

# A Comparison of Two-Level Preconditioning Strategies

Scott MacLachlan  
scott.maclachlan@tufts.edu

Tufts University

March 19, 2008

# The Big Question

How do we solve  $Ax = b$ ?

# The Big Question

How do we solve  $Ax = b$ ?

## Gaussian Elimination

- Recursive procedure to reduce  $A$  to a simpler form
- Use  $i^{\text{th}}$  equation to eliminate  $x_i$  from  $j^{\text{th}}$  equation for  $j > i$

# The Bigger Question

Why do we solve  $Ax = b$ ?

# Discrete PDEs

Given PDE  $L(u) = f$  on domain  $\Omega$ .

## Finite Differences

- Represent  $u$  and  $f$  by pointwise values at nodes  $\xi_i \in \Omega$
- Approximate  $L(u(\xi_i))$  using Taylor series

## Finite Elements

- Represent  $u$  and  $f$  by basis functions for domain and range of  $L$
- Approximate  $L(u)$  by projection onto subsets of these bases

# The Real Question

How do we solve  $Ax = b$ ?

When

- $A$  is large (of dimension  $n > 10^6$ )?
- $A$  is sparse (fixed number of nonzeros per row)?
- $A$  inherits properties from continuum PDE?

# The Real Question

## How do we solve $Ax = b$ ?

When

- $A$  is large (of dimension  $n > 10^6$ )?
- $A$  is sparse (fixed number of nonzeros per row)?
- $A$  inherits properties from continuum PDE?
- we only need an approximate solution?

# The Real Question

## How do we solve $Ax = b$ ?

When

- $A$  is large (of dimension  $n > 10^6$ )?
- $A$  is sparse (fixed number of nonzeros per row)?
- $A$  inherits properties from continuum PDE?
- we only need an approximate solution?

Want **specialized techniques** that take advantage of matrix structure and are **faster than Gaussian elimination**

# Outline

- Polynomial Methods
  - ▶ Preconditioning
  - ▶ Spectral equivalence
  - ▶ Classical preconditioners
- Two-level preconditioners
  - ▶ Projections
  - ▶ Deflation
  - ▶ Balancing Neumann-Neumann
  - ▶ Multigrid
- Some comparisons
  - ▶ Abstract framework
  - ▶ Comparison theorems
  - ▶ What does it mean?

# Preconditioning

The solution of  $A\mathbf{x} = \mathbf{b}$  solves many other linear systems

Consider solving  $M^{-1}A\mathbf{x} = M^{-1}\mathbf{b}$

**Idea:** If  $M^{-1}A$  has nicer properties than  $A$ , easier to solve

# Preconditioning

The solution of  $A\mathbf{x} = \mathbf{b}$  solves many other linear systems

Consider solving  $M^{-1}A\mathbf{x} = M^{-1}\mathbf{b}$

**Idea:** If  $M^{-1}A$  has nicer properties than  $A$ , easier to solve

**Need:** Preconditioner,  $M$ , close to  $A$

**Need:** Cheap computation of  $M^{-1}$

# Iterative Methods

- Want to improve approximation,  $\mathbf{x}^{(0)}$ , to  $\mathbf{x} = A^{-1}\mathbf{b}$
- Residual,  $\mathbf{r}^{(0)}$ , is a measure of the error

$$\mathbf{r}^{(0)} = \mathbf{b} - A\mathbf{x}^{(0)} = A\mathbf{x} - A\mathbf{x}^{(0)} = A(\mathbf{x} - \mathbf{x}^{(0)})$$

- Take  $\mathbf{x}^{(1)} = \mathbf{x}^{(0)} + \omega_0 \mathbf{r}^{(0)}$

# Iterative Methods

- Want to improve approximation,  $\mathbf{x}^{(0)}$ , to  $\mathbf{x} = A^{-1}\mathbf{b}$
- Residual,  $\mathbf{r}^{(0)}$ , is a measure of the error

$$\mathbf{r}^{(0)} = \mathbf{b} - A\mathbf{x}^{(0)} = A\mathbf{x} - A\mathbf{x}^{(0)} = A(\mathbf{x} - \mathbf{x}^{(0)})$$

- Take  $\mathbf{x}^{(1)} = \mathbf{x}^{(0)} + \omega_0 M^{-1}\mathbf{r}^{(0)}$

Error propagation form:  $\mathbf{e}^{(1)} = (I - \omega_0 M^{-1}A)\mathbf{e}^{(0)}$

# Iterative Methods

- Want to improve approximation,  $\mathbf{x}^{(0)}$ , to  $\mathbf{x} = A^{-1}\mathbf{b}$
- Residual,  $\mathbf{r}^{(0)}$ , is a measure of the error

$$\mathbf{r}^{(0)} = \mathbf{b} - A\mathbf{x}^{(0)} = A\mathbf{x} - A\mathbf{x}^{(0)} = A(\mathbf{x} - \mathbf{x}^{(0)})$$

- Take  $\mathbf{x}^{(1)} = \mathbf{x}^{(0)} + \omega_0 M^{-1}\mathbf{r}^{(0)}$

Error propagation form:  $\mathbf{e}^{(1)} = (I - \omega_0 M^{-1}A)\mathbf{e}^{(0)}$   
 $\mathbf{e}^{(2)} = (I - \omega_1 M^{-1}A)\mathbf{e}^{(1)}$

# Iterative Methods

- Want to improve approximation,  $\mathbf{x}^{(0)}$ , to  $\mathbf{x} = A^{-1}\mathbf{b}$
- Residual,  $\mathbf{r}^{(0)}$ , is a measure of the error

$$\mathbf{r}^{(0)} = \mathbf{b} - A\mathbf{x}^{(0)} = A\mathbf{x} - A\mathbf{x}^{(0)} = A(\mathbf{x} - \mathbf{x}^{(0)})$$

- Take  $\mathbf{x}^{(1)} = \mathbf{x}^{(0)} + \omega_0 M^{-1}\mathbf{r}^{(0)}$

Error propagation form:  $\mathbf{e}^{(1)} = (I - \omega_0 M^{-1}A)\mathbf{e}^{(0)}$   
 $\mathbf{e}^{(2)} = (I - \omega_1 M^{-1}A)(I - \omega_0 M^{-1}A)\mathbf{e}^{(0)}$

# Iterative Methods

- Want to improve approximation,  $\mathbf{x}^{(0)}$ , to  $\mathbf{x} = A^{-1}\mathbf{b}$
- Residual,  $\mathbf{r}^{(0)}$ , is a measure of the error

$$\mathbf{r}^{(0)} = \mathbf{b} - A\mathbf{x}^{(0)} = A\mathbf{x} - A\mathbf{x}^{(0)} = A(\mathbf{x} - \mathbf{x}^{(0)})$$

- Take  $\mathbf{x}^{(1)} = \mathbf{x}^{(0)} + \omega_0 M^{-1}\mathbf{r}^{(0)}$

Error propagation form:

$$\begin{aligned}\mathbf{e}^{(1)} &= (I - \omega_0 M^{-1}A)\mathbf{e}^{(0)} \\ \mathbf{e}^{(2)} &= (I - \omega_1 M^{-1}A)(I - \omega_0 M^{-1}A)\mathbf{e}^{(0)} \\ &\vdots \\ \mathbf{e}^{(k)} &= p_{k,0}(M^{-1}A)\mathbf{e}^{(0)}\end{aligned}$$

**Question:** How do we pick coefficients:  $\omega_0, \omega_1, \dots$

# Polynomial Methods

Many ways to choose coefficients of  $p_{k,0}$  (or weights,  $\omega_i$ )

## Uniform

- Fix  $\omega_i = \omega$

## Chebyshev

- Use estimate of field of values of  $M^{-1}A$  to choose weights

## Krylov

- Choose polynomials for optimality in some sense
- Conjugate gradient, GMRES, BiCGSTAB, . . .

Success depends on **closeness** of  $M$  and  $A$

# Spectral Equivalence

Measure closeness of  $M$  and  $A$  by bound

$$\lambda_{\min} \mathbf{x}^T M \mathbf{x} \leq \mathbf{x}^T A \mathbf{x} \leq \lambda_{\max} \mathbf{x}^T M \mathbf{x}, \quad \forall \mathbf{x}$$

$\lambda_{\min}$  and  $\lambda_{\max}$  are called the **spectral equivalence bounds**

Convergence of polynomial methods typically depends on

$$\kappa(M^{-\frac{1}{2}} A M^{-\frac{1}{2}}) = \frac{\lambda_{\max}}{\lambda_{\min}}$$

For example,

$$\|\mathbf{e}^{(k)}\|_A \leq 2 \left( \frac{\sqrt{\kappa} - 1}{\sqrt{\kappa} + 1} \right)^k \|\mathbf{e}^{(0)}\|_A$$

# Spectral Equivalence

Measure closeness of  $M$  and  $A$  by bound

$$\lambda_{\min} \mathbf{x}^T M \mathbf{x} \leq \mathbf{x}^T A \mathbf{x} \leq \lambda_{\max} \mathbf{x}^T M \mathbf{x}, \quad \forall \mathbf{x}$$

$\lambda_{\min}$  and  $\lambda_{\max}$  are called the **spectral equivalence bounds**

Convergence of polynomial methods typically depends on

$$\kappa(M^{-\frac{1}{2}} A M^{-\frac{1}{2}}) = \frac{\lambda_{\max}}{\lambda_{\min}}$$

How do we pick an  $M$  so that  $\kappa(M^{-\frac{1}{2}} A M^{-\frac{1}{2}})$  is small?

# Spectral Equivalence

Measure closeness of  $M$  and  $A$  by bound

$$\lambda_{\min} \mathbf{x}^T M \mathbf{x} \leq \mathbf{x}^T A \mathbf{x} \leq \lambda_{\max} \mathbf{x}^T M \mathbf{x}, \quad \forall \mathbf{x}$$

$\lambda_{\min}$  and  $\lambda_{\max}$  are called the **spectral equivalence bounds**

Convergence of polynomial methods typically depends on

$$\kappa(M^{-\frac{1}{2}} A M^{-\frac{1}{2}}) = \frac{\lambda_{\max}}{\lambda_{\min}}$$

How do we pick an  $M$  so that  $\kappa(M^{-\frac{1}{2}} A M^{-\frac{1}{2}})$  is small?

Start from  $A$

# Classical Preconditioners

Classical choices for  $M$  include

- $M = \alpha I$ , for  $\alpha \approx \|A\|$  (Richardson)
- $M = \alpha \text{diag}(A)$  (Jacobi)
- $M = \text{tril}(A)$  (Gauss-Seidel)
- $M = \hat{L}\hat{U}$ , for  $A = LU$ ,  $\hat{L} \approx L$ ,  $\hat{U} \approx U$  (ILU)

**Question:** How well do these work?

# Classical Preconditioners

Classical choices for  $M$  include

- $M = \alpha I$ , for  $\alpha \approx \|A\|$  (Richardson)
- $M = \alpha \text{diag}(A)$  (Jacobi)
- $M = \text{tril}(A)$  (Gauss-Seidel)
- $M = \hat{L}\hat{U}$ , for  $A = LU$ ,  $\hat{L} \approx L$ ,  $\hat{U} \approx U$  (ILU)

**Question:** How well do these work?

Define **well** by considering  $\sigma \left( M^{-\frac{1}{2}} A M^{-\frac{1}{2}} \right)$

- Lower and upper bounds on  $\sigma \left( M^{-\frac{1}{2}} A M^{-\frac{1}{2}} \right)$  define condition number:  $\kappa \left( M^{-\frac{1}{2}} A M^{-\frac{1}{2}} \right) = \frac{\lambda_{\max}}{\lambda_{\min}}$

# Simple Test Problem

Finite difference discretization of  $-\Delta u = f$  in  $[0, 1]^2$

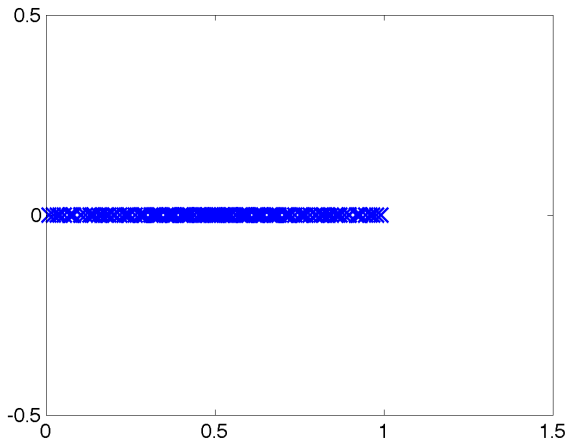
Approximate  $-u_{xx} \approx \frac{1}{h^2} (-u_{i-1,j} + 2u_{i,j} - u_{i+1,j}) + O(h^2)$

$$\frac{1}{h^2} (-u_{i-1,j} + 2u_{i,j} - u_{i+1,j}) + \frac{1}{h^2} (-u_{i,j-1} + 2u_{i,j} - u_{i,j+1}) = f_{i,j}$$

for  $1 \leq i \leq N$ ,  $1 \leq j \leq N$ , along with Dirichlet boundary conditions,  $u_{i,j} = 0$ , for  $i = 0, N + 1$ ,  $j = 0, N + 1$

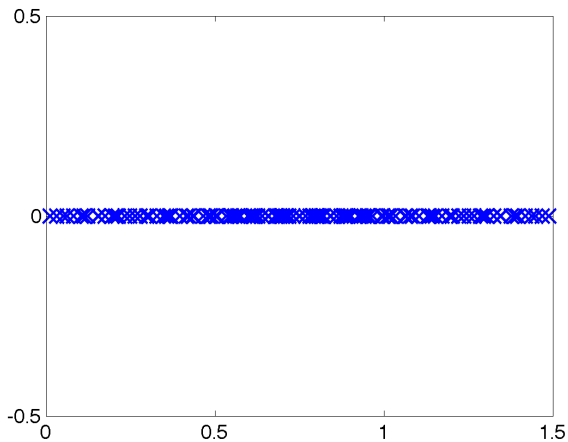
Matrix of dimension  $n = N \times N$ , with 5 nonzero entries per row

# Spectral Pictures



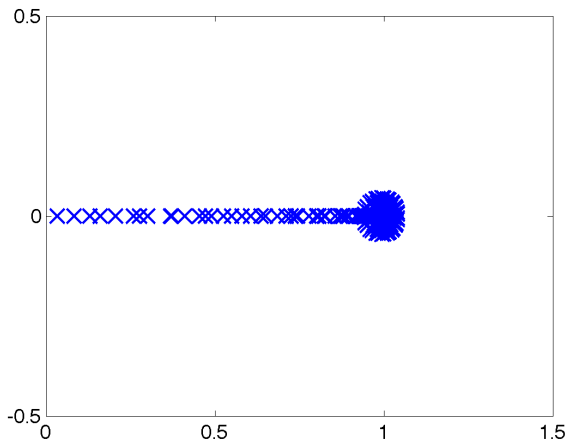
Richardson:  $M = \|A\|$

# Spectral Pictures



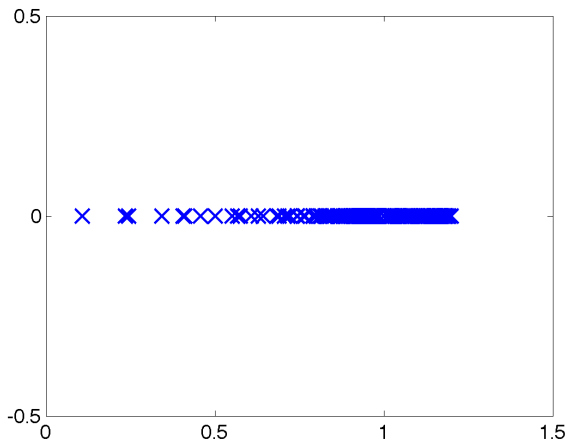
Jacobi:  $M = \frac{3}{4}\text{diag}(A)$

# Spectral Pictures



Gauss-Seidel:  $M = \text{tril}(A)$

# Spectral Pictures



$$\text{ILU: } M = \hat{L}\hat{U}$$

# The Problem

Convergence expected to slow as  $\kappa(M^{-\frac{1}{2}}AM^{-\frac{1}{2}}) = \frac{\lambda_{\max}}{\lambda_{\min}}$  grows

$N$	$\kappa_{\text{JAC}}$	$\kappa_{\text{ILU}}$
16	116.5	9.5
32	440.7	37.3
64	1711.7	146.4
128	6743.7	603.5

We expect performance to degrade as  $N$  increases!

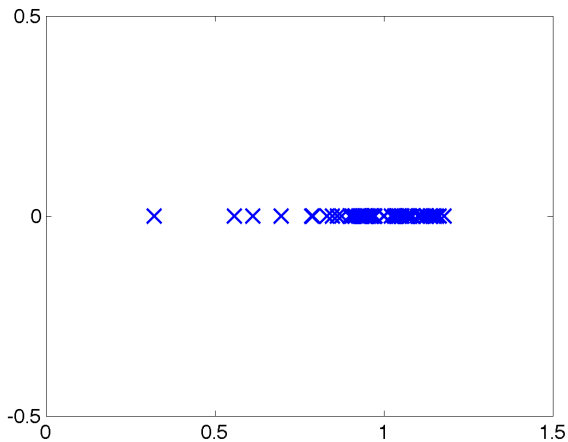
# Optimal Performance

Because we're interested in solving large problems quickly, want

- $\kappa(M^{-\frac{1}{2}}AM^{-\frac{1}{2}})$  bounded, independently of  $N$
- cost of computing  $M^{-1}A$  to be proportional to  $N^2$

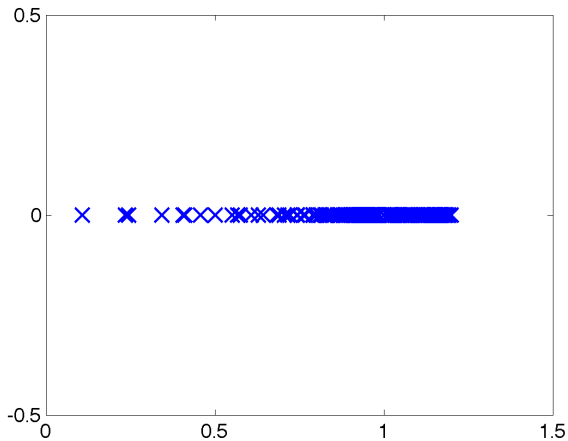
**Question:** Can we modify  $M$  to get optimal performance?

# Spectral Picture Revisited



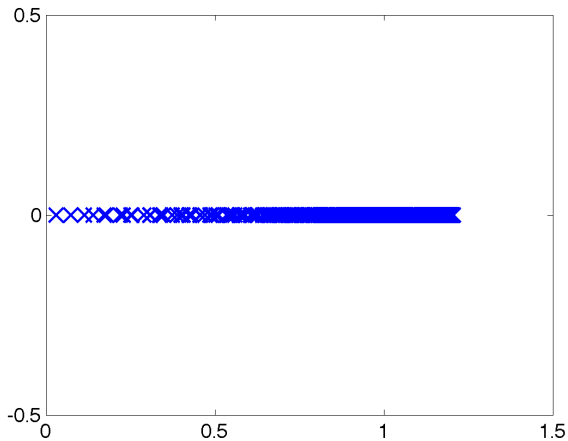
$$\text{ILU: } M = \hat{L}\hat{U}, N = 8$$

# Spectral Picture Revisited



$$\text{ILU: } M = \hat{L}\hat{U}, N = 16$$

# Spectral Picture Revisited



$$\text{ILU: } M = \hat{L}\hat{U}, N = 32$$

# Projection Methods

Only a few modes cause difficulty!

Try and remove these modes by another process. Let

- $Z \in \mathbb{R}^{n \times k}$  have full rank
- $W \in \mathbb{R}^{n \times (n-k)}$  have full rank
- $Z^T A W = 0$

# Projection Methods

Only a few modes cause difficulty!

Try and remove these modes by another process. Let

- $Z \in \mathbb{R}^{n \times k}$  have full rank
- $W \in \mathbb{R}^{n \times (n-k)}$  have full rank
- $Z^T A W = 0$
- $\mathcal{R}(Z)$  contain troublesome modes

Write  $\mathbf{e}^{(k)} = Z\mathbf{v} + W\mathbf{y}$

$$P\mathbf{e}^{(k)} = (I - Z(Z^T A Z)^{-1} Z^T A)\mathbf{e}^{(k)} = W\mathbf{y}$$

$$P^T \mathbf{r}^{(k)} = (I - A Z (Z^T A Z)^{-1} Z^T) A \mathbf{e}^{(k)} = A W \mathbf{y}$$

# Two-Level Preconditioners

Preconditioners that combine  $M^{-1}$  and  $P$  are called **two-level preconditioners**

Complementarity:

- $M^{-1}$  removes most errors
- $P$  removes errors for which  $M^{-1}$  is ineffective

# Two-Level Preconditioners

Preconditioners that combine  $M^{-1}$  and  $P$  are called **two-level preconditioners**

Complementarity:

- $M^{-1}$  removes most errors
- $P$  removes errors for which  $M^{-1}$  is ineffective

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

Applying  $P$  requires solving system with  $Z^T A Z$

- $Z^T A Z$  is called the “coarse-grid” system

# Two Approaches

1. Use  $P$  to remove error in  $\mathbf{x}^{(k)}$  in  $\text{span}(Z)$ :

$$\mathbf{x}^{(k')} = \mathbf{x}^{(k)} + Z(Z^T A Z)^{-1} Z^T A \mathbf{x}^{(k)}$$

$$= \mathbf{x}^{(k)} + Z(Z^T A Z)^{-1} Z^T \mathbf{r}^{(k)}$$

$$\mathbf{e}^{(k')} = P \mathbf{e}^{(k)} = \mathbf{e}^{(k)} - Z(Z^T A Z)^{-1} Z^T A \mathbf{e}^{(k)}$$

2. Use  $P$  to partition exact solution

$$\mathbf{x} = P \mathbf{x} + (I - P) \mathbf{x}$$

Since  $(I - P) \mathbf{x} = Z(Z^T A Z)^{-1} Z^T A \mathbf{x} = Z(Z^T A Z)^{-1} Z^T \mathbf{b}$ ,  
concentrate on computing  $P \mathbf{x}$ , but

$$A P \mathbf{x} = P^T A \mathbf{x} = P^T \mathbf{b}.$$

# Deflation Techniques

Second approach: Apply  $P^T$  directly to residual

Write

$$\mathbf{r}^{(k)} = A\mathbf{e}^{(k)} = AZ\mathbf{v} + AW\mathbf{y},$$

so that

$$P^T\mathbf{r}^{(k)} = P^T A\mathbf{e}^{(k)} = AW\mathbf{y},$$

then precondition:

$$M^{-1}P^T\mathbf{r}^{(k)} = M^{-1}P^T A\mathbf{e}^{(k)} = M^{-1}AW\mathbf{y}$$

**Idea:** exactly eliminate errors in  $\mathcal{R}(Z)$ , by computing  $(I - P)\mathbf{x}$ , then iterate for solution in  $\mathcal{R}(W)$ .

# Deflation Preconditioner

Can write deflation algorithm as classical preconditioner

Solving for  $\tilde{\mathbf{x}} = P\mathbf{x}$ , in

$$M^{-1}P^T A \tilde{\mathbf{x}} = M^{-1}P^T \mathbf{b}$$

This is in the same form as classical preconditioning:

$$M^{-1}A\mathbf{x} = M^{-1}\mathbf{b}$$

Identify deflation preconditioner as

$$\mathcal{P}_{\text{DEF}} = M^{-1}P^T$$

# Balancing Preconditioners

First Approach: Project out errors in  $\mathcal{R}(Z)$  before and after applying  $M^{-1}$

1.  $\mathbf{x}^{(k')} = \mathbf{x}^{(k)} + Z(Z^T AZ)^{-1} Z^T \mathbf{r}^{(k)}$
2.  $\mathbf{x}^{(k'')} = \mathbf{x}^{(k')} + \omega_k M^{-1} \mathbf{r}^{(k')}$
3.  $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k'')} + Z(Z^T AZ)^{-1} Z^T \mathbf{r}^{(k'')}$

Equivalent to solving preconditioned system

$$\mathcal{P}_{\text{BNN}} A \mathbf{x} = \mathcal{P}_{\text{BNN}} \mathbf{b},$$

for

$$\mathcal{P}_{\text{BNN}} = PM^{-1}P^T + Z(Z^T EZ)^{-1}Z^T$$

# Multigrid Preconditioners

First Approach: Apply  $M^{-1}$  before and after projecting out errors in  $\mathcal{R}(Z)$

1.  $\mathbf{x}^{(k')} = \mathbf{x}^{(k)} + M^{-1}\mathbf{r}^{(k)}$
2.  $\mathbf{x}^{(k'')} = \mathbf{x}^{(k')} + Z(Z^T A Z)^{-1} Z^T \mathbf{r}^{(k')}$
3.  $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k'')} + M^{-T} \mathbf{r}^{(k'')}$

Equivalent to solving preconditioned system

$$\mathcal{P}_{\text{MG}} A \mathbf{x} = \mathcal{P}_{\text{MG}} \mathbf{b},$$

for

$$\mathcal{P}_{\text{MG}} = M^{-T} P^T + P M^{-1} + M^{-T} P^T A M^{-1} + Z(Z^T E Z)^{-1} Z^T$$

# Comparing Costs

Three very different implementations of same idea

Compare costs per iteration:

**Deflation:**  $\mathcal{P}_{\text{DEF}} = M^{-1}P^T$

- One multiplication with  $M^{-1}$
- One multiplication with  $P^T$

**Balancing:**  $\mathcal{P}_{\text{BNN}} = PM^{-1}P^T + Z(Z^TEZ)^{-1}Z^T$

- One multiplication with  $M^{-1}$
- Two multiplications with  $P$

**Multigrid:**  $\mathcal{P}_{\text{MG}} = M^{-T}P^T + PM^{-1} + M^{-T}P^TAM^{-1} + \dots$

- One multiplication with  $M^{-1}$ , one with  $M^{-T}$
- One multiplication with  $P$

# Comparing Performance

Comparing performance is more difficult than comparing costs

Actual Performance depends on

- Polynomial method chosen
- Right-hand side,  $\mathbf{b}$
- Initial guess,  $\mathbf{x}^{(0)}$
- Choice of  $M^{-1}$
- Choice of  $Z$

**Instead:** Compare expected performance based on bounds

$$\|\mathbf{e}^{(k)}\|_A \leq 2 \left( \frac{\sqrt{\kappa} - 1}{\sqrt{\kappa} + 1} \right)^k \|\mathbf{e}^{(0)}\|_A$$

# Comparing Bounds

Important factor in

$$\|\mathbf{e}^{(k)}\|_A \leq 2 \left( \frac{\sqrt{\kappa} - 1}{\sqrt{\kappa} + 1} \right)^k \|\mathbf{e}^{(0)}\|_A$$

is  $\kappa(M^{-\frac{1}{2}}AM^{-\frac{1}{2}}) = \frac{\lambda_{\max}}{\lambda_{\min}}$ .

**Goal:** Theoretical comparison of spectra of DEF, BNN, MG

# Comparing DEF and BNN

For any given choice of  $M^{-1}$  and  $Z$ ,

$$\sigma(\mathcal{P}_{\text{DEF}}A) = \{0, \dots, 0, \mu_{k+1}, \dots, \mu_n\}$$

# Comparing DEF and BNN

For any given choice of  $M^{-1}$  and  $Z$ ,

$$\sigma(\mathcal{P}_{\text{DEF}}A) = \{0, \dots, 0, \mu_{k+1}, \dots, \mu_n\}$$

Furthermore, if the same choices of  $M^{-1}$  and  $Z$  are used in BNN,

$$\sigma(\mathcal{P}_{\text{BNN}}A) = \{1, \dots, 1, \mu_{k+1}, \dots, \mu_n\}$$

# Comparing DEF and BNN

For any given choice of  $M^{-1}$  and  $Z$ ,

$$\sigma(\mathcal{P}_{\text{DEF}}A) = \{0, \dots, 0, \mu_{k+1}, \dots, \mu_n\}$$

Furthermore, if the same choices of  $M^{-1}$  and  $Z$  are used in BNN,

$$\sigma(\mathcal{P}_{\text{BNN}}A) = \{1, \dots, 1, \mu_{k+1}, \dots, \mu_n\}$$

So, if  $\mu_{k+1} \leq 1 \leq \mu_n$ , then  $\kappa_{\text{DEF}} = \kappa_{\text{BNN}}$

# Comparing DEF with MG

Direct comparison of DEF and MG is difficult

- MG requires multiplication by both  $M^{-1}$  and  $M^{-T}$
- Preserving symmetry of operations is important
  - ▶ Difficult to just drop one of multiplications

Can show that multigrid with only  $M^{-1}$  or  $M^{-T}$  gives

$$\sigma(\mathcal{P}_{\text{MG1}}A) = \{1, \dots, 1, \mu_{k+1}, \dots, \mu_n\},$$

but numerical stability becomes a problem

# Spectral Two-Level Preconditioners

Take

- $M^{-1}$  to be arbitrary
- $\{\mathbf{v}_j\}, \{\lambda_j\}$  to be the eigenvectors & eigenvalues of  $M^{-1}A$
- $Z = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k]$
- $0 < \lambda_{k+1} \leq \lambda_j \leq \lambda_n < 2$  for all  $k + 1 < j < n$

Then,  $\mathcal{P}_{\text{DEF}}A$  has eigenvalues

$$\begin{cases} 0 & \text{for } j = 1, \dots, k \\ \lambda_j & \text{for } j = k + 1, \dots, n \end{cases}$$

and  $\mathcal{P}_{\text{MG}}A$  has eigenvalues

$$\begin{cases} 1 & \text{for } j = 1, \dots, k \\ \lambda_j(2 - \lambda_j) & \text{for } j = k + 1, \dots, n \end{cases}$$

---

C. Vuik, A. Segal, & J.A. Meijerink, J. Comput. Phys. 1999, **152**:385–403

B. Carpentieri, L. Giraud, & S. Gratton, SISC 2007, **29**:1593-1612

# Spectral Two-Level Preconditioners

Take

- $M^{-1}$  to be arbitrary
- $\{\mathbf{v}_j\}, \{\lambda_j\}$  to be the eigenvectors & eigenvalues of  $M^{-1}A$
- $Z = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k]$
- $0 < \lambda_{k+1} \leq \lambda_j \leq \lambda_n < 2$  for all  $k + 1 < j < n$

Then,  $\mathcal{P}_{\text{DEF}}A$  has condition number

$$\kappa_{\text{DEF}} = \frac{\lambda_n}{\lambda_{k+1}}$$

and  $\mathcal{P}_{\text{MG}}A$  has condition number

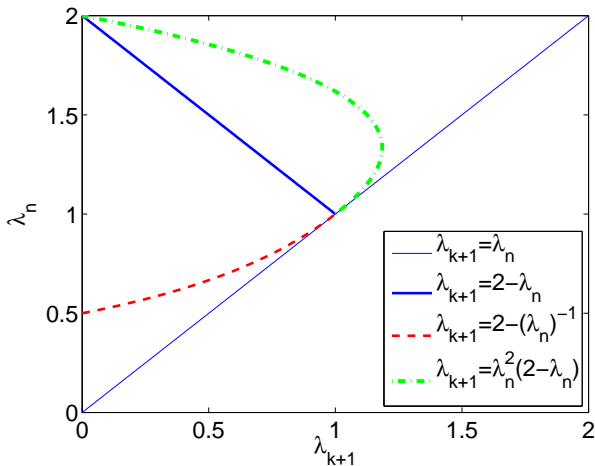
$$\kappa_{\text{MG}} = \frac{1}{\min\{\lambda_{k+1}(2 - \lambda_{k+1}), \lambda_n(2 - \lambda_n)\}}$$

---

C. Vuik, A. Segal, & J.A. Meijerink, J. Comput. Phys. 1999, **152**:385–403

B. Carpentieri, L. Giraud, & S. Gratton, SISC 2007, **29**:1593-1612

# Comparing Spectral Preconditioners



# Relaxation Parameters

Restriction that  $0 < \lambda_{k+1} \leq \lambda_j \leq \lambda_n < 2$  may not be natural.

Imposed by enforcing that  $\mathcal{P}_{\text{MG}}$  be symmetric and definite

- No such restriction for  $\mathcal{P}_{\text{DEF}}$
- Satisfied for many typical choices of  $M^{-1}$
- Easily enforced by adding parameter,  $M^{-1} \leftarrow \alpha M^{-1}$

Addition of **optimal relaxation parameter**,  $\alpha$ , has significant effect on MG performance

For  $M^{-1} = I$ ,  $\alpha = \frac{2}{\lambda_{k+1} + \lambda_n}$  is optimal choice. With this  $\alpha$ ,

$$\kappa_{\text{MG}} = \frac{(\lambda_{k+1} + \lambda_n)^2}{4\lambda_{k+1}\lambda_n} \leq \frac{\lambda_n}{\lambda_{k+1}} = \kappa_{\text{DEF}}$$

# Symmetrizing $M^{-1}$

**Main difficulty:** MG has two classical preconditioning steps, DEF and BNN use just one

**Idea:** Use different  $M^{-1}$  in MG and DEF/BNN

Recall: MG given by

1.  $\mathbf{x}^{(k')} = \mathbf{x}^{(k)} + M^{-1}\mathbf{r}^{(k)}$
2.  $\mathbf{x}^{(k'')} = \mathbf{x}^{(k')} + Z(Z^T AZ)^{-1}Z^T \mathbf{r}^{(k')}$
3.  $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k'')} + M^{-T}\mathbf{r}^{(k'')}$

BNN given by

1.  $\mathbf{x}^{(k')} = \mathbf{x}^{(k)} + Z(Z^T AZ)^{-1}Z^T \mathbf{r}^{(k)}$
2.  $\mathbf{x}^{(k'')} = \mathbf{x}^{(k')} + M^{-1}\mathbf{r}^{(k')}$
3.  $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k'')} + Z(Z^T AZ)^{-1}Z^T \mathbf{r}^{(k'')}$

# Symmetrizing $M^{-1}$

**Main difficulty:** MG has two classical preconditioning steps, DEF and BNN use just one

**Idea:** Use different  $M^{-1}$  in MG and DEF/BNN

Use  $\tilde{M}$  in DEF/BNN, where

$$(I - \tilde{M}^{-1}A) = (I - M^{-T}A)(I - M^{-1}A)$$

Compute:  $\tilde{M}^{-1} = M^{-T}(M + M^T - A)M^{-1}$

# Comparing MG, BNN, and DEF

Compare

- MG using  $M^{-1}$  and arbitrary  $Z$
- BNN using  $\tilde{M}^{-1}$  and same  $Z$
- DEF using  $\tilde{M}^{-1}$  and same  $Z$

Then,

$$\begin{aligned}\sigma(\mathcal{P}_{\text{MG}}A) &= \{1, \dots, 1, \lambda_{k+1}, \dots, \lambda_n\}, \\ \sigma(\mathcal{P}_{\text{BNN}}A) &= \{1, \dots, 1, \lambda_{k+1}, \dots, \lambda_n\}, \\ \sigma(\mathcal{P}_{\text{DEF}}A) &= \{0, \dots, 0, \lambda_{k+1}, \dots, \lambda_n\},\end{aligned}$$

where  $\lambda_{k+1}, \lambda_n$  depend on choices of  $M^{-1}$ ,  $Z$ .

# Consequences

Convergence analysis says the choice of two-level approach  
doesn't matter

Instead, we should concentrate on:

- Choices of  $M^{-1}$ ,  $Z$
- Efficient implementations
- Numerical stability

# Stability

Analysis is based on exactly computing

$$\mathbf{x}^{(k')} = \mathbf{x}^{(k)} + Z(Z^T A Z)^{-1} Z^T \mathbf{r}^{(k)}$$

This requires solution of

$$(Z^T A Z)w = Z^T \mathbf{r}^{(k)}$$

In practice, solve this system by recursion, perturbing eigenvalues of  $\mathcal{P}A$

- For MG, BNN, perturbs unit eigenvalues
- For DEF, perturbs zero eigenvalues

Small non-zero eigenvalues can cause convergence problems

# Conclusions

- Preconditioned polynomial methods best option for solving many linear systems
- Convergence depends on  $\kappa \left( M^{-\frac{1}{2}} A M^{-\frac{1}{2}} \right)$
- Improve performance by incorporating projection,  $P$
- Three variants: Deflation, Balancing, Multigrid
- Parameters important
- For “right” choice of  $M^{-1}$ , performance bounds are identical
- Deflation may have stability issues
- **Important choices** are  $M^{-1}$  and  $Z$

# Acknowledgements

This research was performed in collaboration with

- Jok Tang, Kees Vuik, TU-Delft
- Reinhard Nabben, TU-Berlin

This research was supported by

- the Dutch BSIK/BRICKS project,
- the Deutsche Forschungsgemeinschaft (DFG), Project NA248/2-2, and
- the European Community's Sixth Framework Programme, through a Marie Curie International Incoming Fellowship, MIF1-CT-2006-021927.