Algebraic Multigrid in Lattice QCD computations

Karsten Kahl

Bergische Universität Wuppertal

May 8, 2012
Outline

The Multigrid Principle

Algebraic Multigrid
  Theory of AMG

Algebraic Multigrid for QCD
  Smoother for QCD
  Interpolation for QCD

Bootstrap Algebraic Multigrid
  Least Squares Interpolation
  Bootstrap Setup

Numerical Experiments

Comparison with Inexact Deflation

Outlook
The continuum limit scaling problem — Simple Example

Partial differential equation

\[ \Delta \psi(x) = \varphi, \quad x \in (0, 1)^2 \]
\[ \psi(x) = 0, \quad x \in \partial(0, 1)^2 \]

Discretization by finite differences

\[ Au = f, \quad A \in \mathbb{R}^{N^2 \times N^2}, \quad u, f \in \mathbb{R}^{N^2} \]

- \( A \) symmetric positive definite
- smallest eigenvalue \( \lambda_{\text{min}} \sim h^2 = N^{-2} \)
- largest eigenvalue \( \lambda_{\text{max}} = \text{const} \)

Continuum limit: condition number \( \kappa \underset{h \to 0}{\to} \infty \)
Iterative schemes – Part I – Gauss Seidel

video removed
Iterative schemes – Part II – Scaling Problems
Multigrid in a Nutshell

- **Smooth**
- **Restriction**
- **Prolongation**
- **The Multigrid V-cycle**
- **Finest Grid**
- **First Coarse Grid**
- **Fewer Dofs**

Multigrid in Lattice QCD computations, Karsten Kahl
Multigrid in a Nutshell – Part II

video removed
Iterative schemes – Part III – Multigrid
Components of Algebraic Multigrid

**Given:** Linear system of equations

\[ Au = f \]

**Wanted:** Hierarchy of systems \( A_0 = A \)

\[ A_\ell u_\ell = f_\ell, \quad \ell = 0, \ldots, L \]

**Needed:** Coarse spaces (coarse variables)

\[ \mathbb{C}^{n_0} \supset \mathbb{C}^{n_1} \supset \ldots \supset \mathbb{C}^{n_L} \]

- **Smotherer**
  \[ S_\ell : \mathbb{C}^{n_\ell} \rightarrow \mathbb{C}^{n_\ell} \]
  **High modes**

- **Interpolation**
  \[ P_{\ell+1} : \mathbb{C}^{n_{\ell+1}} \rightarrow \mathbb{C}^{n_\ell} \]
  **Low modes**

Complementarity of Smoother and Interpolation
Coarse variables in Algebraic Multigrid

**Given:** Variables $u_1, \ldots, u_n$

**Wanted:** Coarse variables $v_1, \ldots, v_{n_c}$ with $n_c \ll n$

\[ v_i = \sum_{j \in N_i} \alpha_{ij} u_j \]

**Aggregation**

\[ N_i \cap N_k = \emptyset \text{ for all } k, l \]

**$C - F$ splitting**

\[ \alpha_{ij} = \delta_{ij} \implies P = \begin{pmatrix} W \\ I \end{pmatrix} \]
Generic Multigrid Algorithm

\[ MG_\ell(A_\ell, b_\ell) \]

\[
\text{if } \ell < L \text{ then} \\
\quad u_\ell = 0 \\
\quad \text{for } i = 1, \ldots, \nu_1 \text{ do} \\
\quad \quad u_\ell = S_\ell(A_\ell, u_\ell, b_\ell) \\
\quad \text{end for} \\
\quad u_{\ell+1} = MG(A_{\ell+1}, (P_{\ell+1}^\ell)^\dagger (b_\ell - A_\ell u_\ell)) \\
\quad u_\ell = u_\ell + P_{\ell+1}^\ell u_{\ell+1} \\
\quad \text{for } i = 1, \ldots, \nu_2 \text{ do} \\
\quad \quad u_\ell = S_\ell(A_\ell, u_\ell, b_\ell) \\
\quad \text{end for} \\
\text{else} \\
\quad u_{\ell+1} = A_{\ell+1}^{-1} (P_{\ell+1}^\ell)^\dagger (b_\ell - A_\ell u_\ell) \\
\text{end if} \]
Two-grid convergence strong approximation property

- **Strong approximation property**

\[
\| (I - PA_c^{-1}P^\dagger A)e \|^2_A \leq K \| Ae \|^2_2
\]

- **Smoothing property**

\[
\| Se \|^2_A \leq \| e \|^2_A - \alpha_1 \langle Ae, Ae \rangle_2
\]
\[
\| Se \|^2_A \leq \| e \|^2_A - \alpha_2 \langle ASe, Ae \rangle_2
\]

- **Two-grid convergence**

\[
\| S \left( I - PA_c^{-1}P^\dagger A \right) Se \|^2_A \leq \frac{1 - \alpha_1}{1 + \frac{\alpha_2}{K}} \| e \|^2_A
\]

- **Assumptions fulfilled on all levels ⇒ Multigrid convergence**
Two-grid convergence weak approximation property

- Weak approximation property

\[ \| (I - PR) e \|_2 \leq \frac{K}{\|A\|_2} \| e \|_A \]

- Convergent hermitian positive definite smoother \( S \)

\[ \lambda_{\text{min}}(I - SA) \geq 0 \quad \text{and} \quad \lambda_{\text{min}}(S) \geq \frac{1}{c_S^2\|A\|_2^2} \]

- Two-grid convergence

\[ \| S \left( I - PA_c^{-1} P^\dagger A \right) Se \|_A^2 \leq \left( 1 - \frac{1}{c_0} \right) \| e \|_A \]

with \( c_0 = (2 + Kc_S)^2 \)
Algebraic Multigrid for non-hermitian operators

Different right- and left low modes

\[ \langle Ax, Ax \rangle^{\frac{1}{2}} << \|x\|_2 \quad \text{and} \quad \langle A^\dagger y, A^\dagger y \rangle^{\frac{1}{2}} << \|y\|_2 \]

- Build interpolation \( P \) based on right low modes of \( A \)
- Build restriction \( R \) based on left low modes of \( A \)
- Define Petrov-Galerkin coarse-grid operator

\[ A_c = RAP \]

- Coarse-grid correction error propagator

\[ e = e - P (RAP)^{-1} RAe \]
The continuum limit scaling problem in QCD

Wilson discretization of the Dirac equation

\[ D = \sum_{\mu} \gamma_{\mu} \otimes d_{\mu}^c + a \cdot d_{\mu\mu} + 1 \cdot m \]

Continuum limit scaling problems:

- condition number

  \[ \kappa \sim (a \cdot m)^{-1} m, a \to 0 \to \infty \]

- eigenvalues \( K \) below threshold (e.g. 100 MeV)

  \[ K \sim V = N_s^3 N_t \]

**Scalability** in \( a, m \) and \( V \) needed (possible in optimal complexity?)

(Algebraic) Multigrid
Domain Decomposition

Decomposition of $\mathcal{L}$

- easily parallelizable
- implementation available
- local Dirichlet boundaries
- no theory available

Canonical restrictions

$R_i : \mathcal{L} \rightarrow \mathcal{L}_i$

Block inverses

$B_i := R_i^\dagger (R_i DR_i^\dagger)^{-1} R_i$

RB Schwarz

For $\text{color} = 1, 2$

$r = b - Dz$

$z = z + \left( \sum_{i \text{ with } \mathcal{L}_i \subset \mathcal{L}_{\text{color}}} B_i \right) r$

[1] Hermann Schwarz 1870
Domain Decomposition smoothing properties

- effective reduction of high modes
- no effective reduction of low modes

SAP fulfills the definition of a smoother in AMG
Choices for AMG in QCD Computations

Application of AMG to linear system $D$ of Wilson fermions

$$Du = f, \quad D \in \mathbb{C}^{N^d n_s n_c}, \quad \text{with} \quad D^\dagger = \Gamma_5 D \Gamma_5,$$

where $D$ is non-hermitian pos. real and $\Gamma_5 D$ hermitian indefinite

Define $R_\ell, P_\ell, S_\ell$ such that

- the $\Gamma_5$-’ambiguity’ is hidden to the AMG algorithm
- the $\Gamma_5$-structure is preserved in the multigrid hierarchy

$$D^\dagger_c = \Gamma_5 D_c \Gamma_5$$
Structure-preserving restriction and prolongation

For any right-EV $v$ of $D$ to EW $\lambda \rightarrow \gamma_5 v$ left-EV of $D$ to EW $\bar{\lambda}$

Preserve $\gamma_5$ symmetry $\leftrightarrow$ symmetry of spectrum to real axis

$$P\gamma_5 = \gamma_5 P$$
$$D_c = P^\dagger DP$$
$\implies$
$$D_c^\dagger = \gamma_5 D\gamma_5$$

For $\gamma_5 = \begin{pmatrix} I & 0 \\ -I & 0 \end{pmatrix}$ choose prolongation $P = \begin{pmatrix} P_1 & 0 \\ 0 & P_2 \end{pmatrix}$

- $\gamma_5$-structure transferred to coarse-grid in $D_c$
- no special treatment of $R$ and left-eigenspace necessary
0th and 1st order interpolation...

0th order

- aggregation
- easy to implement
- piece-wise constant

Aggregation interpolation

1st order

- $C/F$ splitting
- next-neighbor interpolation
- linear interpolation

$C/F$ interpolation

Bootstrap Algebraic Multigrid
... for Wilson fermions

In 4D with next neighbor coupling

- next neighbor coupling on coarse grid
- operator complexity

\[ 1 + \frac{k}{|\mathcal{A}|} + \frac{k^2}{|\mathcal{A}|^2} \cdots \]

- additional diagonal coupling (preserved on coarser grids)
- operator complexity

\[ 1 + \frac{81}{9} \frac{1}{16} \left( 1 + \frac{1}{16} \right) \cdots \]
Bootstrap AMG

Bootstrap Algebraic Multigrid Framework

“Continuous updating components of the MG hierarchy using practical tools and measures built from the evolving MG solver”

- Compatible Relaxation and Algebraic Distances
  - Adaptive computation of nested subspaces
  - Quality measurement of grid hierarchies
- Least Squares interpolation
  - Adaptive definition of interpolation
  - Quality and complexity control of the multigrid method
- Bootstrap setup techniques
  - Multigrid discovery of algebraically smooth error
  - Multigrid quality control mechanisms
Least Squares Interpolation

**Given:** test vectors \( u^{(1)}, \ldots, u^{(k)} \in \mathbb{C}^n \) representing low modes

**Wanted:** interpolation \( P \) accurate for test vectors \( u^{(s)} \)

**Least Squares Interpolation**

\[
\mathcal{L}_{C_i}(p_i) = \sum_{s=1}^{k} \omega_s (u_i^{(s)} - \sum_{j \in C_i} (p_{ij})_j u_j^{(s)})^2 \rightarrow \min_{p_i}
\]

- interpolatory points \( C_i \subset C \) for \( i \in \mathcal{F} \)
  - in neighborhood of \( i \)
- weights \( \omega_s \sim \|u^{(s)}\| \|Au^{(s)}\|^{-1} \in \mathbb{R}^+ \)
- test vectors (→ bootstrap)
Bootstrap Setup - Multigrid Eigensolver

Assumption: No a priori information on low modes available!

- smoother action known, initial test vectors

\[ u^{(s)} = S^\eta \tilde{u}^{(s)}, \quad \tilde{u}^{(s)} \text{ random} \]

- observation \((P_\ell = P_1^0 \cdots P_{\ell-1}^0, \quad A_\ell = P_\ell^\dagger A_0 P_\ell, \quad T_\ell = P_\ell^\dagger P_\ell)\)

\[
\frac{\langle v_\ell, v_\ell \rangle_{A_\ell}}{\langle v_\ell, v_\ell \rangle_{T_\ell}} = \frac{\langle P_\ell v_\ell, P_\ell v_\ell \rangle_{A}}{\langle P_\ell v_\ell, P_\ell v_\ell \rangle_2}
\]

Bootstrap Idea

Eigenpairs \((v_\ell, \lambda_\ell)\) of \((A_\ell, T_\ell)\) \[\rightarrow\] Eigenpairs \((P_\ell v_\ell, \lambda_\ell)\) of \(A\) + interpolation error
Bootstrap Setup – Cycling Strategies

- Relax on $Au = 0, u \in U$
- Compute $V$, s.t., $Av = \lambda Tv, v \in V$
- Relax on $Av = \lambda Tv, v \in V$
- Relax on $Au = 0, u \in U$ and $Av = \lambda Tv, v \in V$
Numerical Experiments – Setup Parameters

![Graph showing solver iterations vs total setup smoother iterations for different methods.](image)

- **MP-\(\alpha\)MG(1) +(x, 0)**
- **MP-\(\alpha\)MG(1) +(x, 10)**
- **MP-\(\alpha\)MG(1) +(x, 20)**
- **MP-\(\alpha\)MG(1) +(5, x)**

Solver iterations, \(tol = 10^{-10}\)
Numerical Experiments – Setup Parameters cont.

Numerical Experiments – Setup Parameters cont.

![Graph showing setup time and solve time](image)

- **setup time + solve time**
  - **MP-αMG(1)+(_, 0)**
  - **MP-αMG(1)+(_, 5)**

The graph illustrates the relationship between total setup time and solve time for different methods, with a focus on how the combination of setup and solve times varies with changes in parameters.
Numerical Experiments – Settings

- **FGMRES**
  - restart 25
  - tolerance $10^{-10}$

- **Bootstrap AMG**
  - $V^2$-cycle setup
  - domain decomposition smoother
  - aggregation based coarsening
  - 20 degrees of freedom per aggregate
  - 2-grid method

- **Platform**
  - parallel **MPI-C** implementation by Matthias Rottmann (BUW)
  - 4096 cores on Juropa@JSC
### 24^3 \times 48

<table>
<thead>
<tr>
<th></th>
<th>double precision</th>
<th>single precision</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CGN FGMRES(25) + \alpha\text{MG}(1)</td>
<td>FGMRES(25) + SP-\alpha\text{MG}(1)</td>
</tr>
<tr>
<td>setup time</td>
<td>-</td>
<td>5.01s</td>
</tr>
<tr>
<td>solve time iter</td>
<td>5061</td>
<td>72</td>
</tr>
<tr>
<td>solve time time</td>
<td>27.76s</td>
<td>2.09s (13.3)</td>
</tr>
<tr>
<td>total time</td>
<td>27.76s</td>
<td>7.10s (3.9)</td>
</tr>
</tbody>
</table>


4096 cores on Juropa@JSC
2HEX smeared tree level improved Clover Wilson, $m_\pi = 135\text{MeV}$

<table>
<thead>
<tr>
<th></th>
<th>double precision</th>
<th>single precision</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CGN</td>
<td>FGMRES(25)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+ $\alpha$MG(1)</td>
</tr>
<tr>
<td>setup time</td>
<td>-</td>
<td>28.85s</td>
</tr>
<tr>
<td>solve iter time</td>
<td>96787</td>
<td>645</td>
</tr>
<tr>
<td></td>
<td>2312.22s</td>
<td>58.71s (39.4)</td>
</tr>
<tr>
<td>total time</td>
<td>2312.22s</td>
<td>87.56s (26.4)</td>
</tr>
</tbody>
</table>

$a = 0.092\text{fm}$, $\beta = 3.5$, $(10,30)$ single and double precision setup local lattice: $8^4$, SAP block-size: $2^4$, aggregate-size: $4^4$, 20 test vectors, tol: $10^{-10}$, SAP solver: Minimal Residual(3), coarse solver: GMRES(24)

4096 cores on Juropa@JSC
$64^3 \times 128$ DESY

<table>
<thead>
<tr>
<th></th>
<th>single precision</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CGN</td>
</tr>
<tr>
<td>setup time</td>
<td>-</td>
</tr>
<tr>
<td>solve iter</td>
<td>46771</td>
</tr>
<tr>
<td>solve time</td>
<td>2660.6s</td>
</tr>
<tr>
<td>total time</td>
<td>2660.6s</td>
</tr>
</tbody>
</table>


4096 cores on Juropa@JSC
Weak Scaling: Lattice Sizes from $8^4$ to $64^4$

8$^4$ local lattice, SAP block-size: 2$^4$, aggregate-size: 4$^4$, 20 test vectors

SAP block-size: $3^4$, aggregate-size: $3^4$, 20 test vectors
Error Scaling: $48^3 \times 48$ BMW-c [arXiv:1011.2403], [arXiv:1011.2711]
2HEX smeared tree level improved Clover Wilson, $m_\pi = 136\text{MeV}$

<table>
<thead>
<tr>
<th></th>
<th>FGMRES +MP-$\alpha$MG(1)</th>
<th>FGMRES +MP-$\alpha$MG(1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>coarse GMRES</td>
<td>iter: 12</td>
<td>tol: $10^{-3}$</td>
</tr>
<tr>
<td>setup time</td>
<td>4.51s</td>
<td>4.51s</td>
</tr>
<tr>
<td>solve iter time</td>
<td>461</td>
<td>110</td>
</tr>
<tr>
<td>solve time</td>
<td>12.40s</td>
<td>30.53s</td>
</tr>
<tr>
<td>total time</td>
<td>16.91s</td>
<td>35.04s</td>
</tr>
</tbody>
</table>

$a = 0.116\text{fm}$, $\beta = 3.31$, local lattice: $6^4$, SAP block-size: $3^4$, aggregate-size: $3^4$, 20 test vectors, tol: $10^{-10}$, SAP solver: MINRES(3), 5-500-20-setup

4096 cores on Juropa@JSC
αMG vs. Inexact Deflation

Interpolation / Deflation Subspace

- Aggregation structure
- Spins decoupled $\rightarrow$ coarse $\gamma_5$-symmetry
- Spins coupled $\rightarrow$ no coarse $\gamma_5$-symmetry

Setup

- Bootstrap multigrid eigensolver
- “Inverse” iteration (finest grid eigensolver)

Coarse-grid ($P^\dagger DP$)

- No accuracy requirements (due to multigrid construction)
- High accuracy requirements (due to deflation construction)
- Recursion straight-forward
- Recursion not trivial ($\rightarrow$ MG)

Krylov wrapper

- flexible GMRES
- GCR

SAP

- Smoother in multigrid method
- Additional preconditioner
Summary and Outlook

Summary

▶ development stage of bootstrap AMG for QCD
▶ setup done once per configuration (and mass)
▶ high potential for parallelism in setup and solver
▶ implementation of 0th order interpolation (almost) available

Outlook

▶ optimize setup and its numerous parameters
▶ develop theoretical foundation
▶ implementation of 1st order approach
▶ implementation of full multigrid for 0th order approach