Strathprints Institutional Repository

Preconditioning for nonsymmetry and time-dependence

Eleanor McDonald[1], Sean Hon[2], Jennifer Pestana[3] and Andy Wathen[4]

1 Introduction

Preconditioning, whether by domain decomposition or other methods, is well understood for symmetric (or Hermitian) matrices at least in the sense that guaranteed convergence bounds based on eigenvalues alone describe convergence of iterative methods. Establishing spectral properties of preconditioned operators or matrices is thus all that is required to reliably predict the number of steps of an appropriate Krylov subspace method—it would be Conjugate Gradients (CG) Hestenes and Stiefel [1952] in the case of positive definite matrices and MINRES Paige and Saunders [1975] for indefinite matrices—in the symmetric case. Faster convergence than that predicted by these bounds occurs in rare cases when only few eigenspaces are important; thus in the rare cases that the convergence bounds fail to be descriptive, it is because they overestimate the number of iterations required for convergence—a good thing! Put another way, we know what we're trying to achieve in the construction of preconditioners in the case of symmetric coefficient matrices.

By contrast, in the nonsymmetric case, no generally descriptive convergence bounds are known. In specialist situations, the field of values or other sets can occasionally be usefully employed Loghin and Wathen [2004], but it is known that GMRES can converge in any (monotone) specified manner whatever the eigenvalues for the coefficient matrix; precisely, it is proved in Greenbaum et al. [1996] (and the results extended in Tebbens and Meurant [2014]) that given any set of n eigenvalues and any monotonic convergence curve terminating at or before the nth iteration, then for any b there exists

Mathematical Institute, Oxford University, Radcliffe Observatory Quarter, Oxford, OX2 6GG, England `mcdonalde@maths.ox.ac.uk` · Mathematical Institute, Oxford University, Radcliffe Observatory Quarter, Oxford, OX2 6GG, England `hon@maths.ox.ac.uk` · Department of Mathematics and Statistics, University of Strathclyde,Glasgow, G1 1XH, Scotland `jennifer.pestana@strath.ac.uk` · Mathematical Institute, Oxford University, Radcliffe Observatory Quarter, Oxford, OX2 6GG, England `wathen@maths.ox.ac.uk`

an $n \times n$ matrix B having those eigenvalues and an initial guess x_0 such that GMRES Saad and Schultz [1986] for $Bx = b$ with x_0 as starting vector will give that convergence curve. More negative results than this exist (see for example Tebbens and Meurant [2012]).

Thus, one can for example have an $n \times n$ nonsymmetric matrix with all eigenvalues equal to 1 for which GMRES gives no reduction in the norm of the residual vectors—that is, no convergence—for $n - 1$ iterations. For any of the range of other nonsymmetric Krylov subspace methods, convergence theory is extremely limited. Thus, even though there is often consideration of eigenvalues when considering possible preconditioners even in the nonsymmetric case, this is not well-founded. It is not however foolish, since poor convergence can certainly in general be associated with problems with widely spread eigenvalues!

The important point nevertheless remains that the construction of preconditioners for nonsymmetric problems is of necessity currently heuristic.

In this short paper, we decribe at least one simple and frequently arising situation —that of nonsymmetric real Toeplitz (constant diagonal) matrices— where we can *guarantee* rapid convergence of the appropriate iterative method by manipulating the problem into a symmetric form without recourse to the normal equations. This trick can be applied regardless of the nonnormality of the Toeplitz matrix. We also propose a symmetric and positive definite preconditioner for this situation which is proved to cluster eigenvalues and is by consequence guaranteed to ensure convergence in a number of iterations independent of the matrix dimension. This is described in Section 2 and more fully in Pestana and Wathen [2015].

We then go on to exploit these observations in considering time-stepping problems for ordinary differential equations. The result we establish in this setting is guaranteed convergence of an iterative method for an all-at-once formulation in a number of iterations independent of the number of time-steps. This is described in Section 3.

2 Real nonsymmetric Toeplitz matrices

If B is a real Toeplitz matrix then

$$\underbrace{\begin{bmatrix} a_0 & a_{-1} & \dots & a_{-n+2} & a_{-n+1} \\ a_1 & a_0 & a_{-1} & & a_{-n+2} \\ \vdots & a_1 & a_0 & \ddots & \vdots \\ a_{n-2} & & \ddots & \ddots & a_{-1} \\ a_{n-1} & a_{n-2} & \dots & a_1 & a_0 \end{bmatrix}}_{B} \underbrace{\begin{bmatrix} 0 & 0 & \dots & 0 & 1 \\ 0 & & 0 & 1 & 0 \\ \vdots & \ddots & 1 & 0 & \vdots \\ 0 & \ddots & \ddots & & 0 \\ 1 & 0 & \dots & 0 & 0 \end{bmatrix}}_{Y}$$

is the real *symmetric* matrix

$$\underbrace{\begin{bmatrix} a_{-n+1} & a_{-n+2} & \dots & a_{-1} & a_0 \\ a_{-n+2} & & a_{-1} & a_0 & a_1 \\ \vdots & \ddots & a_0 & a_1 & \vdots \\ a_{-1} & \ddots & \ddots & & a_{n-2} \\ a_0 & a_1 & \dots & a_{n-2} & a_{n-1} \end{bmatrix}}_{\widehat{B}}.$$

Thus the simple trick of reversing the order of the unknowns which is effected by multiplication with Y yields a matrix for which we can get theoretical a priori convergence bounds for MINRES based only on eigenvalues. We comment that the (Hankel) matrix $\widehat{B}$ is most likely indefinite, but it is clearly symmetric. Premultiplication by Y leads to similar conclusions: see Pestana and Wathen [2015].

It is quite likely that MINRES applied to any linear system involving $\widehat{B}$ would converge slowly, but fortunately it is well-known that Toeplitz matrices are well preconditioned by related circulant matrices in many cases (see Chan [1988],Strang [1986],Tyrtyshnikov [1996],Tyrtyshnikov et al. [1997]). Any circulant matrix $C \in \mathbb{R}^{n \times n}$ is diagonalised as $C = U^\star \Lambda U$ by a Fast Fourier Transform (FFT) Cooley and Tukey [1965] in $O(n \log n)$ operations and so matrix multiplication by a vector or solution of equations with a circulant is computationally achieved in $O(n \log n)$ operations. For many Toeplitz matrices which have sufficient decay in the entries in the first row and column moving away from the diagonal it is known that

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

where R is of small rank and E is of small norm. This implies that the eigenvalues of the preconditioned matrix $C^{-1}B$ are clustered around 1 except for a few outliers. Precise statements about the decay of entries are usually expressed in terms of the smoothness of the generating function associated with the Toeplitz matrix which relates to the decay of Fourier coefficients and thus the speed of convergence of Fourier series.

Now, for use with MINRES a symmetric and positive definite preconditioner is required (see Wathen [2015]). Fortunately via the FFT diagonalisation this is easily achieved by taking the absolute value

$$|C| = U^\star |\Lambda| U \tag{1}$$

where $|\Lambda|$ is just the diagonal matrix of absolute values of the eigenvalues for an appropriate (e.g. Strang or Chan) circulant, C. For a nonsymmetric Toeplitz matrix with decay of entries as above, there now follows.

Theorem 2.1 *Pestana and Wathen [2015]*

$$|C|^{-1}\widehat{B} = J + \widehat{R} + \widehat{E}$$

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

The eigenvalues of $|C|^{-1}\widehat{B}$ are thus clustered around ± 1 together with a few outliers and guaranteed rapid convergence follows [Elman et al., 2014, Chapter 4].

A very simple example demonstrates the point: let

$$B = \begin{bmatrix} 1 & 0.01 & & & \\ 1 & 1 & 0.01 & & \\ & \ddots & \ddots & \ddots & \\ & & 1 & 1 & 0.01 \\ & & & 1 & 1 \end{bmatrix} \in \mathbb{R}^{n \times n} \tag{2}$$

with preconditioning via the Strang preconditioner (which simply takes C as B but with an additional 1 in the nth entry of the first row and 0.01 in the first entry of the nth row). The result of (implicitly) reordering/multiplying by Y and preconditioning with $|C|$ are shown in the MINRES iteration counts in Table 1 for a randomly generated right hand side vector. Convergence is accepted when the preconditioned residual vector has norm less than 10^{-10} for the results shown. The eigenvalues of the preconditioned matrix are shown in Table 2.

Table 1 Condition numbers $\kappa(B)$ for the Toeplitz matrix B described in (2) and iteration counts for *MINRES* applied to the symmetrized matrix $\widehat{B}$ with preconditioner $|C|$.

n	$\kappa(B)$	Iterations
10	14	6
100	207	6
1000	2.6×10^6	6

Table 2 Eigenvalues of the Toeplitz matrix as described in (2) preconditioned with absolute value circulant (to 4 decimal places). Repeated eigenvalues are shown in brackets with the number of repeated eigenvalues indicated.

n	Eigenvalues of $|C|^{-1}\widehat{B}$
10	$\{-9.9107, -1.0002, (-1 \times 2), -0.9640, 0.9893, (1 \times 4)\}$
100	$\{-2.2803, -1.0007, (-1 \times 47), -0.2536, 0.9919, (1 \times 49)\}$
1000	$\{-2.1626, -1.0008, (-1 \times 497), -1.8309e\text{-}5, 0.9929, (1 \times 499)\}$

In fact for this example one can prove these and simpler results via consideration of low rank updates and the degree of the minimal polynomial so it is also possible to prove that GMRES will terminate in just a few iterations.

Table 3 Preconditioned *MINRES* convergence for dense nonsymmetric Toeplitz matrices of Wiener class with absolute value circulant preconditioner.

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

For a dense Toeplitz with sufficient decay of entries in the first row and column this is not the case however, so the results presented in Table 3 for random nonsymmetric Toeplitz matrices of so-called Wiener class (see e.g. [Ng, 2004, page 51]) are not explained by any other means as far as we know, but are a demonstration of the theory presented here. The matrices for these numerical experiments were generated by initially selecting independently the entries of two n-vectors, r and c with $r_1 = c_1$ from a normal distribution with mean zero and variance 1 (using the `randn` command in Matlab), then setting $r_i \leftarrow r_i/(i^2), c_i \leftarrow c_i/(i^2)$ and using these vectors as the first row and column of the nonsymmetric Toeplitz matrix, B.

3 Preconditioning for time-dependence

3.1 Theta method

Here, we consider only a scalar linear ordinary differential equation,

$$\frac{\mathrm{d}y}{\mathrm{d}t} = ay + f, \qquad y(0) = y_0$$

on the time interval $[0, T]$. For the solution of systems of ODE and PDE problems via the method of lines, see McDonald et al.. Likewise to begin with for simplicity we consider only the simple two-level θ-method, which gives,

$$\frac{\mathrm{y}^{n+1} - \mathrm{y}^n}{\tau} = \theta a \mathrm{y}^{n+1} + (1 - \theta) a \mathrm{y}^n + f^n, \qquad \mathrm{y}^0 = y_0,$$

where τ is the constant time step with $N\tau = T$. The discrete equations to be solved are

$$(1 - a\theta\tau)\mathrm{y}^{n+1} = (1 + a(1 - \theta)\tau)\mathrm{y}^n + \tau f^n, \qquad n = 0, 1, 2, \ldots, N - 1, \quad (3)$$

with $\mathrm{y}^0 = y_0$.

The usual approach would be to solve the equations (3) sequentially for $n = 0, 1, 2, \ldots$ which is exactly forward substitution for the all-at-once system

$$B \underbrace{\begin{bmatrix} \mathrm{y}^1 \\ \mathrm{y}^2 \\ \mathrm{y}^3 \\ \vdots \\ \mathrm{y}^N \end{bmatrix}}_{\mathrm{y}} = \underbrace{\begin{bmatrix} \tau f^1 + (1 + a(1 - \theta)\tau)\mathrm{y}^0 \\ \tau f^2 \\ \tau f^3 \\ \vdots \\ f^N \end{bmatrix}}_{\mathrm{f}}$$

where

$$B = \begin{bmatrix} 1 - a\theta\tau & & & & \\ -1 - a(1 - \theta)\tau & 1 - a\theta\tau & & & \\ & -1 - a(1 - \theta)\tau & 1 - a\theta\tau & & \\ & & \ddots & \ddots & \\ & & & -1 - a(1 - \theta)\tau & 1 - a\theta\tau \end{bmatrix}.$$
$$(4)$$

However, we can note that the coefficient matrix B, in the all-at-once system is real Toeplitz, hence solution using the idea in the section above is possible. MINRES for $\widehat{B}\mathrm{y} = BY\mathrm{y} = \mathrm{f}, \mathrm{x} = Y\mathrm{y}$ converges in 4 iterations independently of N as can be seen from the results in Tables 4 and 5 below. The parameter values for the presented results are $a = -0.3, \tau = 0.2, \theta = 0.8$; similar behaviour has been observed for many other sets of parameter values. The eigenvalues of the preconditioned matrix for this problem are shown in Table 5.

Table 4 Condition numbers $\kappa(B)$ for a time-dependent linear ODE using the Θ-method, i.e. for B given by (4) and *MINRES* iteration counts with absolute value Strang circulant preconditioner described by (1) applied to the symmetrized matrix $\widehat{B}$.

N	$\kappa(B)$	Iterations
10	10.474	4
100	30.852	4
1000	33.887	4

For such a bidiagonal Toeplitz matrix, with Strang circulant preconditioning, one can show that the minimal polynomial is quadratic, hence this is a rare situation in which it is possible to deduce that GMRES must terminate with the solution after 2 iterations.

Table 5 Eigenvalues of the preconditioned system (to 4 decimal places). Repeated eigenvalues are shown in brackets with the number of repeated eigenvalues indicated.

N	Eigenvalues of $|C|^{-1}\widehat{B}$
10	$\{-0.7206, (-1 \times 4), (1 \times 4), 3.1155\}$
100	$\{-0.4975, (-1 \times 49), (1 \times 49), 2.0157\}$
1000	$\{-0.4966, (-1 \times 499), (1 \times 499), 2.0139\}$

Theorem 3.1 *Let α and $\beta \neq 0 \in \mathbb{C}$. If*

$$
B = \begin{bmatrix}
\alpha & & & & \\
\beta & \alpha & & & \\
& \ddots & \ddots & & \\
& & \beta & \alpha & \\
& & & \beta & \alpha
\end{bmatrix} \in \mathbb{C}^{n \times n}
$$

is preconditioned by

$$
C = \begin{bmatrix}
\alpha & & & & \beta \\
\beta & \alpha & & & \\
& \ddots & \ddots & & \\
& & \beta & \alpha & \\
& & & \beta & \alpha
\end{bmatrix},
$$

the minimal polynomial of the preconditioned system $T = C^{-1}B$ is quadratic provided that both B and C are nonsingular.

Proof. A simple calculation gives

$$
T = C^{-1}B = \begin{bmatrix}
1 & & & & \frac{-\alpha^{n-1}\beta}{\det C} \\
& 1 & & & \frac{\alpha^{n-2}\beta^2}{\det C} \\
& & \ddots & & \vdots \\
& & & 1 & \frac{(-1)^{n-1}\alpha\beta^{n-1}}{\det C} \\
& & & & \frac{\alpha^n}{\det C}
\end{bmatrix},
$$

where

$$
\det C = \begin{cases} \alpha^n + \beta^n & \text{when } n \text{ is odd} \\ \alpha^n - \beta^n & \text{when } n \text{ is even} \end{cases}.
$$

We can now easily show that T satisfies

$$
(T - I)(T - \frac{\alpha^n}{\det C}I) = 0.
$$

Since $(T - I) \neq 0$ and $(T - \frac{\alpha^n}{\det C} I) \neq 0$, $(T - I)(T - \frac{\alpha^n}{\det C} I)$ is the minimal polynomial of the preconditioned system T.

Since the minimal polynomial for the preconditioned coefficient matrix is in this case quadratic we must therefore have that the Krylov subspace is of dimension 2 and so because of its minimisation property, GMRES termination must occur within 2 iterations.

3.2 Multi-step method

In order to examine a slightly more complex system where the minimal polynomial is not as trivial as with the theta method above, we examine also a 2-step BDF time stepping method. We now require two initial conditions y_{-1} and y_0. For the BDF2 method we have

$$\frac{y^{n+1} - \frac{4}{3}y^n + \frac{1}{3}y^{n-1}}{\tau} = \frac{2}{3}a y^{n+1} + \frac{2}{3} f^{n+1}, \qquad y^0 = y_0, \quad y^{-1} = y_{-1}$$

where τ is the constant time step with $N\tau = T$. The discrete equations to be solved are

$$(1 - \frac{2}{3}a\tau)y^{n+1} = \frac{4}{3}y^n - \frac{1}{3}y^{n-1} + \frac{2}{3}\tau f^{n+1}, \qquad n = 0, 1, 2, \ldots, N-1$$

with $y^0 = y_0$ and $y^{-1} = y_{-1}$. The corresponding all-at-once system is

$$B \underbrace{\begin{bmatrix} y^1 \\ y^2 \\ y^3 \\ \vdots \\ y^N \end{bmatrix}}_{\text{y}} = \underbrace{\begin{bmatrix} \frac{2}{3}\tau f^1 + \frac{4}{3}y^0 - \frac{1}{3}y^{-1} \\ \frac{2}{3}\tau f^2 - \frac{1}{3}y^0 \\ \frac{2}{3}\tau f^3 \\ \vdots \\ \frac{2}{3}\tau f^N \end{bmatrix}}_{\text{f}}$$

where

$$B = \begin{bmatrix} 1 - \frac{2}{3}a\tau & & & & \\ -\frac{4}{3} & 1 - \frac{2}{3}a\tau & & & \\ \frac{1}{3} & -\frac{4}{3} & 1 - \frac{2}{3}a\tau & & \\ & \ddots & \ddots & \ddots & \\ & & \frac{1}{3} & -\frac{4}{3} & 1 - \frac{2}{3}a\tau \end{bmatrix}. \tag{5}$$

The coefficient matrix B in (5) has an additional subdiagonal but is still Toeplitz and the method above therefore still applies. Applying MINRES to solve the equation $\widehat{B}y = BY y = f, x = Yy$ with a random starting vector, we see convergence in 6 iterations independently of N as can be seen from the results in Table 6. The parameter values for the presented results are

again chosen as $a = -0.3$ and $\tau = 0.2$ with zero forcing but the behaviour does not change for many other choices of a and τ. As we have used implicit time-stepping we have no restrictions on the value of τ to maintain stability and, as Theorem 3.1 seems to indicate, it is only the lower diagonal Toeplitz structure of B which ensures the number of unique eigenvalues of $C^{-1}B$ so it is not surprising that other parameter values behaviour in the same manner for the symmetrized system. The eigenvalues of the preconditioned matrix in this case are shown in Table 7.

Table 6 Condition numbers $\kappa(B)$ for a time-dependent linear ODE using the BDF2 method, i.e. for B given by (5) and *MINRES* iteration counts with absolute value Strang circulant preconditioner described by (1) applied to the symmetrized matrix $\widehat{B}$.

N	$\kappa(B)$	Iterations
10	29.33	6
100	67.49	6
1000	67.67	6

Table 7 Eigenvalues of the preconditioned system (to 4 decimal places). Repeated eigenvalues are shown in brackets with the number of repeated eigenvalues indicated.

N	Eigenvalues of $|C|^{-1}\widehat{B}$
10	$\{-1.0442, (-1 \times 3), -0.6781, 0.9219, (1 \times 3), 3.3921\}$
100	$\{-1.0610, (-1 \times 48), -0.4410, 0.9424, (1 \times 48), 2.2736\}$
1000	$\{-1.0610, (-1 \times 498), -0.4401, 0.9425, (1 \times 498), 2.2720\}$

This approach for time-dependent problems may not seem of any advantage for such a simple problems as considered here because MINRES requires matrix vector multiplication with B (and Y) as well as solution of a system with $|C|$ at each iteration. Its potential for time-dependent PDEs is however more intriguing (see McDonald et al.).

4 Conclusions

Preconditioning for nonsymmetric linear systems is generally heuristic with no guarantee of the speed of convergence from a priori spectral estimation. This is in stark contrast to the case of real symmetric or complex Hermitian matrices. We have shown that for nonsymmetric real Toeplitz matrices the use of a simple trick gives symmetry so that convergence estimates for MINRES which are based only on eigenvalues rigorously apply. Further, we

propose the use of an absolute value circulant matrix as preconditioner: the action of this preconditioner is effected in $O(n \log n)$ operations via use of the FFT as originally suggested in Strang [1986]. These constructions apply independently of nonnormality and rapid, n-independent convergence is guaranteed and hence observed.

It is further observed how this preconditioning can be applied in the context of time-stepping problems and that convergence is achieved in a small number of iterations independent of the number of time-steps.

References

T.F. Chan. An optimal circulant preconditioner fro Toeplitz systems. *J. Sci. Statist. Comput.*, 9:766–771, 1988.

J.W Cooley and J.W Tukey. An algorithm for the machine calculation of complex Fourier series. *Math. Comput.*, 19:297–301, 1965.

H.C. Elman, D. J. Silvester, and A. J. Wathen. *Finite Elements and Fast Iterative Solvers with Applications in Incompressible Fluid Dynamics.* Numerical Mathematics and Scientific Computation. Oxford Univ. Press, UK, Oxford, 2nd edition edition, 2014.

Anne Greenbaum, V. Ptak, and Z. Strakos. Any nonincreasing convergence curve is possible for GMRES. *SIAM J. Matrix Anal.*, 17(3):465–469, 1996.

M. R. Hestenes and E. Stiefel. Methods of conjugate gradients for solving linear systems. *Journal of Research of the National Bureau of Standards*, 49:409–435, 1952. URL `nvl.nist.gov/pub/nistpubs/jres/049/6/V49.N06.A08.pdf`.

D. Loghin and A. J. Wathen. Analysis of preconditioners for saddle-point problems. *SIAM J. Sci. Comp.*, 25(6):2029–2049, 2004. doi: 10.1137/S1064827502418203. URL `http://epubs.siam.org/doi/abs/10.1137/S1064827502418203`.

E. McDonald, J. Pestana, and A. J. Wathen. Preconditioning and iterative solution of all-at-once systems for evolutionary partial differential equations. In preparation.

Michael K. Ng. *Iterative Methods for Toeplitz Systems (Numerical Mathematics and Scientific Computation).* Oxford University Press, Inc., New York, NY, USA, 2004. ISBN 0198504209.

C. Paige and M. Saunders. Solution of sparse indefinite systems of linear equations. *SIAM J. Numer. Anal*, 12:617–629, 1975.

J. Pestana and A. J. Wathen. A Preconditioned MINRES Method for Nonsymmetric Toeplitz Matrices. *SIAM Journal on Matrix Analysis and Applications*, 36(1):273–288, 2015.

Y. Saad and M. H. Schultz. GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems. *SIAM J. Sci. Stat. Com-*

put., 7(3):856–869, July 1986. ISSN 0196-5204. doi: 10.1137/0907058. URL `http://dx.doi.org/10.1137/0907058`.

G. Strang. A proposal for toeplitz matrix calculations. *Stud. Appl. Math.*, 74:171–176, 1986.

J.D. Tebbens and G. Meurant. Any Ritz value behavior is possible for Arnoldi and for GMRES. *SIAM J. Matrix Anal. Appl.*, 33:958–978, 2012.

J.D. Tebbens and G. Meurant. Prescribing the behavior of early terminating GMRES and Arnoldi iterations. *Numerical Algorithms*, 65:69–90, 2014.

E.E. Tyrtyshnikov. A unifying approach to some old and new theorems on distribution and clustering. *Linear Algebra Appl.*, 232:1–43, 1996.

E.E. Tyrtyshnikov, A.Y. Yeremin, and N.L. Zamarashkin. Clusters, preconditioners, convergence. *Linear Algebra Appl.*, 263:25–48, 1997.

A. J. Wathen. Preconditioning. *Acta Numerica*, 24:329–376, 2015.