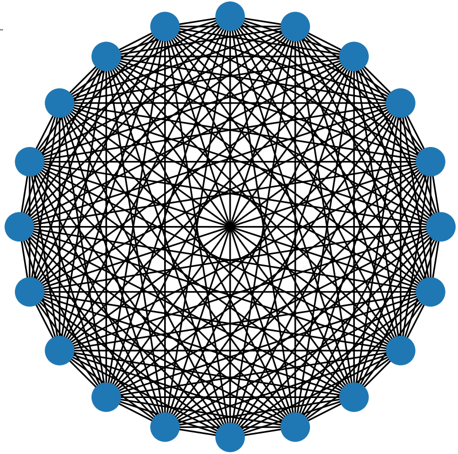
Multigrid Methods — An Overview

Lecture 4: Theory

Luke Olson
Nelder Fellow
Department of Mathematics
Imperial College

home:
Department of Computer Science
University of Illinois at Urbana-Champaign



Overarching Goal

- Consider a matrix problem (s.p.d.) of the form

$$A\boldsymbol{u} = \boldsymbol{f}$$

$A \in \mathbb{R}^{n\times n}$

- Suppose we have a multilevel iteration process $\mathcal{M}$

$$I - \mathcal{M}A = (I - M^T A)^{\nu_{\text{pre}}}(I - P(P^T AP)^{-1}P^T A)(I - MA)^{\nu_{\text{post}}}$$

so that

$$\boldsymbol{e} \leftarrow (I - \mathcal{M}A)\boldsymbol{e}$$

Convergence

The iteration converges for any $\boldsymbol{b}$ and $\boldsymbol{u}_0$

iff

$\rho(I - \mathcal{M}A) < 1.$

Overarching Goal

- Generally we'll want to work with a (matrix) norm $\|\cdot\|$:

$$\rho(I - \mathcal{M}A) \leq \|I - \mathcal{M}A\|$$

- If we consider the error at each step as

$$\boldsymbol{e}_k = (I - \mathcal{M}A)^k \boldsymbol{e}_0$$

then then we term $\frac{\boldsymbol{e}_k}{\boldsymbol{e}_{k-1}}$ the convergence factor from step $k-1$ to step k, and

$$\begin{aligned}\rho &= \lim_{k\to\infty} \left(\max_{\boldsymbol{e}_0} \frac{\|\boldsymbol{e}_k\|}{\|\boldsymbol{e}_0\|} \right)^{1/k} = \lim_{k\to\infty} \left(\max_{\boldsymbol{e}_0} \frac{\|G^k \boldsymbol{e}_k\|}{\|\boldsymbol{e}_0\|} \right)^{1/k} \\ &= \lim_{k\to\infty} \left(\|G^k\| \right)^{1/k} = \rho(G)\end{aligned}$$

the *convergence factor*.
(for initial guess $\boldsymbol{u}_0$ with an error in the principal subspace)

Overarching Goal

- If we return to our multilevel method

$$I - \mathcal{M}A$$

we are seeking a method that yields **a bound on the error reduction in each iteration that is independent of n.**

That is, a fixed number of iterations is needed to a tolerance for any problem size.

- If the *computational complexity* is bounded at $\mathcal{O}(n)$ operator, Then we say the method scales **optimally**.

- $\ldots$ a fixed number of operations, $\mathcal{O}(n)$, to reach a tolerance.

Overarching Goal

- Bounding the convergence can take many forms in many norms.
- Ideally, bounds
  - Predictive; sharp bounds on factors observed in practice
  - Strong dependence on parameters in the method
  - Computable
- Geometric methods have an advantage . . .
  - Fourier analysis for components
  - Reliance on a finite element framework for precise construction of approximation bounds
  - Clear smoothing property focuses error analysis on coarse grid accuracy
- Algebraic methods . . .
  - Components often designed **from** the theory

    Example from last time: Construct interpolation schemes so that P matches $P_{ideal} = \begin{bmatrix} -A_{FF}^{-1}A_{FC} & I \end{bmatrix}$, say spectrally, but is computationally efficient.
  - Can be difficult to compute; can be be unsharp (Today!)

Objectives

- Outline the basic components of algebraic theory.
- Distinguish between sharp bounds and computable bounds.
- Observe this effect in practice.
- Note how this theory is influencing multigrid design and development.

Approach

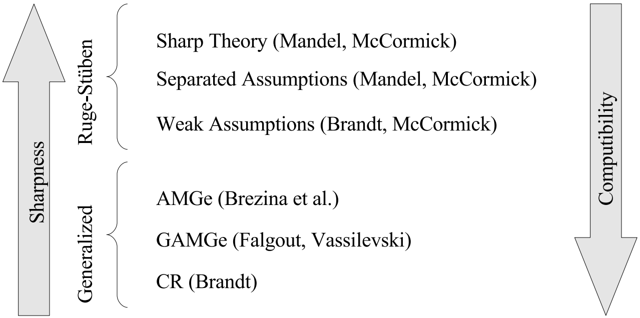
disclaimer

This is *one* slice of algebraic theory.

There are many approaches to multigrid theory. Note just a few here:

- Subspace correction approaches of Xu, et al.
- Extending from two-level to multilevel. Notay et al.
- Multiblock form Vassilevski et al.
- Generalized AMG theory. Falgout, Vassilevski, et al.

Approach



- Sharpness and computability are two competing aspects
- Initial efforts: sharp
- More recent efforts: focus on constructing methods

Setup

- Fine grid $\Omega = \{1, \ldots, n\} = C \cup F$ and coarse grid $\Omega_c = C$.
- Interpolation / restriction:

$$P : \Omega_c \to \Omega \quad \text{and} \quad R : \Omega \to \Omega_c$$

- A is s.p.d., $D = \operatorname{diag}(A)$ — defining an inner product:

$$\begin{aligned} \langle \boldsymbol{u}, \boldsymbol{v} \rangle_A &= \langle A\boldsymbol{u}, \boldsymbol{v} \rangle \\ \langle \boldsymbol{u}, \boldsymbol{v} \rangle_D &= \langle D\boldsymbol{u}, \boldsymbol{v} \rangle, \\ \langle \boldsymbol{u}, \boldsymbol{v} \rangle_{AD^{-1}A} &= \langle D^{-1}A\boldsymbol{u}, A\boldsymbol{v} \rangle \end{aligned}$$

- Define G as **post-relaxation**
(and $\mathcal{G}$ as the affine version $\boldsymbol{u} \leftarrow \mathcal{G}(\boldsymbol{u}, \boldsymbol{f})$):

$$\boldsymbol{u} \leftarrow G\boldsymbol{u} + (I - G)A^{-1}\boldsymbol{f} \qquad \text{or} \quad \boldsymbol{e} \leftarrow G\boldsymbol{e}$$

Setup

Algorithm: AMG Solution Phase

$\boldsymbol{u} \leftarrow \hat{\mathcal{G}}(\boldsymbol{u}, \boldsymbol{f})$	Pre-relax
$\boldsymbol{r}_c \leftarrow R\boldsymbol{r}$	Restrict the residual
$\boldsymbol{e}_c \leftarrow A_c^{-1}\boldsymbol{r}_c$	Coarse grid solve
$\hat{\boldsymbol{e}} \leftarrow P\boldsymbol{e}_c$	Interpolate the error approximation
$\boldsymbol{u} \leftarrow \boldsymbol{u} + \hat{\boldsymbol{e}}$	Correct the fine-grid solution
$\boldsymbol{u} \leftarrow \mathcal{G}(\boldsymbol{u}, \boldsymbol{f})$	Post-relax

Some operators

- In general, consider relaxation as

$$G = I - MA$$

- **Assumption:** M is norm convergent (in A):

$$\|G\|_A < 1$$

- **Assumption:** P is full rank
- **Assumption:** A is s.p.d.

A-orthogonality

- Let the coarse grid correction step be

$$T = I - P(P^T A P)^{-1} P^T A$$

CGC

T is an A-orthogonal projection onto the range of P

- After coarse grid correction, the error is minimized in the energy norm over $\mathcal{R}(P)$.

Focus on V(0,1)

- The A-adjoint of GT is

$$TG^+ \qquad G^+ = I - M^T A$$

- The symmetric V(1,1) cycle is

$$\begin{aligned}(I - MA)(I - P(P^T AP)^{-1} P^T A)(I - M^T A) &= GTG^+ \\ &= GTTG^+\end{aligned}$$

Since $\|GT\|_A = \|TG^+\|_A$ (A-adjoints) we ahve

$$\|GTG^+\|_A = \|GT\|_A^2$$

- Ok, so we can focus focus on the V(0,1) cycle, the other cycles follow.

What we are measuring

- Since $T = I - P(P^TAP)^{-1}P^TA$ and due to our assumptions on G, we will **measure convergence or reduction in** $\|\boldsymbol{e}\|_A$. Note:

$$\|\boldsymbol{e}\|_A^2 = \|(I - T)\boldsymbol{e}\|_A^2 + \|T\boldsymbol{e}\|_A^2$$

- For a V(0,1) cycle, the reduction in $\boldsymbol{e}$ is

$$\|GT\boldsymbol{e}\|_A^2 \leq (1 - \delta^*)\|\boldsymbol{e}\|_A^2$$

- we seek a **sharp** bound in that[1]

$$\|GT\|_A^2 := \sup_{\boldsymbol{e}\neq\boldsymbol{0}} \frac{\|GT\boldsymbol{e}\|_A^2}{\|\boldsymbol{e}\|_A^2} = 1 - \delta^*$$

[1] $\sup = \max$

Sufficient conditions

What should we assume on relaxation and interpolation?

- One idea: assuming relaxation is effective on the range of interpolation.
- There exists $\delta > 0$ such that

$$\|GT\boldsymbol{e}\|_A^2 \leq (1-\delta)\|T\boldsymbol{e}\|_A^2 \text{ for all } \boldsymbol{e}.$$

  Then, since T is an A-orthogonal projector,

$$\|GT\boldsymbol{e}\|_A^2 \leq (1-\delta)\|\boldsymbol{e}\|_A^2 \text{ for all } \boldsymbol{e}$$

- Similarly (norm convergent)

$$\|G\boldsymbol{v}\|_A^2 \leq \|\boldsymbol{v}\|_A^2 \quad \text{for all } \boldsymbol{v} \perp \mathcal{R}(T),$$

- As a result we can combine these into an assumption:

Assumption

Assume there exists $\delta > 0$ such that

$$\|G\boldsymbol{v}\|_A^2 \leq \|\boldsymbol{v}\|_A^2 - \delta\|T\boldsymbol{v}\|_A^2 \quad \text{for all } \boldsymbol{v}.$$

Assumption

Assume there exists $\delta > 0$ such that

$$\|G\boldsymbol{v}\|_A^2 \le \|\boldsymbol{v}\|_A^2 - \delta \|T\boldsymbol{v}\|_A^2 \quad \text{for all } \boldsymbol{v}.$$

- This assumes that relaxation is effective in reducing the error that remains after coarse grid correction.

Theorem

If there exists $\delta > 0$ so that

$$\|G\boldsymbol{e}\|_A^2 \le \|\boldsymbol{e}\|_A^2 - \delta \|T\boldsymbol{e}\|_A^2 \quad \text{for all } \boldsymbol{e},$$

then

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

Sharpness?

Theorem

If there exists $\delta > 0$ so that

$$\|G\boldsymbol{e}\|_A^2 \leq \|\boldsymbol{e}\|_A^2 - \delta \|T\boldsymbol{e}\|_A^2 \quad \text{for all } \boldsymbol{e},$$

then

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

- Is this a **sharp** estimate of the convergence?
- To be sharp, the largest δ, say $\hat{\delta}$

$$\hat{\delta} = \inf_{\boldsymbol{e}:\, T\boldsymbol{e} \neq \boldsymbol{0}} \frac{\|\boldsymbol{e}\|_A^2 - \|G\boldsymbol{e}\|_A^2}{\|T\boldsymbol{e}\|_A^2},$$

should be δ^*.

(Proof)

- Since $T\boldsymbol{e} = \mathbf{0}$ gives $\|GT\boldsymbol{e}\|_A = 0$:

$$\|GT\|_A^2 = \sup_{\boldsymbol{e}:\, T\boldsymbol{e}\neq\mathbf{0}} \frac{\|GT\boldsymbol{e}\|_A^2}{\|\boldsymbol{e}\|_A^2} = \sup_{\boldsymbol{e}:\, T\boldsymbol{e}\neq\mathbf{0}} \frac{\|GT\boldsymbol{e}\|_A^2}{\|T\boldsymbol{e}\|_A^2 + \|(I-T)\boldsymbol{e}\|_A^2}.$$

- Let $\hat{\boldsymbol{e}}$ be the argsup
- Then $T\hat{\boldsymbol{e}}$ is also at the supremum.
- Thus we have an error at the supremum with $(I-T)\hat{ve} = 0$

$$\|GT\|_A^2 = \sup_{\boldsymbol{e}:\, T\boldsymbol{e}\neq\mathbf{0}} \frac{\|GT\boldsymbol{e}\|_A^2}{\|\boldsymbol{e}\|_A^2} = \sup_{\boldsymbol{e}:\, T\boldsymbol{e}\neq\mathbf{0}} \frac{\|G(T\boldsymbol{e} + (I-T)\boldsymbol{e})\|_A^2}{\|T\boldsymbol{e}\|_A^2} = \sup_{\boldsymbol{e}:\, T\boldsymbol{e}\neq\mathbf{0}} \frac{\|G\boldsymbol{e}\|_A^2}{\|T\boldsymbol{e}\|_A^2},$$

And

$$1 - \|GT\|_A^2 = \inf_{\boldsymbol{e}:\, T\boldsymbol{e}\neq\mathbf{0}} \frac{\|T\boldsymbol{e}\|_A^2 - \|G\boldsymbol{e}\|_A^2}{\|T\boldsymbol{e}\|_A^2} = \inf_{\boldsymbol{e}:\, T\boldsymbol{e}\neq\mathbf{0}} \frac{\|\boldsymbol{e}\|_A^2 - \|G\boldsymbol{e}\|_A^2}{\|T\boldsymbol{e}\|_A^2} = \hat{\delta}.$$

Where are we at

- The worst δ is sharp (we'll do an example at the end)

$$\hat{\delta} = \inf_{\boldsymbol{e}\colon T\boldsymbol{e}\neq\mathbf{0}} \frac{\|\boldsymbol{e}\|_A^2 - \|G\boldsymbol{e}\|_A^2}{\|T\boldsymbol{e}\|_A^2},$$

- But this is difficult to compute!
- As a result, the early theory **split**:

For some $g(\boldsymbol{e})$ define δ, α_g, and β_g as in

$$\delta(\boldsymbol{e}) = \underbrace{\frac{\|\boldsymbol{e}\|_A^2 - \|G\boldsymbol{e}\|_A^2}{g(\boldsymbol{e})}}_{\alpha_g(\boldsymbol{e})} \underbrace{\frac{g(\boldsymbol{e})}{\|T\boldsymbol{e}\|_A^2}}_{1/\beta_g(\boldsymbol{e})}$$

- Consider the smallest α_g and the largest β_g:

$$\hat{\alpha}_g = \inf_{\boldsymbol{e}:g(\boldsymbol{e})\neq\mathbf{0}} \alpha_g(\boldsymbol{e}) \quad \hat{\beta}_g = \sup_{\boldsymbol{e}:g(\boldsymbol{e})\neq\mathbf{0}} \beta_g(\boldsymbol{e})$$

A less sharp bound

$$\delta(\boldsymbol{e}) = \underbrace{\frac{\|\boldsymbol{e}\|_A^2 - \|G\boldsymbol{e}\|_A^2}{g(\boldsymbol{e})}}_{\alpha_g(\boldsymbol{e})} \underbrace{\frac{g(\boldsymbol{e})}{\|T\boldsymbol{e}\|_A^2}}_{1/\beta_g(\boldsymbol{e})}$$

- For $\boldsymbol{e}$ such that $g(T\boldsymbol{e}) \neq 0$,

$$\|GT\boldsymbol{e}\|_A^2 \leq \|T\boldsymbol{e}\|_A^2 - \hat{\alpha}_g g(T\boldsymbol{e}) \leq \|T\boldsymbol{e}\|_A^2 - \frac{\hat{\alpha}_g}{\hat{\beta}_g}\|T\boldsymbol{e}\|_A^2 = \left(1 - \frac{\hat{\alpha}_g}{\hat{\beta}_g}\right)\|T\boldsymbol{e}\|_A^2 \quad (1)$$

$$\leq \left(1 - \frac{\hat{\alpha}_g}{\hat{\beta}_g}\right)\|\boldsymbol{e}\|_A^2 \quad (2)$$

- Ok, so this is generally worse than the sharp bound

$$\|GT\|_A = \sqrt{1 - \hat{\delta}} \leq \sqrt{1 - \frac{\hat{\alpha}_g}{\hat{\beta}_g}}$$

(α_g and β_g are not generally simultaneously satisfied)

What about this $g(\cdot)$ thing?

$$\delta(\boldsymbol{e}) = \underbrace{\frac{\|\boldsymbol{e}\|_A^2 - \|G\boldsymbol{e}\|_A^2}{g(\boldsymbol{e})}}_{\alpha_g(\boldsymbol{e})} \underbrace{\frac{g(\boldsymbol{e})}{\|T\boldsymbol{e}\|_A^2}}_{1/\beta_g(\boldsymbol{e})}$$

- Early works, e.g. Ruge-Stüben 1987, use $g(\boldsymbol{e}) = \|\boldsymbol{e}\|_{AD^{-1}A}^2$
- Or the weaker form $g(\boldsymbol{e}) = \|T\boldsymbol{e}\|_{AD^{-1}A}^2$
- Generally offers control of the sharpness
- Choosing $g(\boldsymbol{e}) = \|T\boldsymbol{e}\|_A^2$ naturally leads to $\hat{\beta}_g = 1$ and $\hat{\alpha}_g = \hat{\delta}$ (sharp)

smoothing and approximation assumptions

- Assuming relaxation satisfies a **smoothing assumption**:

$$\|Ge\|_A^2 \leq \|e\|_A^2 - \bar{\alpha}_g g(e) \quad \text{for all } e$$

  for some $\bar{\alpha}_g$ (with $\bar{\alpha}_g < \hat{\alpha}_g$) (as early as Hackbush 1979)
- Practical, but loss of sharpness
- Similarly, can assume (a strong type, for V-cycle convergence)

$$\|Te\|_A^2 \leq \bar{\beta}_g g(e) \quad \text{for all } e.$$

- Since the bound only depends on $g(Te)$ in (1) we assume interpolation satisfies the (weak) **approximation assumption**:

$$\|Te\|_A^2 \leq \bar{\beta}_g g(Te) \quad \text{for all } e.$$

- Stronger assumptions lead to say L^2 boundedness of coarse grid correction

Split Theory

Theorem

if there exists $\bar{\alpha}_g > 0$ *such that*

$$\|G\boldsymbol{e}\|_A^2 \leq \|\boldsymbol{e}\|_A^2 - \bar{\alpha}_g g(\boldsymbol{e}) \quad \textit{for all } \boldsymbol{e} \quad \text{(smoothing)}$$

and there exists $\bar{\beta}_g > 0$ *such that*

$$\|T\boldsymbol{e}\|_A^2 \leq \bar{\beta}_g g(T\boldsymbol{e}) \quad \textit{for all } \boldsymbol{e} \quad \text{(approximation),}$$

then $\|GT\|_A \leq \sqrt{1 - \bar{\alpha}_g / \bar{\beta}_g}$.

Strong and Weak Approximations

- Select $g(\boldsymbol{e}) = \|\boldsymbol{e}\|^2_{AD^{-1}A}$
- Since T is an A-orthogonal projection we have

$$\|T\boldsymbol{e}\|_A = \inf_{\boldsymbol{e}_c} \|\boldsymbol{e} - P\boldsymbol{e}_c\|_A$$

- (**strong approximation**) Assume there is a $\bar{\beta}_s$ such that

$$\inf_{\boldsymbol{e}_c} \|\boldsymbol{e} - P\boldsymbol{e}_c\|^2_A \leq \bar{\beta}_s \|\boldsymbol{e}\|^2_{AD^{-1}A} \quad \text{for all } \boldsymbol{e}.$$

Strong and Weak Approximations

- The weaker version looks like (for some $\hat{\beta}$)

$$\|T\boldsymbol{e}\|_A^2 \le \bar{\beta}\|T\boldsymbol{e}\|_{AD^{-1}A} \quad \text{for all } \boldsymbol{e}.$$

- Weaker, means weaker norm. And we can make this a bit more practical. The range of T is A-orthogonal to the range of P, so

$$\begin{aligned}\|T\boldsymbol{e}\|_A^2 &= \langle AT\boldsymbol{e}, T\boldsymbol{e}\rangle = \langle AT\boldsymbol{e}, T\boldsymbol{e} - P\boldsymbol{e}_c\rangle \\ &\le \|T\boldsymbol{e}\|_{AD^{-1}A}\|T\boldsymbol{e} - P\boldsymbol{e}_c\|_D.\end{aligned}$$

- (**weak approximation**) Assume that

$$\inf_{\boldsymbol{e}_c}\|\boldsymbol{e} - P\boldsymbol{e}_c\|_D^2 \le \bar{\beta}_w\|\boldsymbol{e}\|_A^2 \quad \text{for all } \boldsymbol{e}, \tag{3}$$

  This implies the bound at the top.

Bounds and bounds and bounds

- In McCormick–Ruge–1982, (Strang earlier, in FE) analyze interpolation in terms of the eigenvectors of A
- Set $V_\lambda(A)$ to be the eigenvectors with eigenvalues less than λ and unit A-norm. Choose P such that (for any λ):

$$\sup_{\boldsymbol{e}\in V_\lambda(A)} \inf_{\boldsymbol{e}_c} \|\boldsymbol{e} - P\boldsymbol{e}_c\|_A^2 \le c\lambda^a h^s.$$

  (h is a discretization size)
- Example: Consider a FD scheme A and $\boldsymbol{e} \in V_\lambda(A)$ Then

$$\|\boldsymbol{e}\|_{AD^{-1}A}^2 \le \|D^{-1}\| \|A\boldsymbol{e}\|^2 \le Ch^2\lambda$$

  Here Ch^2 is from the $1/(Ch^2)$ diagonal entries in A

Another form

- Hackbush et al use
$$\|A^{-1} - P(P^T AP)^{-1} P^T\| \le ch^s;$$
- Why is this useful? The strong approximation property from before is
$$\sup_{\boldsymbol{e}\neq\mathbf{0}} \inf_{\boldsymbol{e}_c} \frac{\|\boldsymbol{e} - P\boldsymbol{e}_c\|_A^2}{\|\boldsymbol{e}\|_{AD^{-1}A}^2} \le \bar{\beta}_s$$

  This leads to
$$\sup_{\boldsymbol{e}\neq\mathbf{0}} \inf_{\boldsymbol{e}_c} \frac{\|\boldsymbol{e} - P\boldsymbol{e}_c\|_A^2}{\|\boldsymbol{e}\|_{AD^{-1}A}^2} \le \|A^{1/2}\|^2 \|A^{-1} - P(P^T AP)^{-1} P^T\|^2 \|D^{1/2}\|^2.$$
- Can be related to the strong approximation depending on s. Can be generalized in a different directions.

Opportunity

- What if we expand our notion of the approximation property with $g(\boldsymbol{e})$?
- Consider a more general form, such as

$$g(\boldsymbol{e}) = \|\boldsymbol{e}\|^2_{AB^{-1}A}$$

for some s.p.d. B.

- Then the **same** weak approximation assumption follows, but in a different norm!

$$\inf_{\boldsymbol{e}_c} \|\boldsymbol{e} - P\boldsymbol{e}_c\|^2_B \leq \bar{\beta}_{w,B}\|\boldsymbol{e}\|^2_A \quad \text{for all } \boldsymbol{e} \tag{4}$$

and similar with the strong approximation

$$\inf_{\boldsymbol{e}_c} \|\boldsymbol{e} - P\boldsymbol{e}_c\|^2_A \leq \bar{\beta}_{s,B}\|\boldsymbol{e}\|^2_{AB^{-1}A} \quad \text{for all } \boldsymbol{e}.$$

- Gist: vector $\boldsymbol{e}$ must be approximated (in some way) by the range of interpolation, with accuracy proportional to $\|\boldsymbol{e}\|^2_A$.
- This is for every eigenvector in the BM Principle.

Outlook

- One special case is $B = A$. Then both strong and weak approximations have A-norms.
- $\beta = 1$ automatically satisfies approximation property, but the smoothing property becomes
$$\|G\boldsymbol{e}\|_A^2 \leq \|e\|_A^2 - \bar{\alpha}\|e\|_A^2 \quad \text{for all } \boldsymbol{e},$$
- Instead, look for approximations of B to A.

Back to the basics

- The separated bounds

$$\delta(\boldsymbol{e}) = \underbrace{\frac{\|\boldsymbol{e}\|_A^2 - \|G\boldsymbol{e}\|_A^2}{g(\boldsymbol{e})}}_{\alpha_g(\boldsymbol{e})} \underbrace{\frac{g(\boldsymbol{e})}{\|T\boldsymbol{e}\|_A^2}}_{1/\beta_g(\boldsymbol{e})},$$

  implicitly assume that error that is slow to reduce (*algebraically smooth error*) yields small residuals.
- The approximation assumption, for example,

$$\inf_{\boldsymbol{e}_c} \|\boldsymbol{e} - P\boldsymbol{e}_c\|_A^2 \leq \bar{\beta}_s \|\boldsymbol{e}\|_{AD^{-1}A}^2 \quad \text{for all } \boldsymbol{e}.$$

  then is responsible for reducing these error adequately during coarse grid correction
- Not all schemes exhibit the smoothing property. For example, there are relaxation schemes that target large error with small residuals — e.g. in problems like Maxwell's equation.
- In general, purely algebraic solvers will use standard relaxation schemes that do not assume any (physical) information about the problem.

On measures

- Let's go back to the approximation assumptions. Let the **best** constants be

$$\hat{\beta}_w = \sup_{\boldsymbol{e}\neq\boldsymbol{0}} \inf_{\boldsymbol{e}_c} \frac{\|\boldsymbol{e} - P\boldsymbol{e}_c\|_D^2}{\|\boldsymbol{e}\|_A^2} \qquad \hat{\beta}_s = \sup_{\boldsymbol{e}\neq\boldsymbol{0}} \inf_{\boldsymbol{e}_c} \frac{\|\boldsymbol{e} - P\boldsymbol{e}_c\|_A^2}{\|\boldsymbol{e}\|_{AD^{-1}A}^2}$$

- Element-based AMG (AMGe, and variants) attempts to build the AMG levels based on local components that are optimized.
- Two measures are central to AMGe:

$$M_1(Q, \boldsymbol{e}) = \frac{\|(I - Q)\boldsymbol{e}\|_D^2}{\|\boldsymbol{e}\|_A^2} \quad \text{and} \quad M_2(Q, \boldsymbol{e}) = \frac{\|(I - Q)\boldsymbol{e}\|_A^2}{\|\boldsymbol{e}\|_{AD^{-1}A}^2},$$

- Here, Q is **any** projection onto $\mathcal{R}(P)$ where $Q = PR$ and for R such that $RP = I$.
- If we pick

$$R = (P^T D P)^{-1} P^T D \quad \text{or} \quad R = (P^T A P)^{-1} P^T A$$

then the bounds bounds (constants) at the top are recovered.

On measures

- With
$$M_1(Q,\boldsymbol{e}) = \frac{\|(I-Q)\boldsymbol{e}\|_D^2}{\|\boldsymbol{e}\|_A^2} \quad \text{and} \quad M_2(Q,\boldsymbol{e}) = \frac{\|(I-Q)\boldsymbol{e}\|_A^2}{\|\boldsymbol{e}\|_{AD^{-1}A}^2},$$
the minimization over the coarse space is dropped, and replaced by the direct action of R.
- In a sense, the variation principle is replaced with a direction projection.
- For any $\boldsymbol{e}$ we have
$$\inf_{\boldsymbol{e}_c} \frac{\|\boldsymbol{e} - P\boldsymbol{e}_c\|_D^2}{\|\boldsymbol{e}\|_A^2} \leq M_1(Q,\boldsymbol{e}) \tag{5}$$
and
$$\inf_{\boldsymbol{e}_c} \frac{\|\boldsymbol{e} - P\boldsymbol{e}_c\|_A^2}{\|\boldsymbol{e}\|_{AD^{-1}A}^2} \leq M_2(Q,\boldsymbol{e}). \tag{6}$$
- Bounds ($\sup$) over all $\boldsymbol{e}$ guarantee two level and multilevel convergence

On Measures

- Where is this going?
- AMGe uses a localized version of these measures for the element stiffness matrices
- This leads to (optimal) forms for interpolation based on the element matrices
- Assembling into global forms puts a bound on these measures
- Great methods based on theory. There is a lack of sharpness in these measures.

On Measures

- Take the specific case of a C/F splitting:

$$A = \begin{bmatrix} A_{FF} & -A_{FC} \\ -A_{FC}^T & A_{CC} \end{bmatrix},$$

with the error split at $\boldsymbol{e} = \begin{bmatrix} \boldsymbol{e}_F \\ \boldsymbol{e}_C \end{bmatrix}$.

- Consider interpolation of the form $P = \begin{bmatrix} W \\ I \end{bmatrix}$. If $R = \begin{bmatrix} 0 & I \end{bmatrix}$ Then $Q\boldsymbol{e} = P\boldsymbol{e}_C$,

resulting in

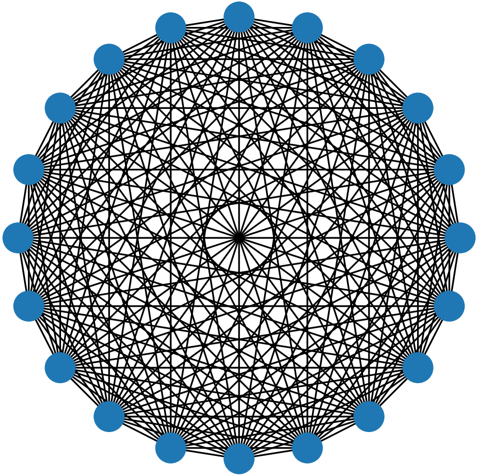
$$\hat{\tau}_w = \sup_{\boldsymbol{e} \neq \boldsymbol{0}} \frac{\|\boldsymbol{e} - P\boldsymbol{e}_c\|_D^2}{\|\boldsymbol{e}\|_A^2}$$

- $\hat{\tau}_w$ is an upper bound to the weak constant $\hat{\beta}_w$.
- Upper bounds on $\hat{\tau}_w$ may result in a large difference between the optimal choice of $\boldsymbol{v}_C$ and $\boldsymbol{e}_C$.

Example 1

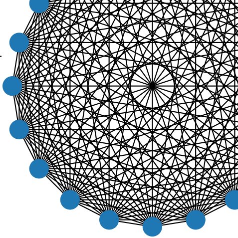
- Consider a graph Laplacian. A dense one.
- $A = (n+1)I - \mathbf{1}\mathbf{1}^T$

$$A = \begin{bmatrix} n & -1 & -1 & \cdots & -1 \\ -1 & n & -1 & \cdots & -1 \\ -1 & -1 & n & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & -1 \\ -1 & -1 & \cdots & -1 & n \end{bmatrix}$$



- Eigenvec/value 1. Eigenspace $\{\boldsymbol{v} \perp \mathbf{1}\}$ with dimension $n-1$ and eigenvalue $n+1$

Example 1



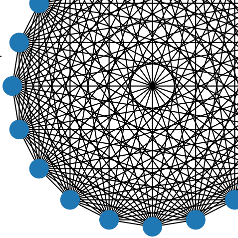
- Coarse grid correction, for vector of $\mathbf{1}$
- Richardson relaxation

$$G = I - \frac{1}{2n}A = \frac{n-1}{2n}I + \frac{1}{2n}\mathbf{1}\mathbf{1}^T$$

- Interpolation of $P = \mathbf{1}$ leads to

$$T = I - P(P^TAP)^{-1}P^TA = I - \frac{1}{n}\mathbf{1}\mathbf{1}^T$$

Example 1



- Then G has eigenvectors/eigenvalues:

$$\begin{array}{rl} \mathbf{1} & 1 - \dfrac{1}{2n} \\ \{\boldsymbol{v} \perp \mathbf{1}\} & \dfrac{n-1}{2n} \end{array}$$

- And T has eigenvectors/eigenvalues:

$$\begin{array}{rl} \mathbf{1} & 0 \\ \{\boldsymbol{v} \perp \mathbf{1}\} & n-1 \end{array}$$

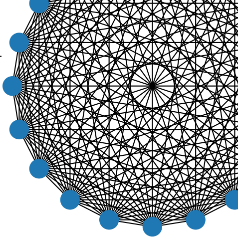
- Then GT has

$$\begin{array}{rl} \mathbf{1} & 0 \\ \{\boldsymbol{v} \perp \mathbf{1}\} & \dfrac{n-1}{2n} \end{array}$$

- Thus (since the spectra is the same as A):

$$\|GT\|_A = \frac{n-1}{2n} < \frac{1}{2} \Rightarrow \delta^* = 1 - \left(\frac{n-1}{2n}\right)^2$$

Example 1



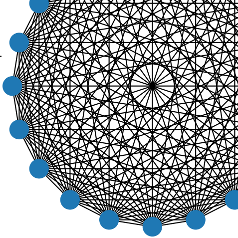
- Can show similar bounds for the split bounds:

$$\hat{\beta}_w = \hat{\beta}_s = \frac{n}{n+1}$$
$$\hat{\alpha}_w = \hat{\beta}_s = \frac{n}{n+1}\hat{\delta}$$

  So both the weak and strong split bounds are **sharp** in this case.

- Highlights bounds that explicitly account for the variational coarse grid correction process, can be sharp

Example 1



- Alternatively, we can show

$$\hat{\tau}_w = \frac{n^2}{n+1}$$

Thus $\hat{\tau}_w$ is larger than $\hat{\beta}_w$ by a factor of n.

- Similarly, the strong form is much sharper:

$$\hat{\tau}_s = \sup_{\boldsymbol{e}\neq\boldsymbol{0}} \frac{\|\boldsymbol{e} - P\boldsymbol{e}_c\|_A^2}{\|\boldsymbol{e}\|_{AD^{-1}A}^2} = \frac{2n^2}{(n+1)^2}$$

Example 2

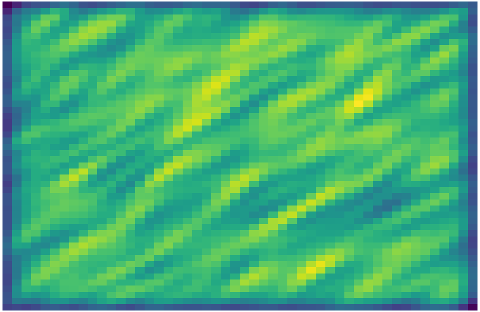
- Consider an anisotropic diffusion problem

$$-\nabla \cdot \kappa \nabla u = f \quad \text{in } \Omega = [0,1]^2,$$

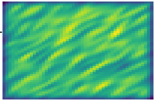
- $\kappa = \Theta K \Theta^T$, where $K = \begin{bmatrix} 1 & 0 \\ 0 & \varepsilon \end{bmatrix}$

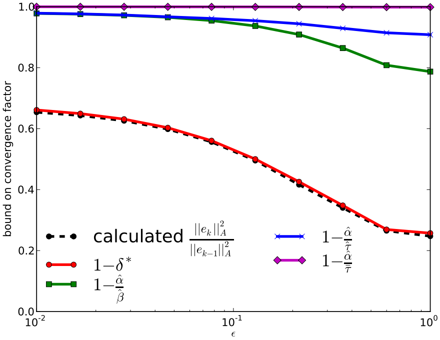
  and $\Theta = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$

- Vary ε

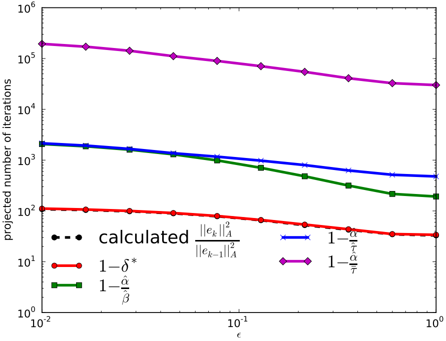


Example 2





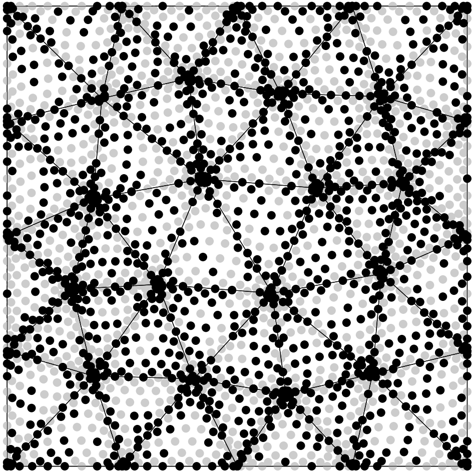
bound on convergence factor
1.0
0.8
0.6
0.4
0.2
0.0
10^{-2}
10^{-1}
10^{0}
ϵ
calculated $\frac{||e_k||_A^2}{||e_{k-1}||_A^2}$
$1-\delta^*$
$1-\frac{\hat{\alpha}}{\hat{\beta}}$
$1-\frac{\hat{\alpha}}{\hat{\tau}}$
$1-\frac{\hat{\alpha}}{\overline{\tau}}$



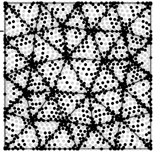
projected number of iterations
10^{6}
10^{5}
10^{4}
10^{3}
10^{2}
10^{1}
10^{0}
10^{-2}
10^{-1}
10^{0}
ϵ
calculated $\frac{||e_k||_A^2}{||e_{k-1}||_A^2}$
$1-\delta^*$
$1-\frac{\hat{\alpha}}{\hat{\beta}}$
$1-\frac{\hat{\alpha}}{\hat{\tau}}$
$1-\frac{\hat{\alpha}}{\overline{\tau}}$

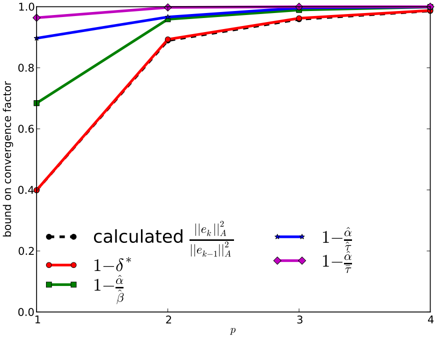
Example 3

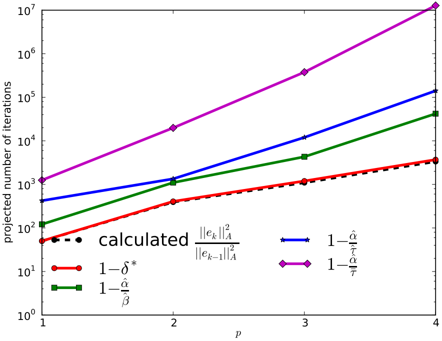
- Consider an higher-order finite elements
- Unstructured mesh. Vary polynomial order from $p = 1, \ldots, 4$.
- AMG convergence known to deteriorate at $p = 3$ or $p = 4$.



Example 3




bound on convergence factor
1.0
0.8
0.6
0.4
0.2
0.0
1
2
3
4
p
calculated $\frac{\|e_k\|_A^2}{\|e_{k-1}\|_A^2}$
$1-\delta^*$
$1-\frac{\hat{\alpha}}{\hat{\beta}}$
$1-\frac{\hat{\alpha}}{\hat{\tau}}$
$1-\frac{\hat{\alpha}}{\bar{\tau}}$


projected number of iterations
10^7
10^6
10^5
10^4
10^3
10^2
10^1
10^0
1
2
3
4
p
calculated $\frac{\|e_k\|_A^2}{\|e_{k-1}\|_A^2}$
$1-\delta^*$
$1-\frac{\hat{\alpha}}{\hat{\beta}}$
$1-\frac{\hat{\alpha}}{\hat{\tau}}$
$1-\frac{\hat{\alpha}}{\bar{\tau}}$

Many more things we could do here

- We could attempt to place bounds on the measures (and $\hat{\tau}_w$)
- We could introduce
$$X = \left(M + M^T - M^T A M\right)^{-1}$$
  the inverse of the symmetric smoother, and use this for B. This results in some relationships to sharpness in the generalized theory.
- We could apply this strategy to the bounds given in Compatible Relaxation.

- More for another day!

Concluding Remarks

- Approximation properties and smoothing properties govern multigrid performance
- Measuring these properties can be a challenge due to sharpness
- Developing methods based on key theoretical properties is important as long as we understand the limitations of the theory.

An incomplete list of some great articles

- McCormick, Ruge, *Multigrid Methods for Variational Problems*, SINUM, 1982.
- McCormick, *Multigrid Methods for Variational Problems: general theory for the V-cycle*, SINUM, 1985.
- Stüben, Trottenberg, *Multigrid methods: fundamental algorithms, model problem analysis and applications*, in Multigrid Methods, Springer, 1982.
- Falgout, Vassilevski, *On Generalizing the Algebraic Multigrid Framework*, SINUM, 2004.
- Vassilevski, *Multilevel Block Factorization Preconditioners*, Springer, 2008.
- Napov, Notay, *When does two-grid optimality carry over to the V-cycle*, NLAA, 2010.
- Napov, Notay, *Comparison of bounds for V-cycle multigrid*, NLAA, 2010.
- Ruge, Stüben, *Algebraic Multigrid*, in Multigrid Methods, SIAM, 1987.

Special Thanks

- Colin Cotter Imperial College
- David Ham Imperial College

More information

`lukeo@illinois.edu`

`http://lukeo.cs.illinois.edu`

`https://github.com/lukeolson/imperial-multigrid`

- **this work** MacLachlan, Olson, *Theoretical Bounds for Algebraic Multigrid Performance: review and analysis*, NLAA, 2014.