Preconditioning for nonsymmetric Toeplitz matrices with application to time-dependent PDEs

Andy Wathen
Oxford University, UK

joint work with
Jen Pestana (University of Strathclyde),
Elle McDonald (Oxford University)

$$\mathbf{Ax} = \mathbf{b}$$

$$\mathbf{Ax} = \mathbf{b}$$

But

$$\mathbf{A}, \mathbf{H}, \mathbf{I}, \mathbf{M}, \mathbf{O}, \mathbf{T}, \mathbf{U}, \mathbf{V}, \mathbf{W}, \mathbf{X}, \mathbf{Y}$$

must be symmetric matrices (Parlett convention)!

$$\mathbf{Bx} = \mathbf{b}$$

$$\mathbf{Bx} = \mathbf{b}$$

This makes a big difference!

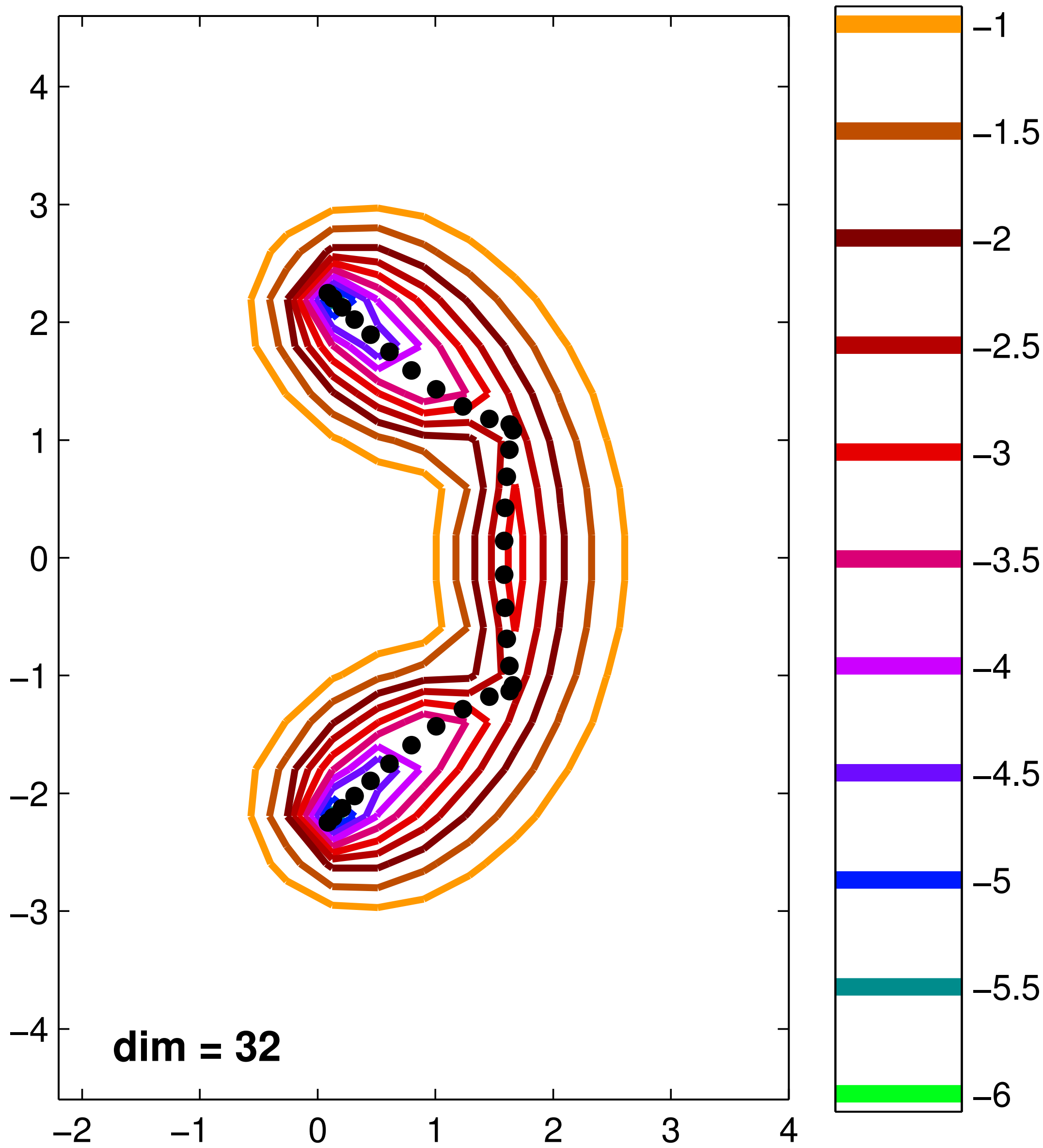
$$\mathbf{A} = \begin{bmatrix} 2 & -1 & 0 & \dots & 0 & 2 \\ -1 & 3 & \ddots & 0 & 1 & 0 \\ 0 & \ddots & \ddots & . & \ddots & \vdots \\ \vdots & \ddots & . & \ddots & \ddots & 0 \\ 0 & 1 & 0 & \ddots & 3 & -1 \\ 2 & 0 & \dots & 0 & -1 & 2 \end{bmatrix}$$ is nice, symmetric

$$\mathbf{A} = \begin{bmatrix} 2 & -1 & 0 & \dots & 0 & 2 \\ -1 & 3 & \ddots & 0 & 1 & 0 \\ 0 & \ddots & \ddots & \cdot & \ddots & \vdots \\ \vdots & \ddots & \cdot & \ddots & \ddots & 0 \\ 0 & 1 & 0 & \ddots & 3 & -1 \\ 2 & 0 & \dots & 0 & -1 & 2 \end{bmatrix}$$ is nice, symmetric

$$\mathbf{G} = \begin{bmatrix} 1 & -1 & 0 & \dots & \dots & \dots & 0 \\ 1 & 1 & -1 & \ddots & & & \vdots \\ 1 & 1 & 1 & -1 & \ddots & & \vdots \\ 1 & 1 & 1 & 1 & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & -1 \\ 0 & \dots & 0 & 1 & 1 & 1 & 1 \end{bmatrix}$$

is one of Nick Trefethen's favourite *non-normal* matrices!

“Test EigTool by running the command `eigtool(gallery('grcar',32)).` ”



Krylov subspace methods

For $\mathbf{Bx} = \mathbf{b}$: from $\mathbf{x}_0$ generate $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k, \ldots\}$ using one matrix×vector product at each iteration:

$$\mathbf{Br}, \mathbf{B}(\mathbf{Br}), \ldots, \mathbf{B}^k\mathbf{r}, \ldots$$

$\Rightarrow$ Krylov subspace methods:

$$\mathbf{r}_k = \mathbf{p}_k(\mathbf{B})\mathbf{r}_0, \qquad \mathbf{r}_k = \mathbf{b} - \mathbf{Bx}_k, \quad \mathbf{p}_k \in \Pi_k, \mathbf{p}_k(0) = 1$$

so if $\mathbf{B} = \mathbf{X\Lambda X}^{-1}$ then $\mathbf{p}_k(\mathbf{B}) = \mathbf{X}\mathbf{p}_k(\mathbf{\Lambda})\mathbf{X}^{-1}$ so

$$\|\mathbf{r}_k\| \leq \|\mathbf{X}\| \; \|\mathbf{p}_k(\mathbf{\Lambda})\| \; \|\mathbf{X}^{-1}\| \; \|\mathbf{r}_0\|$$

and if $\mathbf{B} = \mathbf{B}^T$ so that $\mathbf{X}^{-1} = \mathbf{X}^T$ then this bound on convergence in $\|\cdot\|_2$ depends only eigenvalues

Krylov subspace methods

For self-adjoint problems/symmetric matrices, iterative methods of choice exist: conjugate gradients for SPD, MINRES otherwise

but many possible methods for non-self-adjoint problems/nonsymmetric matrices: GMRES , BICGSTAB , QMR , IDR , . . .

Krylov subspace methods

For self-adjoint problems/symmetric matrices, iterative methods of choice exist: conjugate gradients for SPD, MINRES otherwise

but many possible methods for non-self-adjoint problems/nonsymmetric matrices: GMRES , BICGSTAB , QMR , IDR , . . .

For almost all need *preconditioning*

Preconditioner $\mathbf{P}$ such that

$$\text{“}\mathbf{P}^{-1}\mathbf{B}\mathbf{x} = \mathbf{P}^{-1}\mathbf{b}\text{”}$$

has much faster convergence with the appropriate iterative method than $\mathbf{B}\mathbf{x} = \mathbf{b}$.

Krylov subspace methods

Arises because of convergence guarantees:

- for symmetric matrices: descriptive convergence bounds based on eigenvalues $\Rightarrow$ a priori estimates of iterations for acceptable convergence; good preconditioning ensures fast convergence.
- for nonsymmetric matrices: by contrast, to date there are no generally applicable *and descriptive* convergence bounds even for GMRES ; for any of the other nonsymmetric methods without a minimisation property, convergence theory is extremely limited $\Rightarrow$ no good a priori way to identify what are the desired qualities of a preconditioner

A major theoretical difficulty, but heuristic ideas abound!

The situation is more severe than this:

Theorem (*Greenbaum, Ptak and Strakos, 1996*)

Given any set of eigenvalues and any monotonic convergence curve, then for any $\mathbf{b}$ there exists a matrix $\mathbf{B}$ having those eigenvalues and an initial guess $\mathbf{x_0}$ such that GMRES for $\mathbf{Bx} = \mathbf{b}$ with $\mathbf{x_0}$ as starting vector will give that convergence curve.

In fact more extreme negative results than this exist.

One way to address such questions:
look for (non-standard) inner products in which a problem might be self-adjoint

One way to address such questions:
look for (non-standard) inner products in which a problem might be self-adjoint

such inner products exist for a real nonsymmetric matrix **B** if and only if **B** is diagonalizable and has real eigenvalues

but preconditioners still would have to have self-adjointness in any relevant non-standard inner product!!

Recently (*Pestana & W, 2015*) we have made progress for real nonsymmetric Toeplitz (constant diagonal) matrices regardless of nonnormality with a very simple observation: If **B** is a real Toeplitz matrix then

$$\underbrace{\begin{bmatrix} a_0 & a_{-1} & \cdot & \cdot & a_{1-n} \\ a_1 & a_0 & a_{-1} & \cdot & \cdot \\ \cdot & a_1 & a_0 & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & a_{-1} \\ a_{n-1} & \cdot & \cdot & a_1 & a_0 \end{bmatrix}}_{\mathbf{B}} \underbrace{\begin{bmatrix} 0 & 0 & \cdot & 0 & 1 \\ 0 & \cdot & 0 & 1 & 0 \\ \cdot & 0 & 1 & 0 & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & 0 & \cdot & 0 & 0 \end{bmatrix}}_{\mathbf{Y}}$$

is the real *symmetric* (Hankel) matrix

$$\begin{bmatrix} a_{1-n} & \cdot & \cdot & a_{-1} & a_0 \\ \cdot & \cdot & a_{-1} & a_0 & a_1 \\ \cdot & \cdot & a_0 & a_1 & \cdot \\ a_{-1} & \cdot & \cdot & \cdot & \cdot \\ a_0 & a_1 & \cdot & \cdot & a_{n-1} \end{bmatrix}$$

Thus MINRES can be robustly applied to $\mathbf{BY}$ — it is symmetric but generally indefinite — and its convergence will depend only on eigenvalues.

BUT preconditioning? – needs to be symmetric and positive definite for MINRES

Fortunately it is well known that many Toeplitz matrices are well approximated by related circulant matrices, $\mathbf{C}$ (*Strang, 1986, Chan, 1988, Trefethen, 1990, Tyrtishnikov, 1996/7*) which are diagonalised by an FFT in $O(n \log n)$ work: $\mathbf{C} = \mathbf{F}^{\star}\mathbf{\Lambda}\mathbf{F}$.

For many Toeplitz matrices we have that the Strang or Optimal (Chan) circulant $\mathbf{C}$ satisfy

$$\mathbf{C}^{-1}\mathbf{B} = \mathbf{I} + \mathbf{R} + \mathbf{E}$$

where $\mathbf{R}$ is of small rank and $\mathbf{E}$ is of small norm

$\Rightarrow$eigenvalues clustered around 1 except for a few outliers

For example, the Strang circulant for the standard Toeplitz matrix (as above) is

$$\underbrace{\begin{bmatrix} a_0 & a_{-1} & \cdots & a_{-\lfloor \frac{n}{2} \rfloor} & a_{\lfloor \frac{n-1}{2} \rfloor} & \cdots & a_2 & a_1 \\ a_1 & a_0 & a_{-1} & \cdots & a_{-\lfloor \frac{n}{2} \rfloor} & a_{\lfloor \frac{n-1}{2} \rfloor} & \cdots & a_2 \\ \cdots & a_1 & a_0 & \ddots & \cdots & a_{-\lfloor \frac{n}{2} \rfloor} & \ddots & \vdots \\ a_{\lfloor \frac{n}{2} \rfloor} & \cdots & \ddots & \ddots & \ddots & \cdots & \ddots & a_{\lfloor \frac{n-1}{2} \rfloor} \\ a_{-\lfloor \frac{n-1}{2} \rfloor} & \ddots & \cdots & \ddots & \ddots & \ddots & \cdots & a_{-\lfloor \frac{n}{2} \rfloor} \\ \vdots & \ddots & a_{\lfloor \frac{n}{2} \rfloor} & \cdots & \ddots & a_0 & a_{-1} & \cdots \\ a_{-2} & \cdots & a_{-\lfloor \frac{n-1}{2} \rfloor} & a_{\lfloor \frac{n}{2} \rfloor} & \cdots & a_1 & a_0 & a_1 \\ a_{-1} & a_{-2} & \cdots & a_{-\lfloor \frac{n-1}{2} \rfloor} & a_{\lfloor \frac{n}{2} \rfloor} & \cdots & a_1 & a_0 \end{bmatrix}}_{\mathbf{C}}$$

To ensure a symmetric and positive definite preconditioner for $\mathbf{BY}$ just use the circulant matrix

$$|\mathbf{C}| = \mathbf{F}^{\star}|\mathbf{\Lambda}|\mathbf{F}$$

which is real symmetric and positive definite

Theorem (*Pestana & W, 2015*)

$$|\mathbf{C}|^{-1}\mathbf{BY} = \mathbf{J} + \mathbf{R} + \mathbf{E}$$

where $\mathbf{J}$ is real symmetric and orthogonal with eigenvalues ± 1, $\mathbf{R}$ is of small rank and $\mathbf{E}$ is of small norm

$\Rightarrow$ guaranteed fast convergence because MINRES convergence only depends on eigenvalues which are clustered around ± 1 except for few outliers!

Example (Liesen & Strakoš)

$$\mathbf{B} = \begin{bmatrix} 1 & 0.01 & & & \\ 1 & 1 & 0.01 & & \\ & \ddots & \ddots & \ddots & \\ & & 1 & 1 & 0.01 \\ & & & 1 & 1 \end{bmatrix} \in \mathbb{R}^{n\times n}$$

n	Condition number of **B**	preconditioned MINRES iters
10	14	6
100	207	6
1000	2.6×10^6	6

n	eigenvalues of $\|\mathbf{C}\|^{-1}\mathbf{BY}$ (to 4 decimal places)
10	$\{-9.9107, 0.9893, -0.9640, -1.0002, -1 \times 2, 1 \times 4\}$
100	$\{-2.2803, -0.2536, 0.9919, -1.0007, -1 \times 47, 1 \times 49\}$
1000	$\{-2.1626, -1.0008, -1.8309e-05, 0.9929, -1 \times 497, 1 \times 499\}$

Example

$$\mathbf{G} = \begin{bmatrix} 1 & -1 & 0 & \dots & \dots & \dots & 0 \\ 1 & 1 & -1 & \ddots & & & \vdots \\ 1 & 1 & 1 & -1 & \ddots & & \vdots \\ 1 & 1 & 1 & 1 & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & -1 \\ 0 & \dots & 0 & 1 & 1 & 1 & 1 \end{bmatrix} \quad \text{(Grcar matrix)}$$

n	$\sharp$ eigs $\neq \pm 1$ (to 12 d.p.)	preconditioned MINRES iters
10	8	10
100	8	10
1000	8	10
10000	8?	10

Example

B $=$ random (dense) Toeplitz matrices of Wiener class

n	eigenvalue inclusion	iterations
10	$[-1.018, -0.710] \cup [0.981, 1.804]$	10
100	$[-1.092, -0.856] \cup [0.912, 1.160]$	14
1000	$[-1.154, -0.708] \cup [0.864, 1.381]$	20
10000	$[-1.078, -0.980] \cup [0.922, 1.017]$	12

Block Toeplitz Matrices

If $\mathbf{B} = \begin{bmatrix} \mathbf{B_0} & \mathbf{B_{-1}} & \cdot & \cdot & \mathbf{B_{1-n}} \\ \mathbf{B_1} & \mathbf{B_0} & \mathbf{B_{-1}} & \cdot & \cdot \\ \cdot & \mathbf{B_1} & \mathbf{B_0} & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \mathbf{B_{-1}} \\ \mathbf{B_{n-1}} & \cdot & \cdot & \mathbf{B_1} & \mathbf{B_0} \end{bmatrix}$,

$\mathbf{B_i} \in \mathbb{R}^{k \times k}$ Toeplitz for all $\mathbf{i}$ (BTTB matrix)

then $\mathbf{Y} = \begin{bmatrix} 0 & 0 & \cdot & 0 & \mathbf{Y_k} \\ 0 & \cdot & 0 & \mathbf{Y_k} & 0 \\ \cdot & 0 & \mathbf{Y_k} & 0 & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \mathbf{Y_k} & 0 & \cdot & 0 & 0 \end{bmatrix}$

is such that **BY** and **YB** are real symmetric

If $\mathbf{B} = \begin{bmatrix} \mathbf{B_0} & \mathbf{B_{-1}} & \cdot & \cdot & \mathbf{B_{1-n}} \\ \mathbf{B_1} & \mathbf{B_0} & \mathbf{B_{-1}} & \cdot & \cdot \\ \cdot & \mathbf{B_1} & \mathbf{B_0} & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \mathbf{B_{-1}} \\ \mathbf{B_{n-1}} & \cdot & \cdot & \mathbf{B_1} & \mathbf{B_0} \end{bmatrix}$,

$\mathbf{B_i} \in \mathbb{R}^{k \times k}$ real symmetric for all $\mathbf{i}$ (BT with symmetric blocks)

then $\widehat{\mathbf{Y}} = \begin{bmatrix} 0 & 0 & \cdot & 0 & \mathbf{I_k} \\ 0 & \cdot & 0 & \mathbf{I_k} & 0 \\ \cdot & 0 & \mathbf{I_k} & 0 & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \mathbf{I_k} & 0 & \cdot & 0 & 0 \end{bmatrix}$

is such that $\mathbf{B}\widehat{\mathbf{Y}}$ and $\widehat{\mathbf{Y}}\mathbf{B}$ are real symmetric

In either case we can apply a block circulant preconditioner for MINRES

Thus similar ideas apply for block Toeplitz matrices—higher dimensions—but then one has the same issues as in the symmetric case (rank of $\mathbf{R}$ is greater)

One class of problems which are always non-symmetric: time-dependent problems...

$\langle \boldsymbol{u_t}, \boldsymbol{v} \rangle = -\langle \boldsymbol{u}, \boldsymbol{v_t} \rangle$ so that

$$\boldsymbol{u_t} = \boldsymbol{\mathcal{L}u} + \boldsymbol{f}$$

is non-self-adjoint even if the spatial operator $\boldsymbol{\mathcal{L}}$ is self-adjoint.

Simple example: backwards Euler for a linear(ized) problem

$$\boldsymbol{U}^{n+1} + \Delta t \boldsymbol{\mathcal{L}} \boldsymbol{U}^{n+1} = \boldsymbol{U}^{n} + \Delta t \boldsymbol{f}^{n+1}, \quad n = 0, 1, \ldots, \ell$$

All-at-once (monolithic) discretization:

$$\boldsymbol{\mathcal{A}x} := \begin{bmatrix} A & 0 & \cdot & \cdot & 0 \\ B & A & 0 & \cdot & \cdot \\ \cdot & B & A & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & 0 \\ 0 & \cdot & \cdot & B & A \end{bmatrix} \begin{bmatrix} U^1 \\ U^2 \\ U^3 \\ \vdots \\ U^N \end{bmatrix} = \begin{bmatrix} U^0 + f^1 \\ f^2 \\ f^3 \\ \vdots \\ f^N \end{bmatrix}$$

$\boldsymbol{A} = \boldsymbol{I} + \Delta t \boldsymbol{\mathcal{L}}, \quad \boldsymbol{B} = -\boldsymbol{I}$.
has block Toeplitz coefficient matrix and the ideas presented here apply

If we employ the preconditioner

$$\mathcal{P} := \begin{bmatrix} A & 0 & \cdot & \cdot & B \\ B & A & 0 & \cdot & \cdot \\ \cdot & B & A & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & 0 \\ 0 & \cdot & \cdot & B & A \end{bmatrix}$$

Theorem
$\mathcal{P}^{-1}\mathcal{A}$ is diagonalizable, has $(\ell - 1)n$ eigenvalues equal to 1 and the remaining n eigenvalues are clustered close to 1.

$\Rightarrow$ GMRES terminates in at most $n + 1$ iterations *independent of* ℓ

and $\mathcal{P}^{-1}$ can be applied easily

Moreover: if $\mathcal{A}$ has symmetric blocks (essentially $\mathcal{L}$ must be self-adjoint), then one can symmetrize and use MINRES with preconditioner $|\mathcal{P}|$

Theorem
All but at most $2n$ eigenvalues are ± 1

$\Rightarrow$ MINRES converges in a number of iterations independent of ℓ.

HAPPY BIRTHDAY IAIN

References and Acknowledgement

Pestana, J. & Wathen, A.J., 2015,
'A preconditioned MINRES method for nonsymmetric Toeplitz matrices',
SIAM J. Matrix Anal. Appl. **36**, pp. 273–288.

McDonald, E, Pestana, J. & Wathen, A.J., 2016,
'Preconditioning and iterative solution of all-at-once systems for evolutionary partial differential equations',
submitted to SIAM J. Sci. Comput.

Wathen, A.J., 2015,
'Preconditioning',
Acta Numerica (2015) **24**, pp. 329–376.

This work is partially supported by Award No. KUK-C1-013-04 made by King Abdullah University of Science and Technology (KAUST)