A Greedy Strategy for Coarse-Grid Selection

Scott MacLachlan Yousef Saad

Department of Computer Science and Engineering,
University of Minnesota
{maclach,saad}@cs.umn.edu

Copper Mountain Conference on Iterative Methods
April 4,2006

Multilevel Solvers

Performance shouldn't degrade with increased problem size

Stationary iterative methods:

- Norm of $I - B^{-1}A$ must be bounded uniformly below one

Preconditioned Krylov methods,

- Condition number, $\kappa(B^{-\frac{1}{2}}AB^{-\frac{1}{2}})$, must be uniformly bounded

Multilevel techniques achieve this uniformity by exploiting multiscale structure

Block Factorization

Partition

$$A\mathbf{x} = \left[\begin{array}{cc} A_{ff} & -A_{fc} \\ -A_{cf} & A_{cc} \end{array} \right] \left(\begin{array}{c} \mathbf{x}_f \\ \mathbf{x}_c \end{array} \right) = \left(\begin{array}{c} \mathbf{b}_f \\ \mathbf{b}_c \end{array} \right) = \mathbf{b},$$

then block factor,

$$A = \left[\begin{array}{cc} I & 0 \\ -A_{cf} A_{ff}^{-1} & I \end{array} \right] \left[\begin{array}{cc} A_{ff} & 0 \\ 0 & \hat{A}_{cc} \end{array} \right] \left[\begin{array}{cc} I & -A_{ff}^{-1} A_{fc} \\ 0 & I \end{array} \right],$$

where $\hat{A}_{cc} = A_{cc} - A_{cf} A_{ff}^{-1} A_{fc}$.

Algebraic Recursive Multilevel Solver

Approximate A_{ff} by its ILUT factors, $A_{ff} \approx LU$.
Preconditioner is

$$B = \begin{bmatrix} I & 0 \\ -A_{cf}U^{-1}L^{-1} & I \end{bmatrix} \begin{bmatrix} LU & 0 \\ 0 & S \end{bmatrix} \begin{bmatrix} I & -U^{-1}L^{-1}A_{fc} \\ 0 & I \end{bmatrix},$$

where $S \approx A_{cc} - A_{cf}U^{-1}L^{-1}A_{fc}$.
Coarse-grid problems

- computed using techniques akin to ILUT
- solved recursively

Y. Saad and B. Suchomel, Numer. Linear Algebra Appl. 2002, **9**:359-378

Multigrid

Multigrid Components

Relax: $x^{(1)} = x^{(0)} + D^{-1} r^{(0)}$

- Relaxation

- Use a smoothing process (such as Jacobi or Gauss-Seidel) to eliminate oscillatory errors
- Remaining error satisfies $Ae^{(1)} = r^{(1)} = b - Ax^{(1)}$

Multigrid

Multigrid Components

Relax: $x^{(1)} = x^{(0)} + D^{-1} r^{(0)}$

Restriction

- Relaxation
- Restriction
- Transfer residual to coarse grid
- Compute $P^T r^{(1)}$

Multigrid

Multigrid Components

- Relaxation
- Restriction
- Coarse-Grid Correction

Relax: $x^{(1)} = x^{(0)} + D^{-1} r^{(0)}$

Restriction

Solve: $P^T A P x_c = P^T r^{(1)}$

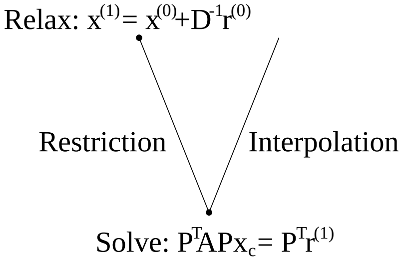
- Use coarse-grid correction to eliminate smooth errors
- Best correction, x_c, in terms of A-norm satisfies

$$P^T A P x_c = P^T r^{(1)}$$

Multigrid

Multigrid Components

- Relaxation
- Restriction
- Coarse-Grid Correction
- Interpolation

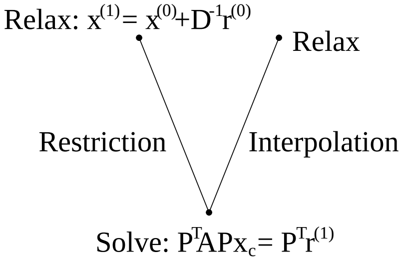


- Transfer correction to fine grid
- Compute $x^{(2)} = x^{(1)} + P x_c$

Multigrid

Multigrid Components

- Relaxation
- Restriction
- Coarse-Grid Correction
- Interpolation
- Relaxation

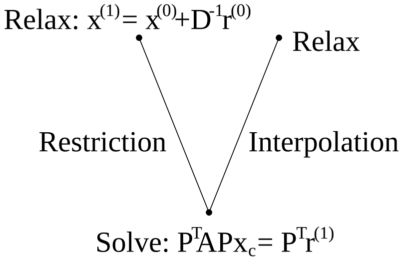


- Relax once again to remove oscillatory error introduced in coarse-grid correction

Multigrid

Multigrid Components

- Relaxation
- Restriction
- Coarse-Grid Correction
- Interpolation
- Relaxation



Direct solution of coarse-grid problem isn't practical

Recursion!

Apply same methodology to solve coarse-grid problem

Algebraic Multigrid (AMG)

- Goal of coarsening is to complement fixed relaxation
- Variational formulation
  - Coarse-grid correction is optimal in A-norm
  - Algebraically smooth error must be in range of interpolation
- Choose coarse-grid, C, and interpolation, P,
  - using only algebraic information
  - with knowledge of (assumed) algebraically smooth errors

A. Brandt, S. McCormick, J. Ruge, in *Sparsity and Its Applications*, 1984

J. Ruge and K. Stüben, in *Multigrid Methods*, 1987

Partitioning

Choice of partition in

ARMS: affects sparsity in ILU
controls size of Schur complement

AMG: influences sparsity in P
determines size of CGO

Good partitioning

- adequately reduces dimension of coarse-scale problem
- allows sparse choices of P or LU without sacrificing accuracy
- enables recursive solve for coarse-scale problem

Goal of partitioning is to enable efficient resolution of coarse-scale errors

Two-level Theory

- Goal is to use theory to inform algorithmic choices
- Solution on a given level depends only on quality of solution on next coarser level
- Multilevel theory can be intricate

Partition

$$A\mathbf{x} = \begin{bmatrix} A_{ff} & -A_{fc} \\ -A_{cf} & A_{cc} \end{bmatrix} \begin{pmatrix} \mathbf{x}_f \\ \mathbf{x}_c \end{pmatrix} = \begin{pmatrix} \mathbf{b}_f \\ \mathbf{b}_c \end{pmatrix} = \mathbf{b}$$

Use two-level analysis to make choices within a multilevel scheme

ARMS Analysis

Let

- $B = \begin{bmatrix} I & 0 \\ -A_{cf}D_{ff}^{-1} & I \end{bmatrix} \begin{bmatrix} D_{ff} & 0 \\ 0 & S \end{bmatrix} \begin{bmatrix} I & -D_{ff}^{-1}A_{fc} \\ 0 & I \end{bmatrix}$
- $\begin{bmatrix} D_{ff} & -A_{fc} \\ -A_{cf} & Acc \end{bmatrix}$ be positive semi-definite
- $\mathbf{x}_f^T D_{ff}\mathbf{x}_f \leq \lambda_{\min}\mathbf{x}_f^T D_{ff}\mathbf{x}_f \leq \mathbf{x}_f^T A_{ff}\mathbf{x}_f \leq \lambda_{\max}\mathbf{x}_f^T D_{ff}\mathbf{x}_f$
- $\nu_{\min}\mathbf{x}_c^T S\mathbf{x}_c \leq \mathbf{x}_c^T \hat{A}_{cc}\mathbf{x}_c \leq \nu_{\max}\mathbf{x}_c^T S\mathbf{x}_c$

Then,

$$\kappa(B^{-\frac{1}{2}}AB^{-\frac{1}{2}}) \leq \left(1 + \sqrt{1 - \frac{1}{\lambda_{\max}}}\right)^2 \frac{\lambda_{\max}^2\nu_{\max}}{\min(\nu_{\min}, \lambda_{\min})}.$$

Y. Notay, Numer. Linear Algebra Appl. 2005, **12**:419-451

Generalized AMG Measure

Let

- Relaxation be given by $I - D^{-1}A$
- Q be a projection onto the range of P
- $\mu(Q, \mathbf{e}) = \frac{\langle D(D+D^T-A)^{-1}D^T(I-Q)\mathbf{e},(I-Q)\mathbf{e}\rangle}{\langle A\mathbf{e},\mathbf{e}\rangle} \leq K$ for $\mathbf{e} \neq \mathbf{0}$
- MG_2 be a two-grid V(0,1)-cycle

Then,

$$\|MG_2\|_A \leq \left(1 - \frac{1}{K}\right)^{\frac{1}{2}}$$

R. Falgout and P. Vassilevski, SIAM J. Numer. Anal. 2004, **42**:1669-1693

Compatible Relaxation

"A general measure for the quality of the set of coarse variables is the convergence rate of the compatible relaxation"

One approach:

- Run relaxation on tentative F-set
- Identify points where compatible relaxation is slow
- Choose subset of these points to add to C

A. Brandt, Elect. Trans. Numer. Anal. 2000, **10**:1-20

O. Livne, Numer. Linear Algebra Appl. 2004, **2**:205-227

J. Brannick, Wednesday 11:00

Compatible Relaxation

Let

- D be symmetric
- $2D - A$ be positive definite
- $\mathbf{x}^T A \mathbf{x} \leq \omega \mathbf{x}^T D \mathbf{x}$
- $\rho_f = \|I - D_{ff}^{-1} A_{ff}\|_{A_{ff}}$

Then,

$$\min_P \max_{\mathbf{e} \neq \mathbf{0}} \mu(Q, \mathbf{e}) \leq \frac{1}{(2-\omega)(1-\rho_f)}$$

For a given F/C partition, the best possible measure depends on the equivalence between D_{ff} and A_{ff}

R. Falgout and P. Vassilevski, SIAM J. Numer. Anal. 2004, **42**:1669-1693

Reduction-based AMG

Let

- Relaxation be fine-grid only, $I - \frac{2}{\lambda_{\max}+1} D_{ff}^{-1} A_{ff}$
- $P = \begin{bmatrix} D_{ff}^{-1} A_{fc} \\ I \end{bmatrix}$
- $\mathbf{x}_f^T D_{ff} \mathbf{x}_f \le \mathbf{x}_f^T A_{ff} \mathbf{x}_f \le \lambda_{\max} \mathbf{x}_f^T D_{ff} \mathbf{x}_f$
- $\begin{bmatrix} D_{ff} & -A_{fc} \\ -A_{cf} & Acc \end{bmatrix}$ be positive semi-definite

Then

$$\|MG_2\|_A \le \left(\frac{1}{\lambda_{\max}} \left(\lambda_{\max} - 1 + \left(\frac{\lambda_{\max} - 1}{\lambda_{\max} + 1} \right)^2 \right) \right)^{\frac{1}{2}}$$

S. MacLachlan, T. Manteuffel, S. McCormick, Numer. Linear Algebra Appl. 2006, to appear

Coarsening

All three bounds depend on equivalence of D_{ff} and A_{ff}

Good partition allows
- effective reduction, $|C| \ll |F|$
- efficient computation of $D_{ff}^{-1}\mathbf{y}_f$ or $D_{ff}^{-1}A_{fc}$
- good equivalence, $\lambda_{\max}$ small

A new approach to Compatible Relaxation

- Identify a property of A_{ff} that guarantees good equivalence
- Choose F so that this is always true

Diagonal Dominance

Jacobi on A_{ff} converges if it is diagonally dominant
Stronger dominance $\rightarrow$ faster convergence

A_{ff} is 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.

Complexity

The problem, $\max\{|F| : A_{ff} \text{ is } \theta\text{-dominant}\}$, is NP-complete. 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 $\hat{\theta}_i \geq \theta$, $i \rightarrow F$
- If $U \neq \emptyset$, then pick $j = \text{argmin}_{i \in U}\{\hat{\theta}_i\}$
    - $j \rightarrow C$
    - Update $\hat{\theta}_i$ for all $i \in U$ with $a_{ji} \neq 0$

Solvers

Two-level analysis gives uniform spectral equivalence of A_{ff} with its diagonal, D_{ff}.

For multilevel solvers,

ARMS: D_{ff} is sparsest possible ILU of A_{ff}

AMG: $D_{ff}^{-1}A_{fc}$ is very simple AMG interpolation operator

Solvers

Two-level analysis gives uniform spectral equivalence of A_{ff} with its diagonal, D_{ff}.

For multilevel solvers,

ARMS: D_{ff} is sparsest possible ILU of A_{ff}

AMG: $D_{ff}^{-1}A_{fc}$ is very simple AMG interpolation operator

Combination of dominance-based partitioning and classical algebraic coarsening leads to robust, efficient multilevel solvers

AMG: V(1,1) cycles, Full Gauss-Seidel, greedy coarsening with second pass, classical AMG interpolation

ARMS: symmetrized ILU, fixed drop and fill thresholds, preconditioned GMRES

PDE Test Problems

Two-dimensional bilinear finite element discretizations of

$$-\nabla \cdot K(x,y)\nabla p(x,y) = 0.$$

Problem 1: $K(x,y) = 1$

Problem 2: $K(x,y) = 10^{-8} + 10(x^2 + y^2)$

Problem 3: $K(x,y) = 10^{-8}$ on 20% of the cells, chosen randomly; $K(x,y) = 1$ otherwise

Problem 4: $K(x,y) = \left[\begin{smallmatrix} 1 & 0 \\ 0 & 0.01 \end{smallmatrix}\right]$

AMG Results

Prob.	Grid	c_A	t_{setup}	t_{solve}	# iters.	ρ
1	512×512	1.33	1.3	0.7	5	0.13
	1024×1024	1.33	5.1	2.5	5	0.14
	2048×2048	1.33	21.9	10.5	5	0.14
2	512×512	1.33	1.3	0.6	5	0.13
	1024×1024	1.33	5.1	2.5	5	0.14
	2048×2048	1.33	21.7	10.4	5	0.14
3	512×512	2.06	2.3	1.2	6	0.35
	1024×1024	2.08	9.6	4.8	6	0.40
	2048×2048	2.10	41.0	19.8	6	0.46
4	512×512	2.39	1.5	1.0	5	0.13
	1024×1024	2.41	6.2	4.1	5	0.20
	2048×2048	2.43	25.8	17.7	5	0.20

ARMS Results

Prob.	Grid	c_B	t_{setup}	t_{solve}	# iters.
1	128×128	2.65	0.2	0.3	28
	256×256	2.67	1.3	2.2	44
	512×512	2.68	11.0	22.5	82
2	128×128	2.39	0.2	0.3	31
	256×256	2.35	0.8	2.9	56
	512×512	2.32	3.0	28.2	97
3	128×128	1.40	0.2	0.3	30
	256×256	1.42	0.7	2.2	45
	512×512	1.42	3.0	22.9	83
4	128×128	1.61	0.2	0.3	26
	256×256	1.62	0.8	2.0	42
	512×512	1.63	3.2	16.2	65

General ARMS Tests

- Test set from Rutherford-Appleton Labs
- 22 Selected problems, from 120K to 3.6M non-zeros
- Compared to ILUTP, fill factors adjusted to match ARMS preconditioner complexities

N. Gould and J. Scott, ACM Trans. Math. Softw. 2004, **30**:300-325

General ARMS Tests

- Test set from Rutherford-Appleton Labs
- 22 Selected problems, from 120K to 3.6M non-zeros
- Compared to ILUTP, fill factors adjusted to match ARMS preconditioner complexities

Results:

- ARMS converged in available memory (2GB + 1 GB swap) on 21 problems
- ILUTP converged for 13 problems, limited to memory or $2\times$ ARMS iteration count
- ILUTP needed fewer iterations for 7 problems
- Equal performance for 4
- ARMS faster for 10

N. Gould and J. Scott, ACM Trans. Math. Softw. 2004, **30**:300-325

Nonsymmetric ARMS

Naïve Approach

- Choose row or column diagonal dominance
- Updates for row dominance require transpose

Nonsymmetric Permutations

- Choose offdiagonals as pivots to maximize dominance
- Simultaneously aim for row and column dominance

Results

- Test problems from earlier paper
- Naïve approach easily solves 31 of 45 problems
- Nonsymmetric permutation approach solves 43 of 45

Y. Saad, SIAM J. Sci. Comp. 2006, **27**:1032-1057

Summary

- Theoretical motivation: fine-scale spectral equivalence
- Choose partition to guarantee good equivalence
- Diagonal dominance is simple, but effective
- Multilevel results show robustness and efficiency

http://www.cs.umn.edu/~maclach/research/selection.pdf

Summary

- Theoretical motivation: fine-scale spectral equivalence
- Choose partition to guarantee good equivalence
- Diagonal dominance is simple, but effective
- Multilevel results show robustness and efficiency

Future Directions

- Symmetric ARMS with IC/MIC versus ILU
- Further explore non-symmetric ARMS
- More complicated measures

http://www.cs.umn.edu/~maclach/research/selection.pdf