A Tuned Preconditioner for Inexact Inverse Iteration

Alastair Spence

Department of Mathematical Sciences
University of Bath, United Kingdom

June 13, 2006
Prague

Joint work with: Melina Freitag (Bath) and Eero Vainikko (Tartu)

Outline

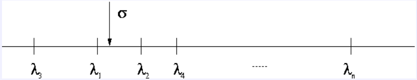
Motivation

- $Ax = \lambda Mx$
- simple (λ_1, x_1): $Ax_1 = \lambda_1 M x_1$
- Large sparse nonsymmetric matrices
- Stability calculations for linearised N-S using Mixed FEM
- Hopf bifurcation: λ complex
- Jacobi-Davidson, Arnoldi,...
- Inverse Iteration with iterative solves for shifted linear systems
- (a)costs of system solves (b)theory for symmetric problems

Outline

Inexact inverse iteration (an inner-outer iteration)

- $Ax = \lambda Mx, \quad (A - \sigma M)^{-1} Mx = \frac{1}{\lambda - \sigma} x$
- Fixed shift, $x^{(0)}$, $c^H x^{(0)} = 1$



for $i = 1$ to $\ldots$ **do**

choose $\tau^{(i)}$

solve

$$\|(A - \sigma M) y^{(i)} - M x^{(i)}\| \leq \tau^{(i)},$$

update eigenvector $x^{(i+1)} = \dfrac{y^{(i)}}{c^H y^{(i)}}$,

update eigenval $\lambda^{(i+1)} =$Ray. Quot.

e-value residual $r^{(i+1)} = (A - \lambda^{(i+1)} M) x^{(i+1)}$.

end for

PDE Example

Consider

$$-\Delta u + 5u_x + 5u_y = \lambda u$$

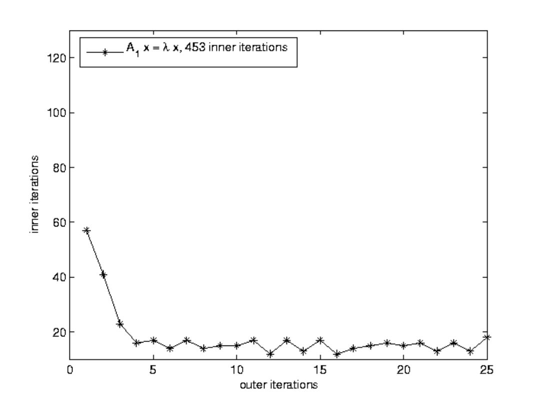
- Finite Difference Discretisation: $A_1 x = \lambda x$
- Finite Element Discretisation: $A_2 x = \lambda M_2 x$

Apply inexact inverse iteration with fixed shift σ and decreasing tolerance:

$$(A_1 - \sigma I) y^{(i)} = x^{(i)}, \quad (A_2 - \sigma M_2) y^{(i)} = M_2 x^{(i)}.$$

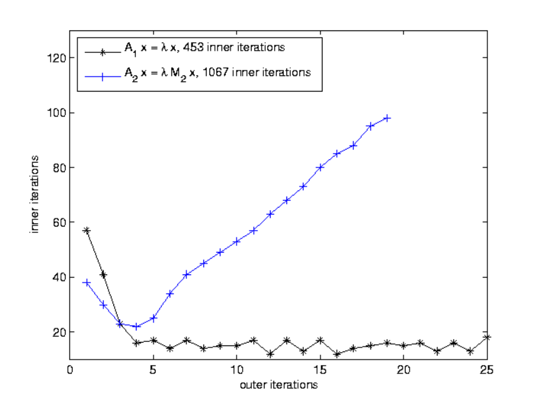
PDE Example: Numerics

Inner v. outer iterations



PDE Example: Numerics

Inner v. outer iterations



Questions

For decreasing solve tolerances

- Since the linear systems are being solved more and more accurately, why isn't the # inner iterations increasing with i for $A_1x = \lambda x$?
- Why is the inner iteration behaviour different for the two discretizations?
- Can we achieve no increase in # inner iterations for $A_2x = \lambda M_2x$? (Yes: 'tuning')

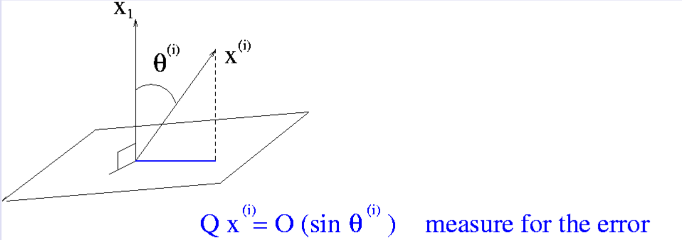
Questions

For decreasing solve tolerances

- Since the linear systems are being solved more and more accurately, why isn't the # inner iterations increasing with i for $A_1 x = \lambda x$?
- Why is the inner iteration behaviour different for the two discretizations?
- Can we achieve no increase in # inner iterations for $A_2 x = \lambda M_2 x$? (Yes: 'tuning')
- What implications are there for preconditioned iterative solvers?
- What can we prove for the symmetric case?

Convergence Analysis

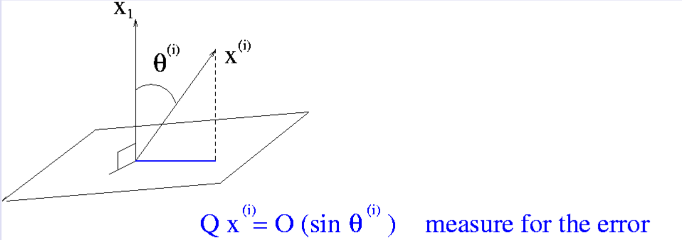
$Ax_1 = \lambda_1 x_1$, A symmetric (see Parlett's book)



- $C|\sin\theta^{(i)}| \leq \|r^{(i)}\| \leq C'|\sin\theta^{(i)}|, \quad r^{(i)} = (A - \lambda^{(i)} M)x^{(i)}$

Convergence Analysis

$Ax_1 = \lambda_1 x_1$, A symmetric (see Parlett's book)



- $C|\sin\theta^{(i)}| \leq \|r^{(i)}\| \leq C'|\sin\theta^{(i)}|, \quad r^{(i)} = (A - \lambda^{(i)}M)x^{(i)}$
- Nonsymmetric: $Ax = \lambda Mx$ [Lai/Lin/Wen-Wei (1997); Golub/Ye (2000); Berns-Müller/Sp (2006)]
- If $\tau^{(i)} = C\|r^{(i)}\|$ then LINEAR convergence

Outline

Krylov solvers for $By = b$

- $B = (A - \sigma I)$ and $b = x$
- later $B = (A - \sigma M)$ and $b = Mx$

Krylov solvers for $By = b$

- $B = (A - \sigma I)$ and $b = x$
- later $B = (A - \sigma M)$ and $b = Mx$
- $\|b - By_k\| = \min\|p_k(B)b\| \leq C\rho^k\|b\|, \quad (0 < \rho < 1).$

Krylov solvers for $By = b$

- $B = (A - \sigma I)$ and $b = x$
- later $B = (A - \sigma M)$ and $b = Mx$
- $\|b - By_k\| = \min\|p_k(B)b\| \leq C\rho^k\|b\|, \quad (0 < \rho < 1).$
- If $\|b - By_k\| \leq \tau$ then

$$k \geq C_1 + C_2 \log \frac{\|b\|}{\tau}$$

- Bound on k increases as τ decreases

Krylov solvers for $By = b$: sophisticated analysis

- for a well-separated eigenvalue
- For $B = (A - \sigma I)$ then $Bx_1 = (\lambda_1 - \sigma)x_1$
-
$$\|b - By_k\| = \min\|p_k(B)b\| \leq \|p_{k-1}(B)q_1(B)b\|$$

Krylov solvers for $By = b$: sophisticated analysis

- for a well-separated eigenvalue
- For $B = (A - \sigma I)$ then $Bx_1 = (\lambda_1 - \sigma)x_1$
- $$\|b - By_k\| = \min\|p_k(B)b\| \leq \|p_{k-1}(B)q_1(B)b\|$$
- $$\|b - By_k\| \leq \|p_{k-1}(B)q_1(B)Qb\|$$

Krylov solvers for $By = b$: sophisticated analysis

- for a well-separated eigenvalue
- For $B = (A - \sigma I)$ then $Bx_1 = (\lambda_1 - \sigma)x_1$
- $$\|b - By_k\| = \min\|p_k(B)b\| \leq \|p_{k-1}(B)q_1(B)b\|$$
- $$\|b - By_k\| \leq \|p_{k-1}(B)q_1(B)Qb\|$$
- to achieve $\|b - By_k\| \leq \tau$ then
$$k \geq C_3 + C_4 \log \frac{\|Qb\|}{\tau}$$
- Bound on k depends on $\frac{\|Qb\|}{\tau}$

Some answers for $\tau^{(i)} = C\|r^{(i)}\| = \mathcal{O}(\sin\theta^{(i)})$

$(A - \sigma I)y^{(i)} = x^{(i)}$

- $b = x^{(i)}$ and $\|Qx^{(i)}\| = \mathcal{O}(\sin\theta^{(i)})$
- Hence $\|Qb\|/\tau = \mathcal{O}(1)$ and the bound on k doesn't increase
- Consistent with numerics for $A_1 x = \lambda x$

Some answers for $\tau^{(i)} = C\|r^{(i)}\| = \mathcal{O}(\sin\theta^{(i)})$

$(A - \sigma I)y^{(i)} = x^{(i)}$

- $b = x^{(i)}$ and $\|Qx^{(i)}\| = \mathcal{O}(\sin\theta^{(i)})$
- Hence $\|Qb\|/\tau = \mathcal{O}(1)$ and the bound on k doesn't increase
- Consistent with numerics for $A_1x = \lambda x$

$(A - \sigma M)y^{(i)} = Mx^{(i)}$

- $b = Mx^{(i)}$ and $\|QMx^{(i)}\| = \mathcal{O}(1)$
- Hence $\|Qb\|/\tau = \mathcal{O}(\sin\theta^{(i)})^{-1}$ and the bound on k increases
- Consistent with numerics for $A_2x = \lambda M_2x$

Outline

For $(A - \sigma M)y^{(i)} = Mx^{(i)}$

Introduce the "tuning matrix" $\mathbb{T}$ and consider (think preconditioning)

$$\mathbb{T}^{-1}(A - \sigma M)y^{(i)} = \mathbb{T}^{-1}Mx^{(i)}$$

For $(A - \sigma M)y^{(i)} = Mx^{(i)}$

Introduce the "tuning matrix" $\mathbb{T}$ and consider (think preconditioning)

$$\mathbb{T}^{-1}(A - \sigma M)y^{(i)} = \mathbb{T}^{-1}Mx^{(i)}$$

- Key Condition: $\mathbb{T}^{-1}Mx^{(i)} = x^{(i)}$

For $(A - \sigma M)y^{(i)} = Mx^{(i)}$

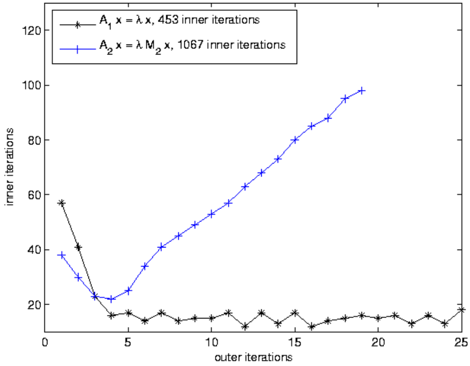
Introduce the "tuning matrix" $\mathbb{T}$ and consider (think preconditioning)

$$\mathbb{T}^{-1}(A - \sigma M)y^{(i)} = \mathbb{T}^{-1}Mx^{(i)}$$

- Key Condition: $\mathbb{T}^{-1}Mx^{(i)} = x^{(i)}$
- Re-arrange to:
$$Mx^{(i)} = \mathbb{T}x^{(i)}$$
- Implement by rank-one change to Identity:
$$\mathbb{T} := I + (Mx^{(i)} - x^{(i)})c^H \quad (c^H x^{(i)} = 1)$$
- So
$$\mathbb{T}x^{(i)} = x^{(i)} + (Mx^{(i)} - x^{(i)})c^H x^{(i)} = Mx^{(i)}$$
- Use Sherman-Morrison to get action of $\mathbb{T}^{-1}$

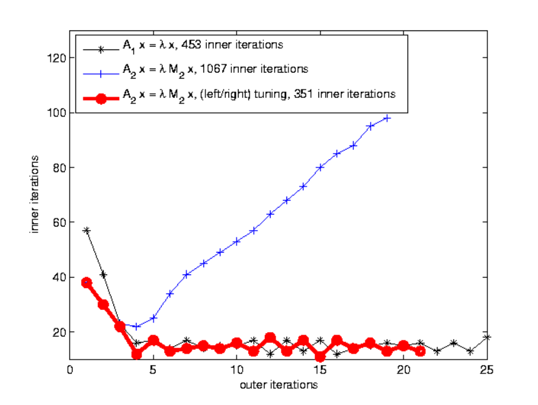
PDE Example: Numerics

Inner v. outer iterations



PDE Example: Numerics

Inner v. outer iterations



Outline

Flow past a circular cylinder (incompressible Navier-Stokes)

- $Re = 25$, $\lambda \approx \pm 10i$
- Mixed FEM: $Q_2 - Q_1$ elements
- Elman preconditioner: 2-level additive Schwarz
- ≈ 54000 degrees of freedom

Flow past a circular cylinder (incompressible Navier-Stokes)

Figure: Fixed Shift

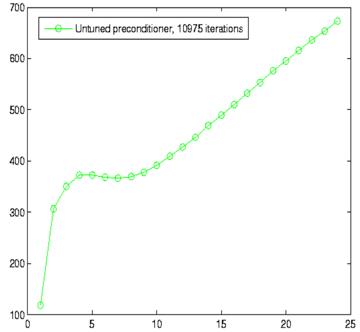
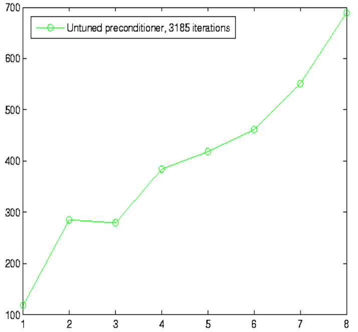


Figure: Rayleigh Quotient Shift



Tuning the preconditioner

- Preconditioned iterates should behave as in unpreconditioned standard EVP case
- right preconditioned system $(A - \sigma M)P_S^{-1}\tilde{y}^{(i)} = Mx^{(i)}$
- theory of Krylov solver for $By = b$ indicates that $Mx^{(i)}$ should be close to eigenvector of $(A - \sigma M)P_S^{-1}$

Tuning the preconditioner

- Preconditioned iterates should behave as in unpreconditioned standard EVP case
- right preconditioned system $(A - \sigma M)P_S^{-1}\tilde{y}^{(i)} = Mx^{(i)}$
- theory of Krylov solver for $By = b$ indicates that $Mx^{(i)}$ should be close to eigenvector of $(A - \sigma M)P_S^{-1}$
- Introduce a tuned preconditioner $\mathbb{P}$ so that we solve

$$(A - \sigma M)\mathbb{P}^{-1}\tilde{y}^{(i)} = Mx^{(i)}$$

Tuning the preconditioner

- Preconditioned iterates should behave as in unpreconditioned standard EVP case
- right preconditioned system $(A - \sigma M)P_S^{-1}\tilde{y}^{(i)} = Mx^{(i)}$
- theory of Krylov solver for $By = b$ indicates that $Mx^{(i)}$ should be close to eigenvector of $(A - \sigma M)P_S^{-1}$
- Introduce a tuned preconditioner $\mathbb{P}$ so that we solve

$$(A - \sigma M)\mathbb{P}^{-1}\tilde{y}^{(i)} = Mx^{(i)}$$

- Remember $Ax_1 = \lambda_1 Mx_1$, so condition $\mathbb{P}x^{(i)} \approx \lambda^{(i)} Mx^{(i)}$?
- Take

$$\mathbb{P}x^{(i)} = Ax^{(i)}$$

since then

$$\mathbb{P}x^{(i)} = \lambda^{(i)} Mx^{(i)} + (Ax^{(i)} - \lambda^{(i)} Mx^{(i)})$$

Implementation of $\boxed{\mathbb{P}x^{(i)} = Ax^{(i)}}$

- Given P_S
- Evaluate $u^{(i)} = Ax^{(i)} - P_S x^{(i)}$

Implementation of $\boxed{\mathbb{P}x^{(i)} = Ax^{(i)}}$

- Given P_S
- Evaluate $u^{(i)} = Ax^{(i)} - P_S x^{(i)}$
- Rank-one update: $\boxed{\mathbb{P} = P_S + u^{(i)} c^H}$

$$\mathbb{P}x^{(i)} = P_S x^{(i)} + u^{(i)} c^H x^{(i)} = P_S x^{(i)} + u^{(i)} = Ax^{(i)}$$

- Use Sherman-Morrison - one extra backsolve per outer iteration

Example

Figure: Fixed Shift

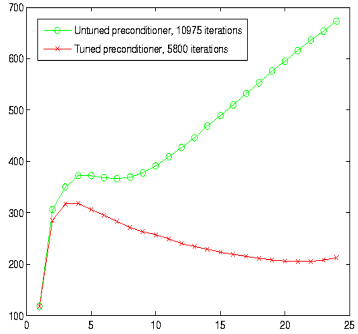
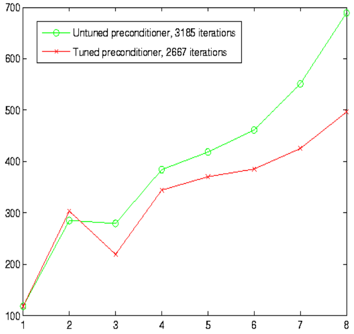


Figure: Rayleigh Quotient Shift



Outline

Convergence rates: symmetric problems - variable shifts

Decreasing tolerance $\tau^{(i)} = C\|r^{(i)}\| = \mathcal{O}(\sin\theta^{(i)})$

1. Rayleigh quotient shift $\sigma^{(i)} = \rho(x^{(i)}) = \dfrac{{x^{(i)}}^T A x^{(i)}}{{x^{(i)}}^T x^{(i)}}$: cubic convergence

Fixed tolerance $\tau^{(i)} = \tau$

1. convergence still possible
2. Rayleigh quotient shift: quadratic convergence

Unpreconditioned MINRES $(A - \sigma I)y = x$

Number of inner solves for each i for $\|x^{(i)} - (A - \sigma^{(i)} I)y^{(i)}\| \leq \tau^{(i)}$

1. $$k \geq C_1 + C_2 \log \frac{\|\mathcal{Q}x^{(i)}\|_2}{|\lambda_1 - \sigma^{(i)}|\tau^{(i)}}$$
2. Interplay between the shift, solve tolerance and right hand side

Preconditioning

Incomplete Cholesky preconditioning

$$A = LL^T + E$$

symmetric preconditioning of $(A - \sigma I)y^{(i)} = x^{(i)}$:

$$L^{-1}(A - \sigma I)L^{-T}\tilde{y}^{(i)} = L^{-1}x^{(i)}, \quad y^{(i)} = L^{-T}\tilde{y}^{(i)}$$

Preconditioning

Incomplete Cholesky preconditioning

$$A = LL^T + E$$

symmetric preconditioning of $(A - \sigma I)y^{(i)} = x^{(i)}$:

$$L^{-1}(A - \sigma I)L^{-T}\tilde{y}^{(i)} = L^{-1}x^{(i)}, \quad y^{(i)} = L^{-T}\tilde{y}^{(i)}$$

Remarks

1. $L^{-1}x^{(i)}$ is not a “good” rhs for $L^{-1}(A - \sigma I)L^{-T}$
2. k increases with i for $\tau^{(i)} = C\|r^{(i)}\|$ even for fixed shift.

Tuning the Preconditioner

1. modify $L \rightarrow \mathbb{L}$

$$\mathbb{L}^{-1}(A - \sigma I)\mathbb{L}^{-T}\tilde{y}^{(i)} = \mathbb{L}^{-1}x^{(i)}, \quad y^{(i)} = \mathbb{L}^{-T}\tilde{y}^{(i)}$$

Tuning the Preconditioner

1. modify $L \rightarrow \mathbb{L}$

$$\mathbb{L}^{-1}(A - \sigma I)\mathbb{L}^{-T}\tilde{y}^{(i)} = \mathbb{L}^{-1}x^{(i)}, \quad y^{(i)} = \mathbb{L}^{-T}\tilde{y}^{(i)}$$

2. minor extra computation cost for $\mathbb{L}$
3. "nice" right hand side $\mathbb{L}^{-1}x^{(i)}$

Tuning the Preconditioner

1. modify $L \to \mathbb{L}$

$$\mathbb{L}^{-1}(A - \sigma I)\mathbb{L}^{-T}\tilde{y}^{(i)} = \mathbb{L}^{-1}x^{(i)}, \quad y^{(i)} = \mathbb{L}^{-T}\tilde{y}^{(i)}$$

2. minor extra computation cost for $\mathbb{L}$
3. "nice" right hand side $\mathbb{L}^{-1}x^{(i)}$
4. Ask that

$$\mathbb{L}\mathbb{L}^{T}x^{(i)} = Ax^{(i)}$$

Tuning the Preconditioner

1. modify $L \to \mathbb{L}$

$$\mathbb{L}^{-1}(A-\sigma I)\mathbb{L}^{-T}\tilde{y}^{(i)} = \mathbb{L}^{-1}x^{(i)}, \quad y^{(i)} = \mathbb{L}^{-T}\tilde{y}^{(i)}$$

2. minor extra computation cost for $\mathbb{L}$
3. "nice" right hand side $\mathbb{L}^{-1}x^{(i)}$
4. Ask that

$$\mathbb{L}\mathbb{L}^{T}x^{(i)} = Ax^{(i)}$$

5.

$$\mathbb{L}^{-1}(A-\sigma I)\mathbb{L}^{-T}\mathbb{L}^{-1}x^{(i)} = \frac{\lambda_1-\sigma}{\lambda_1}\mathbb{L}^{-1}x^{(i)} + C\|r^{(i)}\|$$

How do we achieve $\mathbb{L}\mathbb{L}^T x^{(i)} = Ax^{(i)}$?

1. $x^{(i)}$ is current eigenvector approximation
2. $u^{(i)} = Ax^{(i)} - LL^T x^{(i)}$ (known)
3. $\mathbb{L}$ chosen such that

$$\mathbb{L} = L + \alpha^{(i)} u^{(i)} (L^{-1} u^{(i)})^T$$

with $\alpha^{(i)}$ root of quadratic function

How do we achieve $\mathbb{L}\mathbb{L}^T x^{(i)} = Ax^{(i)}$?

1. $x^{(i)}$ is current eigenvector approximation
2. $u^{(i)} = Ax^{(i)} - LL^T x^{(i)}$ (known)
3. $\mathbb{L}$ chosen such that

$$\mathbb{L} = L + \alpha^{(i)} u^{(i)} (L^{-1} u^{(i)})^T$$

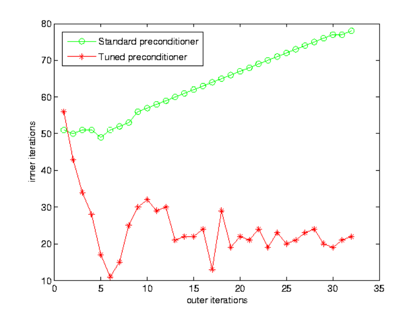
   with $\alpha^{(i)}$ root of quadratic function
4. $\mathbb{L}\mathbb{L}^T x^{(i)} = Ax^{(i)}$.
5. Note: $\mathbb{L}\mathbb{L}^T = LL^T + \frac{1}{{e^{(i)}}^T x^{(i)}} u^{(i)} {u^{(i)}}^T$
6. $\mathbb{L}$ is a rank-one update of L, and $\mathbb{L}\mathbb{L}^T$ is a rank-one update of LL^T.

Example

- SPD matrix from the Matrix Market library (nos5: 3 story building with attached tower)
- seek eigenvalue near fixed shift $\sigma = 100$
- $A \approx LL^T$, incomplete Cholesky factorisation (drop tol. $= 0.1$)
- compare standard and tuned preconditioner

Fixed shift solves

Preconditioning with standard incomplete Cholesky



- total number of inner iterations using standard preconditioner: 2026
- total number of inner iterations using tuned preconditioner: 779

Outline

Comparison of LL^T with $\mathbb{L}\mathbb{L}^T$

Spectral properties of preconditioned matrix

Let

$$L^{-1}(A - \sigma I)L^{-T}w = \mu w$$

$$\mathbb{L}^{-1}(A - \sigma I)\mathbb{L}^{-T}\hat{w} = \xi \hat{w}$$

Theorem - no (i)

If $\sigma \notin \Lambda(A)$ then $\mu, \xi \neq 0$ and

$$\min_{\mu \in \Lambda(L^{-1}(A-\sigma I)L^{-T})} \left| \frac{\mu - \xi}{\xi} \right| \leq |\gamma v^T v|,$$

where $\gamma = 1/(u^T x)$ and $v = L^{-1}u$.

Comparison of LL^T with $\mathbb{L}\mathbb{L}^T$

Spectral properties of preconditioned matrix

Let

$$L^{-1}(A - \sigma I)L^{-T} w = \mu w$$

$$\mathbb{L}^{-1}(A - \sigma I)\mathbb{L}^{-T} \hat{w} = \xi \hat{w}$$

Interlacing property (Golub/Van Loan)

Compare

$$Ds = \mu s$$

with

$$Dt = \xi(I + \gamma z z^T)t$$

Comparison of LL^T with $\mathbb{L}\mathbb{L}^T$

Spectral properties of preconditioned matrix

Let

$$L^{-1}(A - \sigma I)L^{-T}w = \mu w$$

$$\mathbb{L}^{-1}(A - \sigma I)\mathbb{L}^{-T}\hat{w} = \xi \hat{w}$$

Interlacing property (Golub/Van Loan)

Compare

$$Ds = \mu s$$

with

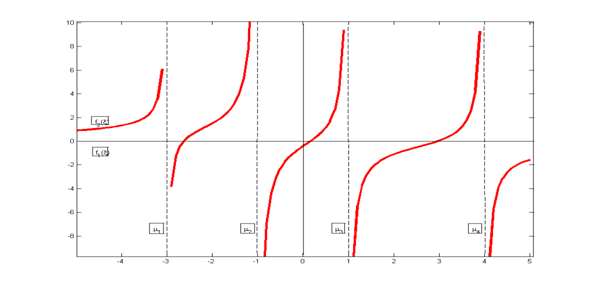
$$Dt = \xi(I + \gamma z z^T)t$$

Interlacing property

- If $\gamma > 0$ eigenvalues are moved towards the origin.
- If $\gamma < 0$ eigenvalues are moved away from the origin.

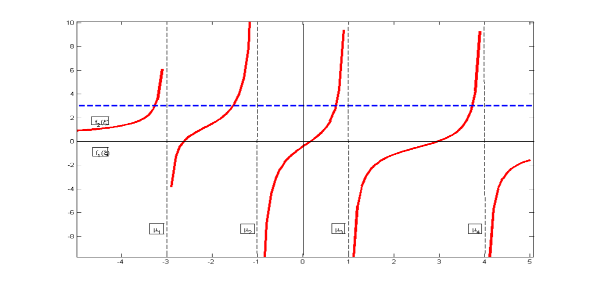
Comparison of LL^T with $\mathbb{L}\mathbb{L}^T$

Rank-one perturbation of a symmetric matrix



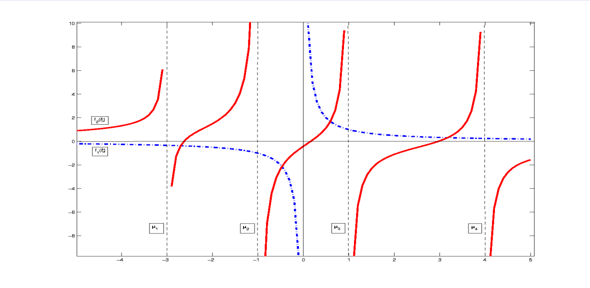
Comparison of LL^T with $\mathbb{L}\mathbb{L}^T$

Rank-one perturbation of a symmetric matrix



Comparison of LL^T with $\mathbb{L}\mathbb{L}^T$

Interlacing property



- μ and ξ interlace each other depending on the sign of γ
- Clustering properties are preserved
- reduced condition number $\kappa_L^1 \leq \kappa_{\mathbb{L}}^1 \leq \kappa_L^1(1 + \gamma v^T v)$

Another approach: Changing the right hand side

Approach by Simoncini/Eldén

Instead of solving

$$L^{-1}(A-\sigma^{(i)}I)L^{-T}\tilde{y}^{(i)} = L^{-1}x^{(i)}, \quad y^{(i)} = \mathbb{L}^{-T}\tilde{y}^{(i)}$$

change the right hand side

$$L^{-1}(A-\sigma^{(i)}I)L^{-T}\tilde{y}^{(i)} = L^{T}x^{(i)}, \quad y^{(i)} = L^{-T}\tilde{y}^{(i)}$$

Comparison

Tuned preconditioner and Simoncini & Eldén approach

Example `nos5.mtx` from Matrix Market. Solves to fixed tolerance $\tau = 0.01$. Rayleigh quotient shift. Quadratic convergence for both methods.

	Simoncini & Eldén		*Tuned preconditioner*	
	Drop Tolerances			
Outer Iteration	0.25	0.1	0.25	0.1
1	67	62	29	26
2	74	66	56	55
3	85	75	71	67
4	63		18	
total	289	203	174	148

Conclusion

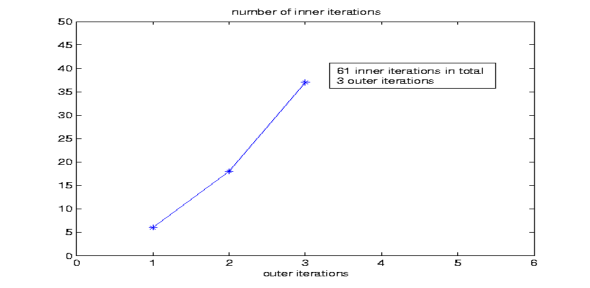
- When preconditioning an eigenvalue problem think of adding the property

$$\mathbb{P}x^{(i)} = Ax^{(i)}$$

  to your favourite preconditioner
- This can be achieved by a simple and cheap rank one modification

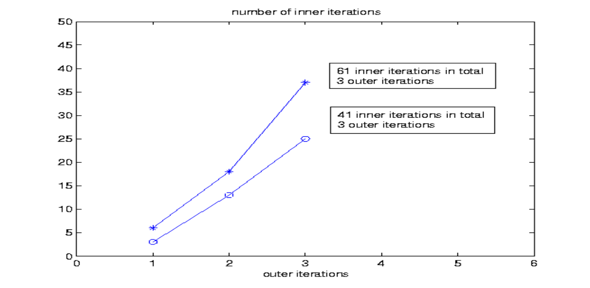
Comparison Simoncini/Eldén with tuning

Standard method



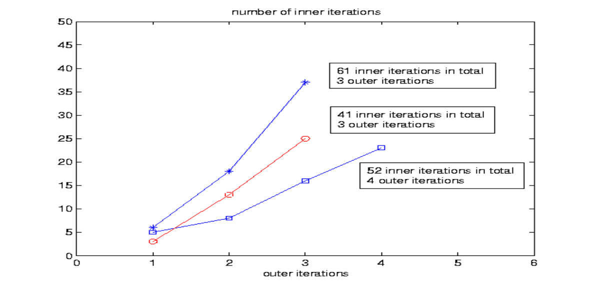
Comparison Simoncini/Eldén with tuning

tuning



Comparison Simoncini/Eldén with tuning

Simoncini/Eldén



M. A. Freitag and A. Spence, *Convergence rates for inexact inverse iteration with application to preconditioned iterative solves*, 2006.
Submitted to BIT.

——, *A tuned preconditioner for inexact inverse iteration applied to Hermitian eigenvalue problems*, 2006.
Submitted to IMA J. Numer. Anal.