Multigrid Methods and Lattice QCD

James Brannick

The Pennsylvania State University

June 19, 2014

Outline

- Lattice QCD – solvers
- Review of Multigrid iterative solvers
- Bootstrap / adaptive Multigrid solvers for the Dirac system

Participants in QCD-Multigrid collaboration

Argonne	J. Osborne	NVIDIA	M. Clark
	B. Smith	PSU	J. Brannick
Boston Univ.	R. Brower	TU Delft	K. Oosterlee
	C. Rebbi	Tufts	S. MacLachlan
Columbia	A. Bassen	UCLA	A. Brandt
	D. Keyes	Univ. Utah	O. Livne
CU Boulder	M. Brezina	Univ. Wuppertal	A. Frommer
	C. Ketelesen		K. Kahl
	T. Manteuffel		Björn Leder
	S. McCormick		M. Rottman
	J. Ruge		
LLNL	R. Falgout		
	C. Schroeder		

Funding: DOE: SciDAC, ASCR; NSF: PetaApps, ITR, Computational Mathematics; DFG: SFB-TR 55

The Dirac PDE

$$\mathcal{D}[\mathcal{A}]\psi = \sum_{\mu} \gamma^{\mu} \partial_{\mu} \psi + m\psi \quad = \quad f$$

- $\gamma_\mu \in \mathbb{C}^{4\times 4}$ satisfy $\{\gamma_\mu, \gamma_\nu\} = \delta_{\mu,\nu} I$ and $\gamma_5 = \prod_{\mu=1}^{4} \gamma_\mu = \gamma_5^*$
- $\psi_{s,c}^{j}:\ s = 1,2,3,4,\ c = 1,2,3,\ j = 1,..,n_f\ (n_f = 1)$

Wilson's discretization

$$D_{x,y} = \delta_{x,y} - \kappa \sum_{\mu=1}^{d} (1-\gamma_\mu) \otimes U_x^{\mu}\, \delta_{x+\mu,y} + (1+\gamma_\mu) \otimes U_{x-\mu}^{\mu *}\, \delta_{x-\mu,y}$$

- Removes spurious zero modes, but breaks chiral symmetry
- Basic building block of chiral Overlap and Domain Wall operators

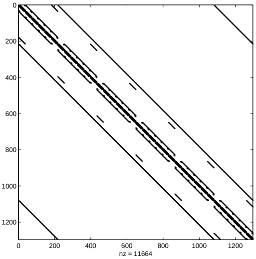
The Dirac-Wilson matrix

$$D_{x,y} = \delta_{x,y} - \kappa \sum_{\mu=1}^{d} (1-\gamma_\mu) \otimes U_x^\mu \, \delta_{x+\mu,y} + (1+\gamma_\mu) \otimes U_{x-\mu}^{\mu *} \, \delta_{x-\mu,y}$$

- $D = I - \kappa D_0$ is positive real for $0 \leq \kappa < \kappa_c$
- nearest neighbor coupling on *hypercubic* lattice embedded in a 4d torus
- 12 variables per grid point
- $n = 12 \cdot n_1 \cdot n_2 \cdot n_3 \cdot n_4$
- $n_i = 16 \ldots 128$
- *Interesting case*: $\kappa \to \kappa_c \Rightarrow$

$$m = \frac{1}{2}\left(\frac{1}{\kappa} - \frac{1}{\kappa_c}\right) \approx 0$$

- Performance of Krylov methods degrades as $\kappa \to \kappa_{cr}$



2d Dirac-Wilson matrix: block-spin form

Wilson system consists of a sum of two parts, a stabilization term

$$A_{xy} = -\frac{1}{2}\sum_{\mu=1}^{d}\left(U_x^{\mu}\,\delta_{x+\mu,y} + U_{x-\mu}^{\mu *}\,\delta_{x-\mu,y}\right) + (d+m)\delta_{x,y}$$

referred to as the Gauge Laplacian, and a central covariant difference approximation of the Dirac system

$$B_{xy} = \frac{1}{2}\sum_{\mu=1}^{d}\gamma_\mu \otimes U_x^{\mu}\,\delta_{x+\mu,y} - \gamma_\mu \otimes U_{x-\mu}^{\mu *}\,\delta_{x-\mu,y}.$$

where for a 2d lattice

$$\gamma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \text{and} \quad \gamma_2 = \begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix}$$

and D can be written as a block matrix

$$D = \begin{pmatrix} A & B \\ -B^* & A \end{pmatrix}$$

Symmetries of the Wilson fermion matrix

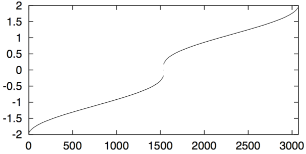
γ_5-*Symmetry*:

$$\Gamma_5 D = D^* \Gamma_5,$$

where

$$\Gamma_5 = I \otimes (\gamma_5), \quad \gamma_5 = \prod_{i=1}^{n_s} \gamma_i$$

- $\lambda \in \mathrm{spec}(D) \Rightarrow \bar{\lambda} \in \mathrm{spec}(D)$
- $Q = \Gamma_5 D$ is hermitian and *maximally indefinite*
  worst case for BiCGSTAB, GMRES, etc...

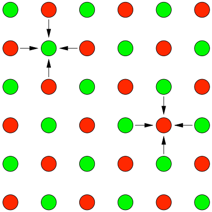


S. M. Pickles, *Algorithms in Lattice QCD*, Ph.D., University of Edinburgh, 1998, UKQCD

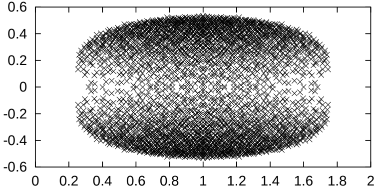
Odd-even system

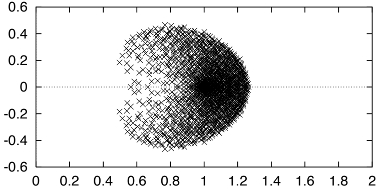
- grid points x are odd or even (= red or green).
- odd-even-ordering yields

$$D_0 = \begin{pmatrix} 0 & D_{eo} \\ D_{oe} & 0 \end{pmatrix}$$



Spectrum of the Dirac-Wilson matrix and the odd-even Schur-complement

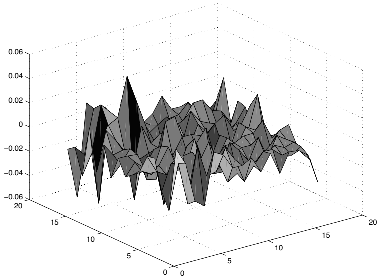


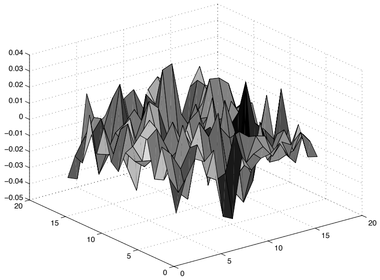


Specta of D and odd-even Schur compl. for 4^4 grid (realistic configuration)

Multigrid for QCD circa 2000

- Gauge field U is not geometrically smooth $\Rightarrow$ near kernel is locally oscillatory





$\Rightarrow$ Constant preserving (algebraic) multigrid methods completely fail

Multigrid for QCD circa 2000

R. Ben-Av et al, *Fermion simulations using parallel transported multigrid*, Phys. Lett. **B253** (1991), pp. 185–192.

R. Ben-Av, M. Harmatz, S. Solomon, and P. G. Lauwers, *Parallel transported multigrid for inverting the dirac operator: Variants of the method and their efficiency*, Nucl. Phys.**B405** (1993), pp. 623–666.

A. Brandt, *Multigrid methods in lattice field computations*, Nucl Phys. Proc. Suppl. **26** (1992), pp. 137–180.

R. C. Brower, T. Ivanenko, A. R. Levi, and K. N. Orginos, *Chronological inversion method for the dirac matrix in hybrid monte carlo*, Nucl. Phys. **B484** (1997), pp. 353–374.

R. C. Brower, E. Myers, C. Rebbi, and K. J. M. Moriarty, *The multigrid method for fermion calculations in quantum chromodynamics*, (1993), Print-87-0335, IAS,PRINCETON.

R. C. Brower, K. J. M. Moriarty, C. Rebbi, and E. Vicari, *Multigrid propagators in the presence of disordered u(1) gauge fields*, Phys. Rev. **D43** (1991), pp. 1974–1977.

R. C. Brower, C. Rebbi, and E. Vicari, *Projective multigrid for propagators in lattice gauge theory*, Phys. Rev. Lett. **66** (1991) pp. 1263–1266.

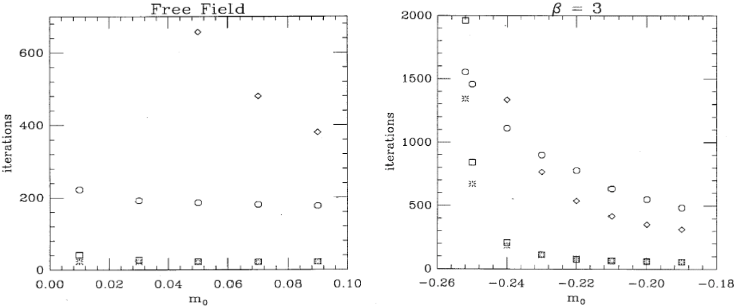
R. C. Brower, et. al., *Projective multigrid for Wilson fermions*, Nucl. Phys. **B366** (1991), pp. 689–705.

P. Hasenfratz, *Prospects for perfect actions*, Nucl. Phys. Proc. Suppl. **63** (1998), pp. 53–58.

A. Hulsebos, J. Smit, and J. C. Vink, *Multigrid inversion of the staggered fermion matrix with u(1) and su(2) gauge fields*, in Juelich 1991, Proceedings, Fermion algorithms (QCD161:W573:1991), pp. 161-168.

Many others ...

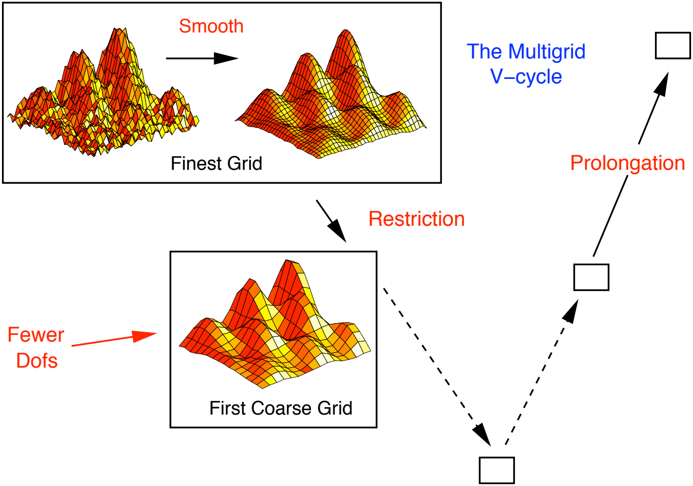
Multigrid for QCD circa 2000



Jacobi (Diamond), CG (circle), MG V-cycle (square), W-cycle (star)

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

Basic Multigrid components



Multigrid Methods

- Geometric Multigrid methods
  - Specialized, e.g., for PDEs they are integrated with the finite element / volume / difference discretization
  - Highly (maximally) efficient iterative solvers
  - Limited applicability, e.g., with respect to the geometry and parameters of the problem
- Single-grid methods: geometric-algebraic methods
  - More generally applicable
  - Slight loss in efficiency vs. GMG resulting from use of auxiliary-grid (mesh) and some algebraic techniques, as needed
  - Intended for grid-based problems
- No-grid methods: algebraic methods
  - Black-box iterative solver for sparse M-matrix systems
  - Convergence and complexity difficult to control in practice

A two-grid method

For a given fine-level system of equations $D\psi = f$ (D HPD) defined on a space V_h, a two-level solver is described in terms of its two main components

1. a smoother M
2. a coarse space V_H related to P the prolongation operator and R the restriction operator

Given an initial guess u^0, a single iteration of a two–grid method is as follows:

1. Fine-level smoothing: $\tilde{u} = u^0 + M^{-1}(f - Du^0)$
2. Coarse-level correction: solve $D_H e_H = r_H$ with $r_H = R(f - D\tilde{u})$
3. Update: $u^1 := \tilde{u} + Pe_H$

Oftentimes, $D_H = RDP$ and $R = P^*$ if D HPD

Two-grid theory

It follows that $u - u^1 = E_{TG}(u - u^0)$ where

$$E_{TG} = (I - \pi_D)(I - M^{-1}D), \quad \pi_D := PD_H^{-1}P^*D$$

A sharp estimate of the convergence of a two-grid method is $\|E_{TG}\|_D^2 = 1 - 1/K(P)$, where

$$K(P) = \sup_v \frac{\|(I - \pi_{\widetilde{M}})v\|_{\widetilde{M}}^2}{\|v\|_D^2} \quad \widetilde{M} := M^*(M^* + M - D)^{-1}M, \quad \|v\|_D^2 := (Dv, v)$$

At least three different approaches are possible in choosing the components of a two–level method:

1. For a fixed P, construct a suitable M – geometric methods
2. For a fixed M, optimize the choice of P – (adaptive) algebraic methods
3. Given certain measures on the suitability of M and P, simultaneously construct both

Weak approximation property

The suitability of V_H is measured via an approximation property. Assuming $\widetilde{M} \approx \|D\| I$ and noting that $\pi_{\widetilde{M}}$ is the $\widetilde{M}$-orthogonal projector onto Range(P), we have

$$\begin{aligned} \|(I - \pi_{\widetilde{M}})v\|^2_{\widetilde{M}} &\leq \|(I - \pi)v\|^2_{\widetilde{M}} \\ &\lesssim \|D\| \|(I - \pi)v\|^2 \quad \text{all} \quad v \in V \end{aligned}$$

Thus,

$$K(P) = \sup_{v \neq 0} \frac{\|(I - \pi_{\tilde{M}})v\|^2_{\tilde{M}}}{\|v\|^2_D} \lesssim \sup_{v \neq 0} \frac{\|D\| \; \|(I - \pi)v\|^2}{\|v\|^2_D},$$

where $\pi := P(P^* P)^{-1} P^*$

Smooth error w satisfies

$$\frac{\|Dw\|}{\|w\|} \approx \min_v \frac{\|Dv\|}{\|v\|}$$

No-grid (algebraic) methods

- "Algebraic" stands for the fact that all the tools of the method are constructed solely on the basis of the original matrix M in a setup phase
- Coarse space is constructed automatically within the algorithm, level by level, in a (hopefully) computationally optimal setup procedure which involves
  1. Picking a set of coarse variables, i.e., set of indices $\Omega_H = \{i_1, \ldots, i_{n_H}\} \rightarrow$ graph theoretic approaches
  2. Definiing $V_H = \text{span}\{\psi_k\}_{k=1}^{n_H}$ such that each ψ_k is supported in Ω_k, for a vector: $\Omega_k \subset \{1, \ldots, n\} \rightarrow$ null space of the system matrix
- Each of the V_H (or V_i obtained recursively) must satisfy certain properties, related to the convergence of the overall algorithm. As subspaces are built "on the fly", multilevel theory for the convergence of such algorithms is very difficult

Operator-dependent interpolation

Setup algorithm

- Given D HPD and $\Omega = \{1, ..., n\}$, select $\Omega_H = \{1, ..., n_H\}$, $n_H < n$
- Compute entries of $P : V_H \mapsto V_h$, $R : V_h \mapsto V_H$, and $D_H = RAP$

  1. Classical AMG: $\Omega_H \subset \Omega$ and
  $$P = \begin{bmatrix} W \\ I \end{bmatrix} \; \} \, \Omega_H \quad ,$$
  where $W \in \mathbb{C}^{m \times n_H}$ with $m = n - n_H$
  2. Aggregation AMG: $\Omega = \cup_{i=1}^{n_H} \Omega_i$,
  $$P_{ji} = \begin{cases} 1 & \text{for} \quad j \in \Omega_i \\ 0 & \text{for} \quad j \notin \Omega_i \end{cases} \qquad i = 1, ..., n_H$$
  In Smoothed Aggregation, additional smoothing step applied to interpolation:
  $$P \leftarrow SP, \quad S = I - \tau D$$

- Typically, for PDEs use constant preserving P, i.e., select P to ensure that there exists v_H such that $Pv_H = 1$ for some vector of coefficients v_H

Multilevel iterative solvers in lattice computations

Solver challenges:

- Systems are nearly singular
- Non-hermitian and positive real or hermitian and maximally indefinite
- Near kernel is unknown: highly oscillatory with oscillations dependent upon fluctuations in background gauge fields $\leftarrow$ heterogeneity of covariant derivatives
- Large near kernel dependent upon on topology

What is needed? A method that can

- Approximate several "arbitrary" kernel components to within desired level of accuracy
- Extract the components from the algebraic problem (suitable smoother required)
- Automatically construct coarse–level basis

Aggregation MG solver for Dirac-Wilson system

Given standard geometric blocking into m^d aggregates and matrix $X := [x^{(1)}, ..., x^{(r)}]$ such that $Dx^{(i)} \approx 0$, interpolation defined as

$$P = \left(\begin{array}{c|c|c} X_1 & & \\ \hline & \ddots & \\ \hline & & X_{n_H} \end{array} \right) \qquad \begin{array}{l} \}\Omega_1 \to Q_1 R_1 \\ \vdots \\ \}\Omega_{n_H} \to Q_{n_H} R_{n_H} \end{array}$$

- $|\Omega_i| = m^d$, $i = 1, ..., n_H$
- $P^* P = Q^* Q = I$
- $P X_H = PR = X$
- $P = SP$, $S = I - \tau D$
  (smoothed aggregation)

Generalized SA, P. Vanek, M. Brezina, J. Mandel, 2001

Basic Adaptive (A)MG Algorithm

For $\ell = 0, ..., J$

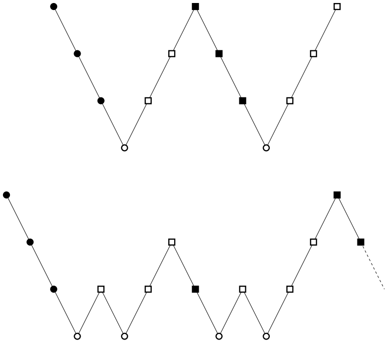
While $\|I - B_\ell^{-1} D_\ell\|_{\text{est}}$ increasing: $x_\ell \leftarrow (I - B_\ell^{-1} D_\ell) x_\ell$

If $\|I - B_\ell^{-1} D_\ell\|_{\text{est}}$ is large

recalibrate interpolation based on (new) x_ℓ

recompute coarse-grid operator

recurse



- ● Relax on $Dx = 0$, x rand, if $\|I - B^{-1} D\|_{\text{est}} > tol$, set $X = [x]$ update P
- ○ Relax on $Dx = 0$
- □ Iterate on $Dx = 0$
- ■ Iterate on $Dx = 0$, if $\|I - B^{-1} D\|_{\text{est}} > tol$, set $X = [X\ x]$ update P

2d Dirac-Wilson system: adaptive Smoothed Aggregation MG solver

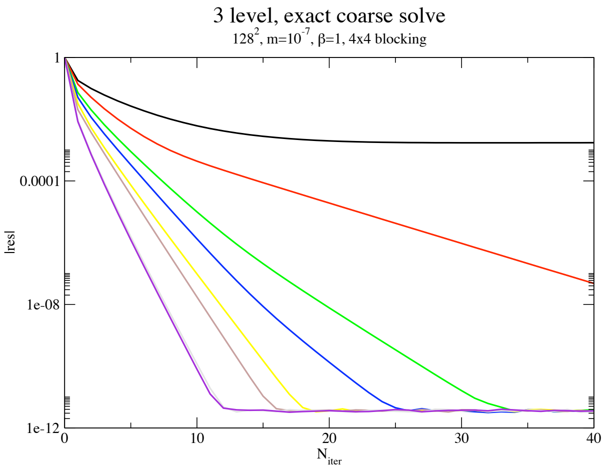
$$D_{x,y} = -\frac{1}{2}\sum_{\mu=1}^{2}(1-\gamma_\mu)\otimes U_x^\mu\,\delta_{x+\mu,y} + (1+\gamma_\mu)\otimes U_{x-\mu}^{\mu *}\,\delta_{x-\mu,y} + (4+m)\delta_{x,y}$$

with $U(x) \in U(1)$ on $2d$ space-time lattice

- Solve $D^* D\psi = D^* f$
- Use adaptive $V(4,4)$-cycle setup with Gauss Seidel smoother based on current hierarchy to test solver
  - Previously found error components quickly reduced and “new” error vector rich in unresolved components of the error
- Augment hierarchy to preserve additional vector space
  - Use SA framework to cut vectors $x_1, x_2, ..., x_r$ into blocks and on each block use QR to define augmented – multiple vector preserving – P
  - Add more vectors until satisfactory solver found, where each vector corresponds to an extra dof per coarse site

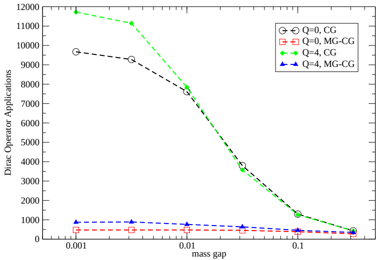
Adaptive Multigrid for Lattice QCD, Br., et al, *PRL*, 2008, arXiv:0707.4018

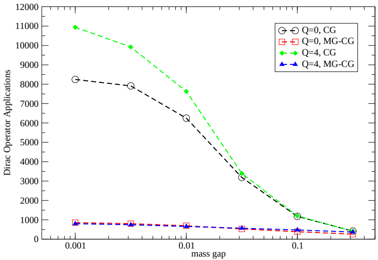
Results for 2d Dirac-Wilson system with $U(1)$ background



Number of iterations versus residual for different number of TVs

Adaptive SA results...





Number of applications of D^*D needed to reduce relative residual to $O(10^{-8})$

- 128×128 lattice, $\beta = 1, 6$, 4×4 blocking, 3 levels, 8 vectors
- Use GS for smoothing with exact solve on coarse grid
- Standard algorithm requires hpd $\implies D^*D$
- Apply MG V-cycle as a preconditioner to CG
- Compare total number of D^*D applications on fine level only with plain CG

J. Brannick, R. Brower, M. Clark, J. Osborn, and C. Rebbi, Adaptive multigrid algorithm for lattice QCD, Phys. Rev. Lett., published 28 January 2008, Issue 4, Volume 100, article 041601

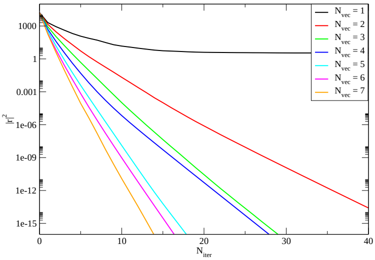
Solving the non-hermitian system

- Solve D directly, instead of D^*D
  - Better sparsity of $D_H = RMP, R \neq P^*$ and less vectors required to define P
  - Reduce the setup cost; for normal equations setup requires equivalent of 3-4 CG inversions
  - $D_H = RDP, R \neq P^*$
- Use variant of MG solver for D^*D for D
  - γ_5 symmerty: $Q = \Gamma_5 D = D^* \Gamma_5 = Q^*$ such that

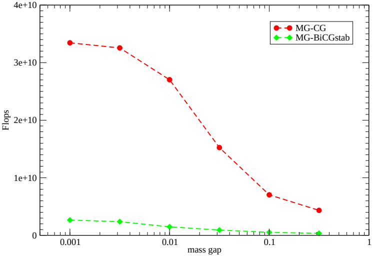
    $$Q = \sum_{i=1}^{N} \lambda_i v_i v_i^* \Rightarrow D = \sum_{i=1}^{n} \lambda_i (\Gamma_5 v_i) v_i^*$$

    so that left and right eigenvectors (singular) vectors are related by $u_i = \Gamma_5 v_i$
  - Coarse-grid operator RDP, with P based on v_i and R based on $u_i = \Gamma_5 v_i$
  - Leads to Galerkin coarse-level operator and preserves γ_5-symmetry on coarse levels
  - Use adaptivity with MinRes smoother to compute the nearly singular vectors and incorporate them into *unsmoothed* aggregation solver
  - Compare with previous results obtained for the normal equations

2d Dirac-Wilson system



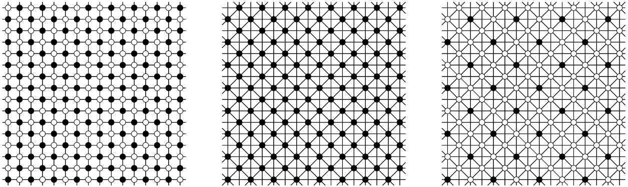
(a) Number of iterations versus residual for different number of TVs



(b) Flops needed in setup and solve to reduce residual to $tol = 10^{-8}$ for various values of the mass

Figure. 2d Dirac Wilson system with $\beta = 1$ on 128×128 lattice

Bootstrap MG solver for Dirac-Wilson system



(c) Odd-even coarsening of the grid

(d) Structure of the even system

(e) Full coarsening of the even grid

Figure : Coarsening of the grid of even points and the odd-even reduction

Given matrix $X = [x^{(1)}, ..., x^{(r)}] = [V \;\; W]$, rows of interpolation, p_i, $i \in \Omega \setminus \Omega_H$, are defined such that

$$\mathcal{L}(p_i) = \sum_{\kappa=1}^{r} \omega_\kappa \left(x^{(\kappa)}_{\{i\}} - \sum_{j \in C_i \cup \{i\}} (p_i)_j \, x^{(\kappa)}_{\{j\}} \right)^2 \mapsto \min,$$

where weights $\omega_\kappa \sim \frac{\|x^{(\kappa)}\|}{\|Dx^{(\kappa)}\|} \in \mathbb{R}^+$ and $r = |\mathcal{V}| + |\mathcal{W}|$

Bootstrap setup - multilevel eigensolver

Assuming no a priori information on low modes available

- Smoother and initial test vectors given

$$v^{(s)} = G^{\eta}\widetilde{v}^{(s)}, \quad \widetilde{v}^{(s)} \text{ random}$$

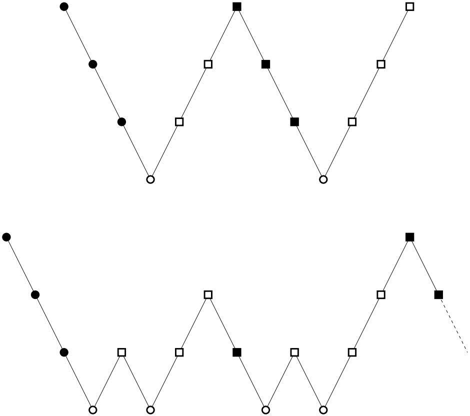
- Observation ($P_\ell = P_1^0 \cdots P_\ell^{\ell-1},\ D_\ell = P_\ell^H D P_\ell,\ T_\ell = P_\ell^H P_\ell$)

$$\frac{\langle w_\ell, w_\ell \rangle_{D_\ell}}{\langle w_\ell, w_\ell \rangle_{T_\ell}} = \frac{\langle P_\ell w_\ell, P_\ell w_\ell \rangle_D}{\langle P_\ell w_\ell, P_\ell w_\ell \rangle_2}$$

Bootstrap Idea

Eigenpairs (w_ℓ, λ_ℓ) of (D_ℓ, T_ℓ) $\longrightarrow$ Eigenpairs $(P_\ell w_\ell, \lambda_\ell)$ of D +interpolation error

Bootstrap multilevel eigensolver – cycling



● Relax on $Dv = 0,\ v \in \mathcal{V}$

○ Compute $\mathcal{W}$ such that $Dw = \lambda Tw, w \in \mathcal{W}$

□ Relax on $(D - \lambda T)w = 0,\ w \in \mathcal{W}$

■ Relax on $Dv = 0,\ v \in \mathcal{V}$ and $(D - \lambda T)w = 0,\ w \in \mathcal{W}$

Solving the non-hermitian system: BAMG

Recall the abstract smoothing property: for smooth error, e, $\frac{\|De\|}{\|e\|} \approx \min_v \frac{\|Mv\|}{\|v\|}$. It is simple to show that

$$\sigma_1 \leq |\lambda| \leq \sigma_n,$$

where $\sigma_1 \leq \sigma_1 \leq ... \leq \sigma_n$ are the ordered singular values of D and λ is any of its eigenvalues. Thus, smooth error dominated by singular vectors (and eigenvectors?), suggesting use of a multilevel SVD solver in bootstrap cycle

Modifications to the setup:

- Kaczmarz smoother for D^* and D in the bootstrap and adaptive cycles to compute left and right singular vectors, respectively
- Weighted LS formulation to compute restriction (left singular vectors) and interpolation (right singular vectors)
- Multilevel singular value solver in the bootstrap approach where on the coarsest level we solve the symmetric eigenvalue problem directly:

$$\begin{pmatrix} & D_L \\ D_L^H & \end{pmatrix} \begin{pmatrix} U & U \\ V & -V \end{pmatrix} = \begin{pmatrix} T_L & \\ & Q_L \end{pmatrix} \begin{pmatrix} U & U \\ V & -V \end{pmatrix} \begin{pmatrix} \Sigma & 0 \\ 0 & -\Sigma \end{pmatrix}$$

  where $D_l = R_l D P_l$, $Q_l = R_l R_l^H$, and $T_l = P_l^H P_l$ with

$$\begin{aligned} P_l &= P_1^0 \cdot \ldots \cdot P_l^{l-1}, \\ R_l &= R_{l-1}^l \cdot \ldots \cdot R_0^1, \quad l = 2, \ldots, L \end{aligned}$$

BAMG variant for 4d Dirac-Wilson system of QCD

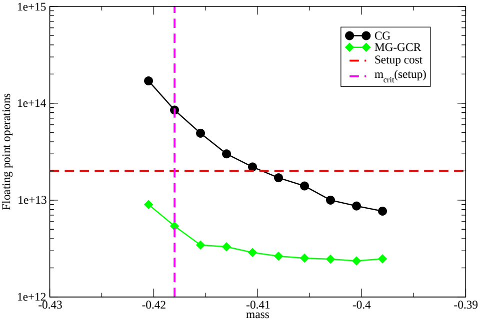


Figure. Flops needed to reduce residual to 10^{-8} versus mass

- $32^3 \times 96$ lattice, $\beta = 6$, 4^4 blocking, 3 levels
- Apply aggregation-based MG solver as preconditioner to GCR

Concluding remarks

Summary

- Bootstrap / Adaptive MG provide effective solvers for Dirac-Wilson systems effectively removing critical slowing down (Rottman)
- Extensions of methodology to chiral models (and other stochastic PDEs) underway with promising results (Kahl)

Outlook

- Parallel version(s) of Bootstrap algorithm under development (e.g., in Hypre)
- Smoothing analysis and two-grid theory in progress, both complicated by fact that D is non-normal matrix