Faster multigrid Chebyshev setup
First cross over of setup + solve faster than red black CGNR
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

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

\[
D_5^{GDW} = \begin{pmatrix}
D_+ & -D_- P_- & 0 & \ldots & 0 & mD_- P_+ \\
-D_- P_+ & \ddots & \ddots & 0 & \cdots & 0 \\
0 & \ddots & \ddots & \ddots & \ddots & 0 \\
\vdots & \ddots & \ddots & \ddots & \ddots & \ddots \\
mD_- P_- & 0 & \ldots & 0 & \ddots & -D_- P_+
\end{pmatrix}
\]

\[
D_+ = (bD_W + 1) \\
D_- = (1 - cD_W)
\]

\[
H_M = \gamma_5 \frac{(b + c)D_W}{2 + (b - c)D_W}
\]

Shamir DWF case: \( b = 1, c = 0 \)

\[ c = 0 \Rightarrow H_{GDW} = \gamma_5 R_5 D_{GDW} = \Gamma_5 D_{GDW} \]
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)
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 $G_{DWF}$, 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 $b$
  \[ P = |\phi_i^{b\pm}\rangle \]
- Coarse space is 4-dimensional
- Coarse space is nearest neighbour - aim for HMC
- Coarse operator
  \[ \hat{H}_{DWF} = P^\dagger \Gamma_5 D_{DWF} P = P^\dagger H_{DWF} P \]
  Then
  \[ \hat{H}_{DWF}^\dagger \hat{H}_{DWF} = (P^\dagger H_{DWF} P)^2 = P^\dagger D_{DWF}^\dagger P P^\dagger D_{DWF} P \]
- 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 $\Re \lambda \geq 0$ in “Hamburger” plot:

- 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 $\bar{z}$ ⇒ real, pos def:

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

HDCR: use $\Gamma_5$ to make the system real *indefinite*. Must make coarsening $\Gamma_5$ *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 $[m^2, 8^2]$
  - 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.
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
• 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

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Fine Matmuls</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>CGNE</td>
<td>3200</td>
<td>44s</td>
</tr>
<tr>
<td>HDCR</td>
<td>650</td>
<td>19s</td>
</tr>
<tr>
<td>HDCR</td>
<td>400</td>
<td>15s</td>
</tr>
<tr>
<td>Chebyshev</td>
<td>2000</td>
<td>26s</td>
</tr>
<tr>
<td>Lanczos</td>
<td></td>
<td>10s</td>
</tr>
<tr>
<td>Ldop calc</td>
<td></td>
<td>10s</td>
</tr>
<tr>
<td>Setup+solve</td>
<td>2650</td>
<td>70s</td>
</tr>
</tbody>
</table>
$48^3 \times 96$ test system

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

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Fine Matmul</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>CGNE</td>
<td>11400</td>
<td>440s</td>
</tr>
<tr>
<td>HDCR</td>
<td>2400</td>
<td>240s</td>
</tr>
<tr>
<td>Chebyshev</td>
<td>2500</td>
<td>100s</td>
</tr>
<tr>
<td>Lanczos</td>
<td></td>
<td>40s</td>
</tr>
<tr>
<td>Ldop calc</td>
<td></td>
<td>20s</td>
</tr>
<tr>
<td>Setup+solve</td>
<td>4900</td>
<td>400s</td>
</tr>
</tbody>
</table>

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. L_s=12 mass 0.00054
- 256 nodes on DOE Summit computer
- single precision, two level multigrid + Lanczos deflation
- Chebyshev smoother with full comms

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Fine Matmuls</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>CGNE</td>
<td>14000</td>
<td>700s</td>
</tr>
<tr>
<td>HDCR</td>
<td>1300</td>
<td>250s</td>
</tr>
<tr>
<td>Chebyshev</td>
<td>2500</td>
<td>100s</td>
</tr>
</tbody>
</table>

- 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

April 17, 2020

Abstract

Critical slowing down for the Krylov Dirac solver presents a major obstacle to further advances in lattice field theory as it approaches the continuum solution. We propose a new multi-grid approach for chiral fermions, applicable to both the 5-d domain wall or 4-d Overlap operator. The central idea is to directly coarsen the 4-d Wilson kernel, giving an effective domain wall or overlap operator on each level. We provide here an explicit construction for the Shamir domain wall formulation with numerical tests for the 2-d Schwinger prototype, demonstrating near ideal multi-grid scaling. The framework is designed for a natural extension to 4-d lattice QCD chiral fermions, such as the Möbius, Zolotarev or Borici domain wall discretizations or directly to a rational expansion of the 4-d Overlap operator. For the Shamir operator, the effective overlap operator is isolated by the use of a Pauli-Villars preconditioner in the spirit of the Kähler-Dirac spectral map used in a recent staggered MG algorithm [1].


- 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 Möbius with $\hat{H}_W$
- BCHW used 2D Schwinger model

$\text{sp}\{(P^\dagger D^\dagger(m_{pv})PP^\dagger D(m_l)P)^n\} = \text{sp}\{(P^\dagger \gamma_5 D(m_{pv})PP^\dagger \gamma_5 D(m_l)P)^n\}$
Implemented D=4 QCD in Grid

Share code between fine Grid Mobius and Coarse Grid Mobius

$$\hat{D}_{GDW}^5 = \begin{pmatrix}
\hat{D}_+ & -\hat{D}_- P_+ & 0 & \cdots & 0 & m\hat{D}_- P_+ \\
-\hat{D}_- P_+ & \ddots & \ddots & 0 & \cdots & 0 \\
0 & \ddots & \ddots & \ddots & 0 & \ddots \\
\vdots & 0 & \ddots & \ddots & \ddots & \ddots \\
m\hat{D}_- P_- & 0 & \cdots & 0 & \ddots & -\hat{D}_- P_+
\end{pmatrix}$$
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 $P^\dagger D(m_{pv})^\dagger PP^\dagger D(m_l)P$
- $V_{11}$ multigrid with BiCGSTAB smoother, BiCGSTAB coarse solver, PrecGCR(16) outer

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Operator</th>
<th>Outer iterations</th>
<th>Fine Matmuls</th>
</tr>
</thead>
<tbody>
<tr>
<td>CG unprec</td>
<td>$D(m_l)^\dagger D(m_l)$</td>
<td>9500</td>
<td>9500</td>
</tr>
<tr>
<td>CGNE</td>
<td>$(M_{ee} - M_{eo} M_{oo}^{-1} M_{oe})$</td>
<td>3200</td>
<td>3200</td>
</tr>
<tr>
<td>CGNE</td>
<td>$(1 - M_{ee}^{-1} M_{eo} M_{oo}^{-1} M_{oe})$</td>
<td>3880</td>
<td>3880</td>
</tr>
<tr>
<td>BiCGSTAB</td>
<td>$D(m_{pv})^\dagger D(m_l)$</td>
<td>4140</td>
<td>4140</td>
</tr>
<tr>
<td>Tuned HDCR</td>
<td>$P^\dagger D(m_l)^\dagger PP^\dagger D(m_l)P$</td>
<td>23</td>
<td>460</td>
</tr>
<tr>
<td>HPD-MG</td>
<td>$P^\dagger D(m_l)^\dagger PP^\dagger D(m_l)P$</td>
<td>27</td>
<td>650</td>
</tr>
<tr>
<td>PV-MG</td>
<td>$P^\dagger D(m_{pv})^\dagger PP^\dagger D(m_l)P$</td>
<td>24</td>
<td>960</td>
</tr>
</tbody>
</table>

- $H_{dwf}$ 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 HDCF
  - Tried reducing $L_s$ in coarse space, but insufficient
Coarse space solver

Converging to $10^{-8}$

<table>
<thead>
<tr>
<th>Coarsening</th>
<th>Algorithm</th>
<th>Operator</th>
<th>Coarse Matmuls</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_{dwf}$</td>
<td>HDCR-CG</td>
<td>$P \dagger D(m_l) \dagger P P \dagger D(m_l) P$</td>
<td>4736</td>
</tr>
<tr>
<td></td>
<td>HDCR-CG(defl)</td>
<td>$P \dagger D(m_l) \dagger P P \dagger D(m_l) P$</td>
<td>668</td>
</tr>
<tr>
<td></td>
<td>BiCGSTAB</td>
<td>$P \dagger D(m_{pv}) \dagger P P \dagger D(m_l) P$</td>
<td>4839</td>
</tr>
<tr>
<td>$H_w$</td>
<td>CG</td>
<td>$P \dagger D(m_l) \dagger P P \dagger D(m_l) P$</td>
<td>4770</td>
</tr>
<tr>
<td></td>
<td>CG defl</td>
<td>$P \dagger D(m_l) \dagger P P \dagger D(m_l) P$</td>
<td>756</td>
</tr>
<tr>
<td></td>
<td>BiCGStab</td>
<td>$P \dagger D(m_{pv}) \dagger P P \dagger D(m_l) P$</td>
<td>1221</td>
</tr>
</tbody>
</table>

- Coarse space is $L_s$ bigger, and even with momest Lanczos deflation clock favours HDCR
- Recursive or SVD deflation may reduce coarse cost for $\hat{D}_{pv} \hat{D}_l$
### Best time comparison $16^3 \times 32$

<table>
<thead>
<tr>
<th>Algo</th>
<th>smoother</th>
<th>outer</th>
<th>fine mat</th>
<th>setup</th>
<th>solve</th>
</tr>
</thead>
<tbody>
<tr>
<td>HDCR</td>
<td>10</td>
<td>23</td>
<td>460</td>
<td>30s + 20s</td>
<td>19s</td>
</tr>
<tr>
<td>MGrid $\hat{D}_l^+ \hat{D}_l$</td>
<td>10</td>
<td>28</td>
<td>560</td>
<td>8s + 160s</td>
<td>76s</td>
</tr>
<tr>
<td>MGrid $\hat{D}^\dagger_{pv} \hat{D}_l$</td>
<td>20</td>
<td>24</td>
<td>960</td>
<td>8s</td>
<td>210s</td>
</tr>
</tbody>
</table>

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

  - Found similar ratio of matrix multiplies to Fine unpreconditioned 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