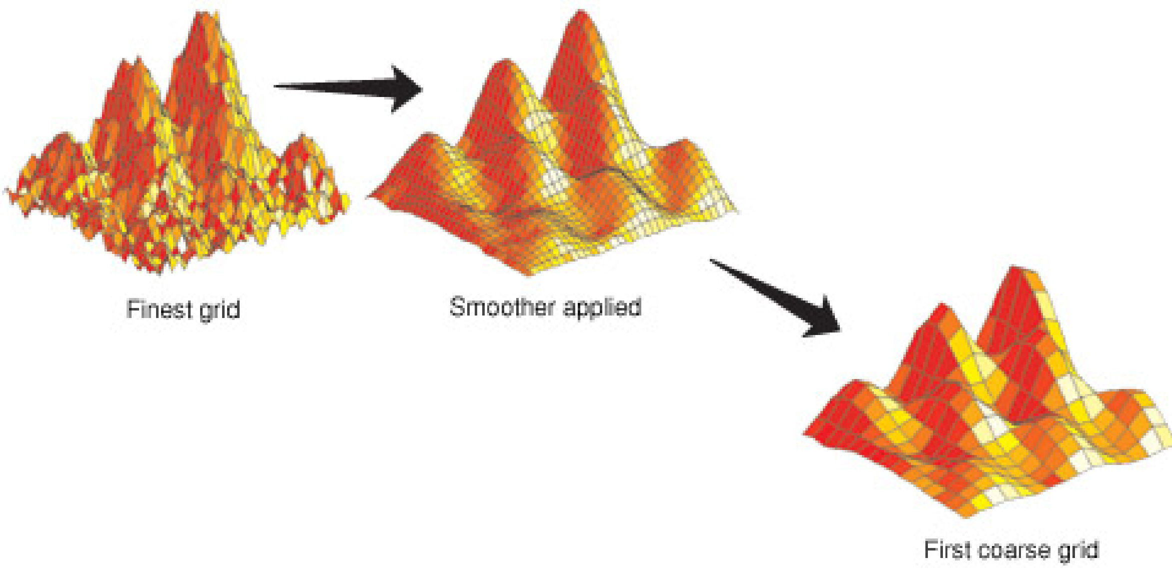
Intro to Multigrid: Failure of stationary solvers

- Iterative solvers (relaxation) effective on high frequency error
- Minimal effect on low frequency error
- Example:
  - Free Laplace operator in $2d$
  - $Ax = 0$, x_0 random
  - Gauss-Seidel relaxation
  - Plot error $= x$

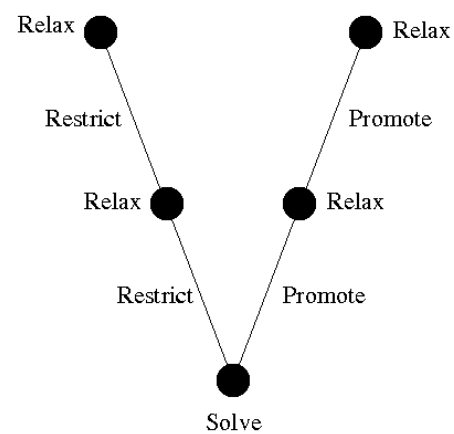


Intro to Multigrid: Coarse Grid Representation I

- Low frequency error components slow down solver
- Low frequency modes are smooth $\Rightarrow$ represent on coarse grid

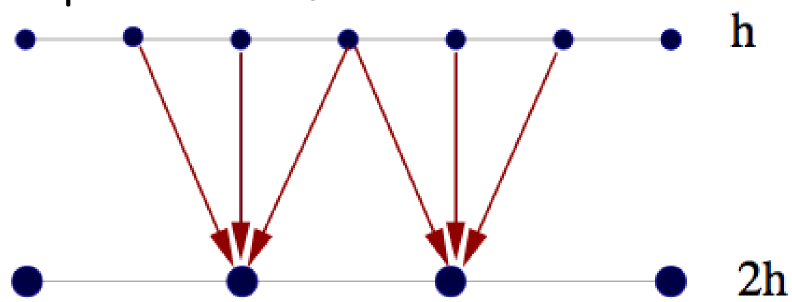


- Fine low frequency $\rightarrow$ Coarse high frequency
  $\Rightarrow$ Relax system on coarse grid
- Prolongate coarse correction to fine grid
- 2 grid scheme much better than relaxation
- Iterate until exact solve is possible (V-cycle)

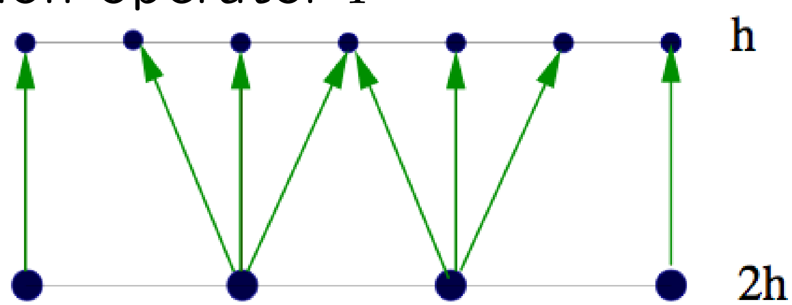


Intro to Multigrid: Coarse Grid Representation II

- Define restriction operator R



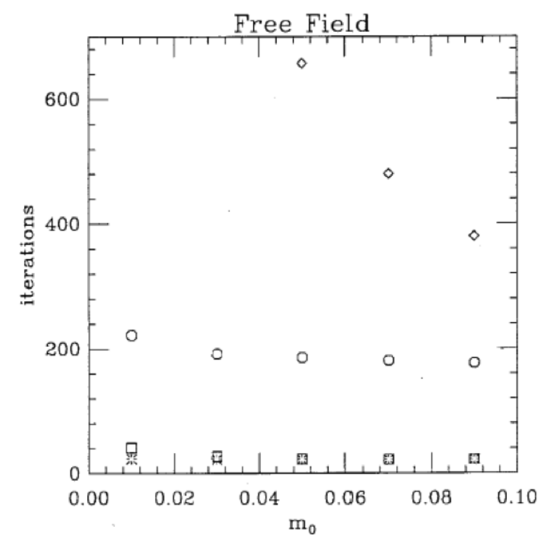
- Define prolongation operator P



- Notice $R = P^{\dagger}$
- Define coarse grid operator $A_c = P^{\dagger} A P$ (Galerkin prescription)
  - Equivalent to rediscretization in free theory

Intro to Multigrid: Geometric MG and QCD

- Classical geometric multigrid
  - $O(N \log N)$ / $O(N)$ work
  - Removes critical slowing down
- Gauge field U is not geometrically smooth
  - Dirac operator low modes oscillatory
  - Geometric MG fails



- 20 years of failures applying MG to QCD
- MG requires low modes preserved in coarse space
  - $(1 - PP^\dagger)\psi_0 = 0$
- **Adaptivity** required for gauge fields

Adaptive Smooth Aggregation MG (αSA) (Brezina et al)

- Adaptively use (s)low modes to define required V-cycle

- Initial Algorithm setup:
  1. Relax on homogenous problem $Ax = 0$, random x_0
     $\Rightarrow$ resulting error vector is a representation of slow modes
  2. Cut vector x into subsets to be aggregated (blocked)
  3. This defines the prolongator such that $(1 - PP^\dagger)x = 0$
  4. Define coarse grid operator $A_c = P^\dagger A P$
  5. Relax on coarse grid $A_c x^c = 0$, $x_0^c = P^\dagger x$, goto 2.

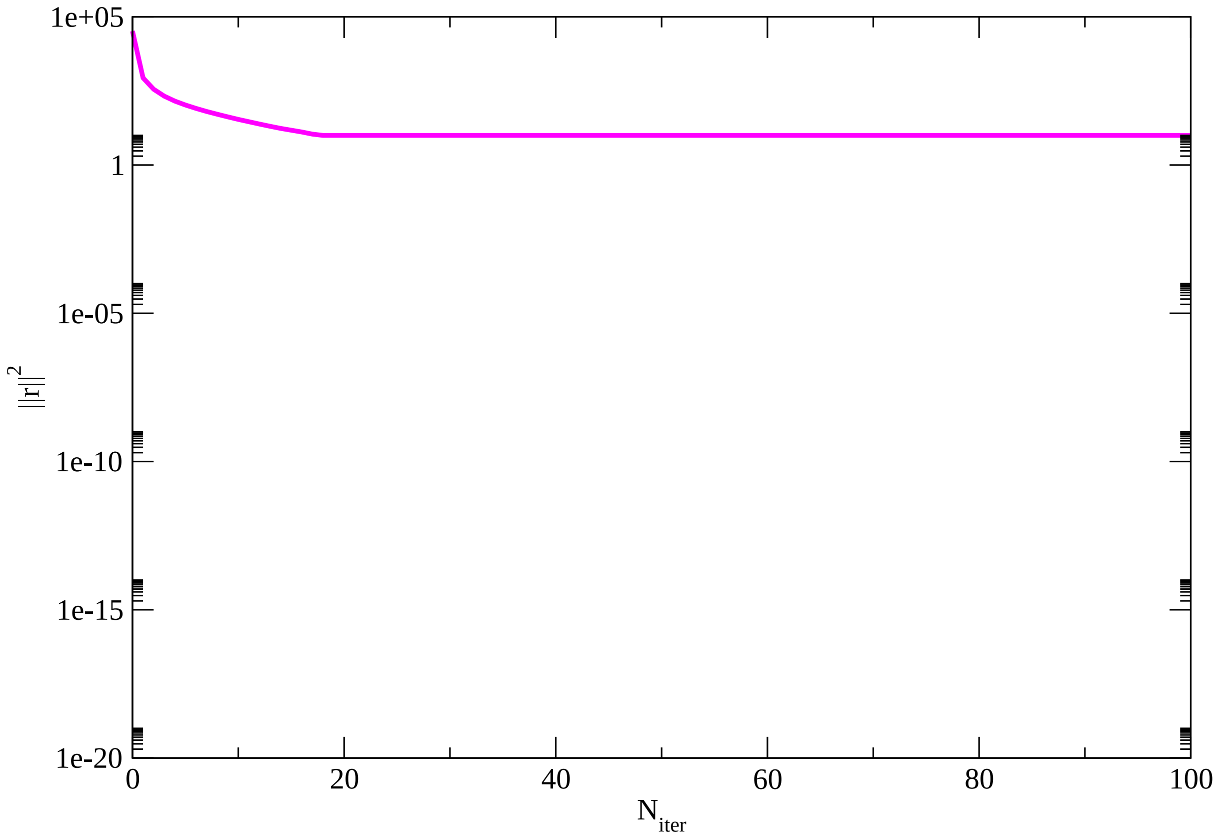
- This defines initial V-Cycle

Adaptive Smooth Aggregation MG (αSA) (Brezina et al)

- In general require more a single vector
- Now repeat setup process replacing relaxation with V-cycle
  - Previously found errors components quickly reduced
  - Error vector rich in new error components
- Augment V-Cycle to preserve additional vector space
  - Cut vectors x_1, x_2 into blocks
  - Orthonormalize blocks to define augmented prolongator
  - Define new coarse grid operator
  - Each vector corresponds to an extra dof per coarse site
- Add more vectors until satisfactory solver found
- Setup cost depends on number of vectors required

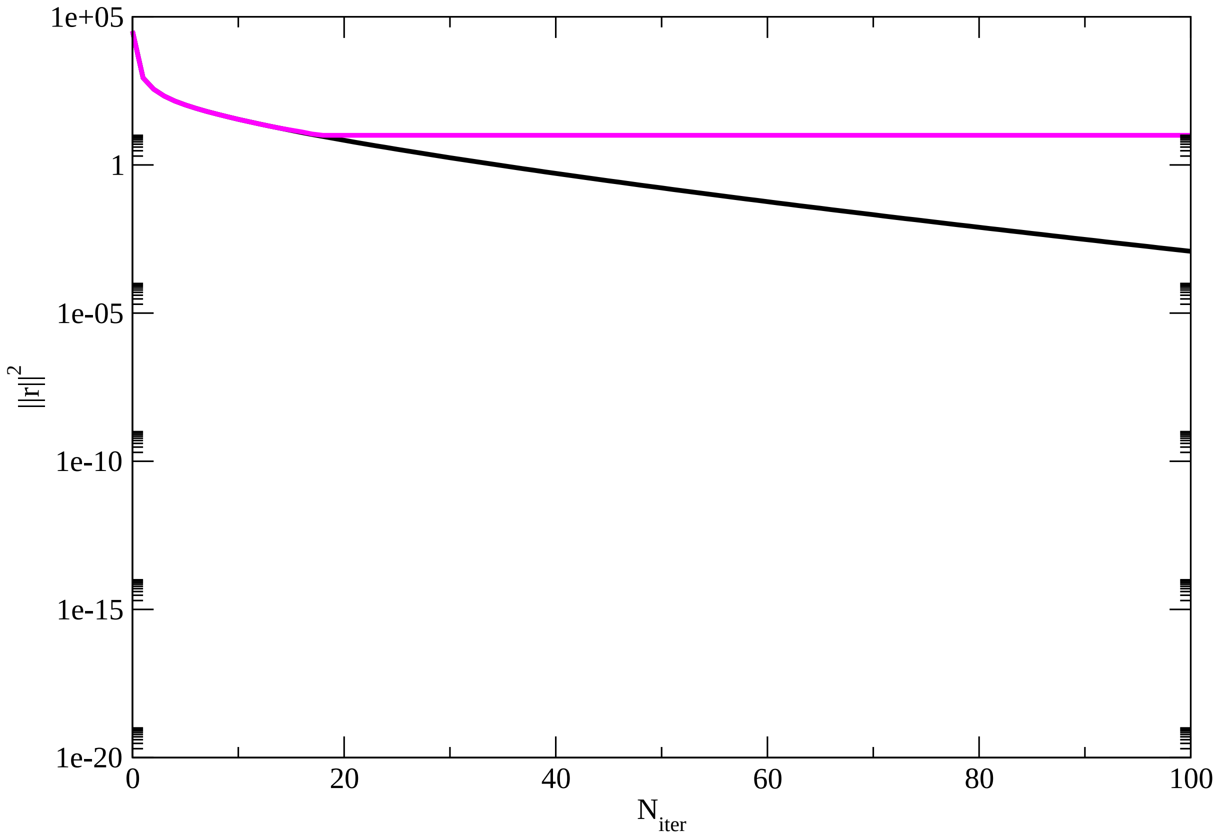
Example Setup: Gauge Laplacian

$2d$ 128×128 lattice, $m = 10^{-6}$



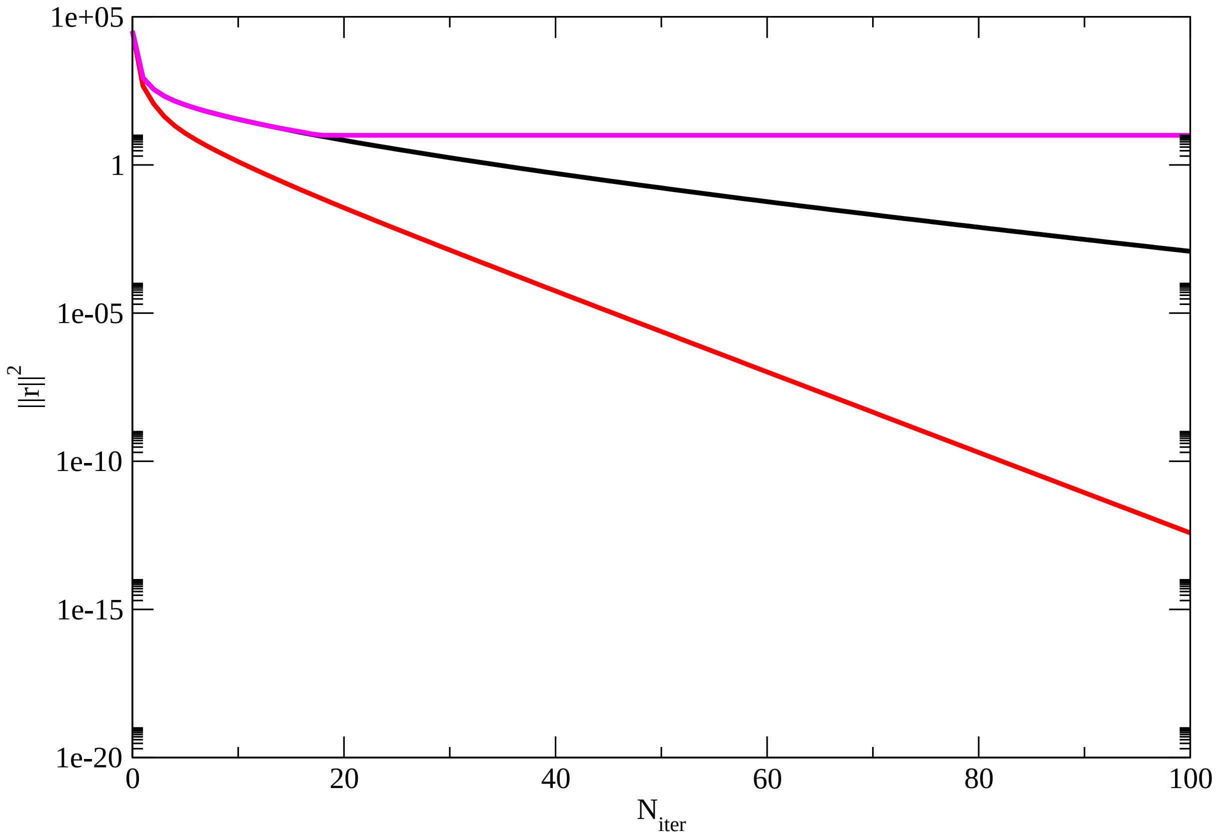
Example Setup: Gauge Laplacian

$2d$ 128×128 lattice, $m = 10^{-6}$



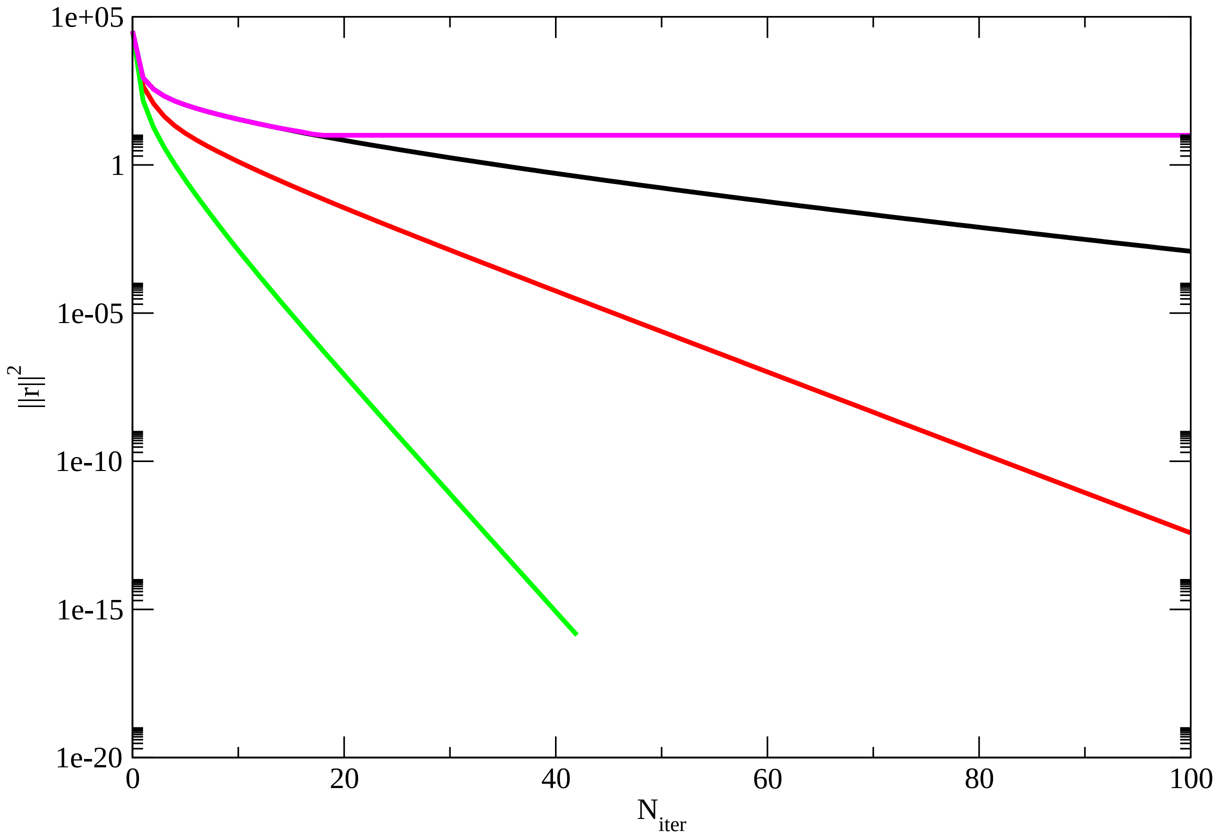
Example Setup: Gauge Laplacian

$2d$ 128×128 lattice, $m = 10^{-6}$



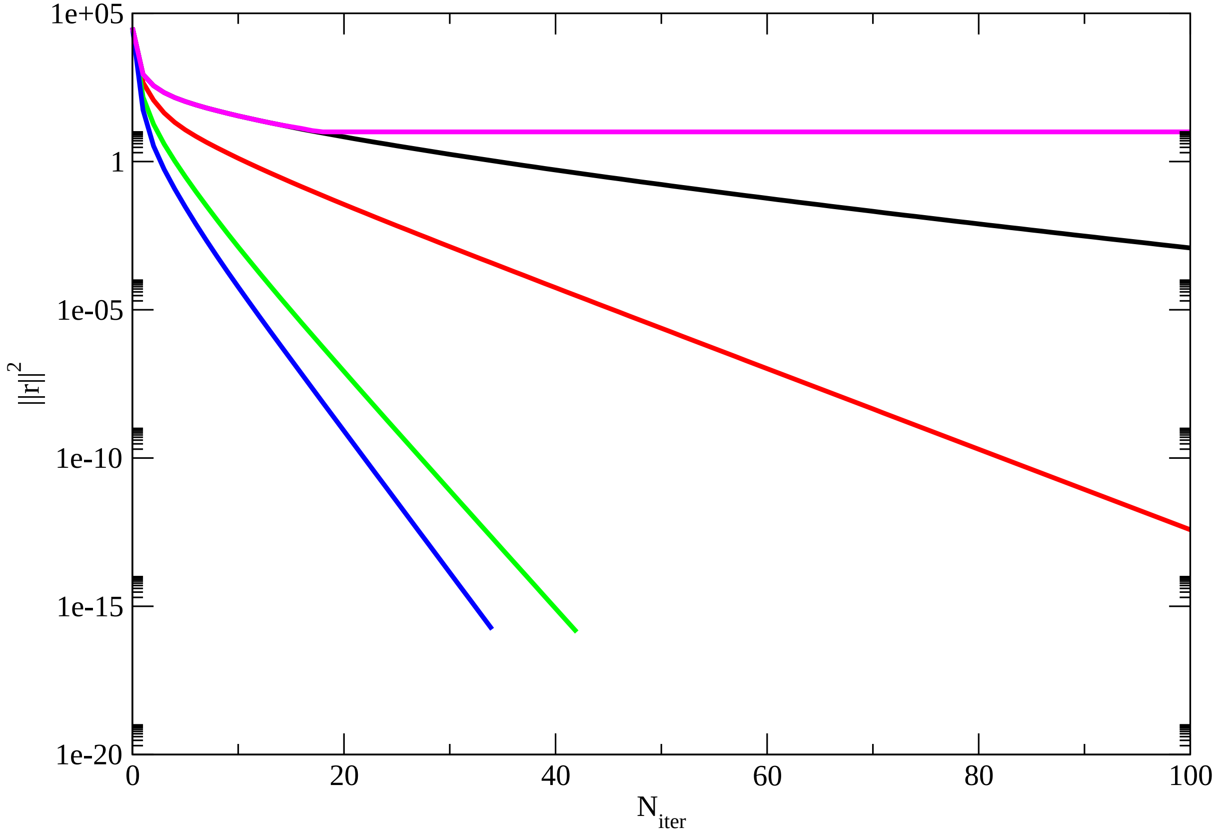
Example Setup: Gauge Laplacian

$2d$ 128×128 lattice, $m = 10^{-6}$



Example Setup: Gauge Laplacian

$2d$ 128×128 lattice, $m = 10^{-6}$



QCD Adaptive Multigrid

- αSA designed for problems without an underlying geometry
  - Uses algebraic "strength of connection" to block system
- QCD has regular geometric lattice with unitary connections
  - Use regular geometric blocking strategy (e.g.,$4^d \times N_s \times N_c$)
  - Coarse operator "looks" like a Dirac operator
- Dirac operator is not Hermitian Positive Definite
  - Multigrid convergence proof requires HPD operator
- Use normal equations $\Rightarrow A = D^\dagger D$
  - Critical slowing down gone in $2d$ (PRL 100:041601)
  - Coarse grid operator $P^\dagger(D^\dagger D)P$ not nearest neighbour
  - $D^\dagger D$ has squared condition number compared to D
  - More vectors to capture null space, cost $\propto N_v^2$
  - Coarse operator very expensive in $4d$

Multigrid on D

- Eigenvector decomposition of $D = |\psi_\lambda\rangle\lambda\langle\widehat{\psi}_\lambda|$
- Consider deflation: bi-orthogonality requires left-right projection

$$
\begin{aligned}
\mathcal{P} &= \left(1 - |\psi_\lambda\rangle\frac{1}{\lambda}\langle\widehat{\psi}_\lambda|\right) \\
&= \left(1 - |\psi_\lambda\rangle\langle\widehat{\psi}_{\lambda'}|D|\psi_\lambda\rangle^{-1}\langle\widehat{\psi}_{\lambda'}|\right) = \left(1 - P(RDP)^{-1}R\right)
\end{aligned}
$$

- Prolongation from right vectors, restriction from left vectors
- Left and right eigenvectors related through γ_5: $\widehat{\psi}_\lambda = \gamma_5\psi_{\lambda*}$

$$
R = P^\dagger\gamma_5 \qquad D_c = P^\dagger\gamma_5 DP
$$

- Leave chirality intact
  - $[\gamma_5, P] = 0$
  - γ_5 cancels out of correction
- Coarse grid operator defined as
  - D_c retains γ_5 Hermiticity
  - D_c is positive real if D is

$2d$ Results

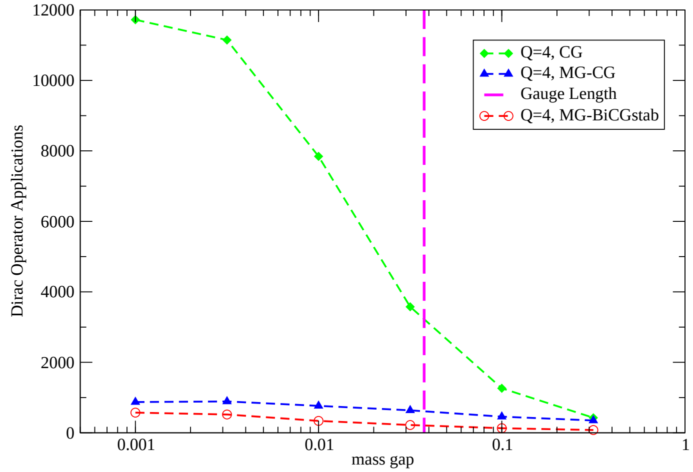
- 128×128 lattice, $\beta = 6, 10$, $\hat{m} = 0.001 - 0.5$

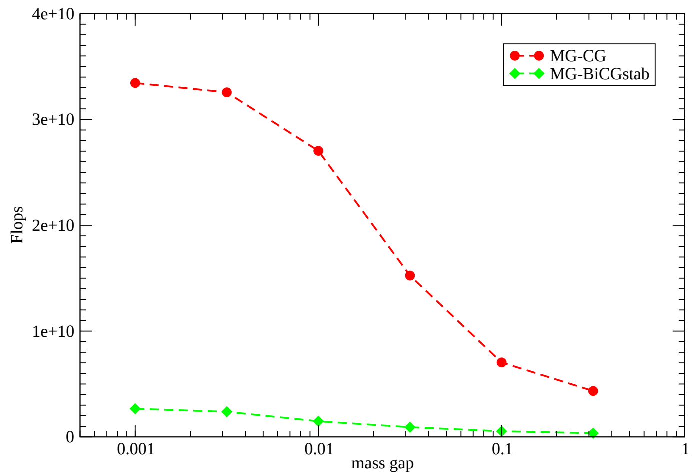
- MG setup run at lightest mass only

- $D^\dagger D$-MG algorithm
  - $4\times4(\times2)$ blocking, 3 levels, $N_v = 8$
  - Under-relaxed MR relaxation
  - Preconditioner for CG

- D-MG algorithm
  - 4×4 blocking, 3 levels, $N_v = 4$
  - Under-relaxed MR relaxation
  - Preconditioner for BiCGstab

- Results
  - Critical slowing down virtually gone
  - Weak dependence on β
  - D-MG superior to $D^\dagger D$-MG

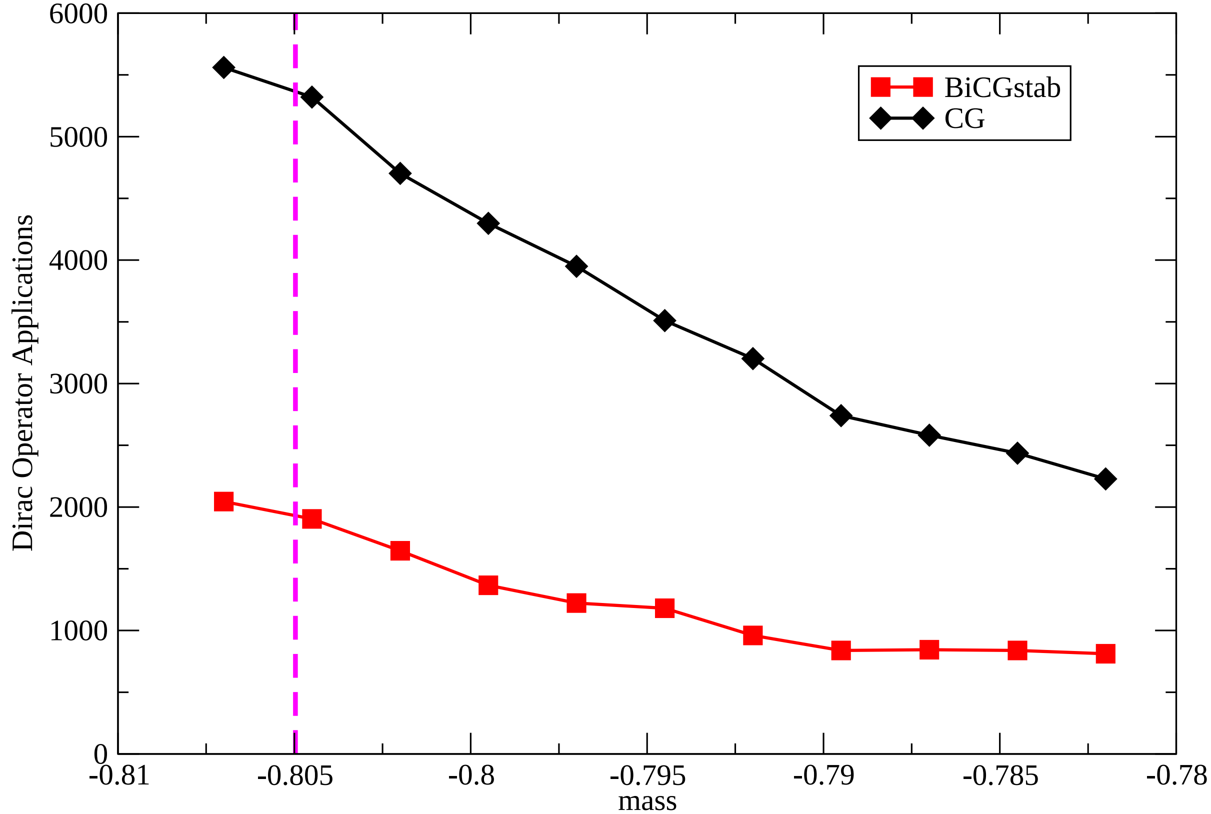




$4d$ Results

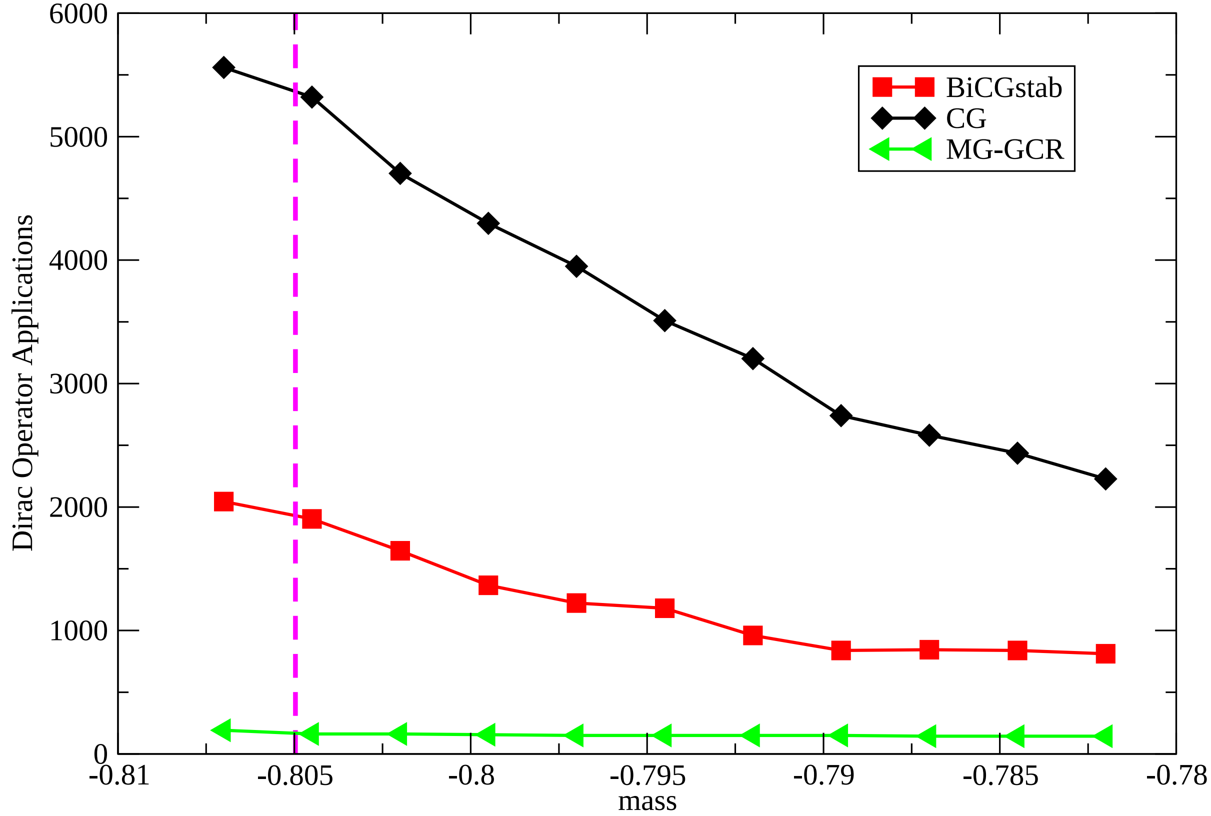
- Apply D-MG algorithm to full problem
- $16^3 \times 32$ lattice
  - $\beta = 6.0$, $m_{crit} = -0.8049$
- Compare against CG, BiCGstab
- D-MG Algorithm
  - $4^4(\times 3 \times 2)$ blocking, 3 levels
  - $N_v = 20$, (c.f. Lüscher)
  - Setup run at m_{crit}
  - Under-relaxed MR smoother
  - Preconditioner for GCR(50)

$4d$ Results



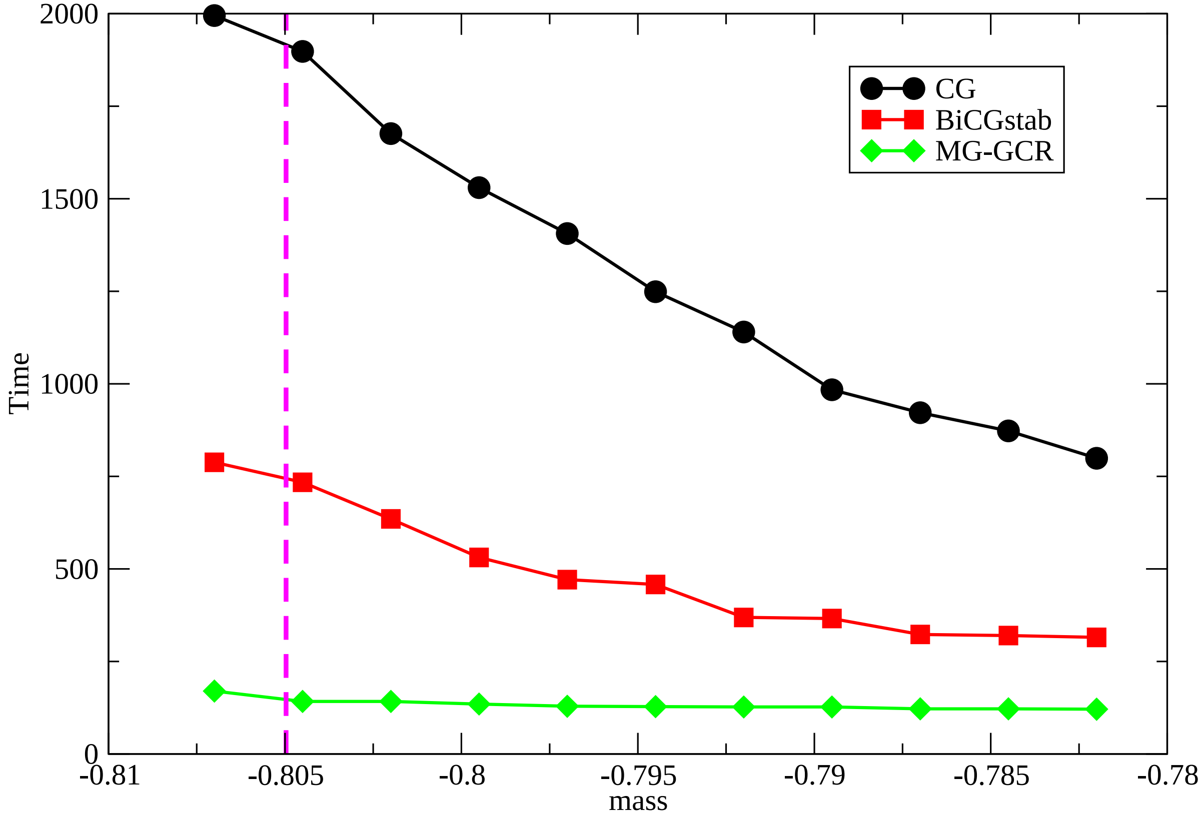
BiCGstab
CG
Dirac Operator Applications
mass
6000
5000
4000
3000
2000
1000
0
-0.81
-0.805
-0.8
-0.795
-0.79
-0.785
-0.78

$4d$ Results



BiCGstab
CG
MG-GCR
Dirac Operator Applications
mass
6000
5000
4000
3000
2000
1000
0
-0.81
-0.805
-0.8
-0.795
-0.79
-0.785
-0.78

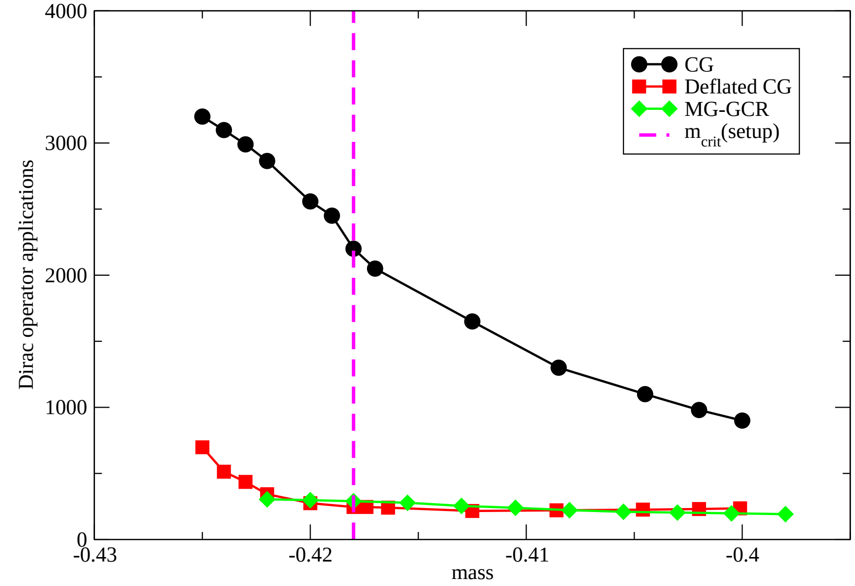
$4d$ Results



2000
1500
1000
500
0
Time
-0.81
-0.805
-0.8
-0.795
-0.79
-0.785
-0.78
mass
CG
BiCGstab
MG-GCR

Adaptive MG vs. Deflation (PRELIMINARY)

- $16^3 \times 64$ anisotropic Wilson lattice
  - $\beta = 5.5$
  - $m_{crit} = -0.4180$
- Adaptive MG
  - $N_v = 20$
- Eigen-vector deflation
  - $N_v = 240$ (Orginos and Stathopoulos)
- Order of magnitude reduction in N_v
- Very similar naïve performance
- Deflation actually $\approx \times 2$ faster but $10 \times N_v$
- Expect MG to be increasingly competitive at large volumes



Conclusions and Future Work

- Now have a variety of algorithms to tame critical slowing down
  - Eigenvector deflation
  - Local mode deflation
  - Adaptive Multigrid
- These have different ranges of applicability
  - All suitable for multiple rhs calculations
  - Local mode deflation also suitable HMC (MG also?)
- Staggered fermions
  - No problems expected here
  - Ideal since very large lattices
- Chiral fermions (overlap, domain wall)
  - Essentially a highly indefinite Wilson operator
  - D-MG requires reformulation

Local Mode Deflation vs. Multigrid

- MG is multiplicative cancellation

$$(PD_c^{-1}P^\dagger)Dx = (PD_c^{-1}P^\dagger)b$$

- Local mode deflation is additive cancellation

$$(1 - DPD_c^{-1}P^\dagger)Dx = (1 - DPD_c^{-1}P^\dagger)b$$

- Expect MG to require less accurate D_c^{-1} solution
  - Local mode deflation D_c^{-1} accuracy 10^{-6}
  - MG D_c^{-1} accuracy?