Preconditioning Weighted Toeplitz Least Squares Problems

Structured Numerical Linear Algebra Problems:
Algorithms and Applications
Cortona, Italy, September 19-24, 2004

Michele Benzi
Emory University
Atlanta, GA

Thanks to:

- NSF (MPS/Computational Mathematics)
- M. Ng (Hong Kong)
- G. Golub (Stanford), V. Simoncini (Bologna)

Outline

- The basic problem
- An example: nonlinear image restoration
- Equivalent formulations
- Preconditioned Krylov methods
- Constraint preconditioning
- HSS preconditioning
- Numerical examples
- Conclusions

Note: Technical Report will be soon made available at
http://www.mathcs.emory.edu/∼benzi.

Basic Problem

Weighted regularized Toeplitz least squares problem:

$$\min_x \|Ax - b\|_2^2$$

where

$$A = \begin{bmatrix} DK \\ \mu L \end{bmatrix} \quad \text{and} \quad b = \begin{bmatrix} Df \\ 0 \end{bmatrix}.$$

- K is $m \times n$, Toeplitz or BTTB, $m \geq n$
- D is $m \times m$, diagonal, nonnegative definite
- f is $m \times 1$, given
- $\mu > 0$ is a regularization parameter
- L is $n \times n$, a smoothing operator (here $L = I_n$)
- We further assume that m, n are large

Motivation

Such problems arise in various applications, including:

- Nonlinear image restoration
- Seismography
- Acoustics
- Linear prediction

See Å. Björck, *Numerical Methods for Least Squares Problems*, SIAM, 1996.

Problem: The weighting matrix D destroys the Toeplitz structure. Note that D can be very ill-conditioned

$\Rightarrow$ fast Toeplitz solvers **do not apply**!

If $D = I$ or is nearly constant, efficient solvers exist.

Example: Nonlinear Image Restoration

Nonlinear image restoration problem:

$$\min_x ||f - s(Kx)||_2$$

- f is the observed image
- x is the original image (unknown)
- K is the blurring operator ($m \times n$, $m \geq n$)
- $s : \mathbf{R}^m \to \mathbf{R}^m$ is a (separable) nonlinear map

Example: Nonlinear Image Restoration

Nonlinear image restoration problem:

$$\min_x ||f - s(Kx)||_2$$

- f is the observed image
- x is the original image (unknown)
- K is the blurring operator ($m \times n$, $m \geq n$)
- $s : \mathbf{R}^m \to \mathbf{R}^m$ is a (separable) nonlinear map

---

Discrete ill-posed problem $\Rightarrow$ Tikhonov regularization:

$$\min_x ||f - s(Kx)||_2^2 + \mu\, ||x||_2^2$$

Example: Nonlinear Image Restoration

Regularized nonlinear least-squares:

$$\min_x \; ||f - s(Kx)||_2^2 + \mu \, ||x||_2^2$$

Example: Nonlinear Image Restoration

Regularized nonlinear least-squares:

$$\min_x \, ||f - s(Kx)||_2^2 + \mu \, ||x||_2^2$$

Gauss-Newton linearization $\Rightarrow$ sequence of weighted linear LS problems of the form

$$\min_x \, ||D(f - Kx)||_2^2 + \mu \, ||x||_2^2$$

with $D = D(k)$ diagonal, positive definite and $f = f(k)$.

Note: $D = D(k)$ is the Jacobian of s evaluated at the current Newton approximation.

Equivalent formulations

Normal Equations: The regularized weighted least squares problem is equivalent to

$$(K^T D^2 K + \mu I)\, x = K^T D^2 f\,, \tag{1}$$

an n-by-n symmetric positive definite linear system.

Note again that the presence of D destroys any structure the problem may have. Also note that D contributes to make (1) more ill-conditioned.

Solving (1) is quite a challenge. Unless the entries of D are nearly constant, standard Toeplitz solvers and preconditioners will fail.

Augmented system formulations

Another equivalent formulation is the following:

$$\begin{bmatrix} D^{-2} & K \\ K^T & -\mu I \end{bmatrix} \begin{bmatrix} y \\ x \end{bmatrix} = \begin{bmatrix} f \\ 0 \end{bmatrix} \tag{2}$$

where the auxiliary variable $y = D(f - K\,x)$ represents a weighted residual.

The $(m+n)\times(m+n)$ coefficient matrix in (2) is symmetric indefinite. This system is equivalent to

$$\begin{bmatrix} D^{-2} & K \\ -K^T & \mu I \end{bmatrix} \begin{bmatrix} y \\ x \end{bmatrix} = \begin{bmatrix} f \\ 0 \end{bmatrix} \tag{3}$$

where the system matrix is now nonsymmetric positive definite: the eigenvalues have positive real part.

Augmented system formulations

Letting $W = D^{-2}$ for simplicity, the augmented matrix can be factored as follows:

$$\begin{bmatrix} W & K \\ K^T & -\mu I \end{bmatrix} = \begin{bmatrix} I & O \\ K^T W^{-1} & I \end{bmatrix} \begin{bmatrix} W & O \\ O & -\Sigma \end{bmatrix} \begin{bmatrix} I & W^{-1}K \\ O & I \end{bmatrix}$$

where $\Sigma = \mu I + K^T W^{-1} K$ is the Schur complement. Note that Σ is precisely the coefficient matrix of the normal equations.

By Sylvester's Law of Inertia, the augmented matrix has m positive and n negative eigenvalues.

Augmented system formulations

The nonsymmetric augmented matrix can be split as

$$\begin{bmatrix} W & K \\ -K^T & \mu I \end{bmatrix} = \begin{bmatrix} W & O \\ O & \mu I \end{bmatrix} + \begin{bmatrix} O & K \\ -K^T & O \end{bmatrix}$$

Since the symmetric part of the matrix is positive definite, the eigenvalues all have positive real part.

Further, we note that the matrix is J-symmetric, i.e., it is symmetric with respect to the indefinite inner product associated with the $(m+n) \times (m+n)$ matrix

$$J = \begin{bmatrix} I_m & O \\ O & -I_n \end{bmatrix}.$$

Preconditioned Krylov methods

Augmented systems from weighted least squares problems belong to the class of saddle point problems.

In recent years, many new methods have been proposed for solving saddle point systems. In most cases, these methods have been designed for large, sparse problems. In particular, many preconditioners have been proposed.

The Toeplitz case has not received much attention. An exception is the paper

X.-Q. Jin, *A preconditioner for constrained and weighted least squares problems with Toeplitz structure*, BIT 36 (1996), pp. 101–109

where circulant-type preconditioners are considered.

Preconditioned Krylov methods

Preconditioning: Find an invertible matrix $\mathcal{P}$ such that Krylov methods applied to the preconditioned system

$$\mathcal{P}^{-1}\mathcal{A}\,\mathbf{x} = \mathcal{P}^{-1}\mathbf{b}$$

will converge rapidly.

Preconditioned Krylov methods

Preconditioning: Find an invertible matrix $\mathcal{P}$ such that Krylov methods applied to the preconditioned system

$$\mathcal{P}^{-1}\mathcal{A}\,\mathbf{x} = \mathcal{P}^{-1}\mathbf{b}$$

will converge rapidly.

Rapid convergence is often associated with a clustered spectrum of $\mathcal{P}^{-1}\mathcal{A}$. However, characterizing the rate of convergence in general is not an easy matter.

Preconditioned Krylov methods

Preconditioning: Find an invertible matrix $\mathcal{P}$ such that Krylov methods applied to the preconditioned system

$$\mathcal{P}^{-1}\mathcal{A}\,\mathbf{x} = \mathcal{P}^{-1}\mathbf{b}$$

will converge rapidly.

Rapid convergence is often associated with a clustered spectrum of $\mathcal{P}^{-1}\mathcal{A}$. However, characterizing the rate of convergence in general is not an easy matter.

To be effective, a preconditioner must significantly reduce the total amount of work:

- $\mathcal{P}$ must be easy to compute
- Evaluating $\mathbf{z} = \mathcal{P}^{-1}\mathbf{r}$ must be cheap

Preconditioned Krylov methods

Available Krylov methods include:

1. Symmetric $\mathcal{A}$:
    - MINRES (Paige & Saunders, SINUM '76)
    - SQMR (Freund & Nachtigal, APNUM '95)
    - Preconditioner must be SPD for MINRES
    - Preconditioner can be symm. indefinite for SQMR

Preconditioned Krylov methods

Available Krylov methods include:

1. Symmetric $\mathcal{A}$:
   - MINRES (Paige & Saunders, SINUM '76)
   - SQMR (Freund & Nachtigal, APNUM '95)
   - Preconditioner must be SPD for MINRES
   - Preconditioner can be symm. indefinite for SQMR
2. Nonsymmetric $\mathcal{A}$:
   - GMRES (Saad & Schultz, SISSC '86)
   - Bi-CGSTAB (van der Vorst, SISSC '91)
   - Preconditioner can be anything

Preconditioned Krylov methods

Available Krylov methods include:

1. Symmetric $\mathcal{A}$:
   - MINRES (Paige & Saunders, SINUM '76)
   - SQMR (Freund & Nachtigal, APNUM '95)
   - Preconditioner must be SPD for MINRES
   - Preconditioner can be symm. indefinite for SQMR
2. Nonsymmetric $\mathcal{A}$:
   - GMRES (Saad & Schultz, SISSC '86)
   - Bi-CGSTAB (van der Vorst, SISSC '91)
   - Preconditioner can be anything

Recent trend: Use GMRES or Bi-CGSTAB with a non-symmetric preconditioner, even when $\mathcal{A}$ is symmetric!

Preconditioners for saddle point systems

Options include:

1. Multigrid methods

Preconditioners for saddle point systems

Options include:

1. Multigrid methods

2. Schur complement-based methods
   - Block diagonal preconditioning
   - Block triangular preconditioning
   - Uzawa preconditioning

Preconditioners for saddle point systems

Options include:

1. Multigrid methods

2. Schur complement-based methods
   - Block diagonal preconditioning
   - Block triangular preconditioning
   - Uzawa preconditioning

3. Constraint preconditioning

Preconditioners for saddle point systems

Options include:

1. Multigrid methods

2. Schur complement-based methods
   - Block diagonal preconditioning
   - Block triangular preconditioning
   - Uzawa preconditioning

3. Constraint preconditioning

4. Hermitian/Skew-Hermitian splitting (HSS)

Preconditioners for saddle point systems

Options include:

1. Multigrid methods

2. Schur complement-based methods
   - Block diagonal preconditioning
   - Block triangular preconditioning
   - Uzawa preconditioning

3. Constraint preconditioning

4. Hermitian/Skew-Hermitian splitting (HSS)

Here we examine methods of type 3 and 4 (methods of type 2 did not work).

Constraint Preconditioning

Consider the symmetric augmented matrix

$$\mathcal{A} = \begin{bmatrix} W & K \\ K^T & -\mu I \end{bmatrix}$$

and the preconditioning matrix

$$\mathcal{P} = \begin{bmatrix} cI & K \\ K^T & -\mu I \end{bmatrix}$$

where c is a constant. For example, c could be the average value of the entries in W, or $c = 1$.

Note that linear systems of the form $\mathcal{P}\mathbf{z} = \mathbf{r}$ must be solved at each iteration. Because $\mathcal{P}$ has a BTTB structure, we can use fast methods to solve $\mathcal{P}\mathbf{z} = \mathbf{r}$.

Constraint Preconditioning

Let K have full rank ($= n$). When $\mu = 0$ (no regularization) we have

$$\mathcal{A} = \begin{bmatrix} W & K \\ K^T & O \end{bmatrix}$$

and the preconditioning matrix becomes, for $c = 1$:

$$\mathcal{P} = \begin{bmatrix} I & K \\ K^T & O \end{bmatrix}.$$

This constraint preconditioner has been studied, in the finite element context, by Axelsson & Gustafsson (1979) and by Ewing et al. (1990).

More recent papers include Lukšan & Vlček (1998), Perugia, Simoncini & Arioli (2000), Keller, Gould & Wathen (2000), and Rozložník & Simoncini (2002).

Constraint Preconditioning

Theorem Let $K \neq I$ have full column rank. The preconditioned matrix is

$$\mathcal{P}^{-1}\mathcal{A} = \begin{bmatrix} I & K \\ K^T & O \end{bmatrix}^{-1} \begin{bmatrix} W & K \\ K^T & O \end{bmatrix} = \begin{bmatrix} W(I-\Pi)+\Pi & O \\ X & I \end{bmatrix}$$

where Π is the orthogonal projector onto $\mathcal{R}(K)$. Hence, $\lambda = 1$ is an eigenvalue of $\mathcal{P}^{-1}\mathcal{A}$ of multiplicity at least $2n$.

The remaining eigenvalues are eigenvalues of the symmetric matrix $(I-\Pi)W(I-\Pi)$.

In the special case $m = n$, we have $\sigma(\mathcal{P}^{-1}\mathcal{A}) = \{1\}$ and the minimum polynomial of $\mathcal{P}^{-1}\mathcal{A}$ has degree 2.

Corollary If $n = m$, GMRES applied to the preconditioned system $\mathcal{P}^{-1}\mathcal{A}\mathbf{x} = \mathcal{P}^{-1}\mathbf{b}$ terminates after at most two steps.

Constraint Preconditioning

More generally, GMRES applied to the preconditioned system $\mathcal{P}^{-1}\mathcal{A}\mathbf{x} = \mathcal{P}^{-1}\mathbf{b}$ terminates after at most $m - n + 2$ steps, regardless of W.

Therefore, if $m-n$ is small (K is "almost square"), constraint preconditioning is a very good choice.

Things, however, can be quite different when regularization is used ($\mu \neq 0$). In this case the constraint preconditioner needs to be regularized as well, and the preconditioned matrix becomes

$$\mathcal{P}^{-1}\mathcal{A} = \begin{bmatrix} I & K \\ K^T & -\mu I \end{bmatrix}^{-1} \begin{bmatrix} W & K \\ K^T & -\mu I \end{bmatrix}.$$

This case has been investigated by Axelsson (1979) and by Axelsson & Neytcheva (2003).

Constraint Preconditioning

When $\mu > 0$, the preconditioned matrix

$$\mathcal{P}^{-1}\mathcal{A} = \begin{bmatrix} I & K \\ K^T & -\mu I \end{bmatrix}^{-1} \begin{bmatrix} W & K \\ K^T & -\mu I \end{bmatrix}$$

has the eigenvalue 1 with multiplicity n, and all the remaining eigenvalues are real.

When $m = n$, the eigenvalues $\lambda \neq 1$ lie in the interval $(0, 1)$.

If K is ill-conditioned (as it will be if regularization is needed), many of the eigenvalues of $\mathcal{P}^{-1}\mathcal{A}$ will be close to zero and the preconditioner quality will deteriorate.

HSS Preconditioner

We start from the splitting

$$\mathcal{A} = \begin{bmatrix} W & K \\ -K^T & \mu I \end{bmatrix} = \begin{bmatrix} W & O \\ O & \mu I \end{bmatrix} + \begin{bmatrix} O & K \\ -K^T & O \end{bmatrix} = \mathcal{H} + \mathcal{S}.$$

The HSS preconditioner is defined as

$$\mathcal{P}_\alpha = \frac{1}{2\alpha}(\mathcal{H} + \alpha I)(\mathcal{S} + \alpha I)$$

where $\alpha > 0$. Note that $\mathcal{H} + \alpha I$ is SPD and that $\mathcal{S} + \alpha I$ is invertible.

See Bai, Golub & Ng (2003) and Benzi & Golub (2004); case $\mu = 0$ analyzed in Simoncini & Benzi (2004).

HSS Preconditioner

Preconditioner action: requires solving

$$(\mathcal{H} + \alpha I)(\mathcal{S} + \alpha I)\,\mathbf{z} = \mathbf{r}$$

at each Krylov subspace iteration, or

$$(\mathcal{H} + \alpha I)\,\mathbf{v} = \mathbf{r} \quad \text{followed by} \quad (\mathcal{S} + \alpha I)\,\mathbf{z} = \mathbf{v}.$$

HSS Preconditioner

Preconditioner action: requires solving

$$(\mathcal{H} + \alpha I)(\mathcal{S} + \alpha I)\,\mathbf{z} = \mathbf{r}$$

at each Krylov subspace iteration, or

$$(\mathcal{H} + \alpha I)\,\mathbf{v} = \mathbf{r} \quad \text{followed by} \quad (\mathcal{S} + \alpha I)\,\mathbf{z} = \mathbf{v}.$$

- The first system is diagonal: cost is $O(m)$.
- The second one is of the form
$$\begin{bmatrix} \alpha I & K \\ -K^T & \alpha I \end{bmatrix} \begin{bmatrix} v \\ w \end{bmatrix} = \begin{bmatrix} g \\ h \end{bmatrix}$$
and can be solved efficiently using fast Toeplitz solvers.

HSS Preconditioner

Theorem (Benzi & Golub, 2004)
Assume W is SPD, K full rank and $\mu \geq 0$. Then for all $\alpha > 0$, the spectral radius of $I - \mathcal{P}_\alpha^{-1}\mathcal{A}$ is less than 1. Therefore the eigenvalues of $\mathcal{P}_\alpha^{-1}\mathcal{A}$ are contained in $D(1,1) = \{z \in \mathbf{C}\,;\, |z-1| < 1\}$.

Theorem (Simoncini & Benzi, 2004)
Assume W is SPD, K full rank and $\mu = 0$. For sufficiently small α, the eigenvalues of $\mathcal{P}_\alpha^{-1}\mathcal{A}$ cluster near zero and two. More precisely, for small $\alpha > 0$,

$$\lambda \in (0, \varepsilon_1) \cup (2 - \varepsilon_2, 2)$$

with $\varepsilon_1, \varepsilon_2 > 0$ and $\varepsilon_1, \varepsilon_2 \to 0$ for $\alpha \to 0$.

Hence, α should be chosen small, but not too small!

HSS Preconditioner

The case $\mu \neq 0$ is more complicated to analyze. We can say something if we choose $\alpha = \mu$.

Theorem (Benzi & Ng, 2004)
Let w_{min}, w_{max} denote the smallest and largest entries of the diagonal matrix W, with $w_{\text{min}} > 0$. Also, let

$$a := \frac{2\mu}{\mu + w_{\text{max}}}$$

and let $\mathcal{P}_\mu$ denote the HSS preconditioner with $\alpha = \mu$. Then

$$\sigma(\mathcal{P}_\mu^{-1}\mathcal{A}) \subset [a, 2) \times (-1, 1) \cap D(1, 1)$$

where $D(1,1) = \{z \in \mathbf{C}\,;\, |z - 1| < 1\}$. If, moreover, the regularization parameter μ satisfies $\mu < w_{\text{min}}$, then the eigenvalues of $\mathcal{P}_\mu^{-1}\mathcal{A}$ are all real and lie in $[a, 2)$.

Note that a is independent of K.

Numerical Examples

- $K = (k_{|i-j|})$ with $k_{|i-j|} = 1/(\sqrt{|i-j|} + 1)$
- W is positive diagonal, random, $\kappa(W) \approx 10^3 - 10^6$
- Regularization parameter $\mu = 10^{-3}$
- CG is CG on normal equations (no prec.)
- GMRES is GMRES on augmented system (no prec.)
- HSS(α): GMRES with HSS preconditioner
- CP = regularized constraint preconditioning
- Stopping criterion: $||r_k|| < 10^{-7}||r_0||$
- Cost per iteration: $O(n \log n)$ for all methods

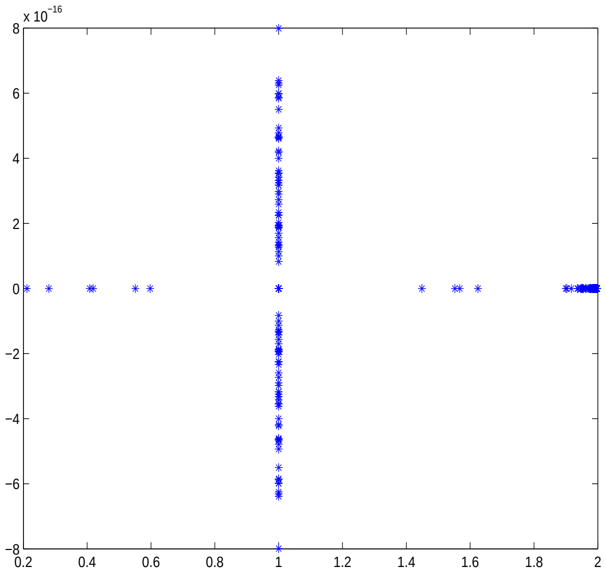
n	CG	GMRES	HSS ($\alpha = \mu$)	HSS ($\alpha = \sqrt{\mu}$)	CP
64	159	48	13	6	3
128	424	66	13	7	3
256	> 1000	90	18	7	3
512	> 1000	132	57	17	3
1024	> 1000	168	72	16	3

Numerical Examples

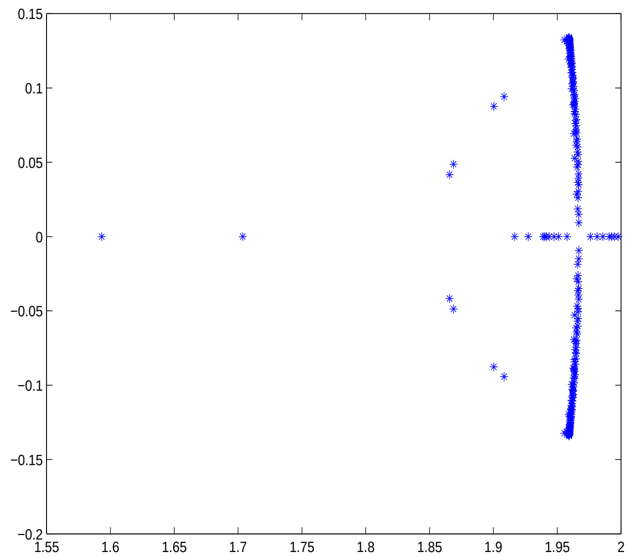
- $K = (k_{|i-j|})$ with $k_{|i-j|} = \frac{1}{2\sqrt{2\pi}} e^{-\frac{|i-j|^2}{8}}$
- W is positive diagonal, random, $\kappa(W) \approx 10^3 - 10^6$
- Regularization parameter $\mu = 10^{-3}$
- CG is CG on normal equations (no prec.)
- GMRES is GMRES on augmented system (no prec.)
- HSS(α): GMRES with HSS preconditioner
- CP = regularized constraint preconditioning
- Stopping criterion: $||r_k|| < 10^{-7}||r_0||$
- Cost per iteration: $O(n \log n)$ for all methods

n	CG	GMRES	HSS ($\alpha = \mu$)	HSS ($\alpha = 6 \cdot 10^{-5}$)	CP
64	761	117	55	43	70
128	> 1000	224	106	74	125
256	> 1000	410	159	95	216
512	> 1000	770	236	127	382
1024	> 1000	> 1000	250	129	655

Problem 1: Spectra of Preconditioned Matrices

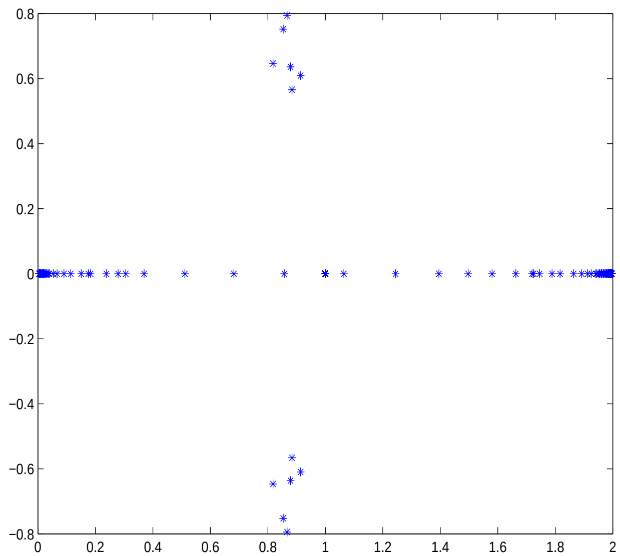


(a) Spectrum for $\alpha = \mu$

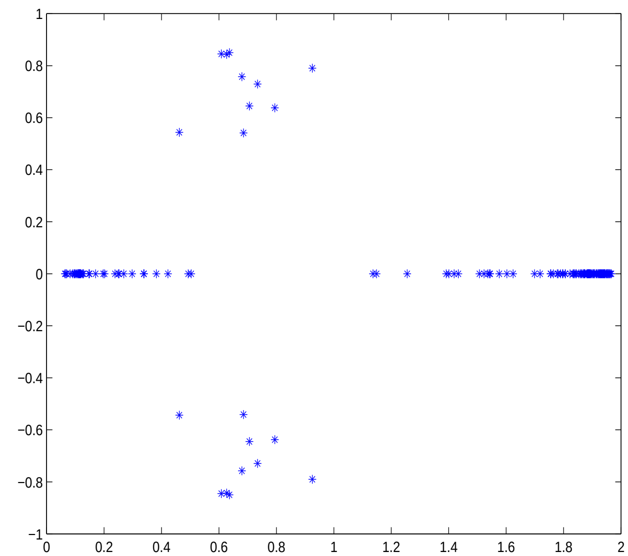


(b) Spectrum for $\alpha = \sqrt{\mu}$

Problem 2: Spectra of Preconditioned Matrices



(a) Spectrum for $\alpha = \mu$



(b) Spectrum for $\alpha = 6 \cdot 10^{-5}$

Numerical Examples

Remarks:

- Preconditioning is absolutely essential for both problems
- Using $\alpha = \mu$ in HSS does not work very well
- "Optimal" value of α is independent of n
- Iteration counts for HSS levels off as n grows
- CP is great on easier problem, very bad on hard problem
- Tests with HSS on image restoration problem (M. Ng) show promise
- Other preconditioners tested but results were poor

Conclusions

- Weighted Toeplitz least squares problems can be hard
- Augmented system formulations allow to decouple Toeplitz part from non-Toeplitz part
- Two methods tested: CP and HSS
- CP best for some problems, but HSS is more robust
- There is plenty of room for improvement!