AMG Preconditioning for the Overlap Operator

James Brannick*, Andreas Frommer*, **Karsten Kahl***, Björn Leder*,
Matthias Rottmann*, Marcel Schweitzer* and Artur Strebel*

*Bergische Universität Wuppertal and *PennState University

June 19, 2014

Motivation

Task: Find solution of $D_N\varphi = \eta$ where

$$\begin{aligned} D_N =& (m_0^N - \frac{m}{2})(1 + \gamma_5 \operatorname{sign}(\overbrace{\gamma_5(D_W - m_0^N)}^{H_W})) + m \\ =& (m_0^N - \frac{m}{2})(1 + \gamma_5 (H_W^\dagger H_W)^{-\frac{1}{2}} H_W) + m \end{aligned}$$

Challenges:

(i) Iteration counts of $\mathcal{O}(1{,}000)$ for $D_N\varphi = \eta$

(ii) Evaluating $(H_W^\dagger H_W)^{-\frac{1}{2}} x$ is quite costly

Outline

Accelerating the solution of $D_N\varphi = \eta$

Challenge (i): Number of iterations for $D_N\varphi = \eta$

Idea: Preconditioning with a “suitable” operator $\widetilde{D}$

Accelerating the solution of $D_N\varphi = \eta$

Challenge (i): Number of iterations for $D_N\varphi = \eta$

Idea: Preconditioning with a "suitable" operator $\widetilde{D}$

Use the concept of auxiliary space preconditioning to define $\widetilde{D}$

- Use a similar (spectrally equivalent) operator that is easier to solve as a preconditioner
- **Examples**
  - Conforming for Nonconforming finite element discretizations
  - Lower order for higher order discretizations
  - Low chirality for high chirality discretization

Accelerating the solution of $D_N\varphi = \eta$

Challenge (i): Number of iterations for $D_N\varphi = \eta$

Idea: Preconditioning with a "suitable" operator $\widetilde{D}$

Use the concept of auxiliary space preconditioning to define $\widetilde{D}$

- Use a similar (spectrally equivalent) operator that is easier to solve as a preconditioner
- **Examples**
  - Conforming for Nonconforming finite element discretizations
  - Lower order for higher order discretizations
  - Low chirality for high chirality discretization

Auxiliary space preconditioning of the Neuberger Overlap operator
Use the kernel operator of D_N, i.e., the Wilson-Dirac operator D_W

$$D_N D_W^{-1}\psi = \eta \text{ with } \varphi = D_W^{-1}\psi$$

- Computing D_W^{-1} is done by DD-αAMG [arXiv:1303.1377]
- D_W^{-1} is cheap, scalable and robust

Why is D_W a good preconditioner for D_N?

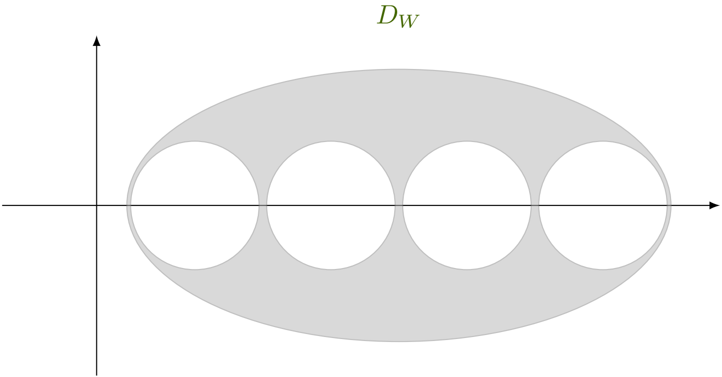
Assuming normality of D_W (i.e. $D_W^\dagger D_W = D_W D_W^\dagger$) we find

Relation between low modes of D_W and D_N

Let λ be a small eigenvalue of D_W, i.e., $D_W x = \lambda x$ with $|\lambda|$ small. W.l.o.g. assume $m = 0$. Then

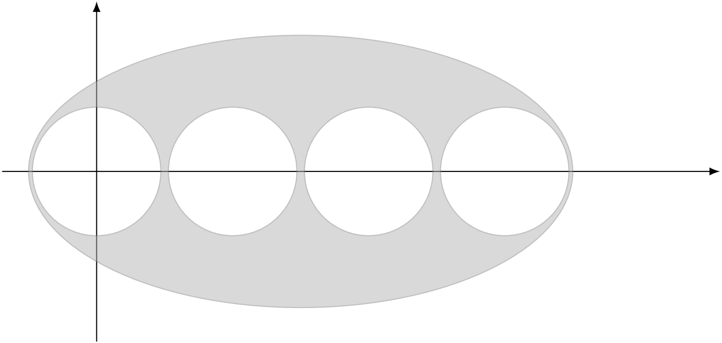
$$
\begin{aligned}
D_N x &= m_0^N \big(1 + \gamma_5 \operatorname{sign}(\gamma_5 (D_W - m_0^N))\big) x \\
&= m_0^N \big(1 + \gamma_5 ((D_W - m_0^N)^\dagger (D_W - m_0^N))^{-\frac{1}{2}} \\
&\quad \cdot (\gamma_5 (D_W - m_0^N))\big) x \\
&= m_0^N x + m_0^N (\lambda - m_0^N) \big((\lambda - m_0^N)(\overline{\lambda - m_0^N})\big)^{-\frac{1}{2}} x \\
&= m_0^N \big(1 + \operatorname{sign}(\lambda - m_0^N)\big) x
\end{aligned}
$$

Overlap construction – assuming normality of D_W



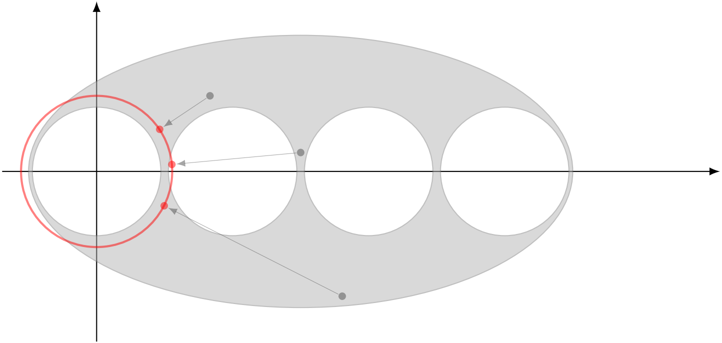
Overlap construction – assuming normality of D_W

$D_W - m_0^N$



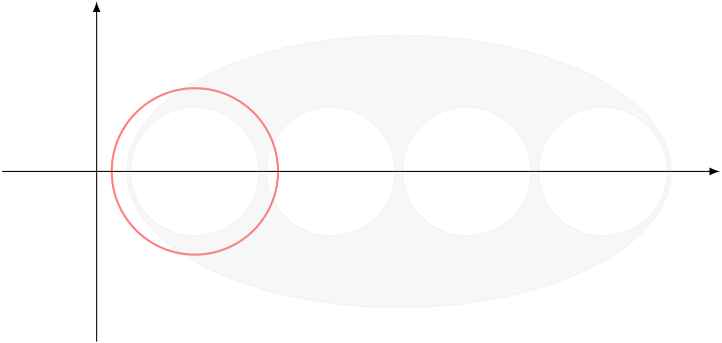
Overlap construction – assuming normality of D_W

$$\gamma_5 \, \mathrm{sign}(\gamma_5(D_W - m_0^N))$$



Overlap construction – assuming normality of D_W

$$D_N = (m_0^N - \tfrac{m}{2})(1 + \gamma_5 \operatorname{sign}(\gamma_5(D_W - m_0^N)) + m$$



What does normality have to do with physics?

- The continuous Dirac operator $\mathcal{D}$ is chiral and γ_5-symmetric

$$\gamma_5 \mathcal{D} - \mathcal{D}\gamma_5 \quad \text{and} \quad (\gamma_5 D)^\dagger = \gamma_5 D$$

- Chirality is equivalent to anti-hermiticity (1-normality)

$$\gamma_5 D + D\gamma_5 = 0 \Leftrightarrow D^\dagger = -D$$

- Ginsparg-Wilson relation is equivalent to $(1,1)$-normality

$$\gamma_5 D - D\gamma_5 = aD\gamma_5 D \Leftrightarrow D^\dagger = -D(1-aD)^{-1}$$

What does normality have to do with physics?

- The continuous Dirac operator $\mathcal{D}$ is chiral and γ_5-symmetric

$$\gamma_5\mathcal{D} - \mathcal{D}\gamma_5 \quad \text{and} \quad (\gamma_5 D)^\dagger = \gamma_5 D$$

- Chirality is equivalent to anti-hermiticity (1-normality)

$$\gamma_5 D + D\gamma_5 = 0 \Leftrightarrow D^\dagger = -D$$

- Ginsparg-Wilson relation is equivalent to $(1,1)$-normality

$$\gamma_5 D - D\gamma_5 = aD\gamma_5 D \Leftrightarrow D^\dagger = -D(1-aD)^{-1}$$

- *Side note:* Higher order Ginsparg-Wilson relations?

Deviation from normality of non-chiral discretizations

Quality of preconditioner depends on normality?!

Measure for the deviation of normality

$$\delta_N := ||D^\dagger D - DD^\dagger||_F,$$

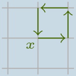
where $||X||_F^2 = \sum_{i,j=1}^n x_{ij}^2, X \in \mathbb{C}^{n\times n}$.

Theorem

The deviation of normality of D_W is given by

$$\delta_N = 16 \sum_x \sum_{\mu>\nu} \mathrm{Re}(\mathrm{tr}(I - Q_x^{\mu,\nu})),$$

where $Q_x^{\mu,\nu}$ is the plaquette defined by

$$Q_x^{\mu,\nu} = U_\nu(x)U_\mu(x+\hat{\nu})U_\nu^H(x+\hat{\mu})U_\mu^H(x) =$$ 

.

Deviation from normality proportional to the Wilson Gauge action!?

Thick Links, smearing and normality

Wilson flow, Stout smearing and the deviation from normality

Anything that changes the gauge action, changes δ_N

The Wilson flow $\mathcal{V}_\tau = \{V_\mu(x,\tau) : x \in \mathcal{L}, \mu = 0,\dots,3\}$ is defined as the solution of the initial value problem

$$\frac{\partial}{\partial\tau} V_\mu(x,\tau) = -\left\{\partial_{x,\mu} S_{\mathrm{W}}(\mathcal{V}_\tau)\right\} V_\mu(x,\tau)\,, \quad V_\mu(x,0) = U_\mu(x).$$

Given a gauge field $\mathcal{U}$, stout smearing modifies the gauge links

$$U_\mu(x) \rightarrow \tilde{U}_\mu(x) = \mathrm{e}^{\epsilon Z_\mu^{\mathcal{U}}(x)} U_\mu(x)$$

where the parameter ϵ is a small positive number and

$$\begin{aligned} Z_\mu^{\mathcal{U}}(x) &= -\frac{1}{2}(M_\mu(x) - M_\mu^H(x)) + \frac{1}{6}\,\mathrm{tr}\left(M_\mu(x) - M_\mu^H(x)\right)\,, \text{ where} \\ M_\mu(x) &= \sum_{\nu\neq\mu} Q_x^{\mu,\nu} + Q_x^{\mu,-\nu}. \end{aligned}$$

Note the dependence of $Z_\mu^{\mathcal{U}}(x)$ on local plaquettes associated with x.

Thick Links, smearing and normality

Wilson flow, Stout smearing and the deviation from normality

Wilson flow and stout smearing

Let $\mathcal{V}_\tau$ be the solution of the Wilson flow. Then

(i) $\mathcal{V}_\tau$ is unique for all $\mathcal{V}_0$ and all $\tau \in (-\infty, \infty)$ and differentiable with respect to τ and $\mathcal{V}_0$.

(ii) The Wilson action $S_W(\mathcal{V}_\tau)$ is monotonically decreasing as a function of τ.

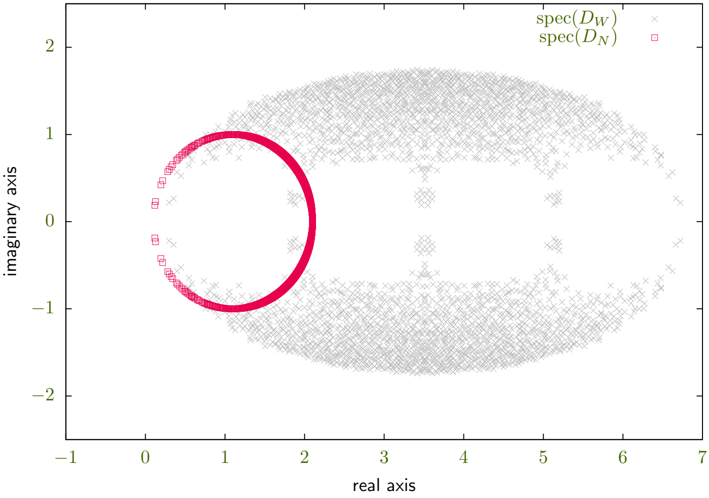
(iii) One step of Lie-Euler integration with step size ϵ for the Wilson flow initial value problem, starting at $\tau = 0$, gives the approximation $\mathcal{V}'_\epsilon = \{V'_\mu(x, \epsilon)\}$ for $\mathcal{V}_\epsilon$ with

$$V'_\mu(x, \epsilon) = \mathrm{e}^{\epsilon Z^{\mathcal{U}}_\mu(x)} U_\mu(x),$$

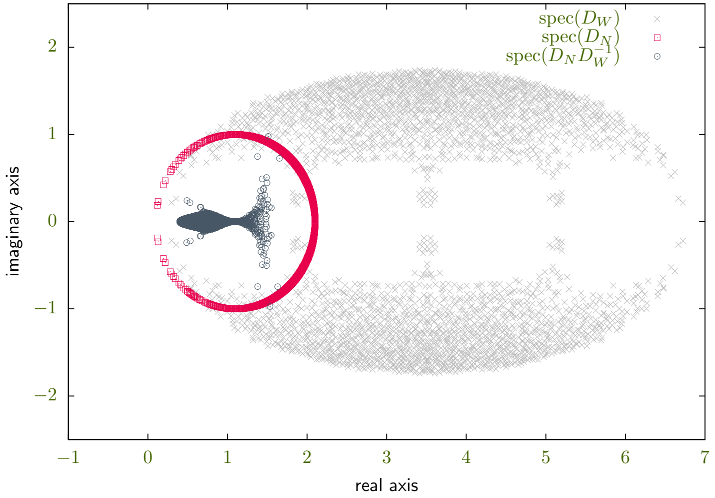
with $Z^{\mathcal{U}}_\mu(x)$ the stout smeared link variables.

For ϵ small enough, stout smearing monotonically decreases the gauge action and thus the deviation from normality.

Numerical illustration of Overlap preconditioning

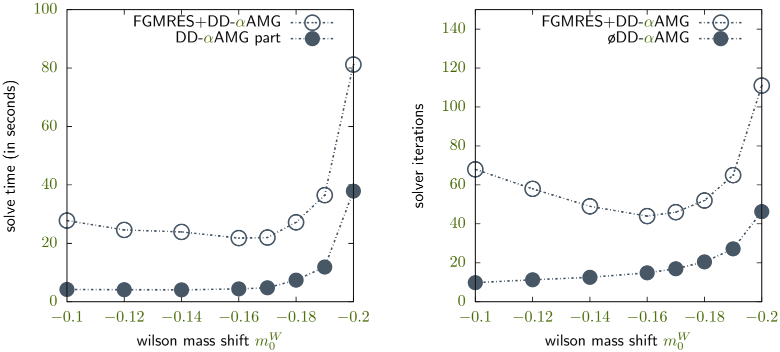


Numerical illustration of Overlap preconditioning



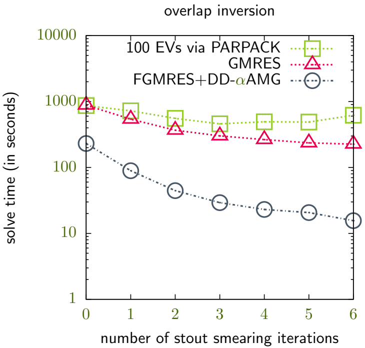
Outline

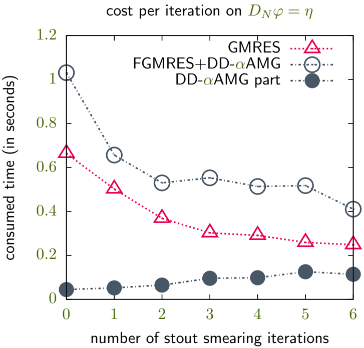
Scanning the Optimal Wilson Preconditioner Mass Shift m_0^W



- 32^4 lat, 3HEX smeared BMW-c cnfg (unpublished), $m_\pi \approx 350$ MeV, $1{,}024$ processes
- overlap tol 10^{-8}, Wilson tol 10^{-2}, sign fct with explicit deflation and relaxed tol
- optimal $m_0^W \approx 0.16$

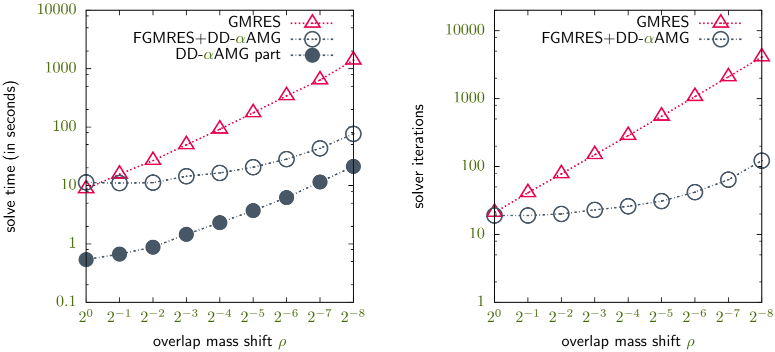
Smearing Study





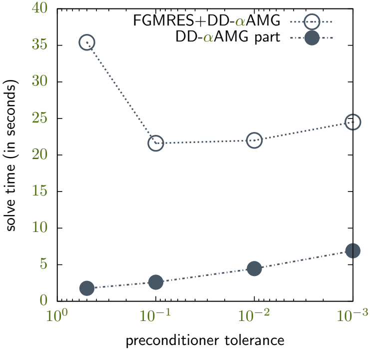
- 32^4 lattice, $1{,}024$ processes
- no smearing $\rightarrow \times 5$ speedup
- cost per iteration for preconditioned method only slightly higher
- preconditioner cost almost negligible

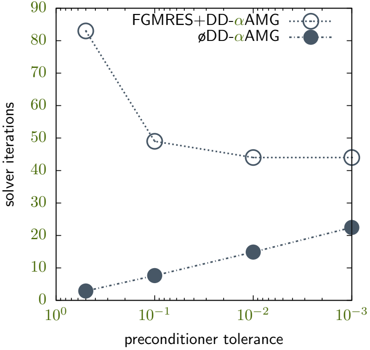
Scaling with the Overlap Mass Shift



- 32^4 lat, 3HEX smeared BMW-c cnfg
- cnfg generated at approx. $\rho = 2^{-6}$ ($m_\pi \approx 350$ MeV)
- smaller masses $\rightarrow$ bigger gain

Influence of the Preconditioner Accuracy





- 32^4 lat, 3HEX smeared BMW-c cnfg
- $tol = 10^{-2}$ optimal in terms of iteration count
- $tol = 10^{-1}$ optimal in terms of solve time

Outline

Solving the inverse square root

Challenge (ii): Evaluating $(H_W^\dagger H_W)^{-\frac{1}{2}} x$

Good convergence without explicit calculation of low modes of H_W?

Idea: Use implicit low mode information via *thick restarts*.

With the Cauchy integral representation

$$f(A) = A^{-\frac{1}{2}} = \int_\Gamma g(t)(tI - A)^{-1}\,\mathrm{d}t$$

and a Lanczos decomposition of $H_W^\dagger H_W$ we can compute the k-th error propagator by numerical quadrature

$$e^{(k)}(T) = c \sum_{i=1}^{l} \rho(T, x_i) \frac{\omega_i}{-\beta(1 - x_i) - T(1 + x_i)}$$

Solving the inverse square root

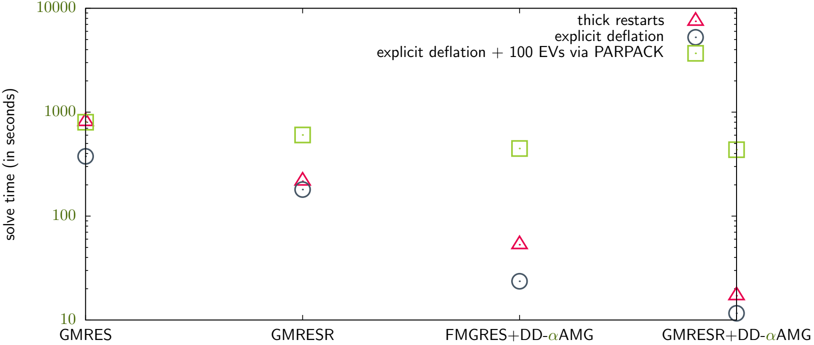
Algorithm Quadrature based restarted Lanczos approximation

```
in: A, b, f, m, t
Compute Lanczos decomposition A V_m^(1) = V_{m+1}^(1) T_{m+1,m}^(1)
f_m^(1) := ||b|| V_m^(1) f(T_{m,m}) e_1
for k = 2, 3, ... do
    Get eigenvalue decomp. of T_{m,m} and extract t smallest eigenvalues
    Extend new search space by m Lanczos steps
    Compute error h_m^(k) := e_m^(k-1)(T_m^(k)) e_{t+1} by adaptive quadrature
    f_m^(k) = f_m^(k-1) + ||b|| V_m^(k) h_m^(k)
end for
```

- **in:** A, b, f, m, t
- Compute Lanczos decomposition $AV_m^{(1)} = V_{m+1}^{(1)} T_{m+1,m}^{(1)}$
- $f_m^{(1)} := ||b|| V_m^{(1)} f(T_{m,m}) e_1$
- **for** $k = 2, 3, \ldots$ **do**
  - Get eigenvalue decomp. of $T_{m,m}$ and extract t smallest eigenvalues
  - Extend new search space by m Lanczos steps
  - Compute error $h_m^{(k)} := e_m^{(k-1)}(T_m^{(k)}) e_{t+1}$ by adaptive quadrature
  - $f_m^{(k)} = f_m^{(k-1)} + ||b|| V_m^{(k)} h_m^{(k)}$
- **end for**

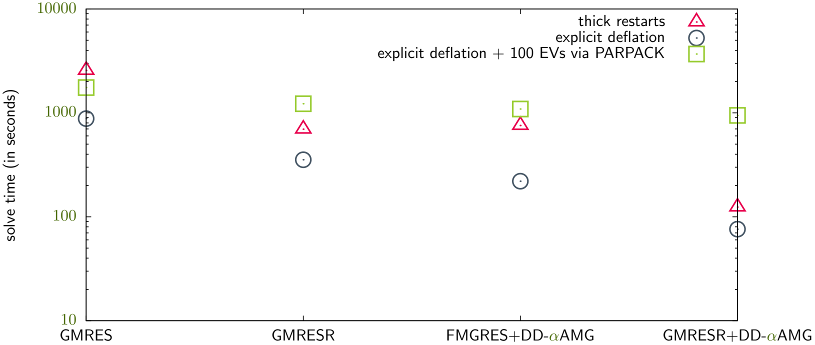
- Implicit deflation only operates on small tri-diagonal matrix
- C-MPI-Code available
- No numerical comparison yet

Thick Restarts vs. Explicit Deflation



- 32^4 lat, 3HEX smeared BMW-c cnfg, $1{,}024$ cores
- GMRESR := FGMRES-64bit + GMRES-32bit
- GMRESR+DD-αAMG := FGMRES-64bit + FGMRES-32bit + DD-αAMG

Thick Restarts vs. Explicit Deflation



- 32^4 lat, no smearing, $1{,}024$ cores
- one RHS: preconditioning + thick restarts
- many RHS: preconditioning + explicit deflation

Outline

Summary & Outlook

Summary:

- Preconditioning overlap equation leads to fewer iterations for the solution of $D_N \varphi = \eta$
- Preconditioner time almost negligible
- Efficiency of preconditioner improves
  - when approaching normality
  - for smaller masses

Outlook:

- Incoorporate solver into production codes of collaborators
- Further optimization of preconditioner
- Overall performance improvement of the method

Acknowledgments

All results computed on JUROPA at
Jülich Supercomputing Centre (JSC)

Partially funded by Deutsche Forschungsgemeinschaft (DFG),
Transregional Collaborative Research Centre 55 (SFB TR 55)

All configurations provided by BMW-c