Introduction to Simulation - Lecture 7

Krylov-Subspace Matrix Solution Methods

Part II

Jacob White

Thanks to Deepak Ramaswamy, Michal Rewienski, and Karen Veroy

Outline

- Reminder about GCR
  - Residual minimizing solution
  - Krylov Subspace
  - Polynomial Connection
- Review Eigenvalues and Norms
  - Induced Norms
  - Spectral mapping theorem
- Estimating Convergence Rate
  - Chebychev Polynomials
- Preconditioners
  - Diagonal Preconditioners
  - Approximate LU preconditioners

Generalized Conjugate Residual Algorithm

With Normalization

$r^0 = b - Ax^0$

For j = 0 to k-1

$p_j = r^j$ } Residual is next search direction

For i = 0 to j-1

$p_j \leftarrow p_j - (Mp_j)^T (Mp_i) p_i$ } Orthogonalize Search Direction

$p_j \leftarrow \frac{1}{\sqrt{(Mp_j)^T (Mp_j)}} p_j$ } Normalize

$x^{j+1} = x^j + (r^j)^T (Mp_j) p_j$ } Update Solution

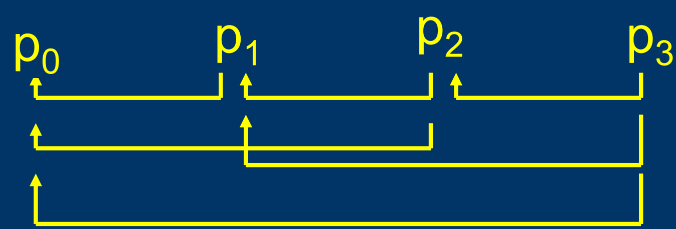
$r^{j+1} = r^j - (r^j)^T (Mp_j) M p_j$ } Update Residual

Generalized Conjugate Residual Algorithm

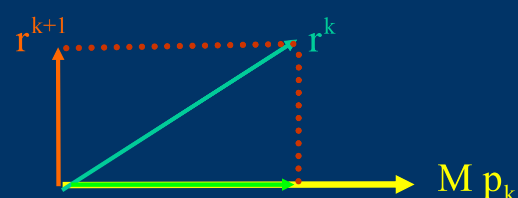
With Normalization

Algorithm Steps by Picture

1) orthogonalize the Mr^i's



2) compute the r minimizing solution x^k



Generalized Conjugate Residual Algorithm

First Few Steps

- First search direction $r^0 = b - Mx^0 = b, \quad p_0 = \frac{r^0}{\left\| Mr^0 \right\|}$
- Residual minimizing solution $x^1 = \left(\left(r^0 \right)^T M p_0 \right) p_0$
- Second Search Direction

$$r^1 = b - Mx^1 = r^0 - \gamma_1 M r^0$$

$$p_1 = \frac{r^1 - \beta_{1,0} p_0}{\left\| M \left(r^1 - \beta_{1,0} p_0 \right) \right\|}$$

Generalized Conjugate Residual Algorithm

First few steps

Continued…

- Residual minimizing solution $x^2 = x^1 + \left(\left(r^1 \right)^T M p_1 \right) p_1$

- Third Search Direction

$$r^2 = b - Mx^2 = r^0 - \gamma_{2,1} M r^0 - \gamma_{2,0} M^2 r^0$$

$$p_2 = \frac{r^1 - \beta_{2,0} p_0 - \beta_{2,1} p_1}{\left\| M \left(r^1 - \beta_{2,0} p_0 - \beta_{2,1} p_1 \right) \right\|}$$

Generalized Conjugate Residual Algorithm

The kth step of GCR

$$\tilde{p}_k = r^k - \sum_{j=0}^{k-1} \left(Mr^k\right)^T \left(Mp_j\right) p_j$$

$$p_k = \frac{\tilde{p}_k}{\left\| M\tilde{p}_k \right\|}$$

Orthogonalize and normalize search direction

$$\alpha_k = \left(r^k\right)^T \left(Mp_k\right)$$

Determine optimal stepsize in kth search direction

$$x^{k+1} = x^k + \alpha_k p_k$$

$$r^{k+1} = r^k - \alpha_k M p_k$$

Update the solution and the residual

Generalized Conjugate Residual Algorithm

Polynomial view

If $\alpha_j \neq 0$ for all $j \leq k$ in GCR, then

1) $\text{span}\left\{p_0, p_1, ..., p_k\right\} = \text{span}\left\{r^0, Mr^0, ..., Mr^k\right\}$

2) $x^{k+1} = \xi_k(M)r^0$, ξ_k is the k$^{\text{th}}$ order poly minimizing $\left\|r^{k+1}\right\|_2^2$

3) $r^{k+1} = b - Mx^{k+1} = r^0 - M\xi_k(M)r^0$
$= \left(I - M\xi_k(M)\right)r^0 \equiv \wp_{k+1}\left(M\right)r^0$

where $\wp_{k+1}\left(M\right)r^0$ is the $\left(k+1\right)^{th}$ order poly minimizing $\left\|r^{k+1}\right\|_2^2$ subject to $\wp_{k+1}\left(0\right)=1$

Krylov Methods

Residual Minimization

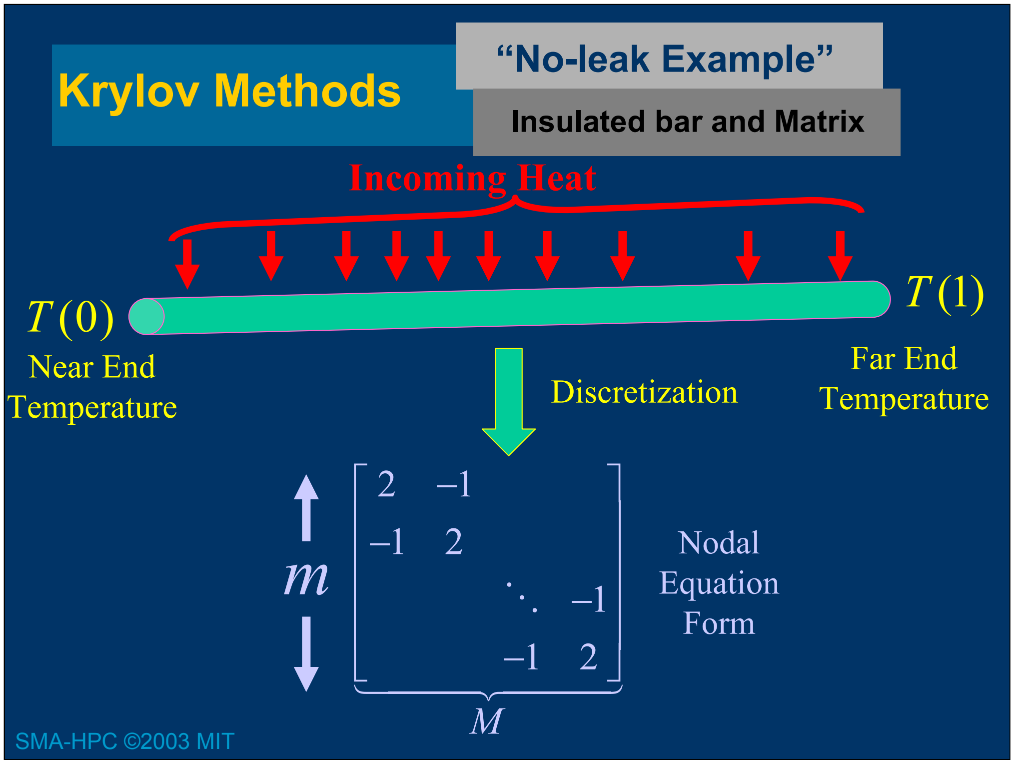
Polynomial View

If $x^{k+1} \in span\left\{r^0, Mr^0, \ldots, Mr^k\right\}$ minimizes $\left\|r^{k+1}\right\|_2^2$

1) $x^{k+1} = \xi_k(M)r^0$, ξ_k is the kth order poly minimizing $\left\|r^{k+1}\right\|_2^2$

2) $r^{k+1} = b - Mx^{k+1} = \left(I - M\xi_k(M)\right)r^0 = \wp_{k+1}\left(M\right)r^0$ where $\wp_{k+1}\left(M\right)r^0$ is the $\left(k+1\right)^{th}$ order poly minimizing $\left\|r^{k+1}\right\|_2^2$ subject to $\wp_{k+1}\left(0\right)=1$

Polynomial Property only a function of solution space and residual minimization

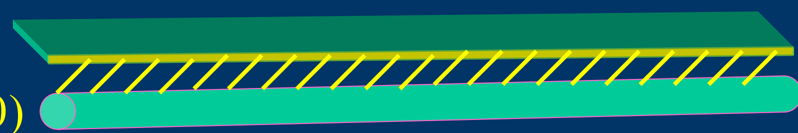

Krylov Methods
"No-leak Example"
Insulated bar and Matrix
Incoming Heat
$T(0)$
Near End Temperature
$T(1)$
Far End Temperature
Discretization
m
$\underbrace{\begin{bmatrix} 2 & -1 & & \\ -1 & 2 & & \\ & & \ddots & -1 \\ & & -1 & 2 \end{bmatrix}}_{M}$
Nodal Equation Form
SMA-HPC ©2003 MIT

Krylov Methods

"leaky" Example

Conducting bar and Matrix

$T(0)$ Near End Temperature



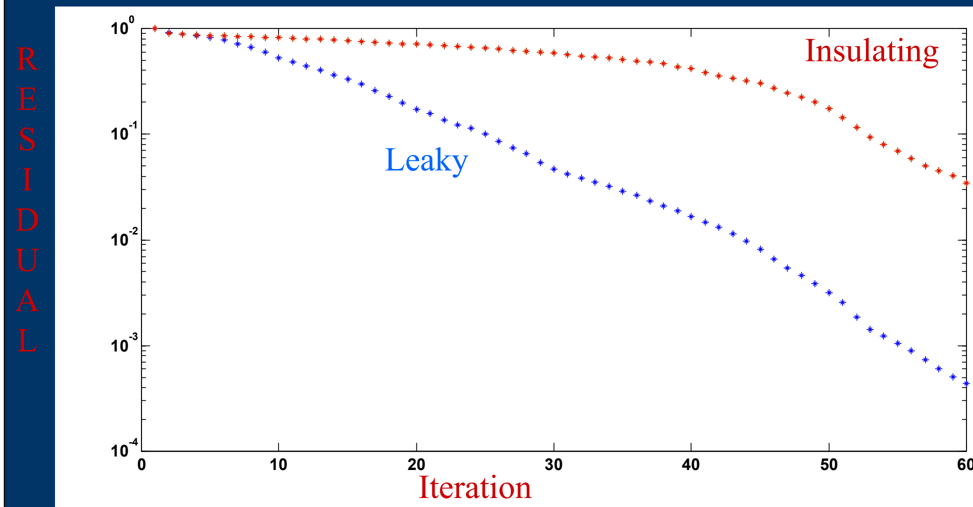
$T(1)$ Far End Temperature

Discretization

$$
m \left\{ \begin{bmatrix} 2.01 & -1 & & \\ -1 & 2.01 & & \\ & & \ddots & -1 \\ & & -1 & 2.01 \end{bmatrix} \right. \quad \text{(}M\text{ columns)}
$$

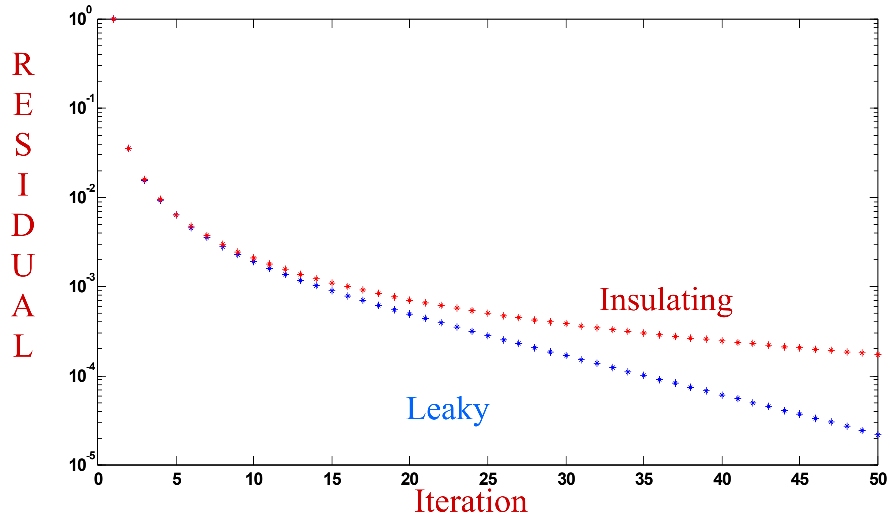
Nodal Equation Form

GCR Performance(Random Rhs)



Plot of *log*(residual) versus Iteration

GCR Performance(Rhs = -1,+1,-1,+1….)



Plot of *log*(residual) versus Iteration

Krylov Methods

Residual Minimization

Optimality of poly

Residual Minimizing Optimality Property

$$\left\| r^{k+1} \right\| \leq \left\| \tilde{\wp}_{k+1}(M) r^0 \right\| \leq \left\| \tilde{\wp}_{k+1}(M) \right\| \left\| r^0 \right\|$$

$\tilde{\wp}_{k+1}$ is any k^{th} order poly such that $\tilde{\wp}_{k+1}(0) = 1$

Therefore

Any polynomial which satisfies the constraints can be used to get an upper bound on $\Rightarrow \dfrac{\left\| r^{k+1} \right\|}{\left\| r^0 \right\|}$

Induced Norms

Matrix Magnification Question

Suppose $y = Mx$

How much larger is y than x?

OR

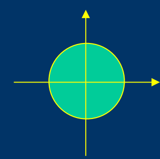
How much does M magnify x?

Induced Norms

Vector Norm Review

L_2 (Euclidean) norm :

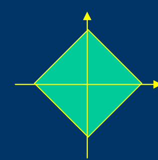
$$\|x\|_2 = \sqrt{\sum_{i=1}^{n} |x_i|^2}$$



$$\|x\|_2 < 1$$

L_1 norm :

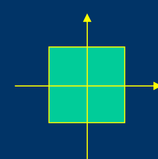
$$\|x\|_1 = \sum_{i=1}^{n} |x_i|$$



$$\|x\|_1 < 1$$

L_∞ norm :

$$\|x\|_\infty = \max_i |x_i|$$



$$\|x\|_\infty < 1$$

Induced Matrix Norms

Standard Induced *l*-norms

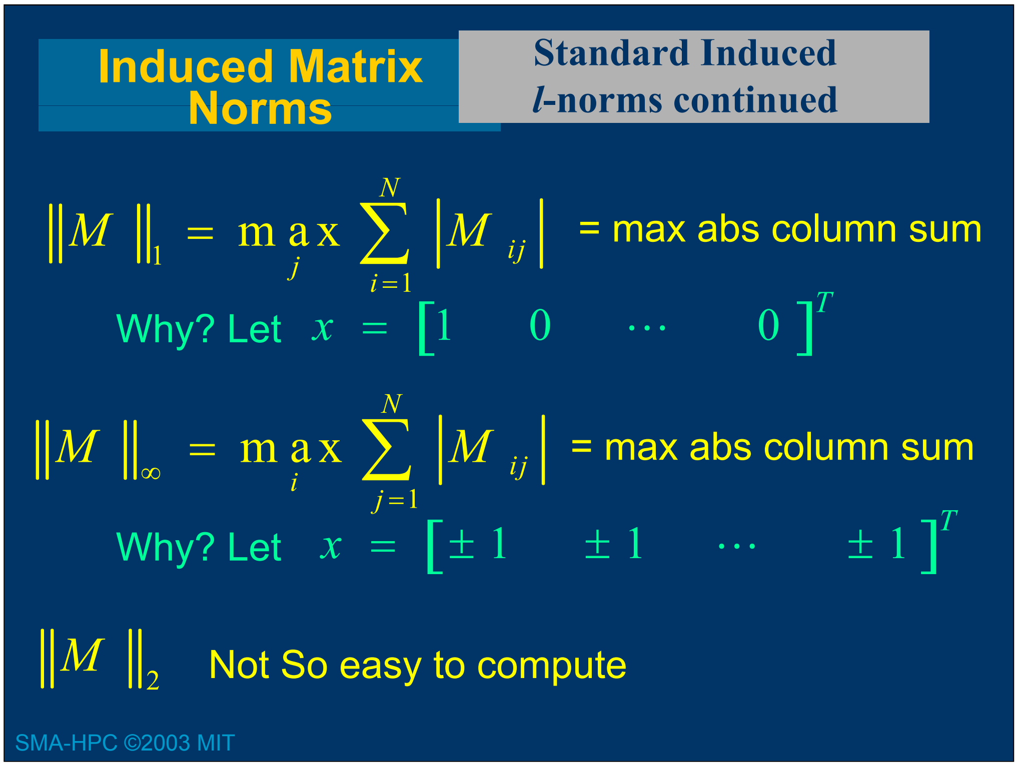
Definition:

$$\|M\|_l \equiv \max_x \frac{\|Mx\|_l}{\|x\|_l} = \max_{\|x\|_l = 1} \|Mx\|_l$$

Examples

$$\|M\|_1 \equiv \max_i \sum_{j=1}^{N} |M_{ij}|$$ Max Column Sum

$$\|M\|_\infty \equiv \max_j \sum_{i=1}^{N} |M_{ij}|$$ Max Row Sum



As the algebra on the slide shows the relative changes in the solution x is bounded by an A-dependent factor times the relative changes in A. The factor

$$\| A^{-1} \| \| A \|$$

was historically referred to as the condition number of A, but that definition has been abandoned as then the condition number is norm-dependent. Instead the condition number of A is the ratio of singular values of A.

$$\text{cond}(A) = \frac{\sigma_{\max}(A)}{\sigma_{\min}(A)}$$

Singular values are outside the scope of this course, consider consulting Trefethen & Bau.

Useful Eigenproperties

Spectral Mapping Theorem

Given a polynomial

$$f(x) = a_0 + a_1 x + \ldots + a_p x^p$$

Apply the polynomial to a matrix

$$f(M) = a_0 + a_1 M + \ldots + a_p M^p$$

Then

$$spectrum(f(M)) = f(spectrum(M))$$

Krylov Methods

Convergence Analysis

Norm of matrix polynomials

$$\left\|\wp_k(M)\right\| = \left\| \underbrace{\begin{bmatrix} \vdots & \cdots & \vdots \\ \vec{u}_1 & \cdots & \vec{u}_N \\ \vdots & \cdots & \vdots \end{bmatrix}}_{\text{eigenvectors of M}} \begin{bmatrix} \wp_k(\lambda_1) & & \\ & \ddots & \\ & & \wp_k(\lambda_N) \end{bmatrix} \begin{bmatrix} \vdots & \cdots & \vdots \\ \vec{u}_1 & \cdots & \vec{u}_N \\ \vdots & \cdots & \vdots \end{bmatrix}^{-1} \right\|$$

$$\leq \underbrace{\text{Cond(U)}}_{\text{condition number of M's eigenspace}} \left\| \begin{bmatrix} \wp_k(\lambda_1) & & \\ & \ddots & \\ & & \wp_k(\lambda_N) \end{bmatrix} \right\|$$

Krylov Methods

Convergence Analysis

Norm of matrix polynomials

$$\left\| \begin{bmatrix} \wp_k(\lambda_1) & & \\ & \ddots & \\ & & \wp_k(\lambda_N) \end{bmatrix} \right\|_2 = \max_{\|x\|=1} \sqrt{\sum_i \left| \wp_k(\lambda_i) x_i \right|^2}$$

$$= \max_i \left| \wp_k(\lambda_i) \right|$$

$$\Rightarrow \left\| \wp_k(M) \right\| \leq cond(V) \max_i \left| \wp_k(\lambda_i) \right|$$

Krylov Methods

Convergence Analysis

Important Observations

1) A residual minimizing Krylov subspace algorithm converges to the exact solution in at most n steps

Proof: Let $\tilde{\wp}_n(x) = (x-\lambda_1)(x-\lambda_2)\ldots(x-\lambda_n)$

where $\lambda_i \in \lambda(M)$. Then, $\max_i \left|\tilde{\wp}_n(\lambda_i)\right| = 0$,

$\Rightarrow \left\|\wp_n(M)\right\| = 0$ and therefore $\left\|r^n\right\| = 0$

2) If M has only q distinct e-values, the residual minimizing Krylov subspace algorithm converges in at most q steps

Proof: Let $\tilde{\wp}_q(x) = (x-\lambda_1)(x-\lambda_2)\ldots(x-\lambda_q)$

Krylov Methods

Convergence for M = M^T

Residual Polynomial

If M = M^T then

1) M has orthonormal eigenvectors

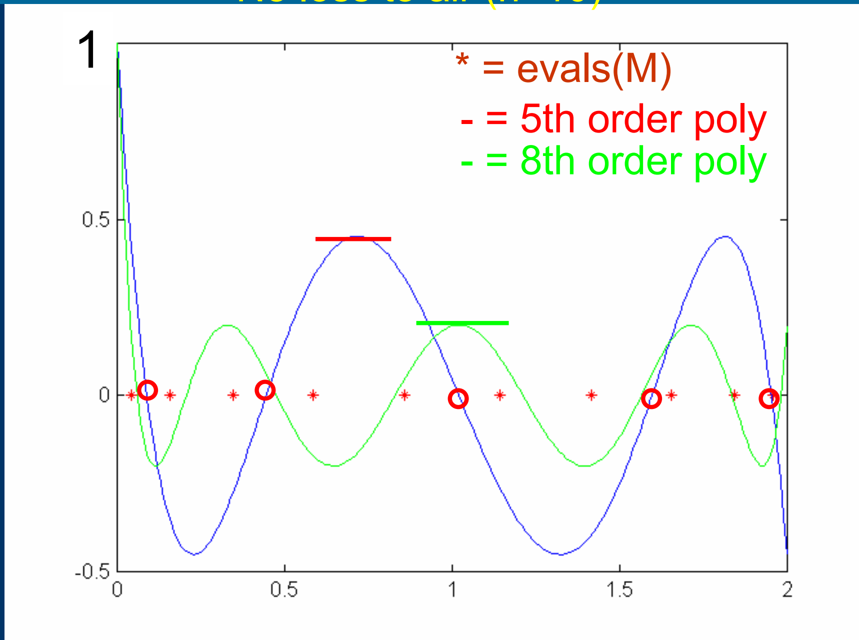
$$\Rightarrow \; cond(V) = \left\| \begin{bmatrix} \vdots & \cdots & \vdots \\ \vec{u}_1 & \cdots & \vec{u}_N \\ \vdots & \cdots & \vdots \end{bmatrix} \right\| \left\| \begin{bmatrix} \vdots & \cdots & \vdots \\ \vec{u}_1 & \cdots & \vec{u}_N \\ \vdots & \cdots & \vdots \end{bmatrix}^{-1} \right\| = 1$$

$$\Rightarrow \; \left\| \wp_k(M) \right\| = \max_i \left| \wp_k(\lambda_i) \right|$$

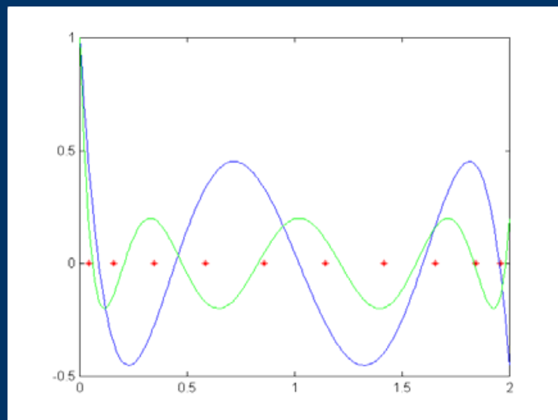
2) M has real eigenvalues

If M is postive definite, then $\lambda(M) > 0$

Residual Poly Picture for Heat Conducting Bar Matrix No loss to air (n=10)



Residual Poly Picture for Heat Conducting Bar Matrix
No loss to air (n=10)



Keep $\left|\wp_k\left(\lambda_i\right)\right|$ as small as possible:

Strategically place zeros of the poly

Convergence for M =M^T

Polynomial Min-Max Problem

Consider $\lambda(M) \in [\lambda_{\min}, \lambda_{\max}],\ \lambda_{\min} > 0$

Then a good polynomial ($\|\tilde{p}_k(M)\|$ is small) can be found by solving the min-max problem

$$\min_{\substack{kth\ order\\ polys\ s.t.\\ \tilde{p}_k(0)=1}} \max_{x \in [\lambda_{\min}, \lambda_{\max}]} \left|\tilde{p}_k(x)\right|$$

The min-max problem is exactly solved by Chebyshev Polynomials

Krylov Methods

Convergence for $M = M^T$

Chebyshev Solves Min-Max

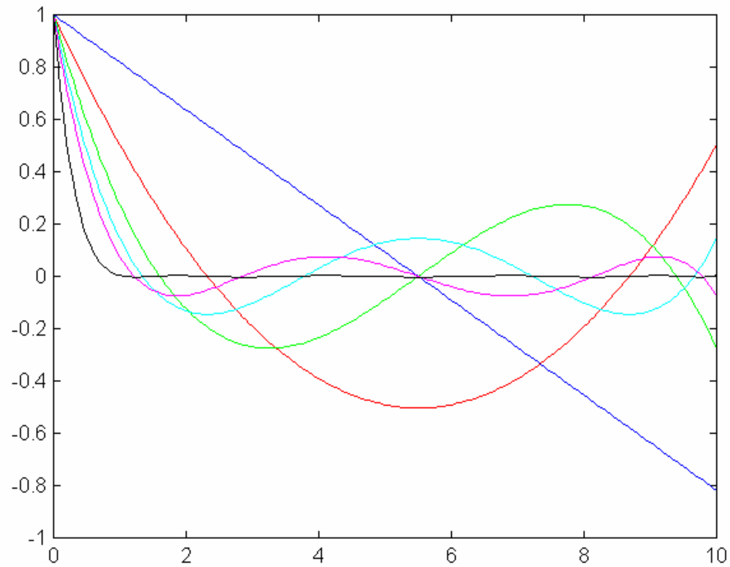
The Chebyshev Polynomial

$$C_k(x) \equiv \cos\left(k \cos^{-1}(x)\right) \quad x \in [-1,1]$$

$$\min_{\substack{kth\ order \\ polys\ s.t. \\ \tilde{\wp}_k(0)=1}} \max_{x \in [\lambda_{\min}, \lambda_{\max}]} \left| \tilde{\wp}_k(x) \right|$$

$$= \max_{x \in [\lambda_{\min}, \lambda_{\max}]} \left| \frac{C_k\left(1 + 2\dfrac{\lambda_{\min} - x}{\lambda_{\max} - \lambda_{\min}}\right)}{C_k\left(1 + 2\dfrac{\lambda_{\min}}{\lambda_{\max} - \lambda_{\min}}\right)} \right|$$

Chebyshev Polynomials minimizing over [1,10]



Krylov Methods

Convergence for $M = M^T$

Chebychev Bounds

$$\min_{\substack{kth\ order \\ polys\ s.t. \\ \wp_k(0)=1}} \max_{x \in [\lambda_{\min}, \lambda_{\max}]} \left| \tilde{\wp}_k(x) \right|$$

$$= \frac{1}{\left| C_k\left(1 - 2\frac{\lambda_{\max}}{\lambda_{\max} - \lambda_{\min}}\right) \right|}$$

$$\leq 2\left(\frac{\sqrt{\frac{\lambda_{\max}}{\lambda_{\min}}} - 1}{\sqrt{\frac{\lambda_{\max}}{\lambda_{\min}}} + 1} \right)^k$$

Krylov Methods

Convergence for $M = M^T$

Chebychev Result

If $\lambda(M) \in [\lambda_{min}, \lambda_{max}],\ \lambda_{min} > 0$

$$\left\| r^k \right\| \leq 2 \left(\frac{\sqrt{\frac{\lambda_{max}}{\lambda_{min}}} - 1}{\sqrt{\frac{\lambda_{max}}{\lambda_{min}}} + 1} \right)^k \left\| r^0 \right\|$$

$$\begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 \end{bmatrix} \qquad \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & N \end{bmatrix}$$

For which problem will GCR Converge Faster?

Krylov Methods

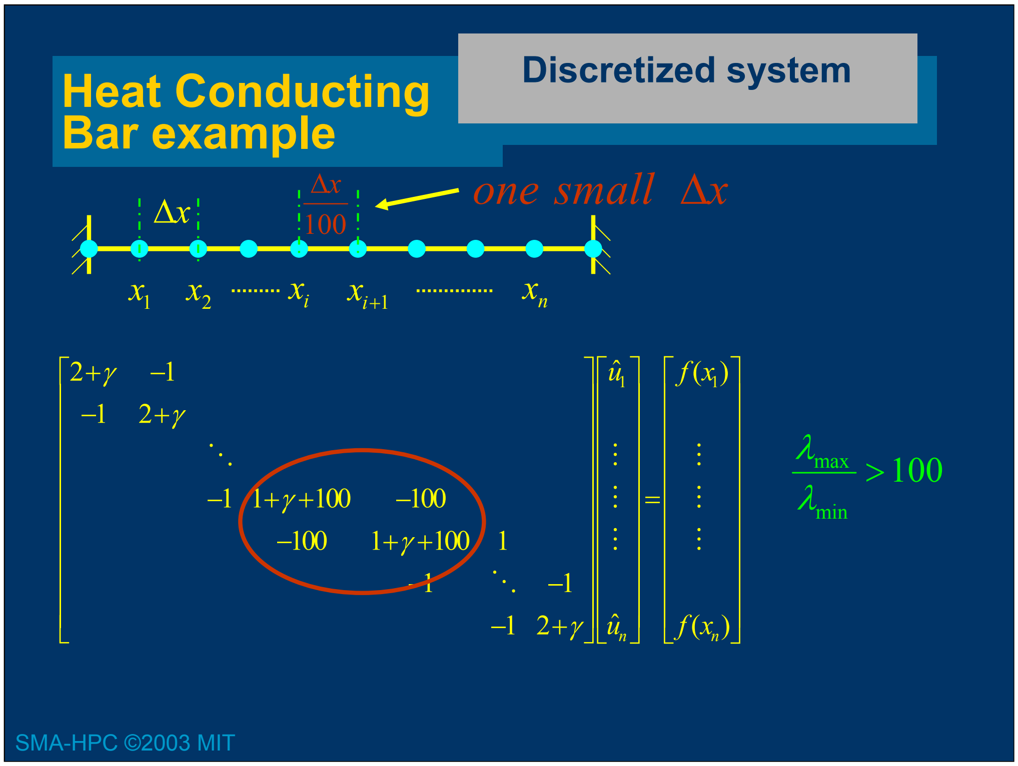
Preconditioning

Diagonal Preconditioners

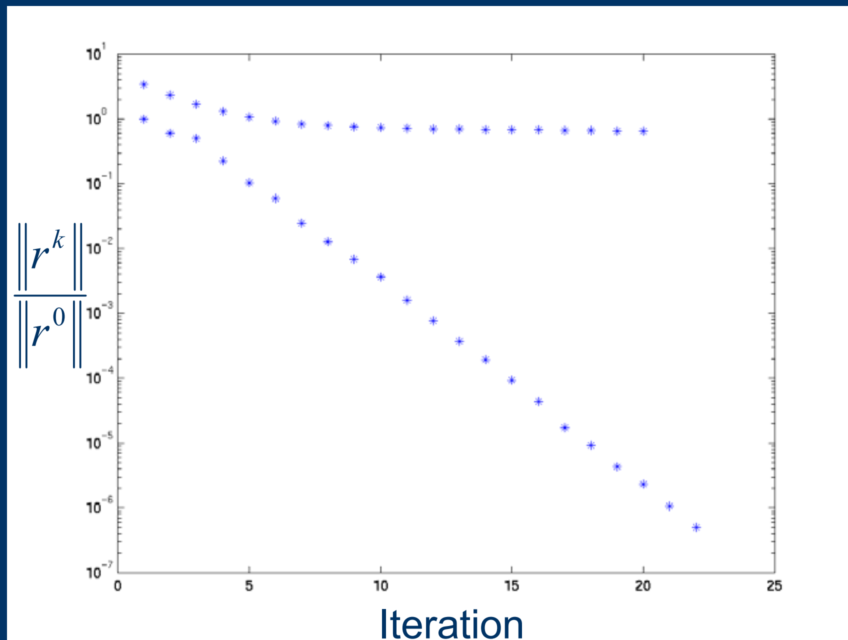
$$\text{Let } M = D + M_{nd}$$

Apply GCR to $\left(D^{-1}M\right)x = \left(I + D^{-1}M_{nd}\right)x = D^{-1}b$

- The Inverse of a diagonal is cheap to compute
- Usually improves convergence

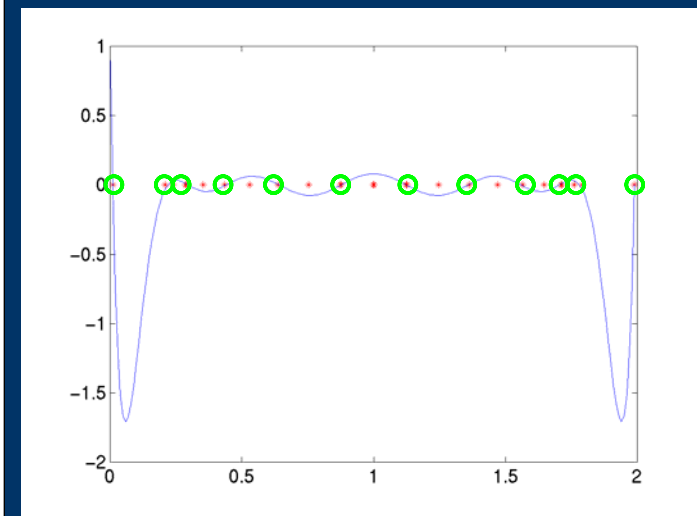

Heat Conducting Bar example
Discretized system
one small Δx
Δx
$\frac{\Delta x}{100}$
x_1 x_2 .... x_i x_{i+1} .... x_n
$$\begin{bmatrix} 2+\gamma & -1 & & & & & \\ -1 & 2+\gamma & & & & & \\ & & \ddots & & & & \\ & & -1 & 1+\gamma+100 & -100 & & \\ & & & -100 & 1+\gamma+100 & 1 & \\ & & & & -1 & \ddots & -1 \\ & & & & & -1 & 2+\gamma \end{bmatrix} \begin{bmatrix} \hat{u}_1 \\ \vdots \\ \vdots \\ \vdots \\ \hat{u}_n \end{bmatrix} = \begin{bmatrix} f(x_1) \\ \vdots \\ \vdots \\ \vdots \\ f(x_n) \end{bmatrix}$$
$\frac{\lambda_{max}}{\lambda_{min}} > 100$
SMA-HPC ©2003 MIT

Which Convergence Curve is GCR?



Heat Conducting Bar example

Preconditioned Matrix Eigenvalues



Residual Minimizing Krylov-subspace Algorithm can eliminate outlying eigenvalues by placing polynomial zeros directly on them.

The World According to Krylov

Heat Flow Comparison Example

Dimension	Dense GE	Sparse GE	GCR
1	$O(m^3)$	$O(m)$	$O(m^2)$
2	$O(m^6)$	$O(m^3)$	$O(m^3)$
3	$O(m^9)$	$O(m^6)$	$O(m^4)$

GCR faster than banded GE in 2 and 3 dimensions

Could be faster, 3-D matrix only m^3 nonzeros.

GCR converges too slowly!

Krylov Methods

Preconditioning

Approximate LU Preconditioners

$$\text{Let } M \approx \tilde{L}\,\tilde{U}$$

Applying GCR to $\left(\left(\tilde{L}\tilde{U}\right)^{-1} M\right) x = \left(\tilde{L}\tilde{U}\right)^{-1} b$

Use an Implicit matrix representation!

Forming $y = \left(\left(\tilde{L}\tilde{U}\right)^{-1} M\right) x$ is equivalent to

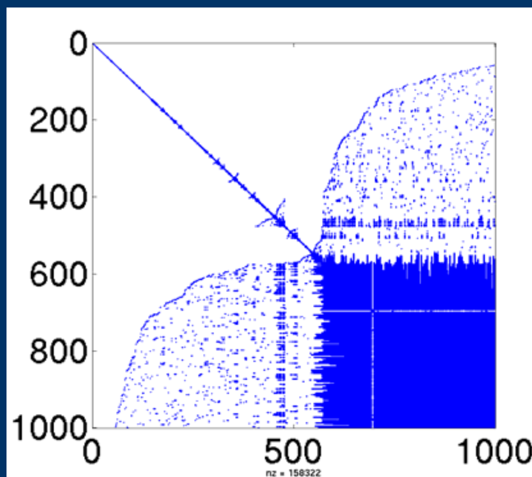
solving $\tilde{L}\tilde{U} y = Mx$

Krylov Methods

Preconditioning

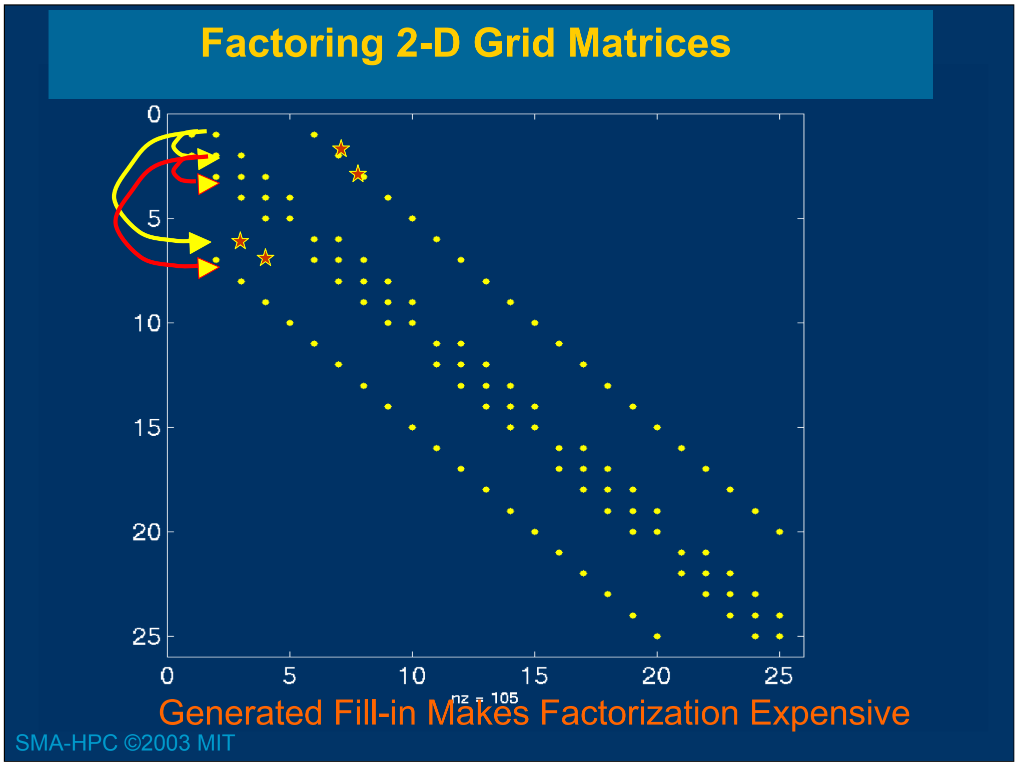
Approximate LU Preconditioners Continued

Nonzeros in an exact LU Factorization



Filled-in LU factorization
Too expensive.

Ignore the fillin!



Factoring 2-D Grid Matrices
nz = 105
Generated Fill-in Makes Factorization Expensive
SMA-HPC ©2003 MIT

Krylov Methods

Preconditioning

Approximate LU Preconditioners Continued

THROW AWAY FILL-INS!

Throw away all fill-ins

Throw away only fill-ins with small values

Throw away fill-ins produced by other fill-ins

Throw away fill-ins produced by fill-ins of other fill-ins, etc.

Summary

- Reminder about GCR
  - Residual minimizing solution
  - Krylov Subspace
  - Polynomial Connection
- Review Norms and Eigenvalues
  - Induced Norms
  - Spectral mapping theorem
- Estimating Convergence Rate
  - Chebychev Polynomials
- Preconditioners
  - Diagonal Preconditioners
  - Approximate LU preconditioners