Adaptive algebraic multigrid for nearly singular matrices

Scott MacLachlan
scott.maclachlan@tufts.edu

Department of Mathematics, Tufts University

July 6, 2009

With Thanks to ...

This research was performed in collaboration with

- James Brannick, Penn State
- Andreas Frommer, Karsten Kahl, University of Wuppertal

This research was supported by

- the European Community's Sixth Framework Programme, through a Marie Curie Incoming International Fellowship, MIF1-CT-2006-021927
- the National Science Foundation, under grants DMS-0811022 and OCI-0749317

Algebraic Multigrid

Family of multigrid algorithms with PDE-independent formulation

Rely on complementarity between

- Relaxation
  - Matrix splittings: Jacobi, Gauss-Seidel, ILU
- Coarse-grid correction
  - Graph-based coarsening
  - Algebraic interpolation operator

A. Brandt, S. McCormick, J. Ruge, in *Sparsity and Its Applications*, 1984
J. Ruge and K. Stüben, in *Multigrid Methods*, 1987

Determining Interpolation

Coarse-grid correction must correct what relaxation doesn't

Two-grid algorithm:

1. Relax: $\mathbf{x} \leftarrow \mathbf{x} + M_1(\mathbf{b} - A\mathbf{x})$
2. Restrict: $\mathbf{b}_c = R(\mathbf{b} - A\mathbf{x})$
3. Coarse-grid solve: $B_c\mathbf{x}_c = \mathbf{b}_c$
4. Interpolate correction: $\mathbf{x} = \mathbf{x} + P\mathbf{x}_c$
5. Relax: $\mathbf{x} \leftarrow \mathbf{x} + M_2(\mathbf{b} - A\mathbf{x})$

Error-Propagation Operator:

$$(I - M_2A)(I - PB_c^{-1}RA)(I - M_1A)$$

Determining Interpolation

Coarse-grid correction must correct what relaxation doesn't

$$(I - M_2 A)(I - P B_c^{-1} R A)(I - M_1 A)$$

Slow-to-converge errors must be in Range(P)

- Known *a priori*
  - elliptic PDEs
  - eigenvalue problem
- Unknown
  - Look for these with adaptive process

Adaptive AMG

Slow-to-converge errors can be found experimentally

- Choose random initial guess, $\mathbf{x}^{(0)}$
- Run Relaxation: $\mathbf{x}^{(k+1)} = (I - MA)\mathbf{x}^{(k)}$

Power method on $(I - MA)$
Converges to slow-to-converge modes of relaxation

Problem: Convergence itself can be slow

- Accelerate adaptive process with coarse-grid correction

M. Brezina et al., SISC 2004, **25**:1896-1920; SISC 2006, **27**:1261-1286

How well does it work?

Bilinear finite element discretizations of $-\nabla \cdot \mathcal{K}\nabla p$

- Problem 1: $\mathcal{K} = 1$, Dirichlet BCs
- Problem 2: $\mathcal{K}(x) = \begin{cases} 10^{-8} & x \in [\frac{1}{3}, \frac{2}{3}]^2, \\ 1 & \text{otherwise.} \end{cases}$

Asymptotic AMG V-cycle convergence factors

	128×128	256×256	512×512	1024×1024
Problem 1	0.115	0.124	0.131	0.137
Problem 2	0.122	0.130	0.136	0.141

Scale matrices symmetrically by D, where $d_{ii} = 10^{5r_i}$ for $\{r_i\}$ uniformly distributed on $[0, 1]$

Asymptotic AMG V-cycle convergence factors

	128×128	256×256	512×512	1024×1024
Problem 1r	0.997	0.996	0.996	0.996
Problem 2r	0.997	0.996	0.996	0.996

M. Brezina et al., SISC 2006, **27**:1261-1286

How well does it work?

Bilinear finite element discretizations of $-\nabla \cdot \mathcal{K} \nabla p$

- Problem 1: $\mathcal{K} = 1$, Dirichlet BCs
- Problem 2: $\mathcal{K}(x) = \begin{cases} 10^{-8} & x \in [\frac{1}{3}, \frac{2}{3}]^2, \\ 1 & \text{otherwise.} \end{cases}$

Asymptotic AMG V-cycle convergence factors

	128×128	256×256	512×512	1024×1024
Problem 1	0.115	0.124	0.131	0.137
Problem 2	0.122	0.130	0.136	0.141

Now use adaptive AMG to calibrate interpolation for scaled matrix

Asymptotic adaptive AMG V-cycle convergence factors

	128×128	256×256	512×512	1024×1024
Problem 1r	0.078	0.077	0.078	0.079
Problem 2r	0.100	0.084	0.111	0.108

M. Brezina et al., SISC 2006, **27**:1261-1286

Lattice Quantum Chromodynamics

- Modeling interactions between fermions on a lattice
- Goal: Solve $H(\mathbf{u}, m)\mathbf{f} = \mathbf{b}$, for multiple source vectors, $\mathbf{b}$, at each step of a Monte Carlo simulation
- Difficulty: $\mathbf{u}$ is a complex unitary field defined on lattice edges, phases chosen randomly based on parameter, m
- As m approaches a $\mathbf{u}$-dependent critical value, H becomes singular
- Structure of near-kernel modes strongly depends on $\mathbf{u}$
  - When $m \rightarrow \infty$, $\mathbf{u} \rightarrow \mathbf{1}$, H looks like a discrete differential operator
  - For each state, new characterization of near-kernel modes

Model Problem: Gauged Laplacian

Finite-difference Laplace Operator:

$$\frac{1}{h^2}\begin{bmatrix} & -1 & \\ -1 & 4 & -1 \\ & -1 & \end{bmatrix}$$

Gauged-Laplace Operator:

$$\frac{1}{h^2}\begin{bmatrix} & -e^{\imath\phi_{k,\ell}} & \\ -e^{-\imath\theta_{k-1,\ell}} & 4-m & -e^{\imath\theta_{k,\ell}} \\ & -e^{-\imath\phi_{k,\ell-1}} & \end{bmatrix}$$

Near-kernel depends on

- m
- $\theta_{k,\ell}$'s, $\phi_{k,\ell}$'s

Effect of conditioning

Performance of adaptive process

- Shifted Laplace Problem: $-\Delta u - 2\pi^2(1 - 2^{-\beta})u = 0$
- Dirichlet BCs
- Discretized with mass matrix, 512×512 grid
- RRQ = relative error in Rayleigh Quotient of prototype

β	1 Cycle		2 Cycles		3 Cycles		4 Cycles	
	RRQ	ρ_{MG}	RRQ	ρ_{MG}	RRQ	ρ_{MG}	RRQ	ρ_{MG}
0	599	0.984	14	0.421	0.4	0.071	0.2	0.071
5	19K	0.9995	438	0.959	6	0.311	0.2	0.103
10	611K	1	14K	0.999	190	0.929	5	0.256

Eigenvector-Approximation Criterion

For each eigenvector, $\mathbf{v}$, of A, interpolation must represent $\mathbf{v}$ with accuracy proportional to its eigenvalue

Consequences:

- If $\mathbf{v}_{\text{min}}$ is known, then AMG is fine
- Work in adaptive process depends on λ_{min}

A. Brandt, Appl. Math. Comput. 1986, **19**:23-56

S. McCormick and J. Ruge, SINUM 1985, **19**:924-929

Understanding Convergence: AMGr

Suppose A is HPD, and that we can partition the grid, $\Omega = F \cup C$, so that $A = \begin{bmatrix} A_{ff} & -A_{fc} \\ -A_{fc}{}^{\star} & A_{cc} \end{bmatrix}$,

$$\mathbf{x}_f^{\star} M_{ff} \mathbf{x}_f \leq \mathbf{x}_f^{\star} A_{ff} \mathbf{x}_f \leq \lambda_{\text{max}} \mathbf{x}_f^{\star} M_{ff} \mathbf{x}_f$$

and that $\begin{bmatrix} M_{ff} & -A_{fc} \\ -A_{cf} & A_{cc} \end{bmatrix}$ is positive semi-definite. Choose

Relaxation: $I - \frac{2}{1+\lambda_{\text{max}}} \begin{bmatrix} M_{ff}^{-1} & 0 \\ 0 & 0 \end{bmatrix} A$

Coarse-grid correction: variational with $P = \begin{bmatrix} M_{ff}^{-1} A_{fc} \\ I \end{bmatrix}$

Then

$$\rho_{\text{MG}} \leq 1 - \left(\frac{2}{\lambda_{\text{max}} + 1} \right)^2$$

M. Ries, U. Trottenberg, G. Winter, JLAA 1983, **49**:1-26

S. MacLachlan, T. Manteuffel, S. McCormick, NLAA 2006, **13**:599-620

Choosing M_{ff}

How to choose M_{ff}? Classic MG principle:

- Need errors relaxation doesn't change to be in Range(P)
- These errors are in Range$\left(\begin{bmatrix} A_{ff}^{-1}A_{fc} \\ I \end{bmatrix}\right)$
- Most important when $\mathbf{x}_c^\star \begin{bmatrix} A_{ff}^{-1}A_{fc} \\ I \end{bmatrix}^\star A \begin{bmatrix} A_{ff}^{-1}A_{fc} \\ I \end{bmatrix} \mathbf{x}_c$ is small

Adaptive AMGr:

1. Find $\mathbf{x}_c$ such that $\mathbf{x}_c^\star \begin{bmatrix} A_{ff}^{-1}A_{fc} \\ I \end{bmatrix}^\star A \begin{bmatrix} A_{ff}^{-1}A_{fc} \\ I \end{bmatrix} \mathbf{x}_c$ is small
2. Choose M_{ff} by fixing $\begin{bmatrix} M_{ff}^{-1}A_{fc} \\ I \end{bmatrix} \mathbf{x}_c = \begin{bmatrix} A_{ff}^{-1}A_{fc} \\ I \end{bmatrix} \mathbf{x}_c$

Define M_{ff} by $M_{ff}^{-1}A_{fc}\mathbf{x}_c = A_{ff}^{-1}A_{fc}\mathbf{x}_c$

Sufficient Conditions

AMGr conditions are sufficient, but not necessary

For Gauged Laplacian:

- A is complex-valued
- $\mathbf{x}_c$ is complex-valued
- M_{ff} is complex-valued

Theory requires $\begin{bmatrix} M_{ff} & -A_{fc} \\ -A_{cf} & A_{cc} \end{bmatrix}$ to be positive semi-definite.

This is impossible!

Question: Can we generalize theory with more appropriate assumptions?

Splitting M_{ff}

M_{ff} plays two roles in AMGr
relaxation and coarse-grid correction

Choose two matrices instead of one:

- M_R for relaxation
- M_P for coarse-grid correction

What are sufficient conditions on M_R and M_P to prove AMGr-type bound?

New AMGr Theorem

Let M_R be HPD and assume there are constants $0 < c_1 \leq c_2 < 2$ such that

$$c_1 \mathbf{x}_f^\star M_R \mathbf{x}_f \leq \mathbf{x}_f^\star A_{ff} \mathbf{x}_f \leq c_2 \mathbf{x}_f^\star M_R \mathbf{x}_f .$$

Let $P_{M_P} = \begin{bmatrix} M_P^{-1} A_{fc} \\ I \end{bmatrix}$, $A_c = P_{M_P}^\star A P_{M_P}$, and let γ be the smallest constant such that

$$\mathbf{x}_c^\star A_c \mathbf{x}_c \leq \frac{1}{1-\gamma^2} \mathbf{x}_c^\star \begin{bmatrix} A_{ff}^{-1} A_{fc} \\ I \end{bmatrix}^\star A \begin{bmatrix} A_{ff}^{-1} A_{fc} \\ I \end{bmatrix} \mathbf{x}_c$$

Then $\rho_{MG} \leq 1 - \frac{1-\gamma^2}{\alpha}$, where $\alpha = \max\left(\frac{1}{c_1(2-c_1)}, \frac{1}{c_2(2-c_2)} \right)$.

J. Brannick, A. Frommer, K. Kahl, S. MacLachlan, 2009, in preparation.

Extension to Full-Grid Smoothing

Subspace decomposition theory extends this bound to

$$\rho_{MG} \leq 1 - \frac{1-\gamma^2}{\kappa}$$

where γ is as before, and κ is given by

$$\mathbf{x}^\star A\mathbf{x} \leq \mathbf{x}^\star \widetilde{M}\mathbf{x} \leq \kappa\mathbf{x}^\star A\mathbf{x}$$

for $\widetilde{M} = M^\star(M^\star + M - A)^{-1}M$

R. Falgout, P. Vassilevski, L. Zikatanov, NLAA 2005, **12**:471-494

New Adaptive AMGr

- Partition $\Omega = F \cup C$ so that A_{ff} is diagonally dominant
- Full-grid Gauss-Seidel Relaxation
- AMGr-style interpolation with $P_{M_P} = \begin{bmatrix} M_P^{-1}A_{fc} \\ I \end{bmatrix}$
- M_P chosen so that $M_P^{-1}A_{fc}\mathbf{x}_c = A_{ff}^{-1}A_{fc}\mathbf{x}_c$ for $\mathbf{x}_c$ given by relaxation on $A\mathbf{x} = \mathbf{0}$

Convergence guaranteed by theory, but bound depends on γ, κ

J. Brannick, A. Frommer, K. Kahl, S. MacLachlan, 2009, in preparation.

Optimal Coarsening

Define A_{ff} to be θ-dominant if, for each $i \in F$,

$$a_{ii} \geq \theta \sum_{j \in F} |a_{ij}|$$

Coarsening Goal: Find largest set F such that A_{ff} is θ-dominant.

The problem, $\max\{|F| : A_{ff} \text{ is } \theta\text{-dominant}\}$, is NP-complete.

S. MacLachlan, Y. Saad, SIAM J. Sci. Comput. 2007, **29**:1825-1853

Greedy Coarsening

Instead,

- Initialize $U = \{1, \ldots, n\}$, $F = C = \emptyset$
- For each point in U, compute $\hat{\theta}_i = \dfrac{a_{ii}}{\displaystyle\sum_{j \in F \cup U} |a_{ij}|}$
- Whenever $\hat{\theta}_i \geq \theta$, $i \to F$
- If $U \neq \emptyset$, then pick $j = \text{argmin}_{i \in U}\{\hat{\theta}_i\}$
  - $j \to C$
  - Update $\hat{\theta}_i$ for all $i \in U$ with $a_{ji} \neq 0$

S. MacLachlan, Y. Saad, SIAM J. Sci. Comput. 2007, **29**:1825-1853

Numerical Results

Finite-Difference Laplacian with Dirichlet BCs, shifted so smallest eigenvalue is $\lambda_{\min}$. V(2,2) cycles

$n_{rel} \setminus \lambda_{\min}$	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}
5	.06	.02	.04	.37	.85	.98
25	.07	.02	.05	.05	.38	.86
50	.07	.02	.05	.05	.17	.66
100	.07	.02	.06	.06	.06	.16
500	.07	.02	.06	.06	.06	.06
exact	.07	.02	.06	.06	.06	.06

Results labeled 'exact' use eigenvector of A with smallest eigenvalue to form M_P.

J. Brannick, A. Frommer, K. Kahl, S. MacLachlan, 2009, in preparation.

Numerical Results

Gauged Laplacian with periodic BCs, shifted so smallest eigenvalue is $\lambda_{\min}$. V(2,2) cycles

$n_{rel} \setminus \lambda_{\min}$	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}
5	.4	.79	.97	.99	.99	.99
25	.32	.53	.83	.98	.99	.99
50	.31	.55	.72	.95	.99	.99
100	.28	.52	.65	.9	.99	.99
300	.32	.48	.53	.54	.61	.89
500	.33	.5	.6	.6	.60	.62
exact	.31	.53	.61	.61	.62	.62

J. Brannick, A. Frommer, K. Kahl, S. MacLachlan, 2009, in preparation.

Numerical Results

Gauged Laplacian with periodic BCs, shifted so smallest eigenvalue is λ_{min}. V(2,2) cycles with CG

$n_{rel} \setminus \lambda_{\text{min}}$	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}
5	9	15	19	21	23	25
25	9	11	14	15	17	18
50	8	11	12	14	15	17
100	8	10	13	14	16	17
300	8	10	10	10	11	13
500	8	10	11	11	11	11
exact	8	10	12	11	12	12
CG	44	75	107	231	343	435

Results labeled 'CG' are for conjugate gradient with no preconditioner

J. Brannick, A. Frommer, K. Kahl, S. MacLachlan, 2009, in preparation.

Summary

- Adaptivity is difficult for nearly singular matrices
- Quantum chromodynamics gives nearly singular disordered systems
- Improved AMGr-type theory, improved AMGr algorithm
- Haven't really "solved" difficulties with near singularity

Summary

- Adaptivity is difficult for nearly singular matrices
- Quantum chromodynamics gives nearly singular disordered systems
- Improved AMGr-type theory, improved AMGr algorithm
- Haven't really "solved" difficulties with near singularity

Future Directions

- Explore question of good AMG(r) preconditioners vs solvers
- Develop more efficient solvers for QCD
- "Solve" difficulties with near singularity