AMG for a Peta-scale Navier Stokes Code

James Lottes

Argonne National Laboratory

October 18, 2007

The Challenge

- Develop an AMG iterative method to solve Poisson

$$-\nabla^2 u = f$$

discretized on highly irregular (stretched, deformed, curved) trilinear FEs.

- System

$$A\mathbf{x} = \mathbf{b}$$

needs solving many many times (for different **b**'s), allowing a very high constant for set-up time.

- Scalability requirements start at

$$P > 32{,}000$$

as a good solution is in place below this level.

Outline

- Notation for iterative solvers and multigrid (3 slides)
- Analysis of two-level multigrid, illustrated on model problem
- AMG iteration design

Iterative Solvers

- Linear system to solve is

$$A\mathbf{x} = \mathbf{b}.$$

- Iteration defined through preconditioner B by

$$\mathbf{x}_k = \mathbf{x}_{k-1} + B(\mathbf{b} - A\mathbf{x}_{k-1}) = (I - BA)\mathbf{x}_{k-1} + B\mathbf{b}.$$

- Error $\mathbf{e}_k \equiv \mathbf{x} - \mathbf{x}_k$ behaves as

$$\mathbf{e}_k = E\mathbf{e}_{k-1} = E^k\mathbf{e}_0, \quad E \equiv I - BA.$$

- Convergence characterized by

$$\rho(E), \qquad \text{and} \qquad \|E\|_A \geq \rho(E).$$

Multigrid Iteration

- Multigrid iteration defined by

$$E_{\text{mg}} = I - B_{\text{mg}}A = (I - PB_cP^tA)(I - BA),$$

- where B_c is a multigrid preconditioner for the coarse operator

$$A_c \equiv P^tAP,$$

  defined in terms of the $n \times n_c$ prolongator P,
- and B is the smoother.

C-F Point Algebraic Multigrid

- Assume n_c coarse variables are a subset of the n variables, so that the prolongation matrix takes the form

$$P \equiv \begin{bmatrix} W \\ I \end{bmatrix}$$

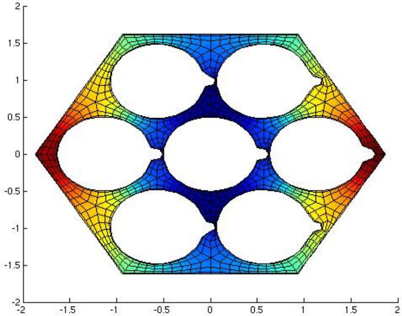
  for some $n_f \times n_c$ W, with $n_f + n_c = n$.
- Let A have the corresponding block form

$$A = \begin{bmatrix} A_{ff} & A_{fc} \\ A_{cf} & A_{cc} \end{bmatrix},$$

- and also let

$$R_f \equiv \begin{bmatrix} I & O \end{bmatrix}, \qquad R_c \equiv \begin{bmatrix} O & I \end{bmatrix}.$$

Model Problem



- Poisson, bilinear FEs, Neumann BCs
- Mesh is 2-D slice from an application mesh
- A is SPD (but not an M-matrix) except
- $A\mathbf{1} = \mathbf{0}$

C-F Points

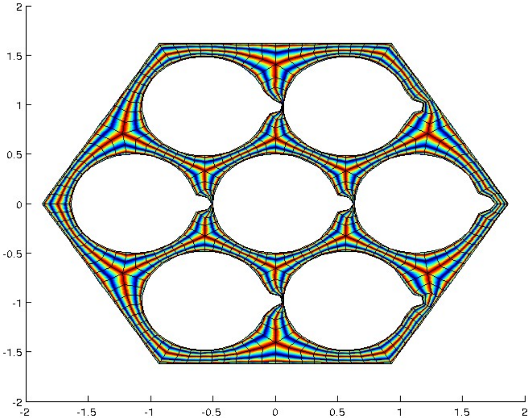


Figure: C-pts in red, F-pts in blue

Prolongation

The energy-minimizing prolongation of Wan, Chan, and Smith [8, 9]

Find P, given its sparsity pattern, and with $R_c P = I$, that

$$\text{minimizes } \operatorname{tr}(P^T A P) \qquad \text{subject to } P\mathbf{1}_{n_c} = \mathbf{1}_n.$$

Smoothers

- Damped Jacobi insufficient, Gauss-Seidel not parallel
- Adams, Brezina, Hu, and Tuminaro [1] recommend Chebyshev polynomial smoothers over Gauss-Seidel
- Sparse Approximate Inverse: Tang & Wan [6]
  - Find B with a given sparsity pattern that minimizes $\|I - BA\|_F$
  - SAI-0: Diagonal B

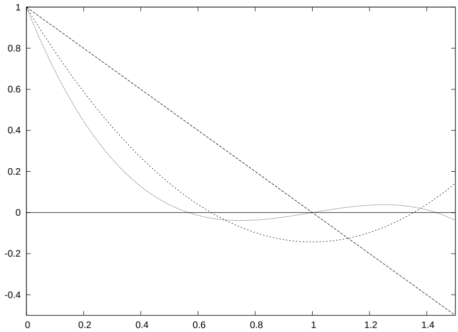
$$B_{ii} = \frac{A_{ii}}{\sum_{k=1}^{n} A_{ik} A_{ik}} \qquad \text{(compare to Jacobi: } B_{ii} = \frac{\omega}{A_{ii}}\text{)}$$

  - SAI-1: Sparsity pattern of A used for M
  - Simple to compute, and parameter-free

Chebyshev Polynomial Smoothing

Chebyshev semi-iterative method to accelerate $I - BA$

$$
\begin{aligned}
B_1 &= a_1 B, & \lambda(I - B_1 A) &= 1 - a_1\lambda(BA) \\
B_2 &= a_1 B + a_2 BAB, & \lambda(I - B_2 A) &= 1 - a_1\lambda(BA) - a_2\lambda^2(BA)
\end{aligned}
$$



- Choose coefficients using Chebyshev polynomials to damp error modes $\lambda \in [\lambda_{\min}, \lambda_{\max}(BA)]$
- $\lambda_{\min}$ open parameter
- B: Jacobi, Diagonal SAI

Smoother Error Propagation Spectra

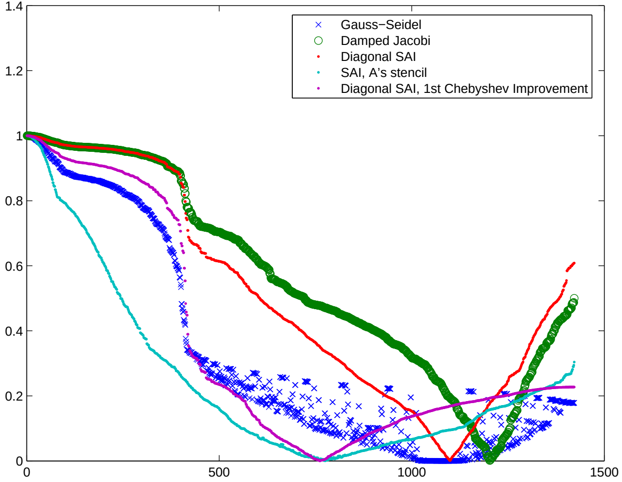


Figure: $|\lambda_i(I - BA)|$ vs. i

Two-grid Error Propagation Spectra

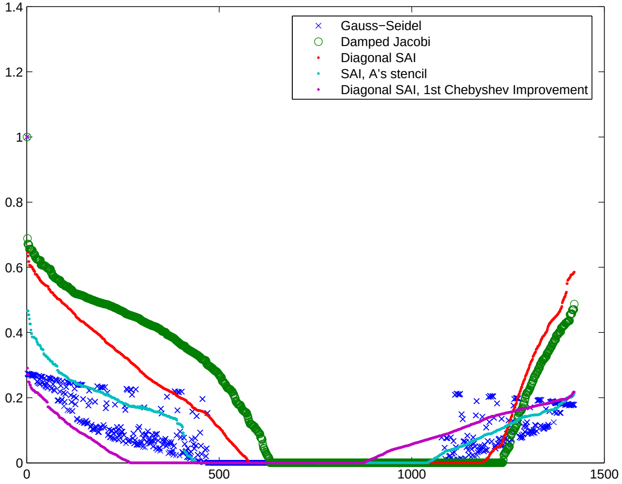


Figure: $|\lambda_i[(I - P(P^tAP)^{-1}P^tA)(I - BA)]|$ vs. i

Change of Basis

- Given invertible T, defining a change of basis by,

$$\mathbf{x} = T\hat{\mathbf{x}},$$

- Transformed system is

$$\hat{A}\hat{\mathbf{x}} = T^t\mathbf{b}, \qquad \hat{A} \equiv T^t A T.$$

- Transformed iteration given by

$$\hat{B} \equiv T^{-1} B T^{-t}, \qquad \hat{P} \equiv T^{-1} P,$$

$$\hat{E}_{\text{mg}} \equiv (I - \hat{P} B_c \hat{P}^t \hat{A})(I - \hat{B}\hat{A}) = T^{-1} E_{\text{mg}} T.$$

- Equivalent in that

$$\lambda(\hat{E}_{\text{mg}}) = \lambda(E_{\text{mg}}), \qquad \|\hat{E}_{\text{mg}}\|_{\hat{A}} = \|E_{\text{mg}}\|_A.$$

On the C-F Point Assumption

- Analysis of a C-F point AMG method can apply to other, more general AMG methods, so long as one can find a T such that

$$\hat{P} = T^{-1}P = \begin{bmatrix} W \\ I \end{bmatrix}.$$

- This T is related to the R and S of Falgout and Vassilevski's "On generalizing the AMG framework" [4] through

$$T^{-1} = \begin{bmatrix} S^t \\ R \end{bmatrix}.$$

Hierarchical Basis

- Two-level hierarchical basis

$$T \equiv \begin{bmatrix} I & W \\ & I \end{bmatrix} = \begin{bmatrix} R_f^t & P \end{bmatrix}, \qquad T^{-1} = \begin{bmatrix} I & -W \\ & I \end{bmatrix}$$

- Chosen so that

$$\hat{P} = R_c^t$$

- Transforms A into

$$\hat{A} = \begin{bmatrix} A_{ff} & A_{ff}W + A_{fc} \\ W^t A_{ff} + A_{cf} & A_c \end{bmatrix}$$

Discrete Fundamental Solutions

- The transformed A,

$$\hat{A} = \begin{bmatrix} A_{ff} & A_{ff}W + A_{fc} \\ W^t A_{ff} + A_{cf} & A_c \end{bmatrix},$$

  is block-diagonal when the coarse basis functions are the discrete fundamental solutions

$$W_\star = -A_{ff}^{-1} A_{fc}.$$

- But $W_\star$ is not sparse, hence not a viable choice. Introduce

$$F \equiv W - W_\star.$$

Introducing F

- In terms of F,

$$\hat{A} = \begin{bmatrix} A_{ff} & A_{ff}F \\ F^t A_{ff} & A_c \end{bmatrix},$$

- and

$$A_c = S_c + F^t A_{ff} F,$$

where

$$S_c \equiv A_{cc} - A_{cf} A_{ff}^{-1} A_{fc}$$

is the Schur complement of A_{ff} in both A and $\hat{A}$.

- Note

$$\|\mathbf{v}\|_{A_c}^2 = \|\mathbf{v}\|_{S_c}^2 + \|F\mathbf{v}\|_{A_{ff}}^2.$$

Exact "Compatible Relaxation"

Theorem

When $B_c = A_c^{-1}$, non-zero eigenvalues of E_{mg} are also eigenvalues of

$$I - \hat{B}_{ff} S_f,$$

where

$$\hat{B}_{ff} = \begin{bmatrix} I & -W \end{bmatrix} B \begin{bmatrix} I & -W \end{bmatrix}^t,$$

is the ff-block of the transformed smoother, and

$$S_f = A_{ff} - A_{ff} F A_c^{-1} F^t A_{ff}$$

is the Schur complement of A_c in $\hat{A}$.

Two-grid Error Propagation Spectra

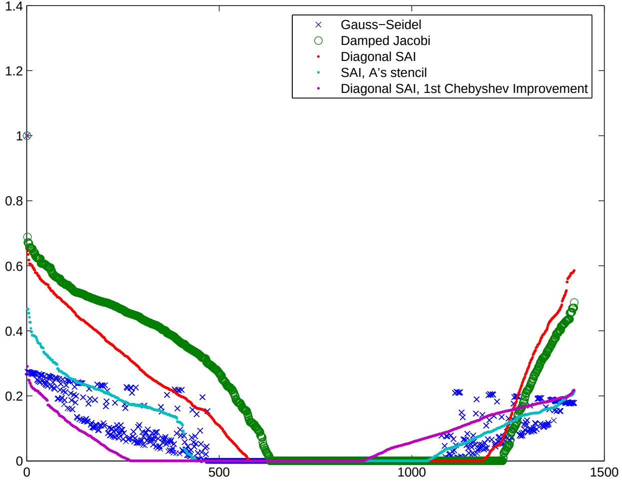


Figure: $|\lambda_i[(I - P(P^tAP)^{-1}P^tA)(I - BA)]|$ vs. i

Exact "Compatible Relaxation" Spectra

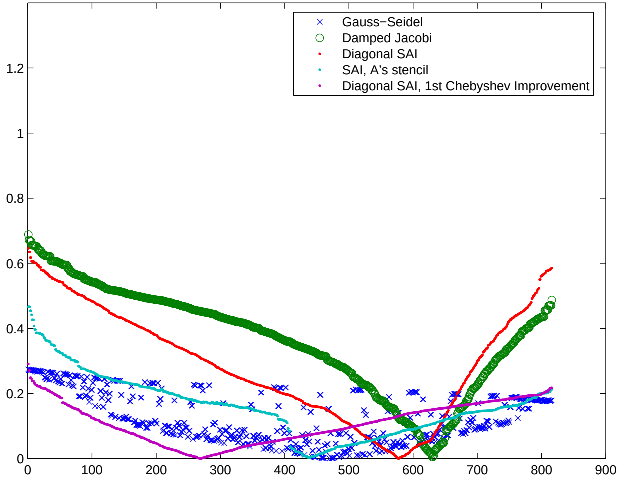


Figure: $|\lambda_i(I - \hat{B}_{ff} S_f)|$ vs. i

Equivalent F-relaxation

Corollary

With an exact coarse grid correction, the spectrum of E_{mg} is left unchanged when the smoother B is replaced by the F-relaxation

$$B_{F\text{-}r} = R_f^t \hat{B}_{ff} R_f ,$$

where again,

$$\hat{B}_{ff} = \begin{bmatrix} I & -W \end{bmatrix} B \begin{bmatrix} I & -W \end{bmatrix}^t .$$

Coarse- and Smoother-space Energies

- Coarse-space operator $A_c = S_c + F^t A_{ff} F$.
- Smoother-space operator $S_f = A_{ff} - A_{ff} F A_c^{-1} F^t A_{ff}$.
- Coarse-space and smoother-space energies

$$\|\mathbf{v}\|_{A_c}^2 = \|\mathbf{v}\|_{S_c}^2 + \|F\mathbf{v}\|_{A_{ff}}^2,$$

$$\|\mathbf{w}\|_{S_f}^2 = \|\mathbf{w}\|_{A_{ff}}^2 - \|F^t A_{ff}\mathbf{w}\|_{A_c^{-1}}^2,$$

  are minimal and maximal for any vectors when $F = O$.
- S_f is close to A_{ff} when F is "small".

Exact "Compatible Relaxation" Spectra

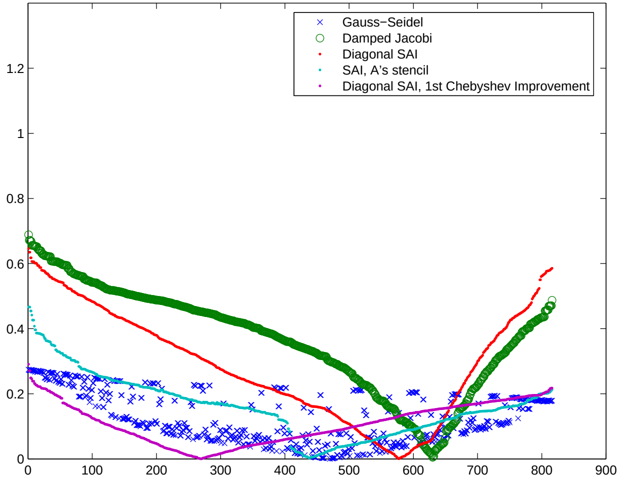


Figure: $|\lambda_i(I - \hat{B}_{ff} S_f)|$ vs. i

Inexact Compatible Relaxation Spectra

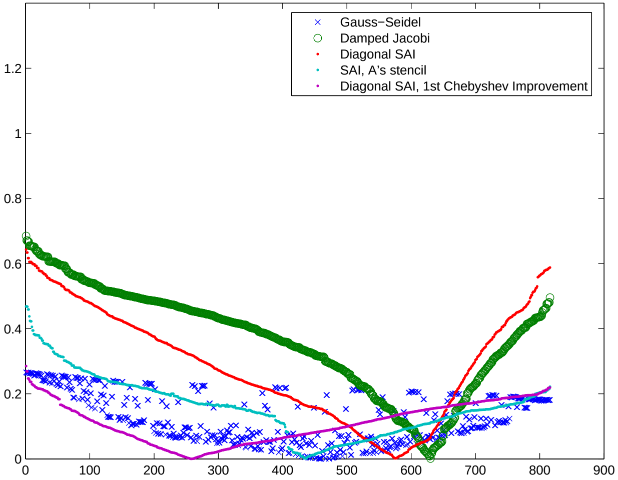


Figure: $|\lambda_i(I - \hat{B}_{ff} A_{ff})|$ vs. i

Measuring F

- Define γ as an energy norm of F,

$$\gamma \equiv \sup_{\mathbf{v}\neq\mathbf{0}} \frac{\|F\mathbf{v}\|_{A_{ff}}}{\|\mathbf{v}\|_{A_c}}.$$

- This quantity appears in, e.g., Falgout, Vassilevski, and Zikatanov [5] in the form

$$\gamma^2 = \sup_{\mathbf{v}\neq\mathbf{0}} \frac{\|\mathbf{v}\|_{A_c}^2 - \|\mathbf{v}\|_{S_c}^2}{\|\mathbf{v}\|_{A_c}^2} < 1$$

as the square of the cosine of the abstract angle between the hierarchical component subspaces.

How Closely A_{ff} Approximates S_f

Lemma

The eigenvalues of $A_{ff}^{-1}S_f$ are real and bounded by

$$0 < 1 - \gamma^2 \leq \lambda(A_{ff}^{-1}S_f) \leq 1.$$

A_{ff} vs. S_f: $1 - \gamma^2 \leq \lambda(A_{ff}^{-1} S_f) \leq 1$

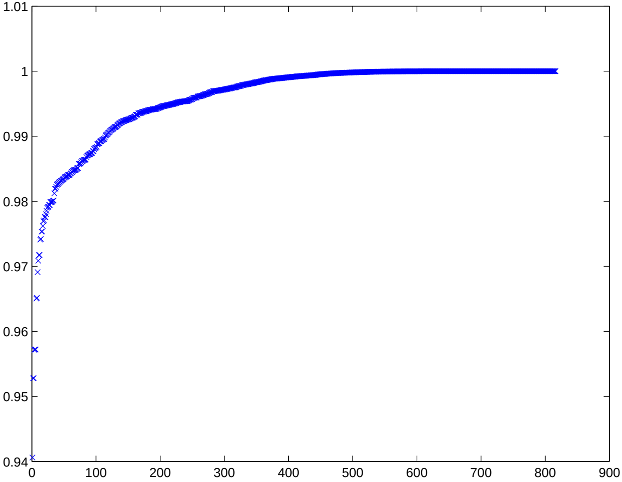


Figure: $\lambda_i(A_{ff}^{-1} S_f)$ vs. i

An Inexact Compatible Relaxation Iteration

Theorem

If $\hat{B}_{ff}$ is symmetric and

$$\rho(I - \hat{B}_{ff} A_{ff}) \leq \rho_f < 1,$$

then

$$\rho(E_{mg}) = \rho(I - \hat{B}_{ff} S_f) \leq \rho_f + \gamma^2 (1 - \rho_f).$$

Symmetric Cycle

Corollary

Define the symmetrized smoother as

$$B_s \equiv B + B^t - BAB^t,$$

and its transformed ff-block as

$$\hat{B}_{ff,s} \equiv \begin{bmatrix} I & -W \end{bmatrix} B_s \begin{bmatrix} I & -W \end{bmatrix}^t.$$

If σ *is given such that*

$$\rho(I - \hat{B}_{ff,s} A_{ff}) \leq \sigma^2 < 1,$$

then

$$\|E_{mg}\|_A^2 = \rho[(I - B^t A) Q (I - BA)] \leq \sigma^2 + \gamma^2 (1 - \sigma^2).$$

Symmetric Cycle with F-relaxation

Corollary

If the smoother is an F-relaxation,

$$B = R_f^t \hat{B}_{ff} R_f,$$

and σ is given such that

$$\|I - \hat{B}_{ff} A_{ff}\|_{A_{ff}} \leq \sigma < 1,$$

then, again,

$$\|E_{mg}\|_A^2 = \rho[(I - B^t A) Q (I - BA)] \leq \sigma^2 + \gamma^2 (1 - \sigma^2).$$

Equivalent to first half of Theorem 4.2 in Falgout, Vassilevski, and Zikatanov's "On two-grid convergence estimates" [5].

AMG Iteration Design

- Central goal: make γ small
- Coarsening heursistic: make the columns of A_{ff}^{-1} decay quickly
- Prolongation: choose sparsity by cutting off $-A_{ff}^{-1}A_{fc}$ according to some tolerance
- Smoother: simple F-relaxation of A_{ff}

Designer F-relaxations, Compatible Relaxation Prediction

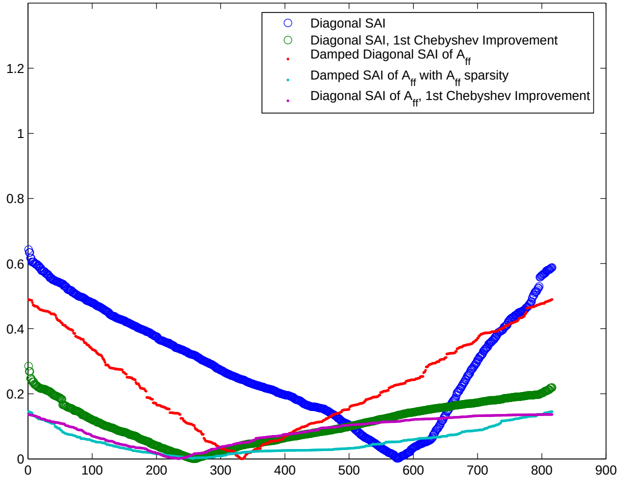


Figure: $|\lambda_i(I - \hat{B}_{ff} A_{ff})|$ vs. i

Designer F-relaxations, Two-level Performance

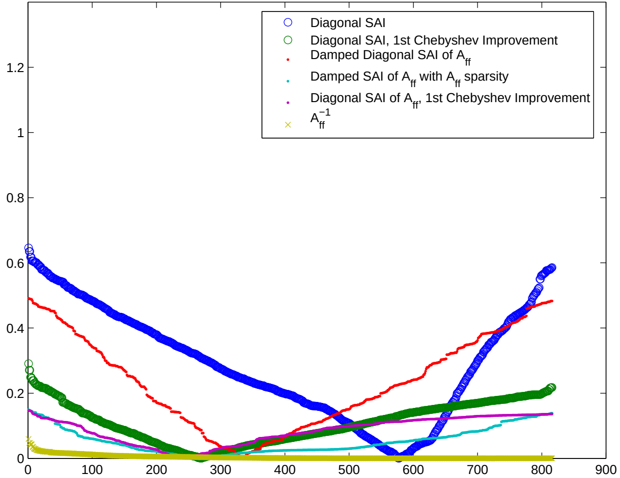


Figure: $|\lambda_i(I - \hat{B}_{ff} S_f)|$ vs. i

M. Adams, M. Brezina, J. Hu, and R. Tuminaro.
Parallel multigrid smoothing: polynomial versus Gauss-Seidel.
Journal of Computational Physics, 188(2):593–610, 2003.

J. Brannick and L. Zikatanov.
Domain Decomposition Methods in Science and Engineering XVI, volume 55 of *Lecture Notes in Computational Science and Engineering*, chapter Algebraic multigrid methods based on compatible relaxation and energy minimization, pages 15–26.
Springer Berlin Heidelberg, 2007.

O. Bröker and M. J. Grote.
Sparse approximate inverse smoothers for geometric and algebraic multigrid.
Applied Numerical Mathematics, 41(1):61–80, 2002.

R. D. Falgout and P. S. Vassilevski.
On generalizing the algebraic multigrid framework.
SIAM Journal on Numerical Analysis, 42(4):1669–1693, 2004.

R. D. Falgout, P. S. Vassilevski, and L. T. Zikatanov.
On two-grid convergence estimates.
Numerical Linear Algebra with Applications, 12(5–6):471–494, 2005.

W.-P. Tang and W. L. Wan.
Sparse approximate inverse smoother for multigrid.
SIAM Journal on Matrix Analysis and Applications, 21(4):1236–1252, 2000.

H. M. Tufo and P. F. Fischer.
Fast parallel direct solvers for coarse grid problems.
J. Par. & Dist. Comput., 61:151–177, 2001.

W. L. Wan, T. F. Chan, and B. Smith.
An energy-minimizing interpolation for robust multigrid methods.
SIAM Journal on Scientific Computing, 21(4):1632–1649, 1999.

J. Xu and L. Zikatanov.

On an energy minimizing basis for algebraic multigrid methods.
Computing and Visualization in Science, 7(3–4):121–127, 2004.

Two-level Analysis

- Assume the coarse grid correction is exact,

$$B_c \equiv A_c^{-1}.$$

- The coarse grid correction

$$Q \equiv I - PA_c^{-1}P^t A,$$

$$\hat{Q} \equiv I - \hat{P}A_c^{-1}\hat{P}^t\hat{A} = I - R_c^t A_c^{-1} R_c \hat{A},$$

  is a projection.
- The error propagation matrix spectrum is

$$\lambda(E_{\mathrm{mg}}) = \lambda[\hat{Q}(I - \hat{B}\hat{A})] = \lambda[(I - \hat{B}\hat{A})\hat{Q}].$$

Proof of First Theorem

- One may calculate

$$\hat{Q} = \begin{bmatrix} I & O \\ -A_c^{-1} F^t A_{ff} & O \end{bmatrix},$$

$$\hat{A}\hat{Q} = \begin{bmatrix} A_{ff} - A_{ff} F A_c^{-1} F^t A_{ff} & \\ & O \end{bmatrix} \equiv \begin{bmatrix} S_f & \\ & O \end{bmatrix},$$

$$\hat{B} \equiv \begin{bmatrix} \hat{B}_{ff} & \hat{B}_{fc} \\ \hat{B}_{cf} & \hat{B}_{cc} \end{bmatrix},$$

$$(I - \hat{B}\hat{A})\hat{Q} = \begin{bmatrix} I - \hat{B}_{ff} S_f & O \\ -A_c^{-1} F^t A_{ff} - \hat{B}_{cf} S_f & O \end{bmatrix}$$

- S_f is the Schur complement of A_c in $\hat{A}$.

More on γ

- One may alternatively define

$$\beta \equiv \sup_{\mathbf{v}\neq\mathbf{0}} \frac{\|F\mathbf{v}\|_{A_{ff}}}{\|\mathbf{v}\|_{S_c}} = \|R_f^t F R_c\|_A.$$

- The two quantities are related through

$$\gamma^2 = \sup_{\mathbf{v}\neq\mathbf{0}} \frac{\|F\mathbf{v}\|_{A_{ff}}^2}{\|\mathbf{v}\|_{S_c}^2 + \|F\mathbf{v}\|_{A_{ff}}^2} = \sup_{\mathbf{v}\neq\mathbf{0}} \frac{\|F\mathbf{v}\|_{A_{ff}}^2/\|\mathbf{v}\|_{S_c}^2}{1 + \|F\mathbf{v}\|_{A_{ff}}^2/\|\mathbf{v}\|_{S_c}^2} = \frac{\beta^2}{1+\beta^2}.$$

How Closely A_{ff} Approximates S_f

Lemma

The eigenvalues of $A_{ff}^{-1} S_f$ are real and bounded by

$$0 < 1 - \gamma^2 \leq \lambda(A_{ff}^{-1} S_f) \leq 1.$$

Proof.

$$A_{ff}^{-1} S_f = I - F A_c^{-1} F^t A_{ff}.$$

$$\lambda(A_{ff}^{-1} S_f) = 1 - \lambda(F A_c^{-1} F^t A_{ff}) = 1 - [\{0\} \cup \lambda(A_c^{-1} F^t A_{ff} F)].$$

$$0 \leq \inf_{\mathbf{v} \neq 0} \frac{\|F\mathbf{v}\|_{A_{ff}}^2}{\|\mathbf{v}\|_{A_c}^2} \leq \lambda(A_c^{-1} F^t A_{ff} F) \leq \sup_{\mathbf{v} \neq 0} \frac{\|F\mathbf{v}\|_{A_{ff}}^2}{\|\mathbf{v}\|_{A_c}^2} = \gamma^2.$$

□