Multigrid Methods for Chiral Fermions

Peter Boyle*, Azusa Yamaguchi (+ input from Daniel Richtmann, Tilo Wettig)
Brookhaven National Laboratory, University of Edinburgh

- Faster multigrid Chebyshev setup
- First cross over of setup + solve faster than red black CGNR
- Detailed comparison of arXiv:1611.06944 and arXiv:2004.07732 in D=4 QCD.
- Aim towards next generation of 2+1+1f HMC simulations

With thanks to USQCD ECP solver call participants (esp. Brower, Clark, Weinberg)

Moebius Domain Wall Fermions

$$D_{ov}(m, L_s) = \left[\frac{1+m}{2} + \frac{1-m}{2}\gamma_5 \tanh L_s \tanh^{-1} H_M\right]$$

$$D^5_{GDW} = \begin{pmatrix} D_+ & -D_-P_- & 0 & \dots & 0 & mD_-P_+ \\ -D_-P_+ & \ddots & \ddots & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \ddots & 0 & \vdots \\ \vdots & 0 & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \ddots & \ddots & -D_-P_- \\ mD_-P_- & 0 & \dots & 0 & -D_-P_+ & \end{pmatrix}$$

$$\begin{aligned} D_+ &= (bD_W + 1) \\ D_- &= (1 - cD_W) \\ H_M &= \gamma_5 \frac{(b+c)D_W}{2+(b-c)D_W} \end{aligned}$$

Shamir DWF case: $b = 1,\ c = 0$

$$c = 0 \Rightarrow H_{GDW} = \gamma_5 R_5 D_{GDW} = \Gamma_5 D_{GDW}$$

Hierarchically deflated conjugate gradient : arXiv:arXiv:1402.2585

Why not HDCG? coarsen $(D_{DWF})_{oo} - (D_{DWF})_{oe}(D_{DWF})_{ee}^{-1}(D_{DWF})_{eo}$

- Significant speed up for valence DWF on BlueGene/Q
- Not as significant as exact eigenvector deflation with 2000 low modes
- Used in UKQCD analysis on small memory machines
- Next-to-next-to-next-nearest neighbour coarse space (81 point stencil)
- Deflate coarse space
- Non-recursive
- Too expensive to set up for use in HMC

Cohen/Brower/Clark/Osborne : coarsen $D_{DWF}^{\dagger} D_{DWF}$ arXiv:1205.2933 (17 point stencil)

Hierarchically deflated conjugate residual : arXiv:1611.06944

PROCEEDINGS OF SCIENCE

Hierarchically deflated conjugate residual

Azusa Yamaguchi[*†]
University of Edinburgh
E-mail: ayamaguc@staffmail.ed.ac.uk

Peter A Boyle
University of Edinburgh
E-mail: paboyle@ph.ed.ac.uk

We present a progress report on a new class of multigrid solver algorithm suitable for the solution of 5d chiral fermions such as Domain Wall fermions and the Continued Fraction overlap. Unlike HDCG [1], the algorithm works directly on a nearest neighbour fine operator. The fine operator used is Hermitian indefinite, for example $\Gamma_5 D_{dwf}$, and convergence is achieved with an indefinite matrix solver such as outer iteration based on conjugate residual. As a result coarse space representations of the operator remain nearest neighbour, giving an 8 point stencil rather than the 81 point stencil used in HDCG. It is hoped this may make it viable to recalculate the matrix elements of the little Dirac operator in an HMC evolution.

- Generate 5D null space $\Gamma_5 D_{DWF} \phi_i \sim 0$
- Coarsen with

  $$\phi_i^{\pm} = 1 \pm \Gamma_5 \phi_i$$

  Restrict to blocks b

  $$\mathbb{P} = |\phi_i^{b\pm}\rangle$$

- Coarse space is 4-dimensional
- Coarse space is nearest neighbour - aim for HMC
- Coarse operator

  $$\hat{H}_{DWF} = \mathbb{P}^\dagger \Gamma_5 D_{DWF} \mathbb{P} = \mathbb{P}^\dagger H_{DWF} \mathbb{P}$$

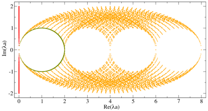
- Then

  $$\hat{H}^\dagger_{DWF} \hat{H}_{DWF} = (\mathbb{P}^\dagger H_{DWF} \mathbb{P})^2 = \mathbb{P}^\dagger D^\dagger_{DWF} \mathbb{P}\mathbb{P}^\dagger D_{DWF} \mathbb{P}$$

- Outer GCR, smoothers and (deflated) coarse solve based on normal equations
- As nearest neighbour it is recursive in principle, but prefer to deflate repeated inner solves

Hierarchically deflated conjugate residual : arXiv:1611.06944

Rationale: Wilson fermions $\mathrm{Re}\lambda \geq 0$ in "Hamburger" plot:



DWF spectrum shifted placing zero in the centre of the first opening.

- Violates the folklore present in numerical analysis of the *half-plane condition*.
  - In the infinite volume the spectrum becomes dense
  - Must approximate $P(z) \to \frac{1}{z}$ over a region in the complex plane *encircling* the pole zero
- *Impossible* to reproduce the phase behaviour around pole with a polynomial

CGNE: multiply by $\bar{z} \Rightarrow$ real, pos def:

$$P(\bar{z}z) \approx \frac{1}{\bar{z}z}; \bar{z}z \in (0, \infty)$$

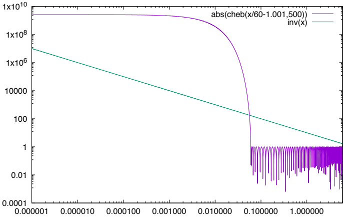
HDCR: use Γ_5 to make the system real *indefinite*. *Must make coarsening* Γ_5 *compatible*

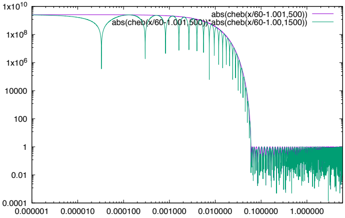
- As $\frac{1}{x}$ is odd, every second term cannot contribute: coarse Krylov space is in effect CGNR krylov space
  - Real spectrum lies in range $\left[m_l^2, 8^2\right]$
  - Coarsening remains nearest neighbour
  - Fine - Coarse - CoarseCoarse - eVectors

Hierarchically deflated conjugate residual

Novel setup scheme:

- Apply Chebyshev low pass filter: grows as x^N
- Inverse iteration costs *multiple* approximate solves per vector
- Use one Chebyshev low pass, then use recursive sequence to generate multiple independent vectors
- O(100-200) fine matrix multiples per new vector.



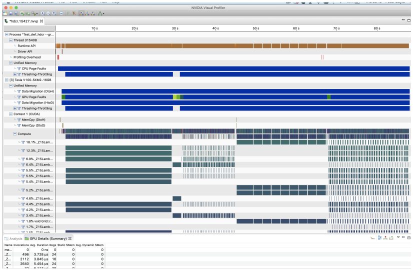


New approach to multigrid setup

Hierarchically deflated conjugate residual

- Significant software effort to keep 4 GPUs busy

  1. Subspace generation
  2. Matrix element calculation
  3. Coarsest space eigenvectors
  4. Solve



16^3 test system

- First test system: $16^3 \times 32 \times 16$. Set mass artificially low 0.00078
- Single node on DOE Summit computer
- Chebyshev smoother with full comms, double precision

Algorithm	Fine Matmuls	Time
CGNE	3200	44s
HDCR	650	19s
HDCR	400	15s
Chebyshev	2000	26s
Lanczos		10s
Ldop calc		10s
Setup+solve	2650	70s

$48^3 \times 96$ test system

- $48^3 \times 96 \times 16$. Ls=24 mass 0.00078
- 128 nodes on DOE Summit computer
- double precision, two level multigrid + Lanczos deflation
- Chebyshev smoother with full comms

Algorithm	Fine Matmuls	Time
CGNE	11400	440s
HDCR	2400	240s
Chebyshev	2500	100s
Lanczos		40s
Ldop calc		20s
Setup+solve	4900	400s

Set up AND solve faster than a single red black preconditioned solve

In principle (slight) win for HMC without subspace reuse across Hasenbusch terms or timesteps

$96^3 \times 192$ test system

- Second test system: $48^3 \times 96 \times 16$. Ls=12 mass 0.00054
- 256 nodes on DOE Summit computer
- single precision, two level multigrid + Lanczos deflation
- Chebyshev smoother with full comms

Algorithm	Fine Matmuls	Time
CGNE	14000	700s
HDCR	1300	250s
Chebyshev	2500	100s

- Still dominated by coarse space (256 evecs)
  - Gain greater at bigger L_s
  - Lanczos or 3 level multigrid is under on-going tuning.
- TODO: change Kernel and study m_{res} vs b

Multigrid for Domain Wall Fermions

Multigrid for Chiral Lattice Fermions: Domain Wall

Richard C. Brower*, M. A. Clark†, Dean Howarth*, and Evan S. Weinberg†

*Boston University, Boston, MA 02215, USA
†NVIDIA Corporation, Santa Clara, CA 95050, USA

April 17, 2020

arXiv:2004.07732

- nice proof the $D(m_{pv})^\dagger D(m_l)$ has half plane complex spectrum
  - Opens new methods for non-hermitian krylov solvers and multigrid for DWF
- **Generate 4D null space** $H_w \phi_i \sim 0$
- Coarsen with

$$\phi_i^{\pm} = 1 \pm \gamma_5 \phi_i$$

- Build 5D coarse Mobius with $\hat{H}_W$
- BCHW used 2D Schwinger model

$$\mathrm{sp}\{(\mathbb{P}^\dagger D^\dagger(m_{pv})\mathbb{P}\mathbb{P}^\dagger D(m_l)\mathbb{P})^n\} = \mathrm{sp}\{(\mathbb{P}^\dagger \gamma_5 D(m_{pv})\mathbb{P}\mathbb{P}^\dagger \gamma_5 D(m_l)\mathbb{P})^n\}$$

Implemented D=4 QCD in Grid

Share code between fine Grid Mobius and Coarse Grid Mobius

$$\hat{D}^5_{GDW} = \begin{pmatrix} \hat{D}_+ & -\hat{D}_- P_- & 0 & \dots & 0 & m\hat{D}_- P_+ \\ -\hat{D}_- P_+ & \ddots & \ddots & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \ddots & 0 & \vdots \\ \vdots & 0 & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \ddots & \ddots & -\hat{D}_- P_- \\ m\hat{D}_- P_- & 0 & \dots & 0 & -\hat{D}_- P_+ & \end{pmatrix}$$

First look at D=4 QCD on 16^3 test system

Compare ignoring cost of coarse space:

- BiCGSTAB on $D(m_{pv})^\dagger D(m_l)$ (20 iterations)
- Coarse BiCGSTAB on $\mathbb{P}^\dagger D(m_{pv})^\dagger \mathbb{P}\mathbb{P}^\dagger D(m_l)\mathbb{P}$
- V_{11} multigrid with BiCGSTAB smoother, BiCGSTAB coarse solver, PrecGCR(16) outer

Algorithm	operator	Outer iterations	Fine Matmuls
CG unprec	$D(m_l)^\dagger D(m_l)$	9500	9500
CGNE	$(M_{ee} - M_{eo}M_{oo}^{-1}M_{oe})$	3200	3200
CGNE	$(1 - M_{ee}^{-1}M_{eo}M_{oo}^{-1}M_{oe})$	3880	3880
BiCGSTAB	$D(m_{pv})^\dagger D(m_l)$	4140	4140
Tuned HDCR	$\mathbb{P}^\dagger D(m_l)^\dagger \mathbb{P}\mathbb{P}^\dagger D(m_l)\mathbb{P}$	23	460
HPD-MG	$\mathbb{P}^\dagger D(m_l)^\dagger \mathbb{P}\mathbb{P}^\dagger D(m_l)\mathbb{P}$	27	650
PV-MG	$\mathbb{P}^\dagger D(m_{pv})^\dagger \mathbb{P}\mathbb{P}^\dagger D(m_l)\mathbb{P}$	24	960

- H_{dwf} and H_w deflation both work
  - Outer iterations for $\hat{D}_l^\dagger \hat{D}_l$ very similar
  - Outer iterations for $\hat{D}_{pv}^\dagger \hat{D}_l$ higher and higher order smoother needed (with BiCGSTAB).
  - Needed to use 20 fine matrix multiplies in smoother for convergence
- H_w set up cost is reduced as 4D setup, but doesn't out balance solve time
- **Coarse space is L_s bigger, and even with Lanczos deflation clock favours HDCR**
  - Tried reducing L_s in coarse space, but insufficient

Coarse space solver

Converging to 10^{-8}

Coarsening	Algorithm	Operator	Coarse Matmuls
H_{dwf}	HDCR-CG	$\mathbb{P}^\dagger D(m_l)^\dagger \mathbb{P}\mathbb{P}^\dagger D(m_l)\mathbb{P}$	4736
	HDCR-CG(defl)	$\mathbb{P}^\dagger D(m_l)^\dagger \mathbb{P}\mathbb{P}^\dagger D(m_l)\mathbb{P}$	668
	BiCGSTAB	$\mathbb{P}^\dagger D(m_{pv})^\dagger \mathbb{P}\mathbb{P}^\dagger D(m_l)\mathbb{P}$	4839
H_w	CG	$\mathbb{P}^\dagger D(m_l)^\dagger \mathbb{P}\mathbb{P}^\dagger D(m_l)\mathbb{P}$	4770
	CG defl	$\mathbb{P}^\dagger D(m_l)^\dagger \mathbb{P}\mathbb{P}^\dagger D(m_l)\mathbb{P}$	756
	BiCGStab	$\mathbb{P}^\dagger D(m_{pv})^\dagger \mathbb{P}\mathbb{P}^\dagger D(m_l)\mathbb{P}$	1221

- **Coarse space is L_s bigger, and even with momest Lanczos deflation clock favours HDCR**
- Recursive or SVD deflation may reduce coarse cost for $\hat{D}^\dagger_{pv}\hat{D}_l$

Best time comparison $16^3 \times 32$

Algo	smoother	outer	fine mat	setup	solve
HDCR	10	23	460	30s + 20s	19s
MGrid $\hat{D}_l^\dagger \hat{D}_l$	10	28	560	8s+160s	76s
MGrid $\hat{D}_{pv}^\dagger \hat{D}_l$	20	24	960	8s	210s

- MGrid is 90% dominated by coarse space.
- Deflating the $\hat{D}_l^\dagger \hat{D}_l$, but not the $\hat{D}_{pv}^\dagger \hat{D}_l$
- Recursive may reduce coarse cost for $\hat{D}_{pv}^\dagger \hat{D}_l$, but greater smoother order is discouraging
- GMRES etc.. possible too

Summary

- Compared two approaches to DWF multigrid arXiv:1611.06944 and arXiv:2004.07732
  - Found similar ratio of matrix multiplies to Fine unpreconditione CG as BCHW. Deflation is working.
  - Possible to deflate with only 4D H_w setup
  - 2^4 blocking and 12 vectors required
  - makes 5D coarse space expense prohibitive; pursuing HDCR
  - If subspace with 4^4 blocking deflated effectively, H_w coarsening would be favourable
- Various failed attempts at using H_w coarsening to accelerate H_{dwf} coarsening
- Demonstrated HDCR for continued fraction overlap (but slow, untuned)
- Aim for 2+1+1f evolution with $b \geq 1, c = 0$ and fast setup multigrid in HMC
  - Implies change of kernel so accompany with change of gauge action and N_f.
- All code was written in Grid, CPU / GPU portable