Preconditioning for Nonsymmetry and Time-dependence

Andy Wathen
Oxford University, UK

joint work with
Jen Pestana and Elle McDonald

Iterative 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 , . . .

Iterative methods

Arises because of convergence guarantees:

- for symmetric matrices: descriptive convergence bounds $\Rightarrow$ 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 $\boldsymbol{b}$ there exists a matrix $\boldsymbol{B}$ having those eigenvalues and an initial guess $\boldsymbol{x_0}$ such that GMRES for $\boldsymbol{Bx = b}$ with $\boldsymbol{x_0}$ 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!!

Convection-diffusion equation:

$$\frac{\partial u}{\partial t} + \mathrm{b} \cdot \nabla u - \epsilon \nabla^2 u = f \quad \text{in } \Omega \times (0, T], \quad \Omega \subset \mathbb{R}^2 \text{ or } \mathbb{R}^3$$

$$u(\mathrm{x}, 0) = u_0(\mathrm{x}), \quad u \text{ given on } \partial\Omega$$

- arises widely e.g. as a subproblem in Navier-Stokes
- is *non-self-adjoint* $\Rightarrow$ nonsymmetric discretization matrix

$\Rightarrow$ convergence of Krylov subspace methods not easily described

so no mathematical idea how to precondition

For steady convection-diffusion

$$\mathbf{b} \cdot \nabla u - \epsilon \nabla^2 u = f$$

the nonsymmetric issue remains even in 1-dimension

$$u' - \epsilon u'' = f$$

(though iterative methods not so crucial here!)

$$u' - \epsilon u'' = f$$

is

$$\left(-\exp(-x/\epsilon)u'\right)' = \frac{1}{\epsilon}\exp(-x/\epsilon)f$$

so continuous problem is self-adjoint in a highly distorted inner product based on this integrating factor (given certain simple boundary conditions)

Discretizations however have matrices which are *not* self-adjoint in any inner product for large enough h/ϵ

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

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

is the real *symmetric* 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 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 Toeplitz matrices are well preconditioned by related circulant matrices, C (*Strang, 1986, Chan, 1988*) which are diagonalised by an FFT in $O(n \log n)$ work: $C = U^{\star} \Lambda U$.

For many Toeplitz matrices we have that the Strang or Optimal (Chan) circulant C satisfy

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

where R is of small rank and E is of small norm

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

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

$$|\boldsymbol{C}| = \boldsymbol{U}^{\star}|\boldsymbol{\Lambda}|\boldsymbol{U}$$

which is real symmetric and positive definite

Theorem (*Pestana & W, 2015*)

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

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

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

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

Similar ideas apply for block Toeplitz matrices—higher dimensions—but theory not so good

Time-dependent Convection-Diffusion

$$\frac{\partial u}{\partial t} + \mathrm{b}.\nabla u - \epsilon \nabla^2 u = f \quad \text{in } \Omega \times (0, T], \quad \Omega \subset \mathbb{R}^2 \text{ or } \mathbb{R}^3$$

$$u(\mathrm{x}, 0) = u_0(\mathrm{x}), \quad u \text{ given on } \partial\Omega$$

Finite elements in space (x), θ time stepping gives

$$M \frac{\mathbf{u}_k - \mathbf{u}_{k-1}}{\tau} + K\Big(\theta \mathbf{u}_k + (1-\theta)\mathbf{u}_{k-1}\Big) = \mathbf{f}_k$$

$M \in \mathbb{R}^{n \times n}$: SPD mass matrix (identity operator, but same sparsity as K)

$K \in \mathbb{R}^{n \times n}$: nonsymmetric discrete steady convection-diffusion matrix

Rearranging:

$$\Big(M + \tau\,\theta\,K\Big)\mathbf{u}_k = \Big(M - \tau\,(1-\theta)\,K\Big)\mathbf{u}_{k-1} + \tau\,\mathbf{f}_k,$$

$$k = 1, 2, \ldots, N$$

$N\tau = T$

Recall for unconditional stability: $\frac{1}{2} \leq \theta \leq 1$

$\theta = 1$: backwards Euler, $\theta = \frac{1}{2}$: Crank-Nicolson

else need $\tau = \mathcal{O}(h^2)$: very small time steps for explicit method

$$\Big(M + \tau\,\theta\,K\Big)\mathbf{u}_k = \Big(M - \tau\,(1-\theta)\,K\Big)\mathbf{u}_{k-1} + \tau\,\mathbf{f}_k,$$

$$k = 1, 2, \ldots, N$$

Standard solution method:

Solve the N separate $n \times n$ nonsymmetric linear systems (sequentially) for $k = 1, 2, \ldots, N$. Here we use a geometric multigrid due to Ramage specifically for convection diffusion

$\Rightarrow r = 5$ V-cycles for solution of each linear system to a relative residual tolerence of 10^{-6}

Hence if we (quite reasonably) regard 1 V-cycle as the main unit of work

$\Rightarrow Nr$ V-cycles sequentially for the overall solution

Alternatively:

Write all timesteps at one go (all-at-once method):

$$\mathcal{A}\begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \\ \vdots \\ \mathbf{u}_N \end{bmatrix} = r.h.s$$

where $\mathcal{A}$ is the matrix

$$\begin{bmatrix} M+\tau\theta K & 0 & 0 & 0 \\ -M+\tau(1-\theta)K & M+\tau\theta K & 0 & 0 \\ 0 & \ddots & \ddots & 0 \\ 0 & 0 & -M+\tau(1-\theta)K & M+\tau\theta K \end{bmatrix}$$

and $r.h.s. = [M-\tau(1-\theta)K\mathbf{u}_0+\tau\,\mathbf{f}_1, \tau\,\mathbf{f}_2, \ldots, \tau\,\mathbf{f}_N]^T$

nonsymmetric!

$$\mathcal{A}=\begin{bmatrix} M+\tau\theta K & 0 & 0 & 0 \\ -M+\tau(1-\theta)K & M+\tau\theta K & 0 & 0 \\ 0 & \ddots & \ddots & 0 \\ 0 & 0 & -M+\tau(1-\theta)K & M+\tau\theta K \end{bmatrix}$$

$\mathcal{A} \in \mathbb{R}^{L\times L}$, $L = Nn$

We propose to solve this huge linear system (for the solution at all time steps) by GMRES (or BICGSTAB) with block diagonal preconditioner

$$\mathcal{P} = \begin{bmatrix} (M+\tau\theta K)_{MG} & 0 & 0 & 0 \\ 0 & (M+\tau\theta K)_{MG} & 0 & 0 \\ 0 & \ddots & \ddots & 0 \\ 0 & 0 & 0 & (M+\tau\theta K)_{MG} \end{bmatrix}$$

where $(M+\tau\theta K)_{MG}$ is one GMG V-cycle exactly as above.

Any other approximate nonsymmetric solver could be used.

Theory: If we used

$$\mathcal{P}_{\text{exact}} = \begin{bmatrix} (\boldsymbol{M}+\boldsymbol{\tau\theta K}) & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & (\boldsymbol{M}+\boldsymbol{\tau\theta K}) & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & (\boldsymbol{M}+\boldsymbol{\tau\theta K}) \end{bmatrix}$$

as preconditioner (no multigrid approximation) then we would have

$$\mathcal{P}_{\text{exact}}^{-1}\boldsymbol{\mathcal{A}} = \begin{bmatrix} \boldsymbol{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \boldsymbol{J} & \boldsymbol{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{J} & \boldsymbol{I} \end{bmatrix},$$

$$\boldsymbol{J} = (\boldsymbol{M}+\boldsymbol{\tau\theta K})^{-1}(-\boldsymbol{M}+\boldsymbol{\tau}(1-\boldsymbol{\theta})\boldsymbol{K})$$

For

$$\mathcal{P}_{\text{exact}}^{-1}\mathcal{A} = \begin{bmatrix} I & 0 & 0 & 0 \\ J & I & 0 & 0 \\ 0 & \ddots & \ddots & 0 \\ 0 & 0 & J & I \end{bmatrix},$$

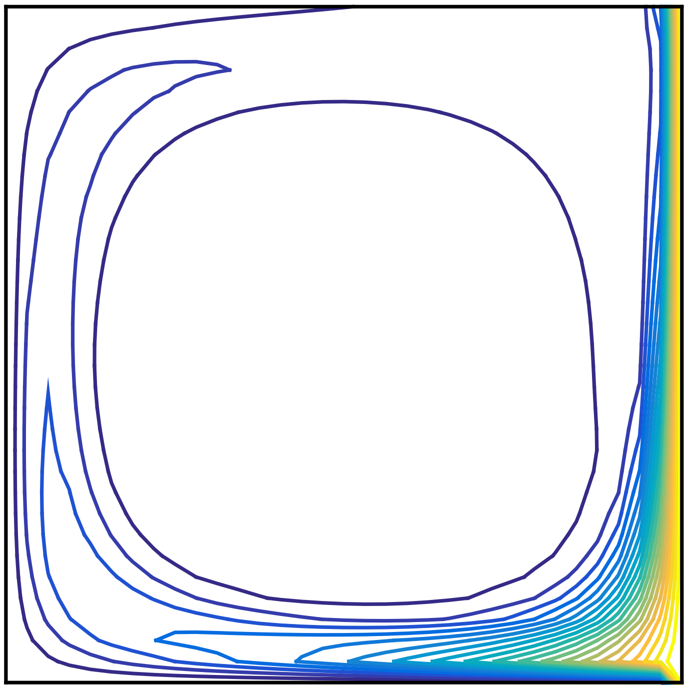
the minimum polynomial is $(1-s)^N$, so GMRES would terminate (in exact arithmetic) in N iterations

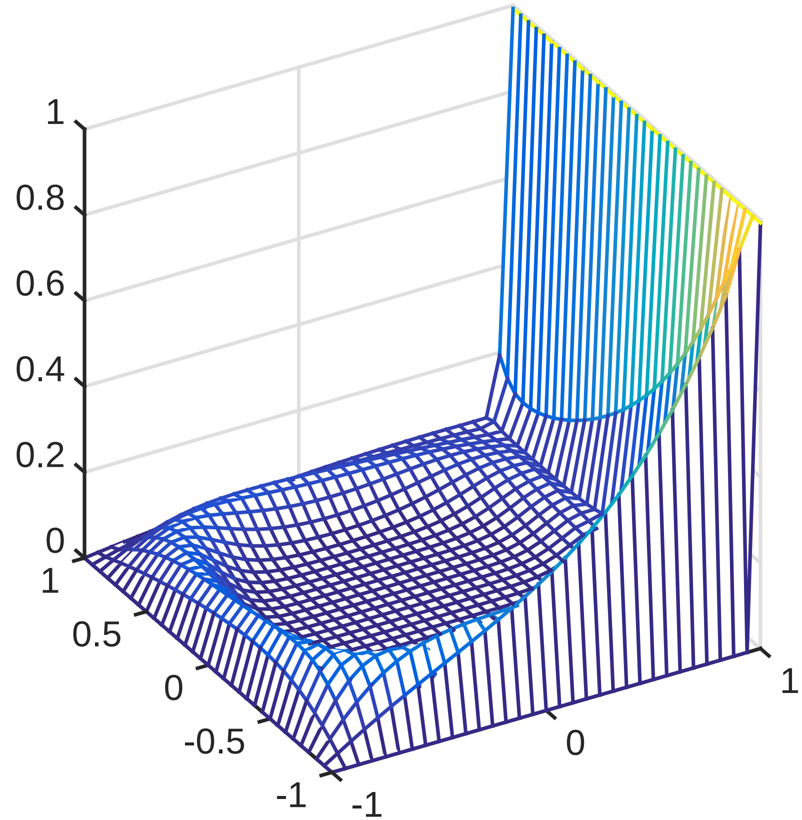
We observe that $(M+\tau\theta K)_{MG}$ is close to $(M+\tau\theta K)$ so that convergence to a tolerance much less than the discretization error is achieved in N iterations also with $\mathcal{P}$ as preconditioner.

Thus: N V-cycles for each of N GMRES iterations—hence N^2 ($> Nr$) overall.

but with N processors, solution with $\mathcal{P}$ is (embarrasingly) parallel—block diagonal $\Rightarrow$ independent computation.

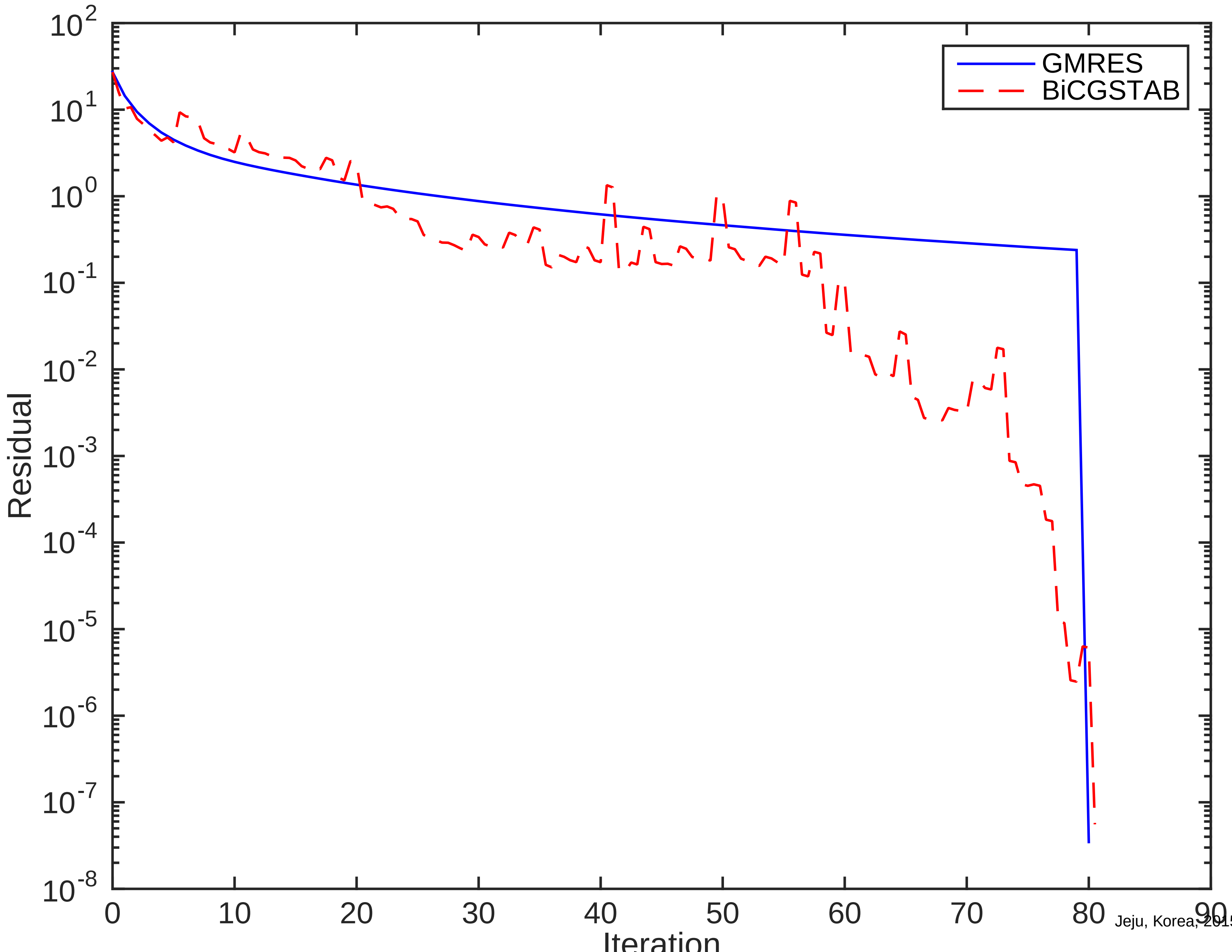
Thus parallel effort is $N < Nr$ (= sequential effort).



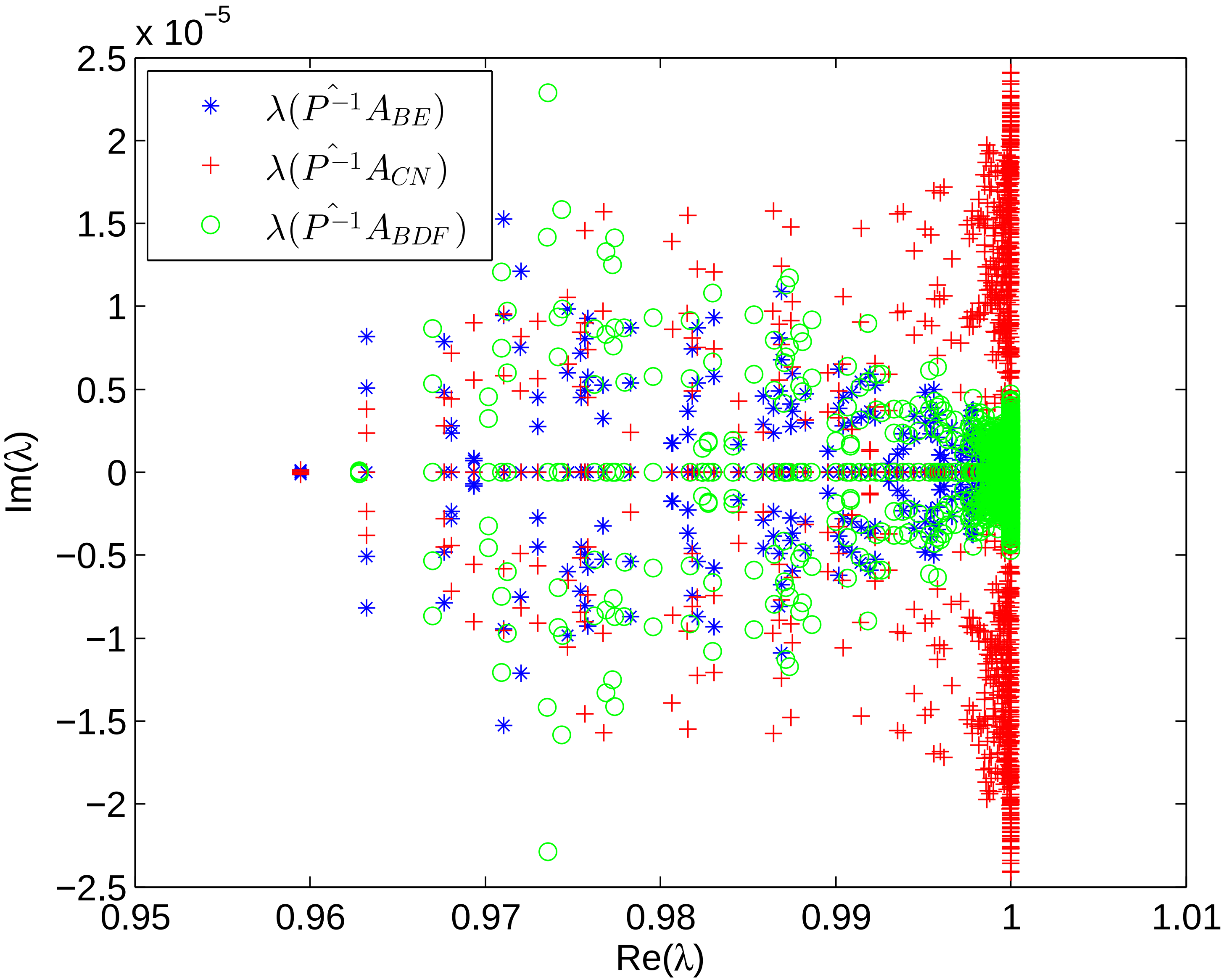


1
0.8
0.6
0.4
0.2
0
1
0.5
0
-0.5
-1
-1
0
1

Convection-diffusion



Convection-diffusion



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.

Pestana, J. & Wathen, A.J., 2013,
'On choice of preconditioner for minimum residual methods for non-Hermitian matrices',
J. Comput. Appl. Math. **249**, pp. 57–68.

McDonald, E.G. & Wathen, A.J.,
'A simple proposal for parallel computation over time of an evolutionary process with implicit time stepping',
in review

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