

# SOME OBSERVATIONS ON PRECONDITIONING FOR NON-SELF-ADJOINT AND TIME-DEPENDENT PROBLEMS

ANDY WATHEN\*

**Abstract.** Numerical Linear Algebra—specifically the computational solution of equations—forms a significant part of Computational Methods for Partial Differential Equations. Here we discuss the contrast between the solution of symmetric systems of equations that arise from self-adjoint problems and non-symmetric systems that arise from non-self-adjoint problems when iterative methods are employed; such methods are the only feasible methods for very large scale computation with PDEs. We then go on to consider non-symmetric all-at-once systems that arise in approximation of time-dependent problems, discuss causality and the parallel-in-time paradigm, suggesting an approach that involves preconditioning initial value problems with time-periodic problems.

**1. Introduction.** Computational Partial Differential Equations is a large and vibrant research area. The computational solution of linear(ized) equations that arise from whatever approximation scheme is employed is commonly a major task involving methods of numerical linear algebra. Sparse direct methods are applicable for many problems, but, in particular for 3-dimensional domain problems, iterative methods usually present the only effective solution approach in combination with preconditioning [32] [8]. Almost always, very slow convergence is observed when an effective preconditioner (matrix approximation) is not used. Simple stationary iterations, including multigrid cycles, can be effectively used for some problems, but more generally applicable are Krylov subspace methods [27] [30]. (Multigrid cycles are really effective parts of many preconditioners: see for example [10, chapter 4]).

For self-adjoint problems, symmetric (or Hermitian) matrices generally arise and a common approach is to use the Conjugate Gradient method (CG) [16] (for definite problems) or the MINRES method [24] (for indefinite problems) with some appropriate symmetric preconditioner. In this situation, a priori descriptive convergence bounds for convergence of the iterative method depend solely on the eigenvalue spectrum of the preconditioned matrix: thus establishing estimates of the eigenvalues is all that is required to reliably predict the number of iterations needed for solution. Fewer iterations than predicted by these bounds can occur, but never more. The mathematics thus indicates what is required of a good preconditioner for a self-adjoint problem (see [32]).

By contrast, for non-self-adjoint problems when non-symmetric (non-Hermitian) matrices arise from approximation, though iterative solution methods are widely used, no generally descriptive convergence bounds are known. There are situations where special techniques can be used, but preconditioning is generally necessarily heuristic for non-symmetric matrices except in rare cases when preconditioning induces symmetry [32] or where self-adjointness in a non-standard inner-product exists [25].

Time-dependent PDE problems which are first order in time necessarily are non-self-adjoint and in model situations even give rise to matrices of the form  $I + F$  where  $F$  is nilpotent of high index. Thus non-diagonalisable matrices and Jordan structure are centrally relevant for such problems. In some sense, such time-dependent problems give rise to matrices that are furthest from the ideal cases that arise for self-adjoint problems as described above. Nevertheless, some effective preconditioners for such problems have been suggested and even theory guaranteeing fast convergence of certain iterations for such problems has been established. We will describe the relevant

---

\*Oxford University Mathematical Institute, Oxford OX2 6GG, UK

structures in this short paper in which we also seek to clarify the symmetric/non-symmetric dichotomy.

**2. The symmetric/non-symmetric dichotomy.** For a linear system  $Bx = c$ , with  $B \in \mathbb{R}^{m \times m}$ , from an initial guess  $x_0$ , a Krylov subspace method generates iterates  $x_1, x_2, \dots, x_k, \dots$  using one matrix-vector product at each iteration,  $k$ , thus giving rise to a (mathematical) basis for a Krylov subspace

$$\{r, Br, B(Br), \dots, B^k r\}.$$

There follows that the residuals  $r_k = c - Bx_k$  satisfy  $r = r_0$  and

$$r_k = p_k(B)r_0 \quad (2.1)$$

where  $p_k$  is a polynomial of degree less or equal to  $k$  satisfying  $p_k(0) = 1$ . Thus if  $B$  is diagonalisable and we have  $B = X\Lambda X^{-1}$  then

$$\|r_k\| \leq \|X\| \|p_k(\Lambda)\| \|X^{-1}\| \|r_0\|, \quad (2.2)$$

and if  $B = B^T$  so that  $X^{-1} = X^T$ , then this bound on convergence in  $\|\cdot\|_2$  depends only on the eigenvalues of  $B$  since the orthogonal eigenvector basis gives  $\|X\|_2 = \|X^{-1}\|_2 = 1$ . In the definite case when  $B$  is also (without loss of generality) positive definite, this type of convergence bound arises most naturally in a norm based on  $B$ , but the same conclusion holds: for symmetric matrices, well distributed (clustered) eigenvalues  $\lambda_i, i = 1, \dots, m$  imply fast convergence because only low degree polynomials are required to ensure that the quantity  $\|p_k(\Lambda)\| \equiv \max_i |p_k(\lambda_i)|$  and thus  $\|r_k\|$  is acceptably small. This is certainly true for CG and MINRES since these methods are based on optimality properties which guarantee that *best* polynomials,  $p_k$ , are computed in each iteration. For details see, for example, [10].

For non-symmetric matrices, there are many possible Krylov subspace iterative methods since—as very nicely indicated in [22]—there is no single method of choice for all problems. The most commonly used method, however, is the GMRES method [28] which, like MINRES in the symmetric case, generates iterates that minimise  $\|r_k\|_2$  in the appropriate Krylov subspace at each iteration.

For a diagonalisable matrix, as above the eigenvalue/eigenvector bound

$$\|r_k\|/\|r_0\| \leq \|X\| \|X^{-1}\| \min_{p \in \Pi_k, p(0)=1} \max_j |p(\lambda_j)|$$

does hold for GMRES, but for non-symmetric matrices the quantity  $\|X\| \|X^{-1}\|$  can be arbitrarily large, rendering the bound less useful. (For non-diagonalisable matrices there is a similar but more complex bound, see [11, 17, 18]).

There are other GMRES convergence bounds, for example based on the field of values  $W(B) = \{x^* B x / x^* x; x \neq 0\}$  [7]:

$$\|r_k\|/\|r_0\| \leq 2 \min_{p \in \Pi_k, p(0)=1} \max_{z \in W(B)} |p(z)|$$

or the  $(\epsilon)$ -pseudospectrum  $\Lambda_\epsilon(B) = \{z \in \mathbb{C} : \|(zI - B)^{-1}\| > \epsilon^{-1}\}$  [31]:

$$\|r_k\|/\|r_0\| \leq \mathcal{L}(\partial\Lambda_\epsilon)/2\pi\epsilon \min_{p \in \Pi_k, p(0)=1} \max_{z \in \Lambda_\epsilon(B)} |p(z)|,$$

where  $\mathcal{L}(\partial\Lambda_\epsilon)$  denotes the length of the  $(\epsilon)$ -pseudospectral boundary, but only in special situations are they of utility.

Thus for self-adjoint problems which lead to symmetric linear(ised) problems, there are rigorous design criteria for preconditioners, namely that the preconditioned system matrices should have nicely distributed (clustered) real eigenvalues—for indefinite problems this requires that preconditioners be positive definite as well as symmetric. For many self-adjoint problems, effective and efficient preconditioners have been derived to satisfy these criteria [1, 3, 27, 32]. The lack of a generally applicable and *descriptive* convergence bound for GMRES (or any of the other proposed methods for non-symmetric systems) means that there is less guidance for what conditions a good preconditioner for a non-symmetric matrix should possess in general.

It is certainly true in any situation that widely spread eigenvalues can lead to very slow iterative convergence and several authors have tried to use eigenvalue clustering as a criterion for constructing preconditioners even for non-symmetric matrices. A note of warning to this approach is, however, provided by a well-developed theory initiated by Greenbaum, Ptak and Strakos in [15] which proves such results as:

**THEOREM 2.1.** *Given any non-increasing positive sequence  $f(0) \geq f(1) \geq \dots \geq f(m-1) > 0$  there exists  $B \in \mathbb{R}^{m \times m}$  and  $c \in \mathbb{R}^m$  with  $\|c - Bx_0\| = f(0)$  such that  $f(k) = \|c - Bx_k\|$ ,  $k = 1, \dots, m-1$ , where  $x_k$  is the iterate at step  $k$  of the GMRES algorithm applied to  $Bx = c$ , with initial iterate  $x_0$ . Moreover, the matrix  $B$  can be chosen to have any eigenvalues.*

Thus even for a non-symmetric matrix with eigenvalues all equal to 1, one can get stagnation of GMRES iteration for  $m-1$  iterations ( $f(0) = f(1) = \dots = f(m-1) > 0$ ) before finally jumping straight to the solution at the  $m^{\text{th}}$  iteration (as must happen because of the Cayley-Hamilton theorem). Some of the time-dependent examples that we consider below give nearly such worst case matrices, however, really effective preconditioners for these problems are available as we shall show.

The above theorem is generally interpreted to mean that eigenvalues do not determine GMRES convergence, but it says nothing about the likelihood of encountering such a matrix. I would summarise that for a symmetric system convergence of iterative methods can be bounded only in terms of eigenvalues; badly distributed eigenvalues can give bad convergence of iterative methods, but for non-symmetric systems even with good eigenvalues one can get arbitrarily bad convergence.

**3. Preconditioning for time-dependent problems.** We start first with the simple linear ordinary differential equation initial value problem

$$y' = ay + f, \quad y(0) = y_0$$

that we discretise with a simple  $\theta$ -method to integrate from 0 up to time  $T$ , giving the equations

$$\frac{y^k - y^{k-1}}{\tau} = \theta a y^k + (1 - \theta) a y^{k-1} + f^{k-1}, \quad y^0 = y_0, \quad (3.1)$$

$k = 1, 2, \dots, \ell$  with  $\ell\tau = T$ ,  $\tau$  being the time-step. As a linear all-at-once system this gives the matrix equation

$$B \underbrace{\begin{bmatrix} y^1 \\ y^2 \\ y^3 \\ \vdots \\ y^\ell \end{bmatrix}}_y = \underbrace{\begin{bmatrix} \tau f^1 + (1 + a(1 - \theta)\tau)y^0 \\ \tau f^2 \\ \tau f^3 \\ \vdots \\ \tau f^\ell \end{bmatrix}}_f, \quad (3.2)$$

where the  $\ell \times \ell$  coefficient matrix  $B$  is

$$\begin{bmatrix} b & & & & \\ c & b & & & \\ & c & b & & \\ & & \ddots & \ddots & \\ & & & c & b \end{bmatrix}, \quad (3.3)$$

with  $b = 1 - a\theta\tau$ ,  $c = -1 - a(1 - \theta)\tau$ . Notice that  $B$  is a bidiagonal Toeplitz (constant diagonal) matrix. Simple forward substitution for such a lower triangular matrix would, of course, deliver the solution and this process would be exactly the usual sequential time evolution: solve (3.1) for  $k = 1$ , then for  $k = 2$ , etc. Thinking of the all-at-once system (3.2) however allows for other solution approaches that might be more amenable to parallel processing. Krylov subspace methods, for example, require only to ‘visit’ the coefficient matrix to perform a matrix-vector product as well as other vector operations; such a product and the further operations are inherently more parallel. This observation leads to possibilities for parallel-in-time computation [12]; the standard forward substitution would require  $O(\ell)$  clock cycles to complete regardless of how many parallel processing units are available, whereas given  $\ell$  parallel processing units, the matrix-vector product could readily be completed in time that does not scale with  $\ell$ .

The coefficient matrix,  $B$  however is not diagonalisable, being a simple scaling and transposition from a single Jordan block, so it is far from having such a nice property as symmetry. Nevertheless, there exists a close matrix approximation (i.e. a possible preconditioner)

$$C = \begin{bmatrix} b & & & & c \\ c & b & & & \\ & c & b & & \\ & & \ddots & \ddots & \\ & & & c & b \end{bmatrix}$$

that is *circulant*; all  $\ell \times \ell$  circulant matrices are diagonalised in  $O(\ell \log(\ell))$  operations by a Fast Fourier Transform (FFT) and linear systems with  $C$  as coefficient matrix can correspondingly be solved in  $O(\ell \log(\ell))$  operations [5]. Moreover, the FFT can be parallelised—for an elementary description of some of the possibilities see [4]—and good implementations are available [9]. Thus with sufficient processors, one may achieve application of the preconditioner—and thus a single preconditioned Krylov subspace iteration—in rather less than  $O(\ell)$  time. This fast solvability is one requirement of a good preconditioner, the other is that  $C$  should be a good approximation to  $B$  so that few iterations are required for convergence. It is clear that  $C$  differs from  $B$  in only one entry, but this does not immediately imply that an appropriate non-symmetric iteration such as GMRES will converge in few iterations for the system (3.2) with  $C$  as preconditioner. The eigenvalues of  $C^{-1}B$  are all 1 except for one, but it is still non-symmetric!

Before we examine this further, let us step back for a second and realise what we have suggested: that an initial value problem on  $[0, T]$  be approximated by a kind of periodic problem on  $[0, T]$ . That is, the preconditioner represents the same discretisation of the same ODE, but with the initial condition  $y(0) = y_0$  replaced by a relationship between  $y(0)$  and  $y(T)$ . In this sense, one might expect that the precise value of the  $(1, \ell)$  entry of  $C$  might not matter too much and this is indeed the case:

by a simple change of phase, a so-called  $\omega$ -circulant

$$C_\epsilon = \begin{bmatrix} b & & & & \epsilon \\ c & b & & & \\ & c & b & & \\ & & \ddots & \ddots & \\ & & & c & b \end{bmatrix}$$

with an arbitrary value  $\epsilon$  can be used with the FFT as has been pointed out by Liu and Wu [19]. The closer that the parameter  $\epsilon$  gets to zero, the closer is the approximation of  $B$  by  $C_\epsilon$  and, indeed, the faster is the convergence of GMRES as shown in [19]. As the value  $\epsilon = 0$  is approached numerical problem start to arise—of course  $B$  well-approximates  $B$ , but then the FFT no longer applies—but small values can be employed nevertheless.

Returning to iterative methods, a proof of rapid convergence of a Krylov subspace method for (3.2) employing a matrix related to  $C$  as a preconditioner is provided by Pestana and the author: the trick they use is to employ the (self-inverse) ‘flip’ matrix

$$Y = \begin{bmatrix} 0 & 0 & \cdot & 0 & 1 \\ 0 & \cdot & 0 & 1 & 0 \\ \cdot & 0 & 1 & 0 & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & 0 & \cdot & 0 & 0 \end{bmatrix}$$

which renders any real Toeplitz matrix as a real symmetric matrix by either pre- or post-multiplication; hence instead of  $Bx = c$  we can solve either of the symmetric systems  $YBx = Yc$  or  $BYy = c$  for  $y$  and recover  $x = Yy$ . In practice, of course, multiplication with  $Y$  can be implemented trivially as a permutation. With such a real symmetric matrix,  $YB$  or  $BY$ , the convergence of MINRES is bounded only in terms of eigenvalues so long as a symmetric and positive definite preconditioner is used as described above. By further using the diagonalisation  $C = F\Lambda F^*$  provided by the FFT for the circulant approximation, one is then able to use the absolute value circulant  $|C| = F|\Lambda|F^*$  as a symmetric and positive definite preconditioner (see [26] for details). The symmetric theory thus applies rigorously in this setup and convergence in a small number of iterative steps is guaranteed for many non-symmetric Toeplitz systems as has been long established for many common types of symmetric Toeplitz system since the early suggestion of this approach by Strang [29] (see for example [23]). The original proofs of Raymond Chan and others [2] for symmetric Toeplitz matrices whose entries decay sufficiently away from the matrix diagonal, established that the eigenvalues of  $C^{-1}B$  are clustered about 1 except for a few outliers. Thus the convergence bound (2.2) guarantees rapid convergence of CG or MINRES for such matrices. For real non-symmetric Toeplitz systems, with the Pestana and Wathen trick, similar results hold when there is sufficient decay, thus rapid convergence of MINRES—in a number of iterations independent of  $\ell$ —is guaranteed also in this case (see [26]). (It is certainly clear that the decay conditions hold for the bi-diagonal matrix  $B$ !)

This rigorously proved fast convergence for MINRES with the flip matrix to ensure symmetry and the absolute value circulant preconditioner gives one guaranteed way to solve the Toeplitz all-at-once systems, but perhaps more importantly it indicates that such systems together with appropriate circulant preconditioners are ‘friendly’ for Krylov methods. (This intuition and the phrase ‘Krylov friendly’ were shared with

us by Tim Kelley). Numerical experiments using GMRES with  $C$  as preconditioner seem to demonstrate that this is also a good strategy for solution, since the required number of iterations for a similar convergence tolerance appears to be about a half of the number of MINRES iterations required with  $|C|$  as preconditioner (see [21]). Of course, this comes with no guarantee and as yet we have no proof—once again we see the limitation of moving to non-symmetric technology!

The above description assumes that the time-step,  $\tau$ , is a constant and that coefficients are constants. The theory is strictly inapplicable when either varying time-steps or variable coefficients arise, but practical computation seems to show that some limited variation can be tolerated without losing all useful properties: averaging along diagonals is a simple expedient, for example, for allowing circulant preconditioners (approximations) to be defined. A contrasting approach that specifically uses variable time-steps and diagonalisation as a parallel-in-time approach has been developed by Gander and Halpern (see, for example, [13]).

The real interest is time-dependent PDEs, where efficient linear solvers are usually required to enable feasible computation. In particular, it is for such large scale problems that effective parallel-in-time computation is desired.

We concentrate here on the simplest parabolic problem

$$\begin{aligned} u_t &= \Delta u + f && \text{in } \Omega \times (0, T], \quad \Omega \subset \mathbb{R}^2 \text{ or } \mathbb{R}^3, \\ u &= g && \text{on } \partial\Omega, \\ u(x, 0) &= u_0(x) && \text{at } t = 0, \end{aligned}$$

but have worked on similar ideas for simple hyperbolic wave equations

$$\begin{aligned} u_{tt} &= u_{xx} + f && \text{in } [a, b] \times (0, T], \\ u(a, t) &= \alpha && \text{and} \quad u(b, t) = \beta, \\ u(x, 0) &= u_0(x) && \text{and} \quad u_t(x, 0) = u'_0(x) \quad \text{at } t = 0, \end{aligned}$$

see [6].

For the parabolic problem, spatial discretisation via Galerkin finite elements (or central finite differences) with mesh size  $h$ ,  $n$  spatial degrees of freedom and time-step  $\tau$ , employing, for simplicity, backwards Euler time stepping, yields

$$M \frac{u_k - u_{k-1}}{\tau} + K u_k = \mathbf{f}_k, \quad k = 1, \dots, \ell,$$

or

$$\mathcal{A}_{BE} \mathbf{x} := \begin{bmatrix} A_0 & & & \\ A_1 & A_0 & & \\ & \ddots & \ddots & \\ & & A_1 & A_0 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_\ell \end{bmatrix} = \begin{bmatrix} M u_0 + \tau \mathbf{f}_1 \\ \tau \mathbf{f}_2 \\ \vdots \\ \tau \mathbf{f}_\ell \end{bmatrix},$$

where  $A_0 = M + \tau K$  is symmetric positive definite and  $A_1 = -M$  is symmetric. Here  $M \in \mathbb{R}^{n \times n}$  is the mass matrix (which is diagonal—a multiple of the identity—for finite differences) and  $K \in \mathbb{R}^{n \times n}$  is the stiffness matrix representing the negative discrete Laplacian - both are symmetric and positive definite. The key thing that we note here is that the all-at-once coefficient matrix  $\mathcal{A}_{BE}$  is a block Toeplitz matrix. Thus following the same line of development as above, we propose to employ the block

circulant preconditioner

$$\mathcal{P}_{BE} := \begin{bmatrix} A_0 & & & A_1 \\ A_1 & A_0 & & \\ & \ddots & \ddots & \\ & & A_1 & A_0 \end{bmatrix}.$$

In [21] it is proved that the preconditioned matrix  $\mathcal{P}_{BE}^{-1}\mathcal{A}_{BE} \in \mathbb{R}^{n\ell \times n\ell}$  is diagonalisable, has  $(\ell - 1)n$  eigenvalues of 1 and  $n$  eigenvalues which cluster around 1 for small  $h$ . Of course  $\mathcal{P}_{BE}^{-1}\mathcal{A}_{BE}$  is non-symmetric, but such eigenvalue knowledge implies rapid Krylov subspace convergence—independently of the number of time-steps,  $\ell$ —if symmetry can be introduced. Fortunately, in this case, since the matrix blocks  $A_0, A_1$  are real symmetric, we can employ the ‘block flip’ matrix

$$\mathcal{Y} = Y \otimes I = \begin{bmatrix} 0 & 0 & \cdot & 0 & I \\ 0 & \cdot & 0 & I & 0 \\ \cdot & 0 & I & 0 & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ I & 0 & \cdot & 0 & 0 \end{bmatrix}$$

where  $I \in \mathbb{R}^{n \times n}$  is the identity matrix and have that  $\mathcal{Y}\mathcal{A}_{BE}$  and  $\mathcal{A}_{BE}\mathcal{Y}$  are real symmetric. The theory now follows very similarly to the case of ODEs and point- rather than block-Toeplitz systems: the result is a guarantee of  $\ell$ -independent convergence for MINRES. Though the number of iterations could still depend on the number of spatial degrees of freedom,  $n$ , the clustering of the  $n$  non-unit eigenvalues about 1 ensures that the number of iterations required to achieve any small tolerance is also essentially independent of  $n$ .

There are many details here that we are skating quickly over - see [21] for precision. The basic point that we want to get across is that even for parabolic PDE problems, the idea of preconditioning an initial value problem with a time-periodic problem (or at least one in which  $u(0)$  is related to  $u(T)$ ) still holds merit. It gives fast convergence apparently also for GMRES without the block flip matrix as well as guaranteed fast convergence for MINRES when flipping is used so that symmetry is ensured. This can lead to an effective and efficient parallel-in-time solver for such problems - see [14] for early parallel numerical results.

It is notable for the parabolic problem that block forward substitution—as in the point case—would correspond precisely to standard sequential time-stepping; sequentially computing  $u_1$  from the initial condition  $u_0$ , then computing  $u_2$  from  $u_1, \dots$ . For the implicit backwards Euler method as above, there would be an  $n \times n$  linear system involving  $A_0$  to solve at each time-step. For explicit Euler with either a lumped mass matrix or with finite differences, the step-by-step computation would be entirely explicit. For finite differences, for example, in the case of forward Euler,

$$\frac{u_k - u_{k-1}}{\tau} + K u_{k-1} = \mathbf{f}_k, \quad k = 1, \dots, \ell,$$

or

$$u_k - u_{k-1} + \tau K u_{k-1} = \tau \mathbf{f}_k, \quad k = 1, \dots, \ell,$$

where we have again used  $K$  for the matrix representing the negative discrete Lapla-

cian, the all-at-once coefficient matrix is

$$\mathcal{A}_{FE}\mathbf{x} := \begin{bmatrix} I & & & \\ A_1 & I & & \\ & \ddots & \ddots & \\ & & A_1 & I \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_\ell \end{bmatrix} = \begin{bmatrix} u_0 + \tau \mathbf{f}_1 \\ \tau \mathbf{f}_2 \\ \vdots \\ \tau \mathbf{f}_\ell \end{bmatrix},$$

where  $A_1$  is here  $-I + \tau K$ . This is the starting point for other parallel-in-time preconditioning approaches (for a recent example, see [33]), as well as other approaches to parallel-in-time computations; see for example [20].

We are not suggesting the use of explicit computation for parabolic equations—the well known severe stability limitations would certainly argue against this—but illustrate this here to show that explicit computation without preconditioning exhibits causality: one needs to have  $u_{j-1}$  before one can compute  $u_j$ . This is also explicitly seen in a Krylov method, since the minimum polynomial,  $m$ , of  $\mathcal{A}_{FE}$  is of degree  $\ell$ , namely  $m(z) = (1 - z)^\ell$ : it would take precisely  $\ell$  iterations to achieve the exact solution on  $[0, \ell\tau = T]$  at least for some starting residual. In fact for a homogeneous initial value problem with zero initial guess (so that the initial residual has only its first block component non-zero), the solution on  $[0, j\tau]$  (but not for any later time) would be correctly calculated at precisely the  $j^{th}$  iteration. In other words in this situation, one would achieve essentially the same as sequential time-stepping when using a Krylov subspace method without preconditioning. This represents one of the worst case situations established as possible through the Greenbaum, Ptak and Strakos theorem above since all of the eigenvalues of  $\mathcal{A}_{FE}$  are clearly 1! In this sense, preconditioning with a periodic problem defeats causality and computation of the solution on  $[0, \ell\tau = T]$  is achievable in many fewer than  $\ell$  iterations/steps.

In fact this observation could easily have been made for the ODE case: the matrix in (3.3) clearly has all of its eigenvalues equal to  $b$ , but since  $B$  is lower bidiagonal,  $B^2$  is lower tridiagonal—that is, lower triangular with only the diagonal and two sub-diagonals non-zero.  $B^j$  is lower triangular with the diagonal and  $j$  sub-diagonals non-zero for  $j = 0, 1, \dots, \ell - 1$  in fact. Thus  $B^{\ell-1}$ , which is a full lower triangular matrix, can not be expressed as a linear combination of lower powers of  $B$  and it follows that the minimum polynomial must be the characteristic polynomial  $(1 - z)^\ell$  in this case. Thus any Krylov subspace method characterised by (2.1) can only terminate with the solution at the  $\ell^{th}$  step at least for some initial residual; for a finite tolerance and a more serendipitous starting guess, convergence could be seen in less iterations. Precisely,  $\frac{1}{b}B = I + F$  where  $F$  is nilpotent of maximal index  $\ell$ . This is evident if we consider the case  $f = 0$ , so that the right hand side vector  $\mathbf{f}$  in (3.2) has only a first entry that is non-zero and if  $x_0 = 0$  so that  $r = r_0 = \mathbf{f}$ , then  $Br$  has only its first two entries non-zero,  $B^2r$  has only its first three entries non-zero and so on. As a consequence the  $j^{th}$  iterate,  $x_j \in \text{span}\{r, Br, \dots, B^{j-1}r\}$  can only have non-zeros in its first  $j$  entries, so all  $\ell$  iterations are needed until a solution iterate can be fully populated with non-zero entries. Given that the solution in this situation is an exponential, hence non-zero for  $y_0 \neq 0$ , we see that without preconditioning the all-at-once solution can not be achieved in less than  $\ell$  iterations; causality applies! Of course, a circulant preconditioner could also be applied to the explicit Euler all-at-once system.

Corresponding analysis for backwards Euler time-stepping along similar lines to the above exposes the same issue; implicit discretisation can not avoid the causality issue. This nilpotency in the linear algebra reflects causality in the differential

equation. It is pretty clear that there is limited scope for diagonal preconditioning or block diagonal preconditioning in the PDE case, since using different diagonal values or blocks will spread out the eigenvalues, and use of the same block will not change this nilpotence structure. A preconditioner which is not so simple is required to defeat causality and so allow parallel-in-time computation - the circulant/block-circulant is just such a preconditioner since the FFT enables rapid treatment of systems involving such circulants.

One other possibility arises with block-circulant preconditioning of block-Toeplitz all-at-once systems: it is not necessary that sequential time-stepping exists as a possibility. That is, it is not necessary to have discretisations for which the all-at-once matrices are block triangular. In other words, application of the time-periodisation idea does not require that sequential time-stepping is possible; one can still build a block circulant approximation for a block Toeplitz matrix that is not block triangular. We do not know if there are useful such discretisations - stability will be a key issue as will be the application of initial conditions. If there are block tridiagonal all-at-once matrices, for example—this would represent a three time-level scheme—then circulant approximations exist with non-zero blocks both in the  $(1, \ell)$  and  $(\ell, 1)$  positions. Such matrices clearly exhibit decay away from the diagonal(!) and fast Fourier techniques are certainly still applicable.

Whether such an observation can lead to new and useful discretisation schemes is an intriguing open question that we leave for the future. What is certainly true is that viewing time-stepping in the rather different all-at-once context leads one to ask different questions and to consider different paradigms as we have done in this article.

### Acknowledgements.

I thank two anonymous referees for their helpful and insightful comments that have led to improvements in this article.

### REFERENCES

- [1] M. BENZI, *Preconditioning techniques for large linear systems: A survey*, J. Comput. Phys., 182 (2002), pp. 418–477.
- [2] R.H. CHAN, *Circulant preconditioners for Hermitian Toeplitz systems*, SIAM J. Matrix Anal. Appl., 10 (4) (1989), pp. 542–550.
- [3] K. CHEN, *Matrix preconditioning techniques and applications*, Cambridge University Press, 2005.
- [4] E. CHU AND A. GEORGE, *Inside the FFT Black Box: Serial and Parallel Fast Fourier Transform Algorithms*, CRC Press, 2019.
- [5] J.W. COOLEY AND J.W. TUKEY, *An algorithm for the machine calculation of complex Fourier series*, Math. Comput., 19 (1965), pp. 297–301.
- [6] F. DANIELI AND A. J. WATHEN, *All-at-once solution of linear wave equations*, accepted for publication in Numer. Linear Alg. Appl. (2021).
- [7] M. EIERMANN, *Fields of values and iterative methods*, Linear Alg. Applic., 180 (1993). pp. 167–197.
- [8] M. FERRONATO, *Preconditioning for Sparse Linear Systems at the Dawn of the 21st Century: History, Current Developments, and Future Perspectives*, International Scholarly Research Notices, 2012.
- [9] [fftw.org](http://fftw.org)
- [10] H. C. ELMAN, D. J. SILVESTER AND A. J. WATHEN, *Finite Elements and Fast Iterative Solvers with Applications in Incompressible Fluid Dynamics*, 2nd edition, Oxford University Press, Oxford, 2014.
- [11] R.W. FREUND, *Quasi-kernel polynomials and convergence results for quasi-minimal residual iterations*. In: Braess, D. and Schumaker, L.L., editors. Numerical Methods in Approximation Theory, Vol. 9. Birkhäuser Verlag (1992), pp. 77–95.
- [12] M. J. GANDER, *50 Years of Time Parallel Time Integration*, In: Carraro T, Geiger M, Körkel S,

- and Rannacher R, editors. Multiple Shooting and Time Domain Decomposition Methods. Springer International Publishing (2014), pp. 69–113.
- [13] M.J. GANDER AND L. HALPERN, *Time Parallelization for Nonlinear Problems Based on Diagonalization*, In: Lee CO, Cai XC, Keyes DE, Kim HH, Klawonn A, Park EJ, et al., editors. Domain Decomposition Methods in Science and Engineering XXIII, Springer International Publishing (2017), pp. 163–170.
  - [14] A. GODDARD AND A. J. WATHEN, *A note on parallel preconditioning for all-at-once evolutionary PDEs*, Electronic Trans. Numer. Anal., 51 (2019), pp. 135–150.
  - [15] A. GREENBAUM, V. PTAK AND Z. STRAKOS, *Any convergence curve is possible for GMRES*, SIAM J. Matrix Anal. Appl. 17 (1996), pp. 465–470.
  - [16] M.R. HESTENES AND E. STIEFEL, *Methods of conjugate gradients for solving linear problems*, J. Res. Nat. Bur. Stand., 49 (1952), pp. 409–436.
  - [17] Z. JIA, *The convergence of Krylov subspace methods for large unsymmetric linear systems*, Acta Math. Sinica, New Series, 14 (1998), pp. 507–518.
  - [18] J. LIESSEN AND Z. STRAKOS, *Krylov Subspace Methods. Principles and Analysis*, Oxford University Press, Oxford, 2013.
  - [19] J. LIU AND S-L. WU, *A fast block  $\alpha$ -circulant preconditioner for all-at-once systems from wave equations*, SIAM J. Matrix Anal. Appl. 41(4) (2021), 19121943.
  - [20] R.D. FALGOUT, S. FRIEDHOFF, T.V. KOLEV, S.P. MACLACHLAN AND J.B. SCHRODER, *Parallel time integration with multigrid*, SIAM J. Sci. Comput., 36(6) (2014), pp. C635–C661.
  - [21] E. McDONALD, J. PESTANA AND A.J. WATHEN, *Preconditioning and iterative solution of all-at-once systems for evolutionary partial differential equations*, SIAM J. Sci. Comput., 40(2) (2018), pp. A1012–A1033.
  - [22] N.M. NACHTIGAL, S.C. REDDY, AND L.N. TREFETHEN, *How fast are nonsymmetric matrix iterations?*, SIAM J. Matrix Anal. Appl., 13 (1992), pp. 778–795.
  - [23] M.K. NG, *Iterative methods for Toeplitz systems*, Oxford University Press, Oxford, 2004.
  - [24] C.C. PAIGE AND M.A. SAUNDERS, *Solution of sparse indefinite systems of linear equations*, SIAM J. Numer. Anal., 12 (1975), pp. 617–629.
  - [25] J. PESTANA AND A.J. WATHEN, *On choice of preconditioner for minimum residual methods for non-Hermitian matrices* J. Comput. Appl. Math., 249, (2013), pp. 57–68.
  - [26] J. PESTANA AND A.J. WATHEN, *A preconditioned MINRES method for nonsymmetric Toeplitz matrices*, SIAM J. Matrix Anal. Appl., 36 (2015), pp. 273–288.
  - [27] Y. SAAD, *Iterative Methods for Sparse Linear Systems*, SIAM, Philadelphia, 2003.
  - [28] Y. SAAD AND M.H. SCHULTZ, *GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems*, SIAM J. Sci. Statist. Comput., 7 (1986), pp. 856–869.
  - [29] G. STRANG, *A proposal for Toeplitz matrix calculations*, Stud. Appl. Math., 74 (1986), pp. 171–176.
  - [30] L.N. TREFETHEN AND D. BAU, *Numerical Linear Algebra*, SIAM, Philadelphia, 1997.
  - [31] L.N. TREFETHEN AND M. EMBREE, *Spectra and pseudospectra: the behavior of nonnormal matrices and operators*, Princeton University Press (2005).
  - [32] A.J. WATHEN, *Preconditioning*, Acta Numerica, 24 (2015), pp. 329–376.
  - [33] Y-L. ZHAO, X-M. GU, M. LI AND H-Y. JIAN, *Preconditioners for all-at-once system from the fractional mobile/immobile advection–diffusion model*, J. Appl. Math. Comput. 65 (2021), pp. 669–691.