Problem

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

$$Au = f, \quad A \in \mathbb{R}^{n \times n}$$

Suppose we have a multilevel iteration process $\mathcal{M}_{\mathrm{MG}}$

$$I - \mathcal{M}_{\mathrm{MG}} A = \left(I - M^\top A\right)^{\nu_{\mathrm{pre}}} \left(I - P\left(P^\top A P\right)^{-1} P^\top A\right) (I - MA)^{\nu_{\mathrm{post}}}$$

so that

$$e \leftarrow (I - \mathcal{M}_{\mathrm{MG}} A) e$$

The iteration converges for any f and u_0 iff $\rho(I - \mathcal{M}_{\mathrm{MG}} A) < 1$.

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

Notation

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., $D = \mathrm{diag}(A)-$ defining an inner product:

$$\langle u, v \rangle_A = \langle Au, v \rangle$$
$$\langle u, v \rangle_D = \langle Du, v \rangle$$
$$\langle u, v \rangle_{AD^{-1}A} = \langle D^{-1} Au, Av \rangle$$

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

$$u \leftarrow Su + (I - S)A^{-1} f \quad \text{or} \quad e \leftarrow Se$$

Algorithm

$$u \leftarrow \hat{\mathcal{S}}(u, f)$$
$$r_c \leftarrow Rr$$
$$e_c \leftarrow A_c^{-1} r_c$$
$$\hat{e} \leftarrow P e_c$$
$$u \leftarrow u + \hat{e}$$
$$u \leftarrow \mathcal{S}(u, f)$$

In general, consider relaxation as

$$S = I - MA$$

Assumptions:

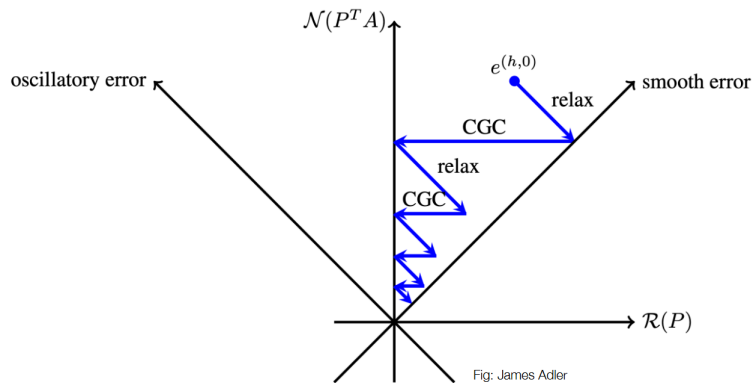- M is norm convergent (in A): $\|S\|_A < 1$
- P is full rank

- A is s.p.d.

Let the coarse grid correction step be

$$T = I - P\left(P^\top A P\right)^{-1} P^\top A$$

T is an A-orthogonal projection onto the range of P: After coarse grid correction, the error is minimized in the energy norm over $\mathrm{range}(P)$

$$\text{if } w \quad \text{s.t.} \quad Aw \in \mathcal{N}\left(P^T\right)$$
$$\left(I - P\left(P^\top A P\right)^{-1} P^\top A\right) w = w$$
$$\|\cdot\| \geq 1$$

Fig: James Adler

Focus on $V(0,1)$

The A-adjoint of ST is

$$TS^+ \quad S^+ = I - M^\top A$$

The symmetric $V(1,1)$ cycle is

$$(I - MA)\left(I - P\left(P^T A P\right)^{-1} P^T A\right)\left(I - M^T A\right) = STS^+$$
$$= STTS^+$$

Since $\|ST\|_A = \|TS^+\|_A$ (A-adjoints) we have

$$\left\|STS^+\right\|_A = \|ST\|_A^2$$

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

Upper Bound

We will measure convergence or reduction in $\|e\|_A$. Note:

$$\|e\|_A^2 = \|(I - T)e\|_A^2 + \|Te\|_A^2$$

For a $V(0,1)$ cycle, the reduction in e is

$$\|STe\|_A^2 \leq (1 - \delta^*)\|e\|_A^2$$

we seek a sharp bound

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

Sufficient conditions

What should we assume on relaxation and interpolation?

- relaxation is effective on the range of coarse grid correction

There exists $\delta > 0$ such that

$$\|STe\|_A^2 \leq (1-\delta)\|Te\|_A^2 \text{ for all } e.$$

Then, since T is an A-orthogonal projector

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

- No side effects on the range of interpolation

$$\|Sv\|_A^2 \leq \|v\|_A^2 \quad \text{for all } v \perp \text{range}(T)$$

Combine these into an assumption

$$\|Sv\|_A^2 \leq \|STv\|_A^2 + \|S(I-T)v\|_A^2 \leq (1-\delta)\|Tv\|_A^2 + \|(I-T)v\|_A^2 = \|v\|_A^2 - \delta\|Tv\|_A^2$$

Theorem
If there exists $\delta > 0$ so that

$$\|Se\|_A^2 \leq \|e\|_A^2 - \delta\|Te\|_A^2 \quad \text{for all } e,$$

then

$$\|ST\|_A^2 \leq 1 - \delta$$

To be sharp, the largest δ, say $\hat{\delta}$

$$\hat{\delta} = \inf_{e:Te \neq 0} \frac{\|e\|_A^2 - \|Se\|_A^2}{\|Te\|_A^2},$$

should be δ^*.

Proof

Since $Te = \mathbf{0}$ gives $\|STe\|_A = 0$:

$$\|ST\|_A^2 = \sup_{e:Te \neq 0} \frac{\|STe\|_A^2}{\|e\|_A^2} = \sup_{e:Te \neq 0} \frac{\|STe\|_A^2}{\|Te\|_A^2 + \|(I-T)e\|_A^2}$$

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

$$\|ST\|_A^2 = \sup_{e:Te \neq 0} \frac{\|STe\|_A^2}{\|e\|_A^2} = \sup_{e:Te \neq 0} \frac{\|S(Te + (I-T)e)\|_A^2}{\|Te\|_A^2} = \sup_{e:Te \neq 0} \frac{\|Se\|_A^2}{\|Te\|_A^2},$$

And

$$1 - \|ST\|_A^2 = \inf_{e:Te \neq \mathbf{0}} \frac{\|Te\|_A^2 - \|Se\|_A^2}{\|Te\|_A^2} = \inf_{e:Te \neq \mathbf{0}} \frac{\|e\|_A^2 - \|Se\|_A^2}{\|Te\|_A^2} = \hat{\delta}$$

The worst δ is sharp

$$\hat{\delta} = \inf_{e:Te \neq 0} \frac{\|e\|_A^2 - \|Se\|_A^2}{\|Te\|_A^2},$$

The early theory split this in two part

For some $g(e)$ define δ, α_g, and β_g as in

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

Consider the smallest α_g and the largest β_g:

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

For e such that $g(Te) \neq 0$,

$$\|STe\|_A^2 \leq \|Te\|_A^2 - \hat{\alpha}_g g(Te) \leq \|Te\|_A^2 - \frac{\hat{\alpha}_g}{\hat{\beta}_g}\|Te\|_A^2 = \left(1 - \frac{\hat{\alpha}_g}{\hat{\beta}_g}\right)\|Te\|_A^2$$

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

Ok, so this is generally worse than the sharp bound

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

(α_g and β_g are not generally simultaneously satisfied)

Early works, e.g. Ruge-Stüben 1987, use $g(e) = \|e\|_{AD^{-1}A}^2$ Or the weaker form $g(e) = \|Te\|_{AD^{-1}A}^2$

Smoothing and Approximation

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

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

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

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

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

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

$$\|Te\|_A = \inf_{e_c} \|e - Pe_c\|_A$$

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

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

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

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

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

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

(weak approximation) Assume that

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