Convergence Theory For Adaptive Smooth Aggregation Multigrid Methods Used in Lattice QCD

Edward C. White, Jr.

LATTICE 2017

The large sparse matrix problem in lattice QCD

- Aggregation-based algebraic multigrid methods have been developed to solve the lattice Wilson-Dirac system $D\psi = \chi$.
- Traditional convergence theory requires D to be Hermitian positive definite (the Wilson-Dirac matrix is not).
- Brezina, Manteuffel, McCormick, Ruge, Sanders, (2010) have developed a general two-level convergence theory for non-symmetric matrices.

The Brezina Theory

- The theory is general in the sense that it applies to matrices from all applications.
- It proves convergence for a smoothed-aggregation AMG method in $O(K_s^2)$ iterations, versus $O(K_s)$ for SPD matrices.
- It does not consider the γ_5-symmetry of the Wilson-Dirac matrix, or the use of spin symmetry in smooth aggregation.

Adaptation to Lattice QCD

- It is possible to rewrite Brezina's proof, using γ_5-symmetry and spin symmetry, to get a lattice QCD version of the proof.
- This results in a substantial improvement over $O(K_s^2)$ iteration convergence.
- The main goal of this talk is to outline the lattice QCD convergence proof, closely following Brezina's method.

Singular Vectors versus Eigenvectors

- Brezina's proof uses singular vectors to build multigrid restriction and interpolation operators.
- In practice, these are more expensive to calculate or estimate than eigenvectors. Can eigenvectors be used instead?
- A second goal of this this talk is to add new insights to this debate.

Why Is It Important?

- The proof developed here applies to two new methods by Frommer, Kahl, Kreig, Leder, Rottmann (2014), and by Brannick and Kahl (2014).
- Convergence theory can help develop newer and better numerical methods.
- The result may tell us more about γ_5 and spin symmetries.

Outline of The Proof

Convert the matrix system $D\psi = \chi$ to Hermitian form.

Construct a multigrid (two-grid) system.
Assume we have interpolation operator P, restriction operator R.

Calculate the error propagator.
It will have the form $error = f(P, R)$.

Assume that the error satisfies an approximation property.
This establishes a bound on the error.

Prove the error decreases as number of iterations increases.
This establishes convergence.

Construction of a Hermitian system

- $D\psi = \chi$ can be written as a Hermitian system.
- D can be replaced by Hermitian matrices $\sqrt{D^H D}$ or $\sqrt{DD^H}$.
- The singular value decomposition of D is $D = U\Sigma V^H$.
- Let $\Gamma_5 = \gamma_5 \otimes I_4$.

Lemma 1

Let D be γ_5-symmetric with $D = U\Sigma V^H$. Then $\Gamma_5 = VU^H$.

Proof:

- By γ_5-symmetry $(\Gamma_5 D)^H = (\Gamma_5 D)$
- $(\Gamma_5 D)^H = \left(V\Sigma^H U^H\right)\Gamma_5^H \qquad (\Gamma_5 D) = \Gamma_5\left(U\Sigma V^H\right)$
- Therefore $\left(V\Sigma^H U^H\right)\Gamma_5^H = \Gamma_5\left(U\Sigma V^H\right)$.
- Noting that $\Sigma = \Sigma^H$, $\Gamma_5 = VU^H$ satisfies this equation.

Corollary:

$$\sqrt{D^H D} = V\Sigma V^H = \Gamma_5 D \qquad \sqrt{DD^H} = U\Sigma U^H = D\Gamma_5$$

Hermitian Systems

The Γ_5 matrix applied to $D\psi = \chi$ yields two Hermitian systems:

$$\Gamma_5 D\psi = \Gamma_5\chi \tag{1}$$

$$D\Gamma_5\xi = \chi, \quad \text{with} \quad \psi = \Gamma_5\xi \tag{2}$$

Multigrid For the System $\Gamma_5 D\vec{\psi} = \Gamma_5\vec{\chi}$

- **Fine grid error:** $r = \Gamma_5\chi - \Gamma_5 D\psi$.
- **Restrict r and $\Gamma_5 D$ to the coarse grid:** Rr, $R\Gamma_5 DP$.
- **Solve for the coarse grid error**: $e_c = (R\Gamma_5 DP)^{-1}\, Rr$.
- **Interpolate e_c back to the fine grid:** Pe_c.
- **Add it to the fine grid solution:** $\psi := \psi + Pe_c$.
- **Rewrite this as:** $\psi := \psi + P(R\Gamma_5 DP)^{-1}\, Rr$.
- $e := \left(I - P(R\Gamma_5 DP)^{-1} R\Gamma_5 D)e\right) := (I - \Pi_1)\, e$.

Multigrid For the System $D\Gamma_5\xi = \chi$

- A similar analysis can be used for the second Hermitian system.

- Error propagator:
  $e := \Gamma_5 \left(I - P\,(RD\Gamma_5 P)^{-1}\; RD\Gamma_5 \right) \Gamma_5^H \,\vec{e} := (I - \Pi_2)\; e.$

- We now have two error propagator equations. The next step is to try to simplify them.

- The following lemma is important in that regard.

Lemma 2 (Frommer (2014))

Let D be γ_5 - symmetric, so that $(\Gamma_5 D)^H = \Gamma_5 D$. Then $D\psi = \lambda\psi$ if and only if $(\Gamma_5\psi)^H D = \bar{\lambda}(\Gamma_5\psi)^H$.

The association of left eigenvectors (eigenvalues) with right eigenvectors (eigenvalues) suggests the association of R with $(\Gamma_5 P)^H$.

Thus, γ_5-symmetry argues that a good choice for R is $R = (\Gamma_5 P)^H$.

Simplifying $(I - \Pi_1)$ and $(I - \Pi_2)$

Use $R \approx (\Gamma_5 P)^H$

Left Hand Side	Right Hand Side
$(I - \Pi_1)$ $(I - \Pi_1)$	$= \left(I - P\left(R\Gamma_5 DP\right)^{-1} R\Gamma_5 D)\right)$ $\approx \left(I - P\left(P^H DP\right)^{-1} P^H D\right)$
$(I - \Pi_2)$ $(I - \Pi_2)$	$= \Gamma_5 \left(I - P\left(RD\Gamma_5 P\right)^{-1} RD\Gamma_5\right) \Gamma_5^H$ $\approx \left(I - R^H\left(RDR^H\right)^{-1} RD\right)$

Lemma 3 (Babich (2010), Frommer (2014))

By exploiting γ_5-symmetry and the spin symmery of aggregates, it can be assumed that $P = R^H$.

Proof:

- In aggregates, group spin 0, spin 1 variables separately from spin 2, spin 3 variables. The two spin groups are treated separately on the coarse grid.

- In the term $\Gamma_5 P$, then, Γ_5 acts as the identity matrix on the spin 0 and spin 1 variables, and as the negative of the identity matrix on spin 2 and spin 3 variables.

- Then each nonzero block in P belongs to an aggregate multiplied by $+1$ or by -1.

Lemma 3 (Babich (2010), Frommer (2014))

- Thus, $\Gamma_5 P = P\Gamma_5^c$, where Γ_5^c is +I on variables of spin 0 or 1, and -I on variables of spin 2 or 3.
- Then on the coarse space each vector in P corresponds to two degrees of freedom. In the coarse grid correction, Γ_5 factors cancel.
- Now R was chosen to be $R = (\Gamma_5 P)^H$, and with the Γ_5 factors cancelling, $R = P^H$.

Further Simplification of $(I - \Pi_2)$

Use $R = P^H$

Approximation	Left Hand Side	Right Hand Side
Old	$(I - \Pi_2)$	$\approx \left(I - R^H \left(RDR^H\right)^{-1} RD\right)$
New	$(I - \Pi_2)$	$\approx \left(I - P\left(P^H DP\right)^{-1} P^H D\right)$
Comparison	$(I - \Pi_1)$	$\approx \left(I - P\left(P^H DP\right)^{-1} P^H D\right)$

A Bit of Intrigue

- The approximations for $(I - \Pi_1)$ and $(I - \Pi_2)$ are the same!
- Therefore, write $(I - \Pi_a) = \left(I - P\left(P^H DP\right)^{-1} P^H D\right)$ as the approximation.
- It can be shown that this expression is the same as it would be for a Hermitian positive definite matrix, except that the equation there would be exact, rather than an approximation.
- This is a very different result from the one that arises is the general nonsymmetric analysis.

Two Assumptions

Assumption	Inequality	Implication
Strong Approximation Property	$\|\| \vec{e} - P\vec{e}_c \|\|^2 \leq \frac{K_s}{\|\|\Gamma_5 D\|\|} \langle \Gamma_5 D\vec{e}, \Gamma_5 D\vec{e} \rangle$	Boundedness of e
Boundedness of Π_a	$\|\| \Pi_a \|\|_{\Gamma_5 D} < C$	Boundedness of P

Furthermore ...

- Π_a is $\Gamma_5 D$ orthogonal and $\Gamma_5^H \Gamma_5 = I$.
- Therefore, the approximation property is equivalent to

$$|| (I - \Pi_a) \vec{e} \, ||^2 \leq \frac{C^2 K_s}{||D||} \langle D\vec{e}, D\vec{e} \rangle \qquad (3)$$

- Assume that the smoother will be the Richardson iteration:

$$\mathcal{G} = \left(I - \frac{1}{||D||} D \right)^{\nu} \qquad (4)$$

where ν is the number of iterations.

The Key Error Inequality

$$|| \, (I \, - \, \Pi_a) \, \mathcal{G}\vec{e} \, ||^2_{\Gamma_5 D} \; = \; \langle (\Gamma_5 D) \, (I - \Pi_a) \, \mathcal{G}\vec{e}, (I - \Pi_a) \, \mathcal{G}\vec{e} \rangle \quad (5)$$

$$\leq \; \frac{C^2 K_s}{||D||} \langle (\Gamma_5 D) \, \mathcal{G}\vec{e}, (\Gamma_5 D) \, \mathcal{G}\vec{e} \rangle \; = \; \frac{C^2 K_s}{||D||} \; || \, (\Gamma_5 D)^{1/2} \; \mathcal{G}\vec{e} \, ||^2_{\Gamma_5 D}$$

Next, (following Brezina), decompose the error in the eigenbasis of $\Gamma_5 D$ to obtain

$$\vec{e} \; = \; \sum_{j=1}^{n} \beta_j \vec{v}_j, \qquad (\Gamma_5 D) \, \vec{v}_j \; = \; \sigma_j \vec{v}_j$$

As A Result

$$\left|\left| (\Gamma_5 D)^{1/2}\, \mathcal{G}\vec{e} \right|\right|^2_{\Gamma_5 D} \tag{6}$$

$$= \left|\left| \sum_{j=1}^{n} \sigma_j \left(1 - \frac{\sigma_j}{||D||}\right)^{\nu} \beta_j \vec{v}_j \right|\right|^2_{\Gamma_5 D}$$

$$= \left[\sup_{\sigma\in[0,||D||]} \sigma \left(1 - \frac{\sigma}{||D||}\right)^{2\nu}\right] \left[\left|\left| \vec{e} \right|\right|^2_{\Gamma_5 D}\right]$$

The Supremum Calculation

The supremum occurs at $\tilde{\sigma} = \frac{||D||}{2\nu+1}$, so that,

$$\left[\sup_{\sigma\in[0,||D||]} \sigma \left(1 - \frac{\sigma}{||D||}\right)^{2\nu}\right]$$

$$= \frac{||D||}{2\nu+1} \left[1 - \frac{||D||}{||D||\,(2\nu+1)}\right]^{2\nu}$$

$$= \frac{||D||}{2\nu+1} \left(\frac{2\nu}{2\nu+1}\right)^{2\nu} \leq \frac{4||D||}{9(2\nu+1)}$$

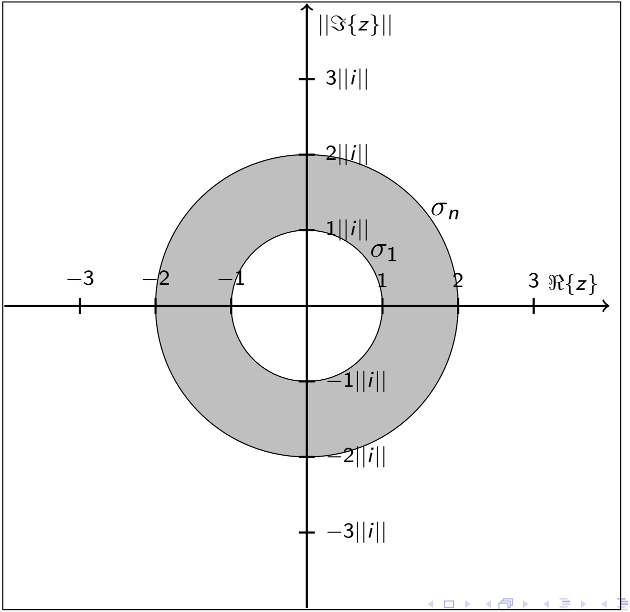
Summary of the convergence results for AMG methods.

Matrix Type	Convergence Result	# of Iterations
General Non Hermitian	$\| \, (I - I_a) \; \mathcal{R}\vec{e} \, \|^2_{\Gamma_5 D} \leq \frac{16 C^2 K_s}{25\sqrt{4\nu+1}} \; \| \, \vec{e} \, \|_{\Gamma_5 D}$	$O(K_s^2)$
γ_5 Symmetric	$\| \, (I - I_a) \; \mathcal{G}\vec{e} \, \|^2_{\Gamma_5 D} \leq \frac{4 C^2 K_s}{9(2\nu+1)} \; \| \, \vec{e} \, \|_{\Gamma_5 D}$	$O(K_s)$
Hermitian Positive Definite	$\| \, (I - I_a) \; \mathcal{G}\vec{e} \, \|^2_{\Gamma_5 D} \leq \frac{4 K_s}{9(2\nu+1)} \; \| \, \vec{e} \, \|_{\Gamma_5 D}$	$O(K_s)$

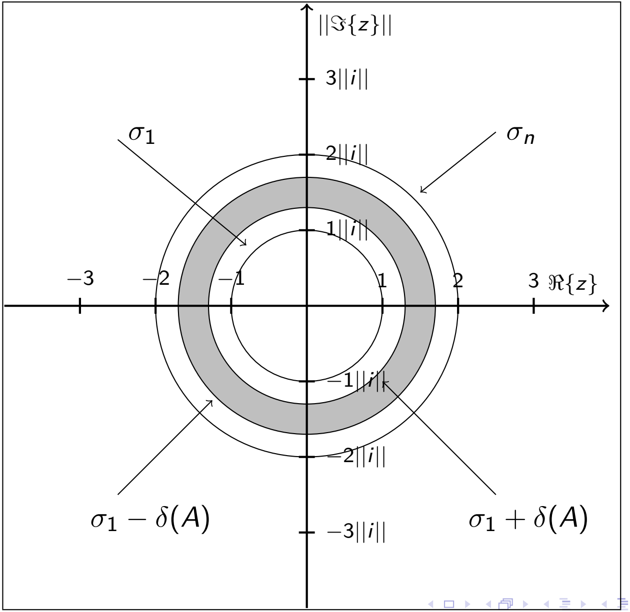
Near Kernel Vectors

- Goal: Build P, R to approximate the near kernel vectors of the Wilson-Dirac matrix.
- This can be done using small eigenvalues corresponding to near-kernel eigenvectors
- Or it can be done using small singular values corresponding to near-kernel singular vectors.

Normal Eigenvalue Inclusion Lehmann (1949)



General Eigenvalue Inclusion (Beattie, Ipsen (2003))



But That Is Not The End of the Story!

- Beattie, Ipsen argue that for general matrices, eigenvalue inclusion bounds may be very inaccurate.
- In fact, they may be too narrow.
- Realistically, the small eigenvalues could be closer to the small singular values than expected.
- Eigenvalue inclusion theorems may not conclusively settle the eigenvalue/singular value debate.

Conclusion

- A two-level convergence proof for aggregation-based multigrid methods has been developed. An extension to three levels is not difficult.
- The proof highlights the importance of γ_5 Hermiticity and spin symmetry from a purely mathematical vantage point.
- The subject of whether eigenvalues or singular values should be used in building multigrid operators has been discussed in the context of eigenvalue inclusion annulus theorems.

References

[1] M. Brezina, T. Manteuffel, S. McCormick, J. Ruge, G. Sanders, Towards Adaptive Smoothed Aggtregation (αSA) For Nonsymmetric Problems, SIAM Journal on Scientific Computing, Volume 32, Issue 1, February 2010, pp. 14-39.

[2] A. Frommer, K. Kahl, S. Kreig, B. Leder, M. Rottmann, Adaptive Aggregation-Based Domain Decomposition Multigrid for the Lattice Wilson–Dirac Operator, SIAM Journal on Sceintific Computing, Volume 36, Issue 4, 2014, pp. A1581 - A1608.

[3] J. Brannick, K. Kahl, Bootstrap Algebraic Multigrid System for the 2D Wilson Dirac System, SIAM Journal on Scientific Computing, Volume 36, Issue 3, 2014, pp. B321-B347.

References

[4] R. Babich, J. Brannick, R.C. Brower, M. A. Clark, T. A. Manteuffel, S. F. McCormick, J.C. Osborn, C. Rebbi, Adaptive Multigrid Algorithm for the Lattice Wilson-Dirac Operator, Physical Review Letters 105 (20) 201602, 2010.

[5] N. Lehmann, Calculation of Eigenvalue Bounds in Linear Problems, Arch. Math. 2, 1949, pp.139-147.

[6] C. Beattie, I.C.F. Ipsen, Inclusion Regions For Matrix Eigenvalues, Linear Algebra and its Applications, Volume 358, Issues 1-3, 2013, pp. 281-291.