PRECONDITIONING AND ITERATIVE SOLUTION OF ALL-AT-ONCE SYSTEMS FOR EVOLUTIONARY PARTIAL DIFFERENTIAL EQUATIONS

ELEANOR MCDONALD *, JENNIFER PESTANA†, AND ANDY WATHEN*

Abstract. Standard Krylov subspace solvers for self-adjoint problems have rigorous convergence bounds based solely on eigenvalues. However, for non-self-adjoint problems, eigenvalues do not determine behavior even for widely used iterative methods. In this paper, we discuss time-dependent PDE problems, which are always non-self-adjoint. We propose a block circulant preconditioner for the all-at-once evolutionary PDE system which has block Toeplitz structure. Through reordering of variables to obtain a symmetric system, we are able to rigorously establish convergence bounds for MINRES which guarantee a number of iterations independent of the number of time-steps for the all-at-once system. If the spatial differential operators are simultaneously diagonalizable, we are able to quickly apply the preconditioner through use of a sine transform, and for those that are not, we are able to use an algebraic multigrid process to provide a good approximation. Results are presented for solution to both the heat and convection diffusion equations.

Key words. evolutionary equations, Toeplitz matrix, circulant preconditioner, iterative methods, block matrices

AMS subject classifications. 65F08, 15B05, 65M22

1. Introduction. It is widely appreciated that self-adjoint problems are, in some respects, easier to solve than problems without natural symmetry. Not least, theoretical understanding is greater than for non-self-adjoint problems, so that, for example, there are linear algebra solution methods—conjugate gradients [22] and MINRES [37]—for large scale symmetric problems for which descriptive and guaranteed convergence bounds based only on eigenvalues exist. For non-symmetric discretized problems there are no generally descriptive convergence bounds, and eigenvalues do not guarantee anything: Greenbaum, Ptàk and Strakoš [18] have proved even for the widely used GMRES method that essentially any convergence curve is possible for a problem regardless of its eigenvalues.

This stark difference means, for example, that one has rigorous theory to guide the design of preconditioners for symmetric problems, but preconditioners for non-symmetric problems must essentially be designed based on heuristics (see [47]). Thus the important multigrid and domain decomposition paradigms are rigorously underpinned and guarantee rapid solvers for symmetric problems, by contrast to non-self-adjoint problems. Further, parallelization must yield the expected benefits for symmetric problems.

One important class of non-self-adjoint problems arise from first order time evolution: an initial value problem for a time-dependent PDE has an adjoint that is a final value problem since

$$\langle u_t, v\rangle = -\langle u, v_t\rangle.$$

This is true regardless of whether the spatial operator is self-adjoint. Via time-stepping (the method of lines), such problems are generally solved one time-step at a time, i.e. in a fully sequential manner. Effective (often parallel) solvers for the spatial partial differential operators at each time step are widely studied and offer practical solution approaches. From this perspective, it can be possible to design solvers that have excellent scalability with respect to the number of spatial degrees of freedom, n, but computational effort must depend on the number of time-steps, ℓ. There has also been significant work on methods that parallelize over time, e.g. [7, 11, 19, 29, 42]. For a review of parallel-in-time methods, see [14]. Our method falls into the class of space-time, or all-at-once, algorithms that solve for all time-steps simultaneously. Such methods include the parareal method [17, 26], space-time multigrid [16, 20, 23] and multigrid-reduction-in-time [12]. Our approach is most closely aligned with methods in which the space-time problem is written as a monolithic linear system, e.g. [1, 16, 20, 23, 28], but our method differs in the way in which this system is solved. Here, we exploit the block Toeplitz structure of the resulting linear system to develop new preconditioners for which the number of Krylov iterations is independent of the number of time-steps ℓ. We note that work by Gander et al [15] presents a complementary all-at-once approach that requires all time-steps to be distinct to ensure diagonalizability. Instead, we consider the case that all time-steps are the same.

*Oxford University Mathematical Institute (mcdonalde@maths.ox.ac.uk, wathen@maths.ox.ac.uk).

†Department of Mathematics and Statistics, University of Strathclyde (jennifer.pestana@strath.ac.uk). This author was supported by Engineering and Physical Sciences Research Council grant EP/I005293.

The approach is based on the block Toeplitz structure of evolutionary problems that allows symmetrization, so that the MINRES method of Paige and Saunders [37], which is designed for symmetric problems, can be correctly applied—convergence then only depends on eigenvalues. After applying block circulant preconditioners to the symmetrized system we prove clustering of eigenvalues so that rapid (and ℓ-independent) convergence is rigorously guaranteed. The relevant computations with circulants are either trivial or almost optimally effected by a fast Fourier transform (FFT). We provide a brief overview to circulant based preconditioning in Section 2.

Our approach is best introduced in terms of a simple application, hence this is described in Section 3. The aspects of symmetrization are covered in Section 4. For non-self adjoint spatial operators, we are still able to obtain eigenvalue estimates based on the LSQR algorithm (also due to Paige and Saunders [38]), which are described in Section 5. Numerical results are presented for the heat and convection-diffusion equations in Section 6 with our conclusions in Section 7.

2. Circulant preconditioning. In order to motivate our block circulant based preconditioner, we first introduce circulant preconditioners for general Toeplitz matrices. Let $T \in \mathbb{R}^{n\times n}$ be the nonsingular Toeplitz matrix and $C \in \mathbb{R}^{n\times n}$ be the nonsingular circulant preconditioner given by

$$T = \begin{bmatrix} t_0 & t_{-1} & \cdots & t_{-n+2} & t_{-n+1} \\ t_1 & t_0 & t_{-1} & & t_{-n+2} \\ \vdots & t_1 & t_0 & \ddots & \vdots \\ t_{n-2} & & \ddots & \ddots & t_{-1} \\ t_{n-1} & t_{n-2} & \cdots & t_1 & t_0 \end{bmatrix}, \quad \text{and} \quad C = \begin{bmatrix} c_0 & c_{n-1} & \cdots & c_2 & c_1 \\ c_1 & c_0 & c_{n-1} & & c_2 \\ \vdots & c_1 & c_0 & \ddots & \vdots \\ c_{n-2} & & \ddots & \ddots & c_{n-1} \\ c_{n-1} & c_{n-2} & \cdots & c_1 & c_0 \end{bmatrix}.$$

For Toeplitz systems, circulant matrices have been popular preconditioners, not least because they can be applied quickly using a fast Fourier transform (FFT). The matrix C has the diagonalization, $C = U\Lambda U^*$ where, if we denote the Fourier matrix by $F = (f_{jk})$, $f_{jk} = \mathrm{e}^{2(j-1)(k-1)\pi \mathrm{i}/n}$, then we have that $U = F/\sqrt{n}$. Also $\Lambda = \mathrm{diag}(Fc_n)$, where c_n is the first column of C. This relationship to the FFT means that the solution of a linear system with a circulant matrix can be performed in $\mathcal{O}(n \log n)$ operations [45].

The idea of preconditioning Toeplitz matrices with a circulant was first introduced independently by Strang in [44] and Olkin in [35]. The so-called *Strang circulant* proposed was constructed by taking the central band of T of width $n/2$ and wrapping the entries around to form a circulant. In this paper, we use the Strang preconditioner, which we find to be very effective for the evolutionary problems we consider. However, many other circulant preconditioners could be applied (see, e.g., the books [5, 32]). One example is the optimal circulant [6], which minimizes the Frobenius norm distance to the given Toeplitz matrix over all possible circulants. A unifying approach to selecting the best possible circulant preconditioner was proposed in [36].

Theoretical convergence bounds for these types of preconditioners have generally been restricted to symmetric (Hermitian) positive definite Toeplitz matrices. For many existing preconditioners—including the Strang and optimal preconditioners—and for wide classes of Toeplitz matrices, the preconditioned system is given by $C^{-1}T = I + R + E$, where R has small rank and E small norm. For non-symmetric systems this is not sufficient to provide descriptive convergence estimates for standard non-symmetric solvers such as GMRES or BiCGSTAB. However [40] provides rigorous convergence bounds for non-symmetric Toeplitz matrices. This is done by reordering the rows or columns of T by pre- or post-multiplying by the Hankel matrix,

$$Y = \begin{bmatrix} & & & 1 \\ & & 1 & \\ & \iddots & & \\ 1 & & & \end{bmatrix}.$$

This results in a symmetric system for any Toeplitz matrix. We extend this method to our block matrix setting in Section 4. We note that other preconditioning methods have been developed for non-symmetric block Toeplitz structures such as those discussed in [24]. That work, however, focusses on small sized blocks and is not motivated by time-dependent problems as is the case here. Furthermore, this method does not include symmetrization techniques that we employ. We note that it is possible to use LSQR or LSMR [13] to

obtain rigorous convergence bounds for non-symmetric Toeplitz matrices, but for scalar Toeplitz problems these methods are typically slower than using symmetrization and MINRES.

3. Motivation and model problem. In order to describe our method, we will begin by considering the solution of the linear diffusion (or heat) equation initial-boundary value problem,

$$\begin{aligned} u_t &= \Delta u + f && \text{in } \Omega \times (0,T], \quad \Omega \subset \mathbb{R}^2 \text{ or } \mathbb{R}^3, \\ u &= g && \text{on } \partial\Omega, \\ u(x,0) &= u_0(x) && \text{at } t = 0. \end{aligned} \tag{1}$$

To solve this system, we discretize in both space and time. For simplicity, we will describe our approach using a finite element discretization in space and a Backward Euler discretization in time. In practice other implicit time stepping schemes and spatial discretization schemes can be used, and this will be discussed in more detail later.

We discretize the spatial domain with a representative mesh size h and take ℓ time steps of size τ such that $\ell\tau = T$. This discretization of (1) gives that

$$M\frac{\mathbf{u}_k - \mathbf{u}_{k-1}}{\tau} + K\mathbf{u}_k = \mathbf{f}_k, \quad k = 1, \ldots, \ell,$$

where $M \in \mathbb{R}^{n\times n}$ is the standard finite element mass matrix, $K \in \mathbb{R}^{n\times n}$ is the stiffness matrix (the discrete Laplacian) and n is the number of spatial degrees of freedom. We assume that M and K are symmetric positive definite matrices. The initial vector $\mathbf{u_0}$ should be obtained from the initial data by a convenient projection. Rearranging, we have that

$$(M + \tau K)\mathbf{u}_k = M\mathbf{u}_{k-1} + \tau\mathbf{f}_k, \quad k = 1, \ldots, \ell. \tag{2}$$

We can solve for all time steps of such a system simultaneously using an 'all-at-once' approach. Conceptually, we construct the following linear system, which defines the solution at all time steps:

$$\mathcal{A}_{BE}\mathbf{x} := \begin{bmatrix} A_0 & & & \\ A_1 & A_0 & & \\ & \ddots & \ddots & \\ & & A_1 & A_0 \end{bmatrix} \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \\ \vdots \\ \mathbf{u}_\ell \end{bmatrix} = \begin{bmatrix} M\mathbf{u}_0 + \tau\mathbf{f}_1 \\ \tau\mathbf{f}_2 \\ \vdots \\ \tau\mathbf{f}_\ell \end{bmatrix} := \mathbf{b}, \tag{3}$$

where $A_0 = M + \tau K$ is symmetric positive definite and $A_1 = -M$ is symmetric negative definite. We note that $\mathcal{A}_{BE}$ is now an immense $n\ell \times n\ell$ matrix; the construction of $\mathcal{A}_{BE}$ only requires copies of A_0 and A_1 and is never done explicitly.

The matrix $\mathcal{A}_{BE}$ is clearly block Toeplitz and we wish to precondition it with the associated block Strang circulant matrix. As $\mathcal{A}_{BE}$ is already lower triangular with just one subdiagonal, the Strang circulant simply consists of wrapping the subdiagonal entry A_1 around to create a circulant. Thus our proposed preconditioner is given by

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

In order to describe the preconditioned system, we make the observation that $\mathcal{P}_{BE}$ is a rank n perturbation of $\mathcal{A}_{BE}$, since $\mathcal{P}_{BE} = \mathcal{A}_{BE} + E_1 A_1 E_\ell^T$, where $E_i = e_i \otimes I_n$ with e_i denoting the i-th column of I_ℓ and $\otimes$ denoting the Kronecker product. We can now examine the eigenvalues of the preconditioned system.

THEOREM 1. *The preconditioned system is equal to* $\mathcal{P}_{BE}^{-1}\mathcal{A}_{BE} = I_{n\ell} - \mathcal{A}_{BE}^{-1}E_1 Z^{-1} E_\ell^T$, *which is a rank* n *perturbation of the identity matrix* $I_{n\ell} \in \mathbb{R}^{n\ell\times n\ell}$, *where* $Z = A_1^{-1} + (\mathcal{A}_{BE}^{-1})_{\ell-1}$ *and* $(\mathcal{A}_{BE}^{-1})_{\ell-1} = E_\ell^T \mathcal{A}_{BE}^{-1} E_1$. *Furthermore,* $\mathcal{P}_{BE}^{-1}\mathcal{A}_{BE}$ *has* $(\ell-1)n$ *eigenvalues equal to* 1 *and* n *eigenvalues equal to the eigenvalues of* $I_n - (\mathcal{A}_{BE}^{-1})_{\ell-1} Z^{-1}$.

Proof. Writing $\mathcal{P}_{BE} = \mathcal{A}_{BE} + E_1 A_1 E_\ell^T$, then by the Sherman-Morrison-Woodbury formula we have that

$$\mathcal{P}_{BE}^{-1} = (\mathcal{A}_{BE} + E_1 A_1 E_\ell^T)^{-1} = \mathcal{A}_{BE}^{-1} - \mathcal{A}_{BE}^{-1} E_1 (A_1^{-1} + E_\ell^T \mathcal{A}_{BE}^{-1} E_1)^{-1} E_\ell^T \mathcal{A}_{BE}^{-1},$$

and thus,

$$\mathcal{P}_{BE}^{-1}\mathcal{A}_{BE} = I_{n\ell} - \mathcal{A}_{BE}^{-1} E_1 (A_1^{-1} + E_\ell^T \mathcal{A}_{BE}^{-1} E_1)^{-1} E_\ell^T.$$

Since $\mathcal{A}_{BE}^{-1} E_1 (A_1^{-1} + E_\ell^T \mathcal{A}_{BE}^{-1} E_1)^{-1} E_\ell^T$ is of rank n, this shows that the preconditioned system is a rank n perturbation of the identity. Noting that the inverse of $\mathcal{A}_{BE}$ will also be block lower triangular and block Toeplitz, and letting $Z = A_1^{-1} + E_\ell^T \mathcal{A}_{BE}^{-1} E_1$, then we have that

$$\begin{aligned}
\mathcal{P}_{BE}^{-1}\mathcal{A}_{BE} &= I_{n\ell} - \mathcal{A}_{BE}^{-1} E_1 Z^{-1} E_\ell^T \\
&= I_{n\ell} - \begin{bmatrix} (\mathcal{A}_{BE}^{-1})_0 & & & \\ (\mathcal{A}_{BE}^{-1})_1 & (\mathcal{A}_{BE}^{-1})_0 & & \\ & \ddots & \ddots & \\ (\mathcal{A}_{BE}^{-1})_{\ell-1} & & (\mathcal{A}_{BE}^{-1})_1 & (\mathcal{A}_{BE}^{-1})_0 \end{bmatrix} \begin{bmatrix} & & Z^{-1} \\ & & \\ & & \end{bmatrix} \\
&= \begin{bmatrix} I_n & & & -(\mathcal{A}_{BE}^{-1})_0 Z^{-1} \\ & I_n & & -(\mathcal{A}_{BE}^{-1})_1 Z^{-1} \\ & & \ddots & \vdots \\ & & & I_n - (\mathcal{A}_{BE}^{-1})_{\ell-1} Z^{-1} \end{bmatrix},
\end{aligned}$$

from which we can easily see that the eigenvalues of $\mathcal{P}_{BE}^{-1}\mathcal{A}_{BE}$ are $(\ell-1)n$ copies of 1 as well as the n eigenvalues of $I_n - (\mathcal{A}_{BE}^{-1})_{\ell-1} Z^{-1}$. ∎

In fact, we can further describe the eigenvalues of $I_n - (\mathcal{A}_{BE}^{-1})_{\ell-1} Z^{-1}$ in terms of the matrices A_0 and A_1.

THEOREM 2. *If μ is an eigenvalue of $A_1^{-1}A_0$ then $\mu \neq \pm 1$ and $\frac{\mu^\ell}{\mu^\ell + (-1)^{\ell-1}}$ is an eigenvalue of $I_n - (\mathcal{A}_{BE}^{-1})_{\ell-1} Z^{-1}$.*

Proof. Firstly, a simple inductive argument can be used to show that $(\mathcal{A}_{BE}^{-1})_{k-1} = (-1)^{k-1}(A_0^{-1}A_1)^{k-1}A_0^{-1}$ for all $k = 1, \ldots, \ell$. Thus we have that

$$\begin{aligned}
I_n - (\mathcal{A}_{BE}^{-1})_{\ell-1} Z^{-1} &= I_n - (\mathcal{A}_{BE}^{-1})_{\ell-1} (A_1^{-1} + (\mathcal{A}_{BE}^{-1})_{\ell-1})^{-1} \\
&= I_n - \left[A_1^{-1} (\mathcal{A}_{BE}^{-1})_{\ell-1}^{-1} + I_n\right]^{-1} \\
&= I_n - \left[(-1)^{\ell-1} (A_1^{-1}A_0)^\ell + I_n\right]^{-1}.
\end{aligned}$$

Now, $A_1^{-1}A_0 = -(I_n + \tau M^{-1}K)$ with M and K both symmetric positive definite. Thus, if μ is an eigenvalue of $A_1^{-1}A_0$ then $\mu \neq \pm 1$, and there exists a nonzero vector $\boldsymbol{x} \in \mathbb{R}^n$ such that

$$\begin{aligned}
A_1^{-1}A_0 \boldsymbol{x} &= \mu \boldsymbol{x} \\
\left[I_n + (-1)^{\ell-1} (A_1^{-1}A_0)^\ell\right]^{-1} \boldsymbol{x} &= \frac{1}{1 + (-1)^{\ell-1}\mu^\ell} \boldsymbol{x} \\
\left[I_n - \left[I_n + (-1)^{\ell-1} (A_1^{-1}A_0)^\ell\right]^{-1}\right] \boldsymbol{x} &= \frac{\mu^\ell}{\mu^\ell + (-1)^{\ell-1}} \boldsymbol{x},
\end{aligned}$$

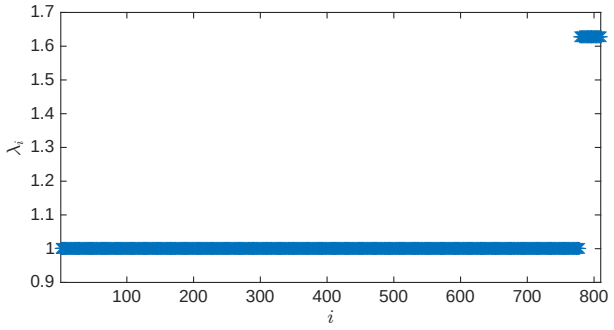
which completes the proof. ∎

This shows that although $\mathcal{P}_{BE}^{-1}\mathcal{A}_{BE}$ has n eigenvalues not equal to one, if μ is large then these eigenvalues can cluster very close to one. In the case of the heat equation, we see that the largest eigenvalues of $A_1^{-1}A_0$ grow with h^{-2}, where h is the grid size, and therefore we see extremely clustered eigenvalues in practice. Figure 1 shows the eigenvalues of $\mathcal{P}_{BE}^{-1}\mathcal{A}_{BE}$ for a small system.

We will now show that $\mathcal{P}_{BE}^{-1}\mathcal{A}_{BE}$ is diagonalizable.

THEOREM 3. *The matrix $\mathcal{P}_{BE}^{-1}\mathcal{A}_{BE}$ is diagonalizable.*

Fig. 1: The eigenvalues of $\mathcal{P}_{BE}^{-1}\mathcal{A}_{BE}$ with $n = 81$, $\ell = 10$ and $\tau = 0.1$. There are 32 eigenvalues approximately equal to 1.6275.



Proof. Recall that $A_1^{-1}A_0 = -(I_n + \tau M^{-1}K)$, with M, K symmetric positive definite. From the proof of Theorem 2 we have that

$$(\mathcal{A}_{BE}^{-1})_{\ell-1}Z^{-1} = [I_n - (I_n + \tau M^{-1}K)^{\ell}]^{-1},$$

which is diagonalizable and has real, negative eigenvalues. Thus, $I_n - (\mathcal{A}_{BE}^{-1})_{\ell-1}Z^{-1}$ is diagonalizable, and has eigenvalues that are real and larger than 1.

Let $I_n - (\mathcal{A}_{BE}^{-1})_{\ell-1}Z^{-1}$ have diagonalization VDV^{-1}. Then $\mathcal{P}_{BE}^{-1}\mathcal{A}_{BE}$ has the diagonalization $\mathcal{V}\mathcal{D}\mathcal{V}^{-1}$,

$$\mathcal{V} = \begin{bmatrix} I & & & V_0 \\ & I & & V_1 \\ & & \ddots & \vdots \\ & & I & V_{\ell-2} \\ & & & V \end{bmatrix}, \quad \text{and} \quad \mathcal{D} = \begin{bmatrix} I & & & \\ & I & & \\ & & \ddots & \\ & & & D \end{bmatrix}$$

where $V_i = (\mathcal{A}_{BE}^{-1})_i Z^{-1} V (D - I_n)^{-1}$. ∎

Theorem 1 shows that GMRES will terminate within $n+1$ iterations, while diagonalizability of $\mathcal{P}_{BE}^{-1}\mathcal{A}_{BE}$ may help us to estimate the rate of convergence. Analogous results to Theorem 1 exist for more complex time-stepping schemes, as we discuss in Section 3.2. However, in these cases it is not obvious whether the preconditioned matrix is diagonalizable, nor when we can expect convergence in fewer steps because of eigenvalue clustering. Furthermore, Theorem 3 will not necessarily be applicable if the preconditioner is applied approximately, such as with a multigrid method.

Although we have now demonstrated that the preconditioned system has a number of non-unit eigenvalues independent of the number of time-steps ℓ, the circulant preconditioner we have proposed is, in principle, just as difficult to invert as the original matrix $\mathcal{A}$. In order to demonstrate an easy, and indeed parallelizable, method of inverting $\mathcal{P}$ we will now consider the matrices in Kronecker product notation.

3.1. Kronecker product form. The block structure of the matrices allows us to describe them in Kronecker product form as

$$\mathcal{A}_{BE} = I_\ell \otimes A_0 + \Sigma \otimes A_1,$$
$$\mathcal{P}_{BE} = I_\ell \otimes A_0 + C_1 \otimes A_1,$$

where

$$\Sigma = \begin{bmatrix} 0 & & & \\ 1 & 0 & & \\ & \ddots & \ddots & \\ & & 1 & 0 \end{bmatrix}, \quad C_1 = \begin{bmatrix} 0 & & & 1 \\ 1 & 0 & & \\ & \ddots & \ddots & \\ & & 1 & 0 \end{bmatrix},$$

and I_ℓ is the identity matrix of dimension $\ell \times \ell$. As described in Section 2 we can apply $C_1 = U\Lambda U^*$ or its inverse to a vector using the FFT. We define the diagonal entries of Λ to be λ_k, $k = 1, \ldots, \ell$, and note that in general they are complex. Furthermore, for this very specific circulant, the eigenvalues are in fact the ℓ roots of unity, so that $\lambda_k = e^{2\pi i k/\ell}$.

The Kronecker product has the property that $(W \otimes X)(Y \otimes Z) = (WY \otimes XZ)$. Using this, and the fact that U is unitary, allows us to rewrite the preconditioner $\mathcal{P}_{BE}$ as

$$\mathcal{P}_{BE} = I_\ell \otimes A_0 + C_1 \otimes A_1 = (U \otimes I_n)[I_\ell \otimes A_0 + \Lambda \otimes A_1](U^* \otimes I_n)$$

and therefore,

$$\mathcal{P}_{BE}^{-1} = (U \otimes I_n)[I_\ell \otimes A_0 + \Lambda \otimes A_1]^{-1}(U^* \otimes I_n).$$

A similar formulation was used in [21] to write a semi-circulant preconditioner.

Applying the inverse of $\mathcal{P}_{BE}$ to a vector requires us to multiply by $U \otimes I_n$ or $U^* \otimes I_n$ and invert the block diagonal matrix $I_\ell \otimes A_0 + \Lambda \otimes A_1$. To apply $U \otimes I_n$ we can first apply a column and row permutation that allows us to instead multiply by the block diagonal matrix $I_n \otimes U$, which has n blocks of size $\ell \times \ell$. Finally, we must reverse the row and column permutation. Since the required permutation, which is a simple reordering of the spatial and temporal degrees of freedom, is known in advance, multiplication by $U \otimes I_n$ or $U^* \otimes I_n$ could be parallelizable over n processors although communication between processors would be required because of the permutations.

The matrix $I_\ell \otimes A_0 + \Lambda \otimes A_1$ is block diagonal and therefore could be inverted in parallel over ℓ processors. This matrix is complex symmetric and therefore a method such as a complex algebraic multigrid, e.g. [25, 27, 33, 41], could be used to approximately perform this step.

3.1.1. Simultaneous diagonalization. For our formulation of the heat equation, the blocks A_0 and A_1 in (3) are symmetric. As we show below, the mass and stiffness matrices M and K also commute. As a result, A_0 and A_1 commute, and so can be simultaneously diagonalized. The property allows us to further simplify the manner in which we apply $\mathcal{P}_{BE}$.

If we let $A_0 = X\Phi X^T$ and $A_1 = X\Psi X^T$ then we have

$$\mathcal{P}_{BE}^{-1} = (U \otimes I_n)(I_\ell \otimes X)[I_\ell \otimes \Phi + \Lambda \otimes \Psi]^{-1}(I_\ell \otimes X^T)(U^* \otimes I_n). \tag{4}$$

Now to apply the inverse of $I_\ell \otimes A_0 + \Lambda \otimes A_1$, we first need to apply $(I_\ell \otimes X)$, which is a block diagonal matrix and could be applied over ℓ separate processors. We then invert $I_\ell \otimes \Phi + \Lambda \otimes \Psi$, which is diagonal and therefore trivial, before applying $(I_\ell \otimes X^T)$, which is again block diagonal. Thus when we have this property, the application of a circulant preconditioner becomes much cheaper.

If we use a finite element formulation to discretize (1) then M and K are simultaneously diagonalizable if we use a uniform square grid. For finite difference methods, the finite element mass matrix is replaced by the identity matrix and therefore will always commute with the diffusion operator K. We note that for the Dirichlet problem discretized by finite elements with uniform grids we are able to compute the diagonalization using sine transforms as we now describe.

For the x and y directions respectively, the i-th element of the j-th normalized eigenvector is given by $V_x(i,j) = \sqrt{\frac{2}{n_x+1}} \sin\left(\frac{ij\pi}{n_x+1}\right)$, $V_y(i,j) = \sqrt{\frac{2}{n_y+1}} \sin\left(\frac{ij\pi}{n_y+1}\right)$, where n_x is the number of interior nodes in the x-direction and n_y is the number of interior nodes in the y-direction. We construct $X_x \in \mathbb{R}^{(n_x+2)\times(n_x+2)}$ and $X_y \in \mathbb{R}^{(n_y+2)\times(n_y+2)}$ by embedding each matrix within an identity matrix such that:

$$X_x = \begin{bmatrix} 1 & & \\ & V_x & \\ & & 1 \end{bmatrix}, \qquad X_y = \begin{bmatrix} 1 & & \\ & V_y & \\ & & 1 \end{bmatrix}.$$

We then form the two-dimensional eigenvectors X by the simple relation $X = X_x \otimes X_y$. As a result, we can apply X to a vector using discrete sine transforms.

We will now examine the effect that more complex time-stepping schemes have on the system.

3.2. Multi step methods. For simplicity, we discretized (1) using a Backward Euler time stepping scheme. However other implicit time stepping schemes could also be used. In this section we describe how the ideas in the previous sections can be extended to a p-step scheme, which means that $\mathcal{A}$ has p subdiagonals.

Define $\mathcal{A}$ to be the following $\ell n \times \ell n$ block lower triangular Toeplitz matrix formed of ℓ blocks of $n \times n$ matrices with $p \le \ell - 1$ subdiagonals, and define $\mathcal{P}$ to be corresponding Strang circulant:

$$\mathcal{A} := \begin{bmatrix} A_0 & & & & & \\ A_1 & A_0 & & & & \\ \vdots & \ddots & \ddots & & & \\ A_p & & \ddots & \ddots & & \\ & \ddots & & A_1 & A_0 & \\ & & A_p & \cdots & A_1 & A_0 \end{bmatrix}, \quad \mathcal{P} := \begin{bmatrix} A_0 & & A_p & \cdots & A_2 & A_1 \\ A_1 & A_0 & & & & A_2 \\ \vdots & \ddots & \ddots & & \ddots & \vdots \\ A_p & & \ddots & \ddots & & A_p \\ & \ddots & & A_1 & A_0 & \\ & & A_p & \cdots & A_1 & A_0 \end{bmatrix}. \tag{5}$$

Define $\Sigma_i \in \mathbb{R}^{\ell \times \ell}$ to be the Toeplitz matrix of zeros except for 1s on the i-th subdiagonal and C_i to be the corresponding Strang circulant with 1s on the i-th subdiagonal and the $(\ell - i)$-th superdiagonal.

By simple computation we can observe that $C_i = (C_1)^i$, and therefore if we diagonalize $C_1 = U\Lambda U^*$ then $C_i = (C_1)^i = (U\Lambda U^*)^i = U\Lambda^i U^*$.

We can write $\mathcal{A}$ and $\mathcal{P}$ in Kronecker form, which gives

$$\begin{aligned} \mathcal{A} &= I_\ell \otimes A_0 + \sum_{i=1}^{p} \Sigma_i \otimes A_i, \\ \mathcal{P} &= I_\ell \otimes A_0 + \sum_{i=1}^{p} C_i \otimes A_i = \sum_{i=0}^{p} U\Lambda^i U^* \otimes A_i. \end{aligned}$$

We make the additional assumption that all A_i commute with each other and are therefore simultaneously diagonalizable. This will occur for any time stepping method if the spatial operators K and M commute. We thus assume that we have the diagonalizations $A_i = X\Delta_i X^T$, X orthogonal. We can now write that

$$\mathcal{P} = \sum_{i=0}^{p} U\Lambda^i U^* \otimes A_i = (U \otimes I_n) \begin{bmatrix} G_1 & & & \\ & G_2 & & \\ & & \ddots & \\ & & & G_\ell \end{bmatrix} (U^* \otimes I_n) = (U \otimes I_n)\mathcal{G}(U^* \otimes I_n), \tag{6}$$

where $\mathcal{G} = \mathrm{diag}(G_1, \ldots, G_\ell)$ and $G_j = \sum_{i=0}^{p} \lambda_j^i A_i = X\left(\sum_{i=0}^{p} \lambda_j^i \Delta_i\right) X^T := X\mathtt{g}_j X^T$. Furthermore,

$$\mathcal{G} = (I_\ell \otimes X)\mathrm{diag}(\mathtt{g}_1, \ldots, \mathtt{g}_\ell)(I_\ell \otimes X^T),$$

where $(I_\ell \otimes X)$ and $(I_\ell \otimes X^T)$ are block diagonal and $\mathrm{diag}(\mathtt{g}_1, \ldots, \mathtt{g}_\ell)$ is diagonal. The point here is that even for multi-step methods, with simultaneous diagonalization of the spatial operators we can apply the inverse of the preconditioner $\mathcal{P}$ using only multiplication with block diagonal matrices and the inversion of a diagonal matrix, which are all extremely cheap to apply.

We also note that, using a similar approach to that in the proof of Theorem 1, we can write the preconditioned system $\mathcal{P}^{-1}\mathcal{A}$ as a rank-np perturbation of the identity. Thus, GMRES converges in at most $np + 1$ steps for this problem.

4. Symmetrized system. Although we have been able to describe the eigenvalues of the preconditioned system and have shown that the number of non-unit eigenvalues is independent of the number of time-steps, this is not generally sufficient to ascertain the convergence rate of non-symmetric solvers such as GMRES. However, if our spatial operators are symmetric and using the ideas developed in [40], we are able to propose a method to rewrite our system as a symmetric one, so that we are able to use eigenvalue analysis to determine convergence estimates.

As stated earlier, the matrix $\mathcal{A}$ in (5) is block Toeplitz with symmetric blocks. We note that we can symmetrize any matrix of this type by pre- or post-multiplication with the following block Hankel matrix,

$$\mathcal{Y} := \begin{bmatrix} & & & I_n \\ & & \iddots & \\ & I_n & & \\ I_n & & & \end{bmatrix} = Y \otimes I_n, \quad \text{where} \quad Y = \begin{bmatrix} & & & 1 \\ & & \iddots & \\ & 1 & & \\ 1 & & & \end{bmatrix}. \tag{7}$$

Pre- or post-multiplication by $\mathcal{Y}$ will symmetrize any block Toeplitz matrix with symmetric blocks, however in general $\mathcal{YA}$ does not equal $\mathcal{AY}$. If we wish to solve the system of equations $\mathcal{A}\mathbf{x} = \mathbf{f}$ then we can solve the equations

$$(\mathcal{YA})\mathbf{x} = \mathcal{Y}\mathbf{f} \text{ or } \mathcal{AY}\mathbf{y} = \mathbf{f},\ \mathbf{y} = \mathcal{Y}\mathbf{x}. \tag{8}$$

However, unlike for the original system we are able to use iterative methods for symmetric systems for which much better convergence estimates exist. We also note that Y and $\mathcal{Y}$ are involutory and thus $\mathcal{Y}^{-1} = \mathcal{Y}$.

In order to use a symmetric matrix solver such as MINRES we require a symmetric positive definite preconditioner. One such matrix is the absolute value preconditioner [40, 46] $|\mathcal{P}|$ defined as,

$$\begin{aligned} |\mathcal{P}| &= (\mathcal{P}^T\mathcal{P})^{1/2} \\ &= \left[(U \otimes I_n)\mathcal{G}^*\mathcal{G}(U^* \otimes I_n)\right]^{1/2} \\ &= (U \otimes I_n)|\mathcal{G}|(U^* \otimes I_n) \\ &= (U \otimes X)\begin{bmatrix} |\mathsf{g}_1| & & \\ & \ddots & \\ & & |\mathsf{g}_\ell| \end{bmatrix}(U^* \otimes X^T), \end{aligned} \tag{9}$$

(10)

where g_j is the diagonal $n \times n$ matrix in (6) and $|\mathsf{g}_j|$ is its elementwise absolute value. We note $|\mathcal{P}|$ is symmetric positive definite and therefore can be used in MINRES with the symmetric form of the equation (8).

4.1. Eigenvalue analysis. We have now described a symmetric positive definite preconditioner for the symmetrized system (8) to be implemented with MINRES. Since eigenvalues provide robust convergence bounds for MINRES, unlike for GMRES, we now wish to determine the eigenvalues of the preconditioned system $|\mathcal{P}|^{-1}\mathcal{YA}$. That, more generally, matrices of the form of $\mathcal{P}$ and $|\mathcal{P}|$ are block circulant will also prove useful later in this section, hence we establish this now.

LEMMA 1. *Let $\mathcal{R} \in \mathbb{R}^{n\ell \times n\ell}$ be any matrix of the form*

$$\mathcal{R} = (U \otimes X)\begin{bmatrix} \mathsf{d}_1 & & & \\ & \mathsf{d}_2 & & \\ & & \ddots & \\ & & & \mathsf{d}_\ell \end{bmatrix}(U^* \otimes X^T),$$

where U and X are as in (4), and $\mathsf{d}_i \in \mathbb{C}^{n\times n}$, $i = 1, \ldots, \ell$ are diagonal matrices. Then $\mathcal{R}$ is block circulant and $\mathcal{RY} = \mathcal{YR}^T$, where $\mathcal{Y}$ is as in (7).

Proof. If $\mathcal{R}_{rs}$ denotes the (r, s) block of $\mathcal{R}$ of size $n \times n$, then

$$\mathcal{R}_{rs} = \sum_{k=1}^{\ell} u_{rk}\overline{u_{sk}} X \mathsf{d}_k X^T. \tag{11}$$

To prove that $\mathcal{R}$ is block circulant we need to look at the definition of each u_{rs}. Now U has as its columns the eigenvectors of a circulant matrix. Thus, $u_{rs} = f_{rs}/\sqrt{\ell}$ where $f_{rs} = \mathrm{e}^{2(r-1)(s-1)\pi\mathrm{i}/\ell}$.

We will first show that $\mathcal{R}$ is block Toeplitz, that is, $\mathcal{R}_{rs} = \mathcal{R}_{(r+1)(s+1)}$ for all $r, s \in [1, \ldots, \ell - 1]$. The scalars $u_{rk}\overline{u_{sk}}$ in (11) satisfy

$$u_{rk}\overline{u_{sk}} = \frac{1}{\ell}\mathrm{e}^{2(r-s)(k-1)\pi\mathrm{i}/\ell} = u_{(r+1)k}\overline{u_{(s+1)k}}.$$

Since $\mathcal{R}_{(r+1)(s+1)} = \sum_{k=1}^{\ell} u_{(r+1)k}\overline{u_{(s+1)k}} X \mathsf{d}_k X^T$, it follows that $\mathcal{R}_{rs} = \mathcal{R}_{(r+1)(s+1)}$. This proves that all diagonals have constant blocks.

If $\mathcal{R}$ is additionally block circulant, then we also require that $\mathcal{R}_{r\ell} = \mathcal{R}_{(r+1)1}$ for all $r \in [1, \ldots, \ell - 1]$. To show this, note that $\mathcal{R}_{r\ell} = \sum_{k=1}^{\ell} u_{rk}\overline{u_{\ell k}} X \mathsf{d}_k X^T$, with

$$u_{rk}\overline{u_{\ell k}} = \frac{1}{\ell}\mathrm{e}^{2(r-\ell)(k-1)\pi\mathrm{i}/\ell} = \frac{1}{\ell}\mathrm{e}^{2r(k-1)\pi\mathrm{i}/\ell} = \frac{1}{\ell}\mathrm{e}^{2r(k-1)\pi\mathrm{i}/\ell}\mathrm{e}^{-2\pi\mathrm{i}(1-1)(k-1)/\ell} = u_{(r+1)k}\overline{u_{1k}}.$$

Since $\mathcal{R}_{(r+1)1} = \sum_{k=1}^{\ell} u_{(r+1)k}\overline{u_{1k}} X \mathsf{d}_k X^T$, it follows that $\mathcal{R}_{r\ell} = \mathcal{R}_{(r+1)1}$ for all $r \in [1, \ldots, \ell-1]$, from which we see that $\mathcal{R}$ is block circulant.

Finally, we prove the symmetrization property $\mathcal{R}\mathcal{Y} = \mathcal{Y}\mathcal{R}^T$. The (r,s) block of $\mathcal{R}\mathcal{Y}$ is

$$(\mathcal{R}\mathcal{Y})_{rs} = \mathcal{R}_{r(\ell-s+1)} = \sum_{k=1}^{\ell} u_{rk}\overline{u_{(\ell-s+1)k}} X \mathsf{d}_k X^T,$$

while

$$(\mathcal{Y}\mathcal{R}^T)_{rs} = (\mathcal{R}^T)_{(\ell-r+1)s} = \left(\mathcal{R}_{s(\ell-r+1)}\right)^T = \sum_{k=1}^{\ell} \overline{u_{sk}} u_{(\ell-r+1)k} X \overline{\mathsf{d}_k} X^T.$$

Since, for all $r, s, k \in [1, \ldots, \ell]$,

$$u_{rk}\overline{u_{(\ell-s-1)k}} = \frac{1}{\ell} \mathrm{e}^{2(r+s-\ell-1)(k-1)\pi \mathrm{i}/\ell} = u_{sk}\overline{u_{(\ell-r+1)k}},$$

we see that $(\mathcal{R}\mathcal{Y})_{rs} = \overline{(\mathcal{Y}\mathcal{R}^T)_{rs}} = (\mathcal{Y}\mathcal{R}^T)_{rs}$, since $\mathcal{Y}$ and $\mathcal{R}$ are real. ∎

In our eigenvalue analysis, it will prove useful to relate $\mathcal{P}$ in (5) and $|\mathcal{P}|$ in (10). To do this we introduce the real orthogonal matrix

$$\widetilde{\mathcal{P}} = (U \otimes X) \begin{bmatrix} \mathrm{sgn}(\mathsf{g}_1) & & & \\ & \mathrm{sgn}(\mathsf{g}_2) & & \\ & & \ddots & \\ & & & \mathrm{sgn}(\mathsf{g}_\ell) \end{bmatrix} (U^* \otimes X^T),$$

where $\mathrm{sgn}(\mathsf{g}_j) = \mathsf{g}_j |\mathsf{g}_j|^{-1}$. Then,

$$|\mathcal{P}|\widetilde{\mathcal{P}} = \widetilde{\mathcal{P}}|\mathcal{P}| = \mathcal{P}. \tag{12}$$

Since they share the same eigenvector matrix $U \otimes X$ the matrices $\mathcal{P}$, $|\mathcal{P}|$ and $\widetilde{\mathcal{P}}$ all commute and are block circulant (see Lemma 1).

Additionally, under conditions that are met for all our numerical experiments, $\widetilde{\mathcal{P}}$ has a real, orthogonal square root, as we now show.

LEMMA 2. *Assume that $A_0, \ldots, A_p$ have real eigenvalues and that $\sum_{i=0}^{p} A_i$ has positive eigenvalues. When ℓ is even, additionally assume that $\sum_{i=0}^{p} (-1)^i A_i$ has positive eigenvalues. Then $\widetilde{\mathcal{P}}$ has a real, orthogonal matrix square root.*

Proof. The proof proceeds in two parts. We first show that if $\widetilde{\mathcal{P}}$ has unit determinant then $\widetilde{\mathcal{P}}$ has a real, orthogonal matrix square root. Then, we prove that $\det(\widetilde{\mathcal{P}}) = 1$.

We begin the proof of the first part by showing that any matrix in $SO(n)$ (the group of real orthogonal matrices with unit determinant) has a real orthogonal square root. To do this we use the fact that the exponential of a skew-symmetric matrix belongs to $SO(n)$ (the group of orthogonal matrices with unit determinant) and every matrix in $SO(n)$ has a skew-symmetric matrix logarithm [4]. Thus, if $B \in SO(n)$ then $B = \mathrm{e}^F$ for some skew-symmetric F, and $\mathrm{e}^{F/2}$ is a real orthogonal square root of B.

We wish to apply this result to $\widetilde{\mathcal{P}}$. First, note that (12) shows that $\widetilde{\mathcal{P}}$ is real. Additionally, using the definition of the sign function, it is clear that $\widetilde{\mathcal{P}}$ is orthogonal. Thus, all that remains is to show that $\det(\widetilde{\mathcal{P}}) = 1$.

We treat the more difficult case that ℓ is even first. The matrix C_1 has as its eigenvalues the roots of unity $\lambda_k = \mathrm{e}^{2\pi k \mathrm{i}/\ell}$, $k = 1, \ldots, \ell$. If ℓ is even, $\lambda_{\ell/2} = -1$, $\lambda_\ell = 1$ and $\lambda_k = \overline{\lambda_{\ell-k}}$, $k = 1, \ldots, \ell/2-1$. It follows that for $j = 1, \ldots, \ell/2-1$,

$$(\mathsf{g}_{\ell-j})^* = \sum_{i=0}^{p} (\overline{\lambda_{\ell-j}})^i \Delta_i = \sum_{i=0}^{p} \lambda_j^i \Delta_i = \mathsf{g}_j.$$

Thus,

$$\det(\widetilde{\mathcal{P}}) = \prod_{k=1}^{\ell} \det(\mathrm{sgn}(\mathsf{g}_k)) = \det(\mathrm{sgn}(\mathsf{g}_{\ell/2}))\det(\mathrm{sgn}(\mathsf{g}_\ell)) \prod_{k=1}^{\ell/2-1} \det(\mathrm{sgn}(\mathsf{g}_k)\mathrm{sgn}(\mathsf{g}_k^*)). \tag{13}$$

Using the assumptions of the lemma, and the definition of the sign function, we find that $\det(\operatorname{sgn}(\mathsf{g}_{\ell/2})) = 1$, $\det(\operatorname{sgn}(\mathsf{g}_\ell)) = 1$ and $\operatorname{sgn}(\mathsf{g}_k)\operatorname{sgn}(\mathsf{g}_k^*) = \operatorname{sgn}(\mathsf{g}_k)(\operatorname{sgn}(\mathsf{g}_k))^* = I_n$. Thus, when ℓ is even, (13) shows that $\det(\widetilde{\mathcal{P}}) = 1$, so that $\widetilde{\mathcal{P}}$ has a real, orthogonal matrix square root.

If ℓ is odd then $\lambda_\ell = 1$ and $\lambda_k = \overline{\lambda_{\ell-k}}$, $k = 1, \ldots, (\ell-1)/2$. The proof that $\det(\widetilde{\mathcal{P}}) = 1$ then follows similarly, except that C_1 does not have an eigenvalue at -1. Thus, when ℓ is odd, $\widetilde{\mathcal{P}}$ also has a real, orthogonal matrix square root. ∎

We remark that the conditions of Lemma 2 are generally easy to check. When K and M in (2) are positive definite, then all that is required is to compute sums involving the scalar coefficients that define the time-stepping scheme. The conditions are met for all numerical experiments involving the heat equation in Section 6.

We want to look at the eigenvalues of the preconditioned system $|\mathcal{P}|^{-1}\mathcal{Y}\mathcal{A}$ and we can easily see that these will be the same as the eigenvalues of the matrix $|\mathcal{P}|^{-1/2}\mathcal{Y}\mathcal{A}|\mathcal{P}|^{-1/2}$ by a similarity transform. The matrix $\mathcal{Y}$ of (7) comprises ℓ blocks, and we write $\mathcal{Y}_p$ for the corresponding matrix with p blocks.

THEOREM 4. *Let* $V = [E_{\ell-p+1}, \ldots, E_\ell] \in \mathbb{R}^{n\ell\times np}$ *and*

$$W = \begin{bmatrix} A_p & \ldots & A_2 & A_1 \\ & A_p & & A_2 \\ & & \ddots & \vdots \\ & & & A_p \end{bmatrix}, \tag{14}$$

$W \in \mathbb{R}^{np\times np}$. *Then for* $|\mathcal{P}|$ *and* $\mathcal{A}$ *as defined as in* (10) *and* (5) *respectively,*

$$|\mathcal{P}|^{-1/2}\mathcal{Y}\mathcal{A}|\mathcal{P}|^{-1/2} = Q - Z\Theta Z^T,$$

where $Q = \mathcal{Y}\widetilde{\mathcal{P}}$ *is orthogonal and symmetric, the symmetric matrix* $\mathcal{Y}_pW \in \mathbb{R}^{np\times np}$ *has the eigenvalue decomposition* $\mathcal{Y}_pW = S\Theta S^T$ *and* $Z = |\mathcal{P}|^{-1/2}VS \in \mathbb{R}^{n\ell\times np}$ *has full rank.*

Proof. Firstly we see from (5) that we can write $\mathcal{P} = \mathcal{A} + UWV^T$, where $U = [E_1, \ldots E_p] \in \mathbb{R}^{n\ell\times np}$. Thus, $\mathcal{A} = \mathcal{P} - UWV^T$ and we have

$$|\mathcal{P}|^{-1/2}\mathcal{Y}\mathcal{A}|\mathcal{P}|^{-1/2} = |\mathcal{P}|^{-1/2}\mathcal{Y}\mathcal{P}|\mathcal{P}|^{-1/2} - |\mathcal{P}|^{-1/2}\mathcal{Y}UWV^T|\mathcal{P}|^{-1/2}.$$

Now $\mathcal{Y}U = \mathcal{Y}[E_1 \ldots E_p] = [E_\ell \ldots E_{\ell-p+1}] = V\mathcal{Y}_p$. Thus,

$$|\mathcal{P}|^{-1/2}\mathcal{Y}UWV^T|\mathcal{P}|^{-1/2} = |\mathcal{P}|^{-1/2}V\mathcal{Y}_pWV^T|\mathcal{P}|^{-1/2} = (|\mathcal{P}|^{-1/2}VS)\Theta(|\mathcal{P}|^{-1/2}VS)^T.$$

Since $|\mathcal{P}|$, V and S have full rank, $Z = |\mathcal{P}|^{-1/2}VS$ has rank np.

The matrix $|\mathcal{P}|^{-1/2}$ is symmetric and so, by Lemma 1, $|\mathcal{P}|^{-1/2}\mathcal{Y} = \mathcal{Y}|\mathcal{P}|^{-1/2}$. Additionally, $\mathcal{P}$ and $|\mathcal{P}|^{1/2}$ commute. It follows that

$$|\mathcal{P}|^{-1/2}\mathcal{Y}\mathcal{P}|\mathcal{P}|^{-1/2} = \mathcal{Y}\mathcal{P}|\mathcal{P}|^{-1} = \mathcal{Y}\widetilde{\mathcal{P}} = Q.$$

Since $\mathcal{Y}$ and $\widetilde{\mathcal{P}}$ are orthogonal, Q is also orthogonal. Additionally, $Q = |\mathcal{P}|^{-1/2}\mathcal{Y}\mathcal{A}|\mathcal{P}|^{-1/2} + Z\Theta Z^T$ is the sum of symmetric matrices, and so must be symmetric. ∎

LEMMA 3. *Assume that the conditions of Lemma 2 hold. Then, the matrix* Q *has the same eigenvalues as* $\mathcal{Y}$*, which has* $\lfloor \ell/2 \rfloor n$ *eigenvalues equal to* -1 *and* $\lceil \ell/2 \rceil n$ *eigenvalues equal to* 1.

Proof. Firstly we want to show that Q and $\mathcal{Y}$ are similar, and therefore have the same eigenvalues. Lemma 1 shows that $\widetilde{\mathcal{P}}^{1/2}$ is block circulant and symmetrized by $\mathcal{Y}$. Additionally, since $\widetilde{\mathcal{P}}$ is orthogonal, $\widetilde{\mathcal{P}}^{1/2}$ is as well. Thus,

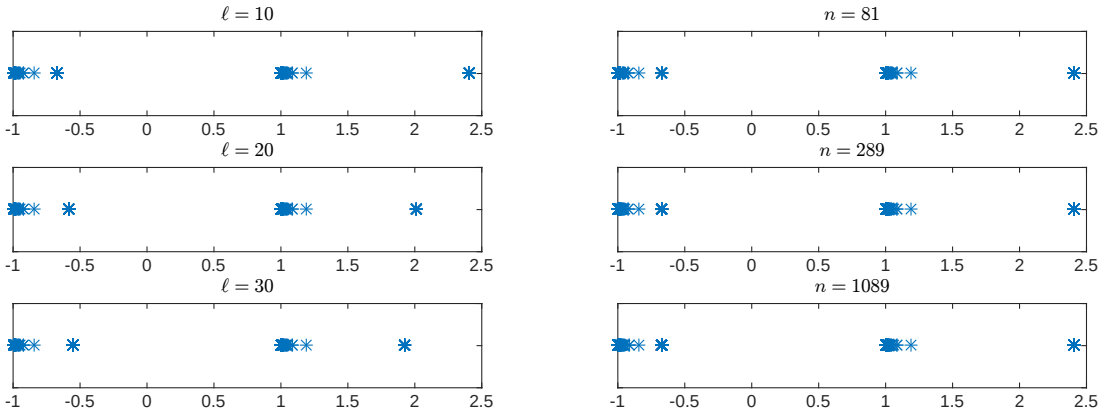
$$Q = \widetilde{\mathcal{P}}\mathcal{Y} = \widetilde{\mathcal{P}}^{1/2}\widetilde{\mathcal{P}}^{1/2}\mathcal{Y} = \widetilde{\mathcal{P}}^{1/2}\mathcal{Y}(\widetilde{\mathcal{P}}^{1/2})^T = \widetilde{\mathcal{P}}^{1/2}\mathcal{Y}\widetilde{\mathcal{P}}^{-1/2}.$$

Therefore Q and $\mathcal{Y}$ will have the same eigenvalues.

It is left to determine the eigenvalues of $\mathcal{Y}$. Firstly we note that $\mathcal{Y}E_j = E_{\ell-j+1}$. Therefore we have

$$\mathcal{Y}(E_j - E_{\ell-j+1}) = E_{\ell-j+1} - E_j = -(E_j - E_{\ell-j+1}),$$

Fig. 2: Eigenvalues of the preconditioned system $|\mathcal{P}|^{-1}\mathcal{YA}$ for varying grid and time step sizes. In the left figure, $n = 81$, and in the right figure $\ell = 10$. In all cases $\tau = 0.1$.



so -1 will be an eigenvalue associated with an eigenvector equal to one of the columns of $(E_j - E_{\ell-j+1})$. This gives the required algebraic multiplicity of the eigenvalue -1.

Similarly, the columns of

$$\mathcal{Y}(E_j + E_{\ell-j+1}) = E_{\ell-j+1} + E_j$$

give the form of the eigenvectors corresponding to unit eigenvalues. If ℓ is odd then for $j = \lceil \ell/2 \rceil$ we have

$$\mathcal{Y}E_{\lceil \ell/2 \rceil} = E_{\lceil \ell/2 \rceil},$$

so that the remaining n eigenvalues are 1. Thus, we obtain the stated multiplicity of the unit eigenvalue. ∎

THEOREM 5. *Assume that the conditions of Lemma 2 hold, and that $\lfloor \ell/2 \rfloor > p$. Then, the geometric multiplicity of the eigenvalue 1 of $|\mathcal{P}|^{-1/2}\mathcal{YA}|\mathcal{P}|^{-1/2}$ is at least $(\lceil \ell/2 \rceil - p)n$, while the geometric multiplicity of the eigenvalue -1 is at least $(\lfloor \ell/2 \rfloor - p)n$. This leaves at most $2np$ eigenvalues that are not ± 1.*

Proof. We know from Theorem 4 and Lemma 3 that Q is symmetric with $\lfloor \ell/2 \rfloor n$ eigenvalues equal to -1 and $\lceil \ell/2 \rceil n$ eigenvalues equal to 1. Thus, Q has diagonalization $Q = V_Q \Lambda_Q V_Q^T$, where Λ_Q has diagonal entries 1 or -1.

Accordingly,

$$V_Q^T |\mathcal{P}|^{-1/2}\mathcal{YA}|\mathcal{P}|^{-1/2} V_Q = \Lambda_Q - H,$$

where $H = V_Q^T Z \Theta Z^T V_Q$ is a Hermitian matrix of rank np. By Corollary 3 in [2], at most np copies of the each distinct eigenvalue of Q can be perturbed by H. It follows that $V_Q^T |\mathcal{P}|^{-1/2}\mathcal{YA}|\mathcal{P}|^{-1/2} V_Q$, and hence $|\mathcal{P}|^{-1/2}\mathcal{YA}|\mathcal{P}|^{-1/2}$ have the required eigenvalue multiplicities. ∎

Having shown that the preconditioned system has at most $2np$ eigenvalues that are not ± 1, we know that MINRES will converge in at most $2np + 2$ steps, which is independent of the number of time steps. In practice, we do not see nearly this many steps, as the eigenvalues that are not ± 1 are also closely clustered in our numerical experiments for the heat equation, and this eigenvalue clustering can be linked to the convergence rate of MINRES. Figure 2 shows the eigenvalues of the preconditioned system $|\mathcal{P}|^{-1}\mathcal{YA}$ for the same grid sizes with varying numbers of time steps. We can see that the eigenvalues remain extremely well clustered as the number of time steps increases.

In Figure 2 we also show the eigenvalues of the preconditioned system for a fixed number of time-step sizes and various spatial grid sizes. It is evident that although the eigenvalues become more spread out as n increases, the eigenvalues remain well clustered, with only one cluster of eigenvalues away from ± 1.

5. Non-symmetric systems. Throughout the previous sections we have assumed that all A_i are symmetric, as without this property $\mathcal{Y}$ would not symmetrize the system. However, for cases where the A_i are not symmetric we can also form the normal equations and solve the system using LSQR. We note that we could also use this method when the A_i are symmetric. We now analyse the eigenvalues of the normal equations of the preconditioned system.

THEOREM 6. *The matrix* $(\mathcal{P}^{-1}\mathcal{A})^T(\mathcal{P}^{-1}\mathcal{A})$ *has* $(\ell-2p)n$ *eigenvalues equal to* 1*,* np *eigenvalues less than or equal to* 1*, and* np *eigenvalues greater than or equal to* 1*.*

Proof. Let $\mathcal{P} = \mathcal{A}+UWV^T$ where $U = [E_1, \dots E_p] \in \mathbb{R}^{n\ell\times np}$, $V = [E_{\ell-p+1}, \dots E_\ell] \in \mathbb{R}^{n\ell\times np}$ and $W \in \mathbb{R}^{np\times np}$ is as in (14). Using the Sherman-Morrison-Woodbury formula as described in Theorem 1, we find that $\mathcal{P}^{-1}\mathcal{A} = I_{n\ell} - \mathcal{A}^{-1}UZ^{-1}V^T$, where $Z = W^{-1} + V^T\mathcal{A}^{-1}U \in \mathbb{R}^{np\times np}$. If we partition $\mathcal{A}^{-1}$ as

$$\mathcal{A}^{-1} = \begin{bmatrix} \mathcal{A}_{11}^{-1} & 0 \\ \mathcal{A}_{21}^{-1} & \mathcal{A}_{22}^{-1} \end{bmatrix} \text{ then } \mathcal{P}^{-1}\mathcal{A} = I_{n\ell} - \begin{bmatrix} 0 & \mathcal{A}_{11}^{-1}Z^{-1} \\ 0 & \mathcal{A}_{21}^{-1}Z^{-1} \end{bmatrix},$$

where $\mathcal{A}_{11}^{-1} \in \mathbb{R}^{np\times np}$, $\mathcal{A}_{21}^{-1} \in \mathbb{R}^{(\ell-p)n\times np}$, and $\mathcal{A}_{22}^{-1} \in \mathbb{R}^{(\ell-p)n\times(\ell-p)n}$. We can now write that

$$(\mathcal{P}^{-1}\mathcal{A})^T(\mathcal{P}^{-1}\mathcal{A}) = \begin{bmatrix} I_{(\ell-p)n} & -\mathcal{A}_{11}^{-1}Z^{-1} \\ -Z^{-T}\mathcal{A}_{11}^{-T} & Z^{-T}\mathcal{A}_{11}^{-T}\mathcal{A}_{11}^{-1}Z^{-1} + (I_{np} - Z^{-T}\mathcal{A}_{21}^{-T})(I_{np} - \mathcal{A}_{21}^{-1}Z^{-1}) \end{bmatrix}.$$

From here we can see that the upper $(\ell-p)n$ principle submatrix is the identity and we can use the Cauchy Interlacing theorem (see for example Chapter 10 of [39]) to relate the eigenvalues of $(\mathcal{P}^{-1}\mathcal{A})^T(\mathcal{P}^{-1}\mathcal{A})$ to the eigenvalues of the identity. The theorem tell us that if we let λ_i be the i-th eigenvalue of $(\mathcal{P}^{-1}\mathcal{A})^T(\mathcal{P}^{-1}\mathcal{A})$ with $\lambda_1 \le \lambda_2 \le \cdots \le \lambda_{\ell n}$, then $\lambda_i \le \sigma_i(I) = 1 \le \lambda_{np+i}$, which gives that the eigenvalues λ_1 to λ_{np} must be less than or equal to 1, the eigenvalues λ_{np+1} to $\lambda_{(\ell-p)n}$ must be equal to 1 and eigenvalues $\lambda_{(\ell-p)n+1}$ to $\lambda_{\ell n}$ must be greater than or equal to 1. ∎

Now since $|\mathcal{P}|^2 = \mathcal{P}^T\mathcal{P} = \mathcal{P}\mathcal{P}^T$, we have

$$(\mathcal{P}^{-1}\mathcal{A})^T(\mathcal{P}^{-1}\mathcal{A}) = \mathcal{A}^T(\mathcal{P}\mathcal{P}^T)^{-1}\mathcal{A} = \mathcal{A}^T(|\mathcal{P}|)^{-2}\mathcal{A} = (|\mathcal{P}|^{-1}\mathcal{A})^T(|\mathcal{P}|^{-1}\mathcal{A}).$$

Thus, the eigenvalues of the normal equations when using either $\mathcal{P}$ or $|\mathcal{P}|$ as the preconditioner are the same. We also note that $\mathcal{A}^T(|\mathcal{P}|)^{-2}\mathcal{A}$ has the same eigenvalues as $\mathcal{Y}\mathcal{A}(|\mathcal{P}|)^{-2}\mathcal{A}\mathcal{Y}$, since this is a similarity transform with $\mathcal{Y}^{-1} = \mathcal{Y}$. It follows that the eigenvalues of $(|\mathcal{P}|^{-1}\mathcal{A}\mathcal{Y})^T(|\mathcal{P}|^{-1}\mathcal{A}\mathcal{Y})$ are the same as the eigenvalues of $(\mathcal{P}^{-1}\mathcal{A})^T(\mathcal{P}^{-1}\mathcal{A})$, and that the singular values of $|\mathcal{P}|^{-1}\mathcal{A}\mathcal{Y}$ are the same as those of $\mathcal{P}^{-1}\mathcal{A}$.

Therefore we have again shown that using a block circulant based preconditioner results in a number of non-unit eigenvalues independent of the number of time-steps. However, the values of the non-unit eigenvalues can depend on both the number of time-steps ℓ and the number of spatial degrees of freedom n. This means that despite the guarantee of termination, iteration counts can increase as ℓ increases as seen in some of the results in the following section. We find that this is particularly pronounced for the convection-diffusion equation, for which this method is unlikely to be practical.

6. Numerical results. In this section, we present numerical results for an implementation of the method described in the previous sections within the IFISS [8, 9, 43] framework. Since GMRES can require large amounts of storage due to the orthogonalization process, we have also used the BiCGSTAB method as an alternative iterative method for solving non-symmetric systems. We note, however, that none of the termination theory applies with this method; it is simply shown as a potentially practical alternative. When applying the AMG preconditioner, which is nonlinear, we applied right-preconditioned flexible GMRES (FGMRES); neither GMRES nor FGMRES allowed restarting. We also used the standard Matlab implementations of MINRES, LSQR and BiCGSTAB. All methods were stopped with a relative residual tolerance of 10^{-6} and used a random initial guess. The finite element discretization used $Q1$ finite elements over the domain $\Omega = [0,1]\times[0,1]$ for the heat equation and $\Omega = [-1,1]\times[-1,1]$ for the convection diffusion equation. For the algebraic multigrid preconditioner, we used AGMG [30, 31, 33, 34] with default settings, which can be applied to complex matrices. This applies a single K-cycle (sometimes referred to as a non-linear AMLI cycle); details can be found [33]. Note that adjusting the number of AMG cycles did not affect the iteration numbers obtained.

Note that for use with GMRES, we employ $\mathcal{P}_{MG}$ and not $|\mathcal{P}_{MG}|$ (which would in this case be awkward to compute). We have no rate of convergence guarantees for this approximate non-symmetric solver, but we observe rapid convergence as seen in Tables 1, 2 and 3. These observations are perhaps not a complete surprise given the supporting rigorous theory in the corresponding symmetric case.

6.1. Heat equation. Our first example is the heat equation as defined in (1) with the initial conditions

$$u_0 = x(x-1)y(y-1)$$

with no external forcing (i.e. $f = 0$). We used both the Backward Euler and the 2-step Backward Differentiation Formula (BDF2) for the time-stepping method, with time step size equal to $\tau = 1/\ell$.

The results presented in Table 1 are for the Backward Euler time-stepping method and show that for all methods, iteration numbers are essentially independent of the number of time steps. Mesh independent convergence is observed for MINRES and GMRES, but not for LSQR. FGMRES with the AMG preconditioner $\mathcal{P}_{MG}$ performs well for coarse discretisations, but there is some iteration growth as the mesh is refined. Although this particular AMG algorithm is not accurately approximating the diagonal blocks in $I_\ell \otimes A_0 + \Lambda \otimes A_1$ (cf. Section 3.1), we would expect better performance from a tailored AMG algorithm. Similar results are observed for the BDF2 method (see Table 2), with iteration counts for GMRES and MINRES with $|\mathcal{P}|$ robust with respect to the number of time steps and mesh width.

We note that using the symmetrization method within MINRES results in higher iteration numbers than seen when applying GMRES to the non-symmetric system. For practical purposes it may, therefore, be advantageous to use GMRES even though there is then no theoretical guarantee of fast convergence. We include results for both iterative methods for comparison. We also notice that whilst the LSQR method has comparable iterations counts to MINRES for small values of ℓ, for larger numbers of time-steps LSQR requires a significant increase in iterations.

Table 1: Iteration numbers for the heat equation using the Backward Euler method. (— indicates iterations above the maximum of 300 or that GMRES stagnated.)

n	ℓ	DoF	GMRES $\mathcal{P}^{-1}\mathcal{A}$	MINRES $\lvert\mathcal{P}\rvert^{-1}\mathcal{Y}\mathcal{A}$	LSQR $\mathcal{P}^{-1}\mathcal{A}$	FGMRES $\mathcal{P}_{\text{MG}}^{-1}\mathcal{A}$
81	2^4	1296	3	12	10	3
	2^6	5184	3	13	16	3
	2^8	20736	3	15	27	3
	2^{10}	82944	3	15	52	3
	2^{12}	331776	3	15	90	3
	2^{14}	1327104	3	14	157	3
289	2^4	4624	3	11	10	8
	2^6	18496	3	13	14	8
	2^8	73984	3	15	27	8
	2^{10}	295936	3	19	56	8
	2^{12}	1183744	3	18	130	7
	2^{14}	4734976	3	16	—	7
1089	2^4	17424	3	10	9	8
	2^6	69696	3	13	13	8
	2^8	278784	3	14	24	8
	2^{10}	1115136	3	18	50	8
	2^{12}	4460544	3	20	128	7
	2^{14}	17842176	3	19	—	6
4225	2^4	67600	3	10	7	15
	2^6	270400	3	11	12	16
	2^8	1081600	3	13	21	16
	2^{10}	4326400	3	18	44	16
	2^{12}	17305600	3	20	113	17
	2^{14}	69222400	2	19	—	16

Table 2: Iteration numbers for the heat equation using the BDF2 method. (— indicates iterations above the maximum of 300 or that GMRES stagnated.)

n	ℓ	DoF	GMRES $\mathcal{P}^{-1}\mathcal{A}$	MINRES $\lvert\mathcal{P}\rvert^{-1}\mathcal{Y}\mathcal{A}$	LSQR $\mathcal{P}^{-1}\mathcal{A}$	FGMRES $\mathcal{P}_{\text{MG}}^{-1}\mathcal{A}$
81	2^4	1296	3	14	13	3
	2^6	5184	3	17	22	3
	2^8	20736	3	19	44	3
	2^{10}	82944	3	20	97	3
	2^{12}	331776	3	20	177	3
	2^{14}	1327104	3	18	265	3
289	2^4	4624	3	13	12	7
	2^6	18496	3	16	21	8
	2^8	73984	3	19	43	8
	2^{10}	295936	3	21	106	7
	2^{12}	1183744	3	24	—	7
	2^{14}	4734976	3	22	—	6
1089	2^4	17424	3	13	11	8
	2^6	69696	3	15	20	8
	2^8	278784	3	18	39	8
	2^{10}	1115136	3	22	98	7
	2^{12}	4460544	3	24	288	7
	2^{14}	17842176	3	25	—	6
4225	2^4	67600	3	11	10	15
	2^6	270400	3	13	17	16
	2^8	1081600	3	18	33	16
	2^{10}	4326400	3	21	83	17
	2^{12}	17305600	3	24	245	17
	2^{14}	69222400	3	25	—	16

6.2. Convection diffusion equation. The convection diffusion test problem is given by Example 6.1.4 in [10] and is known as the double glazing problem. The wind is described by $\mathbf{w} = (2y(1-x^2), -2x(1-y^2))$. Dirichlet boundary conditions are imposed everywhere on the boundary, with $u = 1$ on the boundary where $x = 1$ and zero on all other boundaries. The initial vector $\mathbf{u}_0$ was zero everywhere except the boundaries where it satisfies the boundary conditions. Streamline-Upwind Petrov Galerkin (SUPG) stabilization [3] was used to stabilize the system. For this problem we used Backward Euler time-stepping with time-step size $\tau = 1/\ell$.

As this is a non-symmetric system and the spatial operators do not commute, we were not able to use the simultaneous diagonalization method described in Section 3.1.1. However, we were still able to apply the absolute value preconditioner, although this did require computing ℓ diagonalizations. We therefore also used the AGMG preconditioner with both the FGMRES and BiCGSTAB methods. For the exact preconditioner, we used the backslash operator in Matlab i.e. an elimination (direct) method was used for the relevant block systems.

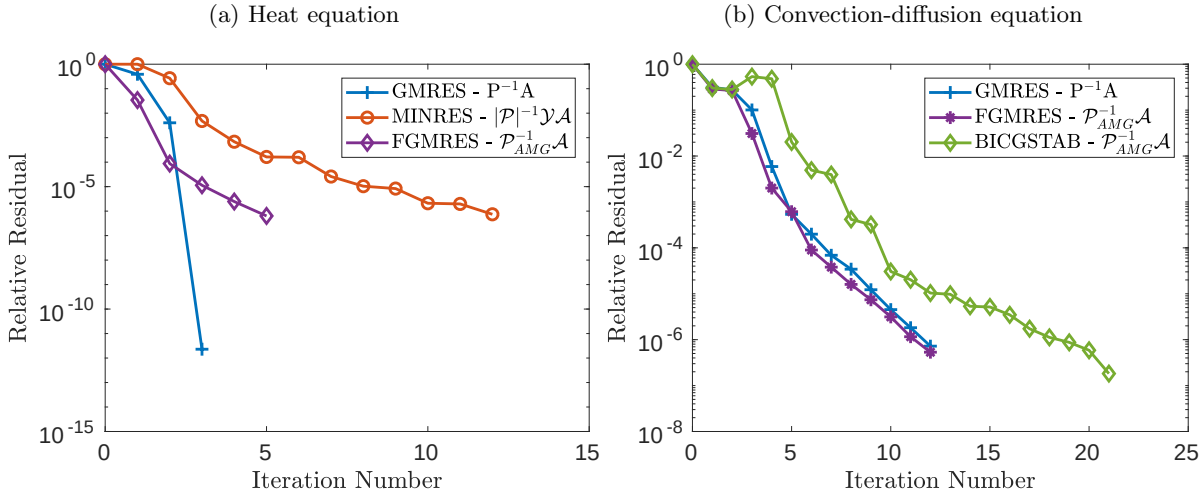
We can see iteration numbers for GMRES that are independent of the number of time-steps and essentially also independent of the grid size. The results for FGMRES and BiCGSTAB with the AMG preconditioner show similar trends; though the iteration counts increase for the largest spatial grid, this method allows solution of these problems for all numbers of time steps. As for the heat equation, we could expect more robust performance from an AMG algorithm better suited to our problem. For the LSQR method, although we are able to prove that the number of non-unit eigenvalues of the normal equations is independent of ℓ the values taken by the outlying eigenvalues can become large as ℓ increases; we therefore see that the number of LSQR iterations grows quite rapidly and so this method is unlikely to be practical. There is essentially no growth in the number of iterations for the GMRES, FMGRES and BiCGSTAB methods to which our analysis does not apply, with the exception of the the finest grid for which the AMG component of the preconditioner seems less effective.

Table 3: Iteration numbers for the convection diffusion equation (- indicates iterations above the maximum of 300).

n	ℓ	DoF	GMRES $\mathcal{P}^{-1}\mathcal{A}$	LSQR $\mathcal{P}^{-1}\mathcal{A}$	FGMRES $\mathcal{P}_{\text{MG}}^{-1}\mathcal{A}$	BICGSTAB $\mathcal{P}_{\text{MG}}^{-1}\mathcal{A}$
81	2^4	1296	12	63	12	21
	2^6	5184	12	137	12	19
	2^8	20736	12	262	12	19
	2^{10}	82944	12	—	12	20
	2^{12}	331776	12	—	12	20
	2^{14}	1327104	12	—	12	19
289	2^4	4624	13	71	12	17
	2^6	18496	13	206	12	21
	2^8	73984	13	—	12	21
	2^{10}	295936	13	—	12	21
	2^{12}	1183744	13	—	12	21
	2^{14}	4734976	13	—	12	20
1089	2^4	17424	12	72	12	21
	2^6	69696	13	226	12	21
	2^8	278784	13	—	12	21
	2^{10}	1115136	13	—	12	21
	2^{12}	4460544	13	—	12	21
	2^{14}	17842176	13	—	12	21
4225	2^4	67600	12	66	22	98
	2^6	270400	12	217	22	83
	2^8	1081600	12	—	23	97
	2^{10}	4326400	12	—	23	106
	2^{12}	17305600	12	—	23	168
	2^{14}	69222400	12	—	23	120

In order to further investigate the convergence properties of the proposed methods in practice, in Figure 3 we have plotted the convergence curves for each, with the exception of LSQR for which convergence was significantly slower. For the heat equation, we see that GMRES with the exact preconditioner exhibits rapid residual norm reduction at the third iteration while the other methods converge at comparable rates. For convection-diffusion, we do not see this drop off in the GMRES convergence curve with the exact preconditioner. This is likely due to the small number of distinct eigenvalues for the preconditioned system for the heat equation as compared with the convection-diffusion equation. We see that BiCGSTAB behaves differently to GMRES however there is no associated theory for convergence of the preconditioner with this method. Note as well that, since BiCGSTAB requires two matrix-vector products and two preconditioner solves at each iteration, its cost per iteration is roughly double that of GMRES and MINRES. All methods converge fairly well in these computations, but the theory only guarantees this for MINRES.

Fig. 3: Convergence of each of the methods ($n = 1089$, $\ell = 2^{10}$).

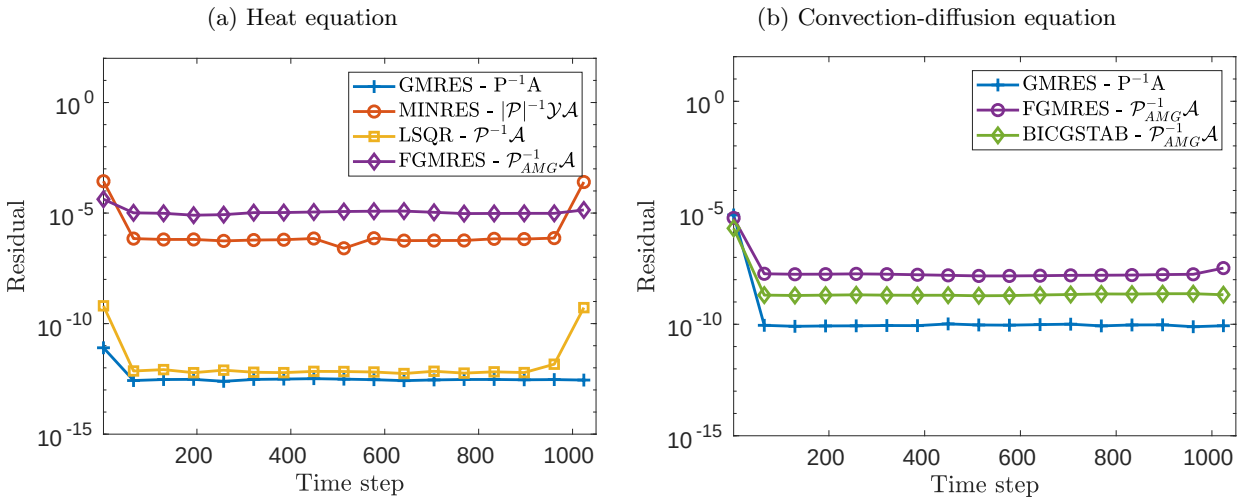


When calculating the solution of a time-dependent problem in a sequential manner, an error at a given time-step is typically propagated forward at subsequent time-steps. As the all-at-once method computes the solution at all time-steps simultaneously, the error in the solution at each individual time-step may have a different distribution than when calculated sequentially.

Figure 4 shows the residual of the linear system at each time-step when calculated by each method. For the sequential methods, the LU factorization of the matrix in (2) was calculated and then used to evaluate the solution at each step. We also note that this method has essentially solved the problem to machine precision, although the error grows slightly at later time-steps. For the heat equation, the all-at-once GMRES methods have essentially constant residuals after the first time step. Interestingly, for the heat equation, the residuals for the symmetrized MINRES method are symmetric over the time interval i.e. the residual at $t_i = i\tau$ equals the residual at $t_{\ell-i+1} = (\ell - i + 1)\tau$. However, this is not replicated for the convection-diffusion problem. Again note that BiCGStab requires roughly twice the work per iteration of GMRES and MINRES.

7. Conclusions. We have presented a method of preconditioning an all-at-once system of evolutionary equations with constant time-steps based on circulant methods for Toeplitz matrices. For symmetric systems, such as the heat equation, on a regular grid we can use simultaneous diagonalization to efficiently apply a block circulant or its absolute value as a preconditioner. We can also rewrite the system as a symmetric one through the use of a block Hankel matrix. This allows us to use MINRES and to provide an eigenvalue analysis, which guarantees convergence in a maximum number of iterations independent of the number of time-steps. In practice we observe much better convergence even than predicted by this eigenvalue analysis. For non-symmetric systems, we can also provide eigenvalue analysis for the preconditioned normal equations. For both symmetric and non-symmetric systems an algebraic multigrid process can also be

Fig. 4: Residual of the solution at each time-step ($n = 1089$, $\ell = 2^{10}$).



employed to approximate the preconditioner; this provides an inexpensive alternative. Although we cannot prove convergence bounds when AMG is used in this way, we nevertheless see promising results for both symmetric and non-symmetric spatial operators with our approach. Due to the block diagonal structures present in the application of the preconditioners, we believe that parallel-in-time implementations may be possible however investigation of this would require further research.

REFERENCES

[1] A. O. H. Axelsson and J. G. Verwer, *Boundary value techniques for initial value problems in ordinary differential equations*, Math. Comp., 45 (1985), pp. 153–171.

[2] J. H. Brandts and R. Reiss da Silva, *Computable eigenvalue bounds for rank-k perturbations*, Linear Algebra Appl., 432 (2010), pp. 3100–3116.

[3] A. N. Brooks and T. J. Hughes, *Streamline upwind/Petrov-Galerkin formulation for convection dominated flows with particular emphasis on the incompressible Navier-Stokes equations*, Comput. Methods Appl. Mech. Engrg., 32 (1982), pp. 199–259.

[4] J. R. Cardoso and F. S. Leite, *Exponentials of skew-symmetric matrices and logarithms of orthogonal matrices*, J. Comput. Appl. Math., 233 (2010), pp. 2867–2875.

[5] R. H.-F. Chan and X.-Q. Jin, *An Introduction to Iterative Toeplitz Solvers*, SIAM, Philadelphia, PA, USA, 2007.

[6] T. Chan, *An optimal circulant preconditioner for Toeplitz systems*, SIAM J. Sci. Statist. Comput., 9 (1988), pp. 766–771.

[7] A. J. Christlieb, C. B. Macdonald, and B. W. Ong, *Parallel high-order integrators*, SIAM J. Sci. Comput., 32 (2010), pp. 818–835, doi:10.1137/09075740X.

[8] H. Elman, A. Ramage, and D. Silvester, *Algorithm 866: IFISS, a Matlab toolbox for modelling incompressible flow*, ACM Trans. Math. Software, 33 (2007), pp. 2–14.

[9] H. Elman, A. Ramage, and D. Silvester, *IFISS: A computational laboratory for investigating incompressible flow problems*, SIAM Rev., 56 (2014), pp. 261–273.

[10] H. 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 University Press, Oxford, UK, 2nd ed., 2014.

[11] M. Emmett and M. L. Minion, *Toward an efficient parallel in time method for partial differential equations*, Commun. Appl. Math. Comput. Sci., 7 (2012), pp. 105–132, 10.2140/camcos.2012.7.105.

[12] 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 (2014), pp. C635–C661, doi:10.1137/130944230.

[13] D. C.-L. Fong and M. Saunders, *LSMR: An iterative algorithm for sparse least-squares problems*, SIAM J. Sci. Comput., 33 (2011), pp. 2950–2971.

[14] M. J. Gander, *50 years of time parallel time integration*, in Multiple Shooting and Time Domain Decomposition Methods, T. Carraro, M. Geiger, S. Körkel, and R. Rannacher, eds., Springer International Publishing, Switzerland, 2015, pp. 69–113.

[15] M. J. Gander, L. Halpern, J. Ryan, and T. T. B. Tran, *A Direct Solver for Time Parallelization*, Springer International

Publishing, 2016, pp. 491–499.
[16] M. J. Gander and M. Neumüller, *Analysis of a new space-time parallel multigrid algorithm for parabolic problems*, SIAM J. Sci. Comput., 38 (2016), pp. A2173–A2208, doi:10.1137/15M1046605.
[17] M. J. Gander and S. Vandewalle, *Analysis of the parareal time-parallel time-integration method*, SIAM J. Sci. Comput., 29 (2007), pp. 556–578, doi:10.1137/05064607X.
[18] A. Greenbaum, V. Pták, and Z. Strakoš, *Any nonincreasing convergence curve is possible for GMRES*, SIAM J. Matrix Anal. Appl., 17 (1996), pp. 465–469.
[19] S. Güttel, *A parallel overlapping time-domain decomposition method for odes*, in Domain decomposition methods in science and engineering XX, vol. 91 of Lect. Notes Comput. Sci. Eng., Springer, Heidelberg, 2013.
[20] W. Hackbusch, *Parabolic multi-grid methods*, in Proceedings of the Sixth International Symposium on Computing Methods in Applied Sciences and Engineering, VI, R. Glowinski and J.-L. Lions, eds., North-Holland, Amsterdam, 1984, pp. 189–197.
[21] L. Hemmingsson, *A semi-circulant preconditioner for the convection-diffusion equation*, Numer. Math., 81 (1998), pp. 211–248, doi:10.1007/s002110050390.
[22] M. R. Hestenes and E. Stiefel, *Methods of conjugate gradients for solving linear systems*, J. Res. Nat. Bur. Stand., 49 (1952), pp. 409–435, nvl.nist.gov/pub/nistpubs/jres/049/6/V49.N06.A08.pdf.
[23] G. Horton and S. Vandewalle, *A space-time multigrid method for parabolic partial differential equations*, SIAM J. Sci. Comput., 16 (1995), pp. 848–864, doi:10.1137/0916050.
[24] T. K. Huckle and D. Noutsos, *Preconditioning block Toeplitz matrices*, Electron. Trans. Numer. Anal., 29 (2007), pp. 31–45.
[25] D. Lahaye, H. De Gersem, S. Vandewalle, and K. Hameyer, *Algebraic multigrid for complex symmetric systems*, IEEE Trans. Magn., 36 (2000), pp. 1535–1538.
[26] J.-L. Lions, Y. Maday, and G. Turinici, *A parareal in time discretization of PDEs*, C.R. Acad. Sci. Paris, Serie I, 332 (2001), pp. 661–668, doi:10.1016/S0764-4442(00)01793-6.
[27] S. MacLachlan and C. Oosterlee, *Algebraic multigrid solvers for complex-valued matrices*, SIAM J. Sci. Comp., 30 (2008), pp. 1548–1571.
[28] Y. Maday and E. M. Rønquist, *Parallelization in time through tensor-product space–time solvers*, Comptes Rendus Mathematique, 346 (2008), pp. 113–118.
[29] W. L. Miranker and W. Liniger, *Parallel methods for the numerical integration of ordinary differential equations*, Math. Comp., 21 (1967), pp. 303–320.
[30] A. Napov and Y. Notay, *Aggregation-based algebraic multigrid for convection-diffusion equations*, SIAM J. Sci. Comput., 34 (2012), pp. A2288–A2316.
[31] A. Napov and Y. Notay, *An algebraic multigrid method with guaranteed convergence rate*, SIAM J. Sci. Comput., 34 (2012), pp. A1079–A1109.
[32] M. K. Ng, *Iterative Methods for Toeplitz Systems*, Oxford University Press, Oxford, UK, 2004.
[33] Y. Notay, *AGMG software and documentation; see http://homepages.ulb.ac.be/ ynotay/AGMG*.
[34] Y. Notay, *An aggregation-based algebraic multigrid method*, Electron. Trans. Numer. Anal., 37 (2010), pp. 123–146.
[35] J. A. Olkin, *Linear and nonlinear deconvolution problems*, PhD thesis, Rice University, 1986.
[36] I. Oseledets and E. Tyrtyshnikov, *A unifying approach to the construction of circulant preconditioners*, Linear Algebra Appl., 418 (2006), pp. 435–449, doi:10.1016/j.laa.2006.02.037.
[37] C. Paige and M. Saunders, *Solution of sparse indefinite systems of linear equations*, SIAM J. Numer. Anal, 12 (1975), pp. 617–629.
[38] C. C. Paige and M. A. Saunders, *LSQR: An algorithm for sparse linear equations and sparse least squares*, ACM Trans. Math. Software, 8 (1982), pp. 43–71, doi:10.1145/355984.355989.
[39] B. Parlett, *The symmetric eigenvalue problem*, SIAM, Philadelphia, PA, USA, classics ed., 1998.
[40] J. Pestana and A. J. Wathen, *A preconditioned MINRES method for nonsymmetric Toeplitz matrices*, SIAM J. Matrix Anal. Appl., 36 (2015), pp. 273–288.
[41] S. Reitzinger, U. Schreiber, and U. van Rienen, *Algebraic multigrid for complex symmetric matrices and applications*, J. Comput. Appl. Math., 155 (2003), pp. 405–421.
[42] D. Sheen, I. Sloan, and V. Thomée, *A parallel method for time discretization of parabolic equations based on Laplace transformation and quadrature*, IMA J. Numer. Anal., 23 (2003), pp. 269–299.
[43] D. Silvester, H. Elman, and A. Ramage, *Incompressible Flow and Iterative Solver Software (IFISS) version 3.2*, May 2012. http://www.manchester.ac.uk/ifiss/.
[44] G. Strang, *A proposal for Toeplitz matrix calculations*, Stud. Appl. Math., 74 (1986), pp. 171–176.
[45] C. Van Loan, *Computational Frameworks for the Fast Fourier Transform*, SIAM, Philadelphia, PA, USA, 1992.
[46] E. Vecharynski and A. V. Knyazev, *Absolute value preconditioning for symmetric indefinite linear systems*, SIAM J. Sci. Comput., 35 (2013), pp. A696–A718.
[47] A. J. Wathen, *Preconditioning*, Acta Numer., 24 (2015), pp. 329–376, doi:10.1017/S0962492915000021.