Modern and classical theory for adaptive algebraic multigrid

Scott MacLachlan
scott.maclachlan@tufts.edu

Department of Mathematics, Tufts University

July 11, 2008

With Thanks to ...

This research was performed in collaboration with

- James Brannick, Ludmil Zikatanov, 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.

Big Picture

AMG is a nice algorithm

- Efficiently solves many problems
- Good algorithmic and parallel scalability
- Somewhat mature technology

Big Picture

AMG is a nice algorithm

- Efficiently solves many problems
- Good algorithmic and parallel scalability
- Somewhat mature technology

AMG isn't perfect

- Sensitive to parameter choices
- Requires some expert knowledge
- Convergence isn't well understood

Big Picture

AMG is a nice algorithm

- Efficiently solves many problems
- Good algorithmic and parallel scalability
- Somewhat mature technology

AMG isn't perfect

- Sensitive to parameter choices
- Requires some expert knowledge
- Convergence isn't well understood

When AMG works, it is often the best solver

Multigrid

Multigrid Components

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

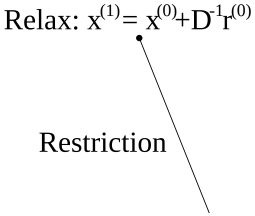
- Relaxation

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

Multigrid

Multigrid Components

- Relaxation
- Restriction



- Transfer residual to coarse grid
- Compute $R\mathbf{r}^{(1)}$

Multigrid

Multigrid Components

- Relaxation
- Restriction
- Coarse-Grid Correction

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

Restriction

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

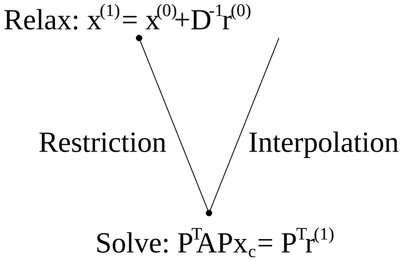
- Use coarse-grid correction to eliminate smooth errors
- In general, solve $A_c \mathbf{x}_c = R\mathbf{r}^{(1)}$
- Best correction, $\mathbf{x}_c$, in terms of A-norm satisfies

$$P^T A P \mathbf{x}_c = P^T \mathbf{r}^{(1)}$$

Multigrid

Multigrid Components

- Relaxation
- Restriction
- Coarse-Grid Correction
- Interpolation

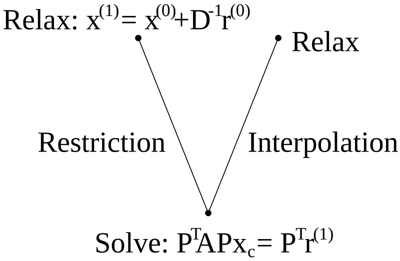


- Transfer correction to fine grid
- Compute $\mathbf{x}^{(2)} = \mathbf{x}^{(1)} + P\mathbf{x}_c$

Multigrid

Multigrid Components

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

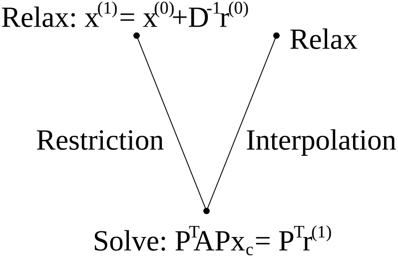


- Relax once again to remove oscillatory error introduced in coarse-grid correction

Multigrid

Multigrid Components

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



Direct solution of coarse-grid problem isn't practical

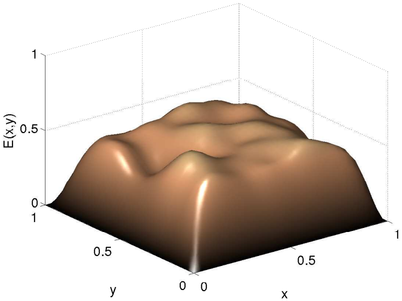
Recursion!

Apply same methodology to solve coarse-grid problem

Geometric Multigrid

For homogeneous operators, relaxation is predictable

- Jacobi/Gauss-Seidel relaxation
- Regular coarsening
- Linear interpolation



Fully explained by local mode (Fourier) analysis

Limitations of Geometric Approach

Geometric multigrid requires several assumptions on

- Problem geometry
- Form of operator
- Performance of relaxation

Limitations of Geometric Approach

Geometric multigrid requires several assumptions on

- Problem geometry
- Form of operator
- Performance of relaxation

These assumptions may be difficult to satisfy

- Heterogeneous coefficients
- Unstructured geometry
- Time-dependence
- Monte-Carlo simulations

Try to generalize algorithm to allow for heterogeneous coefficients and unstructured grids

Multigrid Without Grids

The essence of multigrid has nothing to do with grids!

Complementarity is key:

- Fix choice of relaxation
- For any A, some errors are slow to converge
- These errors must be corrected some other way

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

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

Multigrid Without Grids

The essence of multigrid has nothing to do with grids!

Complementarity is key:

- Fix choice of relaxation
- For any A, some errors are slow to converge
- These errors must be corrected some other way

Coarse-grid correction:

$$\mathbf{x} \leftarrow \mathbf{x} + PB_c^{-1}R\mathbf{r}$$
$$\mathbf{e} \leftarrow \mathbf{e} - PB_c^{-1}R\mathbf{r}$$

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

Variational Coarsening

Coarse-grid correction,

$$I - PB_c^{-1}RA,$$

can only correct errors in the range of P

Choosing $R = P^T$ and $B_c = P^TAP$ exactly eliminates errors in this space.

Complementarity is key:

- Errors reduced by relaxation and coarse-grid correction
- Errors that relaxation reduces slowly must be in range(P)

R. Nicolaides, Math. Comp. 1977, **31**:892-906

Multigrid Theory

Goal: Quantify complementarity

- Ensure sufficient reduction of all modes

Many approaches to MG Theory

- Local Fourier analysis (local mode analysis)
- Variational theory (finite elements)
- Algebraic theory (AMG, AMGe, AMGr)

A. Brandt, Math. Comp. 1977, **31**:333-390

S. McCormick, J. Ruge, SIAM J. Numer. Anal. 1982, **19**:924-929

J. Bramble, J. Pasciak, J. Wang, J. Xu, Math. Comp. 1991, **57**:23-45

M. Brezina et al., SIAM J. Sci. Comput. 2000 **22**:1570-1592

S. Brenner, Math. Comp. 2002, **71**:507-525

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

Predictive Theory

Our interest is in providing sharp bounds on expected MG performance

- Convergence by itself is interesting
- Convergence + bound may be informative
- Convergence + sharp bound allows evaluation of algorithmic options

Aim is to design algorithms for difficult problems, based on theoretical insight.

For strongly heterogeneous problems, only algebraic theory applies

Multigrid Iteration Operator

Consider a two-level multigrid cycle,

- Pre-relaxation: $\mathbf{x}^{(k')} = \mathbf{x}^{(k)} + M\mathbf{r}^{(k)}$
- Coarse-grid correction: $\mathbf{x}^{(k'')} = \mathbf{x}^{(k')} + PB_c^{-1}R\mathbf{r}^{(k')}$
- Post-relaxation: $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k'')} + M^T\mathbf{r}^{(k'')}$

Error propagation operator for V(1,1) cycle is

$$\begin{aligned} E_{TG} &= (I - M^T A)(I - PB_c^{-1}RA)(I - MA) \\ &= GTG' \end{aligned}$$

For V(0,1) cycle, have

$$GT = (I - M^T A)(I - PB_c^{-1}RA)$$

Classical Algebraic Theory

Suppose that, for all **e**,

$$\|G\mathbf{e}\|_A^2 \leq \|\mathbf{e}\|_A^2 - \delta\|T\mathbf{e}\|_A^2$$

for some fixed $\delta > 0$. Then,

$$\|GT\|_A^2 \leq 1 - \delta.$$

Proof:

$$\begin{aligned}\|GT\mathbf{e}\|_A^2 &\leq \|T\mathbf{e}\|_A^2 - \delta\|T^2\mathbf{e}\|_A^2 \\ &= (1-\delta)\|T\mathbf{e}\|_A^2 \leq (1-\delta)\|\mathbf{e}\|_A^2\end{aligned}$$

S. McCormick, SIAM J. Numer. Anal. 1985, **22**:634-643

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

J.W. Ruge and K. Stüben, in *Multigrid Methods*, SIAM 1987, pp.73-130

Sufficient Conditions

Showing $\|G\mathbf{e}\|_A^2 \le \|\mathbf{e}\|_A^2 - \delta\|T\mathbf{e}\|_A^2$ is not natural

Often separate into

Smoothing: $\|G\mathbf{e}\|_A^2 \le \|\mathbf{e}\|_A^2 - \alpha\|\mathbf{e}\|_X^2$

Approximation: $\|T\mathbf{e}\|_A^2 \le \beta\|\mathbf{e}\|_X^2$ or $\|T\mathbf{e}\|_A^2 \le \beta\|T\mathbf{e}\|_X^2$

for $\delta = \frac{\alpha}{\beta}$, $X = AD^{-1}A$

Can show smoothing property holds for

- Gauss-Seidel if A is SPD
- weighted Jacobi if A is SPD, small enough ω

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

J.W. Ruge and K. Stüben, in *Multigrid Methods*, SIAM 1987, pp.73-130

Smoothing Property

Key in showing that $\|G\mathbf{e}\|_A^2 \le \|\mathbf{e}\|_A^2 - \alpha\|\mathbf{e}\|_X^2$ is:
Lemma: Given A spd,

$$\|G\mathbf{e}\|_A^2 \le \|\mathbf{e}\|_A^2 - \alpha\|\mathbf{e}\|_X^2 \quad \forall\mathbf{e}$$

for $G = I - M^T A$ if and only if

$$\alpha\mathbf{e}^T D^{-1}\mathbf{e} \le \mathbf{e}^T(M^T + M - MAM^T)\mathbf{e} \quad \forall\mathbf{e}$$

or

$$\alpha\mathbf{e}^T(M^T + M - MAM^T)^{-1}\mathbf{e} \le \mathbf{e}^T D\mathbf{e} \quad \forall\mathbf{e}$$

Spectral equivalence between D and "symmetrized smoother"

Mandel, McCormick, Ruge, SIAM J. Numer. Anal. 1988, **25**:91-110

Approximation Property

Satisfy $\|T\mathbf{e}\|_A^2 \leq \beta \|T\mathbf{e}\|_X^2$ by satisfying

$$\min_{\mathbf{y}} \|\mathbf{e} - P\mathbf{y}\|_D^2 \leq \beta \|\mathbf{e}\|_A^2$$

for all $\mathbf{e}$.

This requires that

- Interpolation be accurate when $\|e\|_A$ is small
- Measure accuracy in weaker norm than A-norm

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

J.W. Ruge and K. Stüben, in *Multigrid Methods*, SIAM 1987, pp.73-130

Generalized AMGe

Define, for $\tilde{M} = (M^T + M - MAM^T)^{-1}$

$$\mu(P) = \sup_{\mathbf{e}} \frac{\langle \tilde{M}(I - \pi(P))\mathbf{e}, (I - \pi(P))\mathbf{e} \rangle}{\langle A\mathbf{e}, \mathbf{e} \rangle}$$

$$K(P) = \sup_{\mathbf{e}} \frac{\langle \tilde{M}(I - \pi_{\tilde{M}}(P))\mathbf{e}, (I - \pi_{\tilde{M}}(P))\mathbf{e} \rangle}{\langle A\mathbf{e}, \mathbf{e} \rangle}$$

where

- $\pi(P)$ is a general projection onto $R(P)$
- $\pi_{\tilde{M}}(P)$ is the $\tilde{M}$-orthogonal projection

Then

$$\|GT\|_A^2 = 1 - \frac{1}{K(P)} \leq 1 - \frac{1}{\mu(P)}$$

M. Brezina et al., SIAM J. Sci. Comp. 2000, **22**:1570-1592

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

Falgout, Vassilevski, Zikatanov, Num. Lin. Alg. Appl. 2005, **12**:471-494

Sharp Theory & Bounds

Since $K(P)$ provides a sharp two-level convergence bound, can we accurately bound it?

Lemma: Assume $\|\tilde{M}\| \leq c_{\tilde{M}}\|A\|$ and

$$\|A\|\|(I - \pi(P))\mathbf{e}\|^2 \leq \hat{K}\|\mathbf{e}\|_A^2$$

for all $\mathbf{e}$ and some $\hat{K}$, then

$$K(P) \leq \hat{K}.$$

Smoothing and Approximation

Question: How different are new assumptions?

Classical smoothing property is equivalent to $\alpha \mathbf{e}^T \tilde{M} \mathbf{e} \leq \mathbf{e}^T D \mathbf{e}$
Modern smoothing property requires that $\|\tilde{M}\| \leq c_{\tilde{M}} \|A\|$

These are equivalent up to $\frac{\|D\|}{\|A\|}$

Smoothing and Approximation

Question: How different are new assumptions?

Classical smoothing property is equivalent to $\alpha \mathbf{e}^T \tilde{M} \mathbf{e} \leq \mathbf{e}^T D \mathbf{e}$
Modern smoothing property requires that $\|\tilde{M}\| \leq c_{\tilde{M}} \|A\|$

These are equivalent up to $\frac{\|D\|}{\|A\|}$

Classical approximation property satisfied by

$$\min_{\mathbf{y}} \|\mathbf{e} - P\mathbf{y}\|_D^2 \leq \beta \|\mathbf{e}\|_A^2$$

Modern approximation property is

$$\|A\| \|(I - \pi(P))\mathbf{e}\|^2 \leq \hat{K} \|\mathbf{e}\|_A^2$$

Modern approximation property implies classical one

Where does this leave us?

Motivation for looking at sharp theory was adaptive AMG for heterogeneous problems

Adaptive AMG characterized by

- Interpolation chosen to fit prototypical errors
- Prototypes chosen based on performance of relaxation

Typical interpolation is very disordered

- Difficult to prove modern or classical approximation property

Reduction-based AMG

Simplify choices within AMG framework

Write $A = \begin{pmatrix} A_{ff} & A_{fc} \\ A_{fc}^T & A_{cc} \end{pmatrix}$ and choose

- $M = \sigma \begin{pmatrix} D_S^{-1} & 0 \\ 0 & 0 \end{pmatrix}$
- $P = \begin{pmatrix} -D_P^{-1} A_{fc} \\ 0 \end{pmatrix}$

Natural framework for CR and greedy coarsening

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

Generalized AMGr

Theorem: Assume that A is Hermitian and positive definite and that

- $\lambda \mathbf{e}_f^T D_s \mathbf{e}_f \leq \mathbf{e}_f^T A_{ff} \mathbf{e}_f \leq \Lambda \mathbf{e}_f^T D_s \mathbf{e}_f$
- $\sigma = \frac{2}{\lambda + \Lambda}$
- $\mathbf{e}_c^T S_A \mathbf{e}_c \leq \mathbf{e}_c^T P^T A P \mathbf{e}_c \leq \frac{1}{1-\gamma^2} \mathbf{e}_c^T S_A \mathbf{e}_c$

Then,

$$\|GT\|_A^2 \leq 1 - \frac{4\lambda\Lambda}{(\Lambda + \lambda)^2}(1 - \gamma^2)$$

Can prove this in two different ways

- Mimic original AMGr proof
- Based on sharp estimate

Implications for Adaptive AMG

Consider coarsening by compatible relaxation or greedy algorithm

Guarantee that $\lambda \mathbf{e}_f^T D_s \mathbf{e}_f \leq \mathbf{e}_f^T A_{ff} \mathbf{e}_f \leq \Lambda \mathbf{e}_f^T D_s \mathbf{e}_f$ for reasonable λ, Λ

Then need to choose D_P so that $\mathbf{e}_c^T P^T A P \mathbf{e}_c \leq \frac{1}{1-\gamma^2} \mathbf{e}_c^T S_A \mathbf{e}_c$
Highlights role of adaptivity

- Choose P so that $P^T A P$ approximates S_A
- Maximize CBS bound, γ (minimize $\frac{1}{1-\gamma^2}$)

Choose prototypical errors to represent low-energy modes of A

Adaptive AMGr

Algorithm that directly aims to satisfy generalized AMGr theory

- Partition based on compatible relaxation or greedy algorithm
- Choose smoothing to be F-relaxation or full
- Use AMGr, choose $P = \begin{pmatrix} -D_P^{-1} A_{fc} \\ 0 \end{pmatrix}$
- Choose D_P so that $D_P^{-1} A_{fc} \mathbf{e}_c = A_{ff}^{-1} A_{fc} \mathbf{e}_c$ for prototypical error $\mathbf{e}_c$

Numerical results

Gauge Laplacians from lattice QCD

$\lambda_{\min}$	10^{-3}	10^{-5}	10^{-7}
32×32	0.619	0.636	0.636
64×64	0.646	0.672	0.672
128×128	0.590	0.705	0.707
256×256	0.641	0.657	0.657

Summary

- Goal: efficient solvers for heterogeneous problems
- Goal: predictive theory for adaptive AMG
- Classical theory splits naturally into smoothing and approximation
- Modern theory splits into nearly equivalent conditions
- Reduction-based framework provides more algorithmic clues
- Lead to improved adaptive AMGr-type algorithms