NUMERICAL LINEAR ALGEBRA WITH APPLICATIONS
Numer. Linear Algebra Appl. 2001; **8**:111–125

Reliable preconditioned iterative linear solvers for some numerical integrators

D. Bertaccini*,†

Università di Roma "La Sapienza", Dipartimento di Matematica "G. Castelnuovo", P.le A. Moro 2, 00185 Roma, Italy.

SUMMARY

Implicit time-step numerical integrators for ordinary and evolutionary partial differential equations need, at each step, the solution of linear algebraic equations that are unsymmetric and often large and sparse. Recently, a block preconditioner based on circulant approximations for the linear systems arising in the boundary value methods (BVMs) was introduced by the author. Here, some circulant approximations are compared and a further new type is considered. Numerical experiments are presented to check the effectiveness of the various approximations that can be used in the underlying block preconditioner. Copyright © 2001 John Wiley & Sons, Ltd.

KEY WORDS: circulant preconditioning; unsymmetric Toeplitz matrices; linear multistep formulae; numerical integration of ODEs and evolutionary PDEs

1. INTRODUCTION

The solution of the large and sparse linear systems of algebraic equations, arising at each integration step of a numerical integrator for differential equations based on implicit formulae, constitutes a crucial part of its computational cost (see e.g. References [1–4]).

Let us consider the numerical integration of a solution (supposed unique) of the system of ordinary differential equations (ODEs) $y' = f(t, y)$, $y(t): \mathbb{R} \to \mathbb{R}^m$, f suitably smooth, under given suitably initial and/or boundary conditions. If we use an implicit scheme based on a set of k-step implicit linear multistep formulae (LMFs), at each integration step, we need to solve some linear algebraic equations. In each subintervals $[x_j, x_{j+1}]$ of a chosen mesh $\{x_0, \ldots, x_N\}$, such linear systems can be written as

$$M Y = b, \quad Y = (y_0, \ldots, y_s)^{\mathrm{T}}, \quad M = A \otimes I_m - h_j B \otimes J \tag{1}$$

where b is a suitable vector, A, B are the matrices whose entries are related to the time-step integrator formulae, J is the Jacobian matrix of the ODEs, and h_j is the integration step

*Correspondence to: D. Bertaccini, Università di Roma "La Sapienza", Dipartimento di Matematica "G. Castelnuovo", P.le A. Moro 2, 00185 Roma, Italy.
†E-mail: bertaccini@na-net.ornl.gov

Received 18 March 2000
Revised 5 September 2000

size. A and B have usually a Toeplitz pattern except for small rank perturbations (see References [3, 5, 6] for details). An $n \times n$ matrix $A_n = (a_{j,k})$ is said to be Toeplitz if $a_{j,k} = a_{j-k}$, $j, k = 1, \ldots, n$, i.e. A_n is constant along its diagonals, quasi Toeplitz if is a small rank perturbation of a Toeplitz matrix.

Notice that the matrices of some schemes not based on linear multistep formulae can be reduced to the quasi-Toeplitz (block) pattern using suitable transformations. This is the case of several classes of implicit Runge–Kutta methods (see Reference [5]).

When J, the Jacobian matrix of the underlying system of the ODEs, and/or when A and B are large and sparse, M defined as (1) turns out to be large and sparse. In those cases, the solution of the underlying linear system via a direct method is usually computationally very expensive, so an iterative method should be used. Unfortunately, unpreconditioned iterations of Krylov subspace methods such as GMRES [7] and BiCGStab [8] often show very slow convergence or even diverge. In order to speed up the convergence rate, in References [6, 9] we proposed a block preconditioner. In Reference [10] we analyzed the spectrum of the eigenvalues of the underlying matrices in connection to the stability of the methods.

The underlying preconditioner is based on a circulant approximation of the matrices A, B arising in (1). An $n \times n$ matrix $\breve{A}_n$ is said to be circulant if it is Toeplitz and its diagonals satisfy $\breve{a}_{n-j} = \breve{a}_{-j}$, $j = 1, \ldots, n-1$. The circulant matrices $\breve{A}_n$ are diagonalized by the Fourier matrix $F = (F_{j,k})$, $F_{j,k} = \mathrm{e}^{2\pi \mathrm{i} jk/n}/\sqrt{n}$, $j, k = 0, \ldots, n-1$, i is the imaginary unit, see Reference [11]. For the previous arguments, it follows that such matrices are easily and efficiently invertible using the Fast Fourier Transform (FFT), see Reference [12]. From 1986, there has been intensive work on circulant preconditioners and their spectral properties, see e.g. References [12–21].

The block preconditioner we mentioned above, used in connection with several Krylov methods for linear systems, has found to be effective for several differential problems and formulae, see References [6, 9] and [22].

In this paper we will compare the spectral, stability and convergence properties of our block preconditioner when using various circulant approximations as the Strang's [17] called also *natural* circulant, the T. Chan's [14] also called *optimal* and the P-circulant [6, 9, 10] for linear systems such as (1). Moreover, we introduce a further type of circulant approximation that can be used in the block preconditioners for (1). Such approximation is derived from the Strang's natural circulant to reduce as much as possible the drawbacks (e.g. a severe limitation on the differential problems it can integrate, see References [9, 10] and the next sections) and to retain the advantages (e.g. fast convergence for preconditioned iterations, see Reference [22]).

In Section 2 we describe the systems of linear algebraic equations arising in time-step integrators based on implicit LMFs. In Section 3 the block preconditioner for (1) and the circulant approximations above mentioned are described with merits and drawbacks. Section 4 is devoted to compare the main properties of the preconditioners of Section 3. Finally, in Section 5 we will present some numerical results and a few final remarks are given in Section 5.

2. THE LINEAR SYSTEMS

Often, in applied sciences, we have to solve linear systems as (1), where A, B are (small rank perturbations of) Toeplitz matrices. This is the case of the linear systems that arise when

we approximate a differential problem by using a linear multistep formulae-based numerical integrator. For the sake of simplicity, we consider a linear initial value problem

$$\begin{aligned} y'(t) &= f(t, y(t)) := Jy(t) + g(t), \quad t \in (t_0, T] \\ y(t_0) &= \eta \end{aligned} \tag{2}$$

where $y(t), g(t) : \mathbb{R} \to \mathbb{R}^m$, $J \in \mathbb{R}^{m \times m}$, $\eta \in \mathbb{R}^m$. The Jacobian matrix of (2) will be supposed diagonalizable. To approximate the solution of (2), we consider a generalization of a linear multistep technique (see e.g. References [3, 4]) known as boundary value methods (BVMs) (see e.g. References [5, 23] and references therein). A BVM approximates the solution of (2) by means of a discrete boundary value problem. The latter is obtained by using a k-step linear multistep formula of order p over a mesh that, for simplicity, can be supposed uniform, i.e. $t_j = t_0 + jh$, $j = 0, \ldots, s$, $h = (T - t_0)/s$:

$$\sum_{i=-v}^{k-v} \alpha_{i+v} y_{n+i} = h \sum_{i=-v}^{k-v} \beta_{i+v} f_{n+i}, \quad n = v, \ldots, s - k + v \tag{3}$$

where y_n is the discrete approximation to $y(t_n)$, $f_n = f(t_n, y_n) \equiv Jy_n + g_n$, $g_n = g(t_n)$, and the values

$$y_0, \ldots, y_{v-1}, \quad y_{s-k+v+1}, \ldots, y_s \tag{4}$$

are given. We observe that the IVP (2) provides only the initial value y_0. It is possible to avoid to supply the other conditions in (4) by coupling the *main method* (3) with other difference schemes of order p, called *additional methods*, which lead to a set of equations independent of those in (3). All the above-mentioned difference equations define the use of a BVM on problem (2). Notice that, for consistency with the differential problem, the coefficients of the formulae must satisfy the basic conditions (see e.g. References [4, 5]) $\rho(1) = 0$, $\rho'(1) = \sigma(1)$ where $\rho(z)$ and $\sigma(z)$ denotes the two characteristic polynomials associated with the given method, i.e.

$$\rho(z) = z^v \sum_{j=-v}^{k-v} \alpha_{j+v}\, z^j, \qquad \sigma(z) = z^v \sum_{j=-v}^{k-v} \beta_{j+v}\, z^j \tag{5}$$

and we consider the scaling $\sigma(1) = 1$ for (3).

The discrete problem generated by the application of the BVM to the problem (2) can be reduced to a linear system of algebraic equations

$$\begin{aligned} M\,Y &= e_1 \otimes \eta + h(B \otimes I)g, \\ e_1 &= (1, 0, \ldots, 0)^{\mathrm{T}} \in \mathbb{R}^{s+1}, \quad Y = (y_0, \ldots, y_s)^{\mathrm{T}}, \quad g = (g_0, \ldots, g_s)^{\mathrm{T}} \\ M &= A \otimes I_m - hB \otimes J \end{aligned} \tag{6}$$

where $A, B \in \mathbb{R}^{s+1\times s+1}$ are small rank perturbations of Toeplitz matrices, i.e.

$$A = \begin{pmatrix} 1 & \cdots & 0 & & & \\ \alpha_0^{(1)} & \cdots & \alpha_k^{(1)} & & & \\ \vdots & \vdots & \vdots & & & \\ \alpha_0^{(v-1)} & \cdots & \alpha_k^{(v-1)} & & & \\ \alpha_0 & \cdots & \alpha_k & & & \\ & \alpha_0 & \cdots & \alpha_k & & \\ & & \ddots & \ddots & \ddots & \\ & & & \alpha_0 & \cdots & \alpha_k \\ & & & \alpha_0^{(s-k+v+1)} & \cdots & \alpha_k^{(s-k+v+1)} \\ & & & \vdots & \vdots & \vdots \\ & & & \alpha_0^{(s)} & \cdots & \alpha_k^{(s)} \end{pmatrix} \tag{7}$$

and B is defined similarly, but with β_js instead of α_js, and all zeros in its first row. The entries $\alpha_i^{(j)}$ in (7) are the coefficients of the additional methods. Notice that the matrices A, B differ from their Toeplitz counterparts

$$A_{\mathrm{T}} = \begin{pmatrix} \alpha_v & \ldots & \alpha_{k-1} & \alpha_k & 0 & \ldots & 0 \\ \vdots & \alpha_v & \ddots & \alpha_{k-1} & \alpha_k & \ddots & \vdots \\ \alpha_0 & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \ddots & \ddots & \ddots & \ddots & \alpha_{k-1} & \alpha_k \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \alpha_{k-1} \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \ldots & \ldots & 0 & \alpha_0 & \ldots & \alpha_v \end{pmatrix} \tag{8}$$

only in their first v and last $k-v$ rows, independently on s.

The matrix M in (6) turns out to be large and sparse when either $s \gg k$ or J is large and sparse.

When the ODE (2) is non-linear, the discrete nonlinear problem corresponding to the approximation of a BVM can be reduced, at each step, to discrete problems in the form (6), see e.g. Reference [3].

3. THE BLOCK PRECONDITIONER

In order to obtain the preconditioner, let us consider the matrix

$$P = \check{A} \otimes I_m - h\check{B} \otimes \hat{J} \tag{9}$$

where $\hat{J}$ can be a suitable approximation of the Jacobian matrix of the IVP (2), $\check{A}$, $\check{B}$, are matrices approximating A, B as (6) which entries derive from coefficients of main method (3).

Some circulant approximations for such matrices have been already considered. We mention the optimal, the P-circulant and the natural circulant (see References [6, 9, 10, 22]).

3.1. Circulant approximations

The *optimal circulant* is defined as the minimizer of the functional $\|C - T\|_F$, where T is a given Toeplitz matrix,

$$T = (t_{ij}), \quad t_{ij} = t_{i-j}, \quad i, j = 0, \ldots, n-1 \tag{10}$$

C ranges in the set of circulant matrices, and $\|\cdot\|_F$ denotes the Frobenius norm. If $c_0, \ldots, c_{n-1}$ are the entries of the first row of C, and T is the Toeplitz matrix (10), we have (see Reference [14]):

$$c_j = \frac{(n-j)t_j + jt_{j-n}}{n}, \quad j = 0, \ldots, n-1 \tag{11}$$

C is Hermitian and definite positive if T is (see Reference [19]).

If T is the Toeplitz matrix (10), $s_0, \ldots, s_{n-1}$ are the entries of the first row of a $n \times n$ *Strang's natural circulant* matrix S for T, where

$$s_j = \begin{cases} t_j, & 0 < j \leqslant \lfloor \frac{n}{2} \rfloor \\ t_{j-n}, & \lfloor \frac{n}{2} \rfloor < j < n \end{cases} \tag{12}$$

The circulant matrices as above minimize $\|S - T\|_1$, $\|S - T\|_\infty$, S chosen in the set of all circulant matrices, see Reference [13].

Let us consider circulant approximations for the (quasi) Toeplitz matrix T written as

$$C = (c_{ij}), \quad c_{ij} = c_{i-j}, \quad c_j = s_{1,j} t_j + s_{2,j} t_{n-j}, \quad j = 0, \ldots, n-1 \tag{13}$$

and $s_{1,j} = s_{1,j}(n)$, $s_{2,j} = s_{2,j}(n)$, $j = 0, \ldots, n-1$ are piecewise linear function of j. It is easy to check that all circulant approximations above mentioned can be written as (13). Consider the following coefficients for (13):

$$s_{1,j} = 1 + \frac{j}{n}, \quad s_{2,j} = \frac{j}{n}, \quad j = 0, \ldots, n-1 \tag{14}$$

The circulant approximation (13), (14) will be called *P-circulant* for brevity. The P-circulant, the natural and the optimal circulant are equivalent for T as $n \to \infty$ in the sense of the (linear) approximation processes (see Reference [16]).

We recall that the eigenvalues ϕ_j, $j = 0, \ldots, s$ of a circulant matrix $\breve{A}$ can be written as a linear combination of the entries $\tilde{\alpha}_j$, $j = 0, \ldots, s$ of its first row (see Reference [11])

$$\phi_l = \sum_{j=0}^{s} \tilde{\alpha}_j \varepsilon^{jl}, \quad l = 0, \ldots, s, \quad \varepsilon = \mathrm{e}^{2\pi \mathrm{i}/(s+1)} \tag{15}$$

Using (15), we can give an explicit expression of the eigenvalues of the circulant approximations as above for A, B as (7) as a function of the coefficients of (3). Indeed, the eigenvalues of the optimal circulant for A are

$$\phi_l = \sum_{j=-v}^{k-v} \alpha_{j+v} \left(1 - \frac{|j|}{s+1}\right) \varepsilon^{jl}, \ l = 0, \ldots, s, \ \varepsilon = \mathrm{e}^{2\pi \mathrm{i}/(s+1)} \tag{16}$$

Similarly, for the natural circulant we have

$$\phi_l = \sum_{j=-v}^{k-v} \alpha_{j+v} \varepsilon^{jl}, \quad l = 0, \ldots, s \tag{17}$$

and finally for P-circulants

$$\phi_l = \sum_{j=-v}^{k-v} \alpha_{j+v} \left(1 + \frac{j}{s+1}\right) \varepsilon^{jl}, \quad l = 0, \ldots, s \tag{18}$$

see Reference [10] for details. The eigenvalues of the circulant matrix approximating B can be derived from (16)–(18) by changing α_j with β_j, $j = 0, \ldots, k$.

The T. Chan's circulant approximation (11) for A, B in (6) can be effective only for a small step number k of the underlying LMF (see References [6, 9] and Section 5), but can be very ill conditioned even if the original matrix A is well conditioned, see the analysis in Reference [10].

The Strang's is in some way the most natural circulant approximation that can be used for A, B in (6). Notice that $\breve{A}_{\mathrm{NAT}} - A$, $\breve{B}_{\mathrm{NAT}} - B$ are small rank matrices for all k and s, $\breve{A}_{\mathrm{NAT}}$, $\breve{B}_{\mathrm{NAT}}$ natural circulant approximations for A, B. If the differential problem has a Jacobian matrix J whose eigenvalues have negative real part, bounded away from zero and moderate imaginary part, the natural block preconditioner (9) for (6) can give fast preconditioned iterations (see Reference [22] and Section 5). Unfortunately, as observed in References [9, 10], there is a severe restriction on the possible range of the differential problems it can integrate. Indeed, differently to M as (6), the preconditioner (9) based on the natural circulant is singular when the Jacobian matrix is singular and can be ill conditioned even when M and J are well conditioned and non-singular, see Sections 4.2 and 5. Notice that, especially in case of non-trivial non-linear problems, usually one cannot know in advance information on the eigenvalues of the Jacobian. The above restrictions can often be avoided by using the *Modified Strang circulant* we introduce in the sequel, or the P-circulant (see References [6, 9]). Indeed, we have proved in Reference [10] that P-circulants are well conditioned for various A-stable ($A_{v,k-v}$-stable, a generalization of A-stability for BVMs [5]) methods. Unfortunately, a P-circulants-based block preconditioner may require a few more preconditioned iterations for some problems with respect to the previously mentioned circulant approximations (see e.g. Reference [22] and Section 5).

3.1.1. Modified Strang's circulant. In Reference [10] we have observed that the natural preconditioner has the first eigenvalue equal to zero, i.e. ϕ_0 as (17) is zero for all LMFs as (3) (use (5) and (17) for $l = 0$). Let us introduce an approximation based on the natural circulant for a given matrix by substituting its null eigenvalue ϕ_0, e.g. with $\tilde{\phi}_0 = \chi$, $\mathrm{Re}(\chi) > 0$. The matrix $\breve{A}_{\mathrm{MS}}$ that results will be called *Modified Strang's circulant* or, for short, *MS-circulant*, is a rank-one correction of the natural circulant. Indeed, notice that

$$\breve{A}_{\mathrm{NAT}} = F^* D_{\mathrm{NAT}} F, \quad D_{\mathrm{NAT}} = \mathrm{diag}(\phi_0, \ldots, \phi_s)$$

$$F = (F_{j,k}), \quad F_{j,k} = \mathrm{e}^{2\pi \mathrm{i} jk/n}/\sqrt{n}, \quad j, k = 0, \ldots, n-1$$

i.e. D_{NAT} is a diagonal matrix whose entries ϕ_j, $j=0,\ldots,s$ are defined as in (17),

$$\begin{aligned}\breve{A}_{\mathrm{MS}} &= F^{*}\mathrm{diag}(\hat{\phi}_0,\phi_1,\ldots,\phi_s)F = F^{*}[D_{\mathrm{NAT}}+\mathrm{diag}(\hat{\phi}_0-\phi_0,0,\ldots,0)]F\\ &= F^{*}D_{\mathrm{NAT}}F+E=\breve{A}_{\mathrm{NAT}}+E,\quad E=F^{*}\mathrm{diag}(\hat{\phi}_0-\phi_0,0,\ldots,0)F\end{aligned}\tag{19}$$

and E is a rank-one circulant matrix. We have experienced that the most effective choices for $\tilde{\phi}_0=\chi$ are usually either $\hat{\phi}_0=1/(s+1)$, i.e. the same value of the first eigenvalue of a P-circulant, or $\hat{\phi}_0=\mathrm{Re}(\phi_s)$. The latest choice is based on symmetry reasons and it is the one we will use for the numerical tests in Section 5.

Similar techniques based on the perturbation of the eigenvalue with smallest modulus of a circulant matrix were already used e.g. in References [15, 20, 24] and, recently, also in Reference [25].

The block preconditioner (9) based on MS-circulants is robust enough to integrate various classes of differential problems with steady-state solutions, increasing and decreasing modes. Moreover, the MS-circulant preconditioned iterations often converge very fast, as the Strang's circulant-based one does for the problems with only decreasing modes, e.g., when J as in (1) has negative and safely bounded away from zero eigenvalues, see Section 5.

4. SPECTRUM, INVERTIBILITY OF THE PRECONDITIONER AND CONVERGENCE

4.1. Spectrum of the circulant approximations

In Reference [10] we have proved some properties of circulant approximations for A, B as (6). More precisely, we have shown that, for several families of A-stable ($A_{v,k-v}$-stable) boundary value methods, the P-circulant and the natural circulant approximations to the matrices A, B as (7) have eigenvalues in the right-half plane and those of the former have strictly positive real part. Notice that this property is shared also by A_{T}, B_{T} as (8) (see Reference [5]) and by MS-circulants. Indeed, if the natural circulant approximation for A as (7) has the spectrum of eigenvalues in the right-half plane, then the related MS-circulant has uniformly bounded and positive real part eigenvalues. This is a consequence of the result stated for P-circulant and for the natural circulant. To this end, it is enough to observe that the eigenvalue of the Strang's circulant for A that is equal to zero, i.e. ϕ_0 (the other eigenvalues have positive real part), was set to χ, $\mathrm{Re}(\chi)>0$ for MS-circulants.

The results stated above will be useful in the investigation of the invertibility of the block preconditioner (9).

4.2. Invertibility of the block preconditioner

Let us give some sufficient conditions for the uniform invertibility with respect to the stepsize $h>0$ of the block preconditioner (9).

In References [6, 22] it has been observed that, under suitable hypotheses, the block preconditioner based on the natural circulant is non-singular.

Proposition 4.1. *If the Jacobian matrix J of* (2) *has eigenvalues μ_r such that* $\mathrm{Re}(\mu_r)<-\delta<0$, $r=1,\ldots,m$, *then the preconditioner* (9) *based on natural circulants* (12) *is invertible for A-stable* ($A_{v,k-v}$*-stable*) *formulae* (3).

The block preconditioner (9) based on the MS-circulants (or on the P-circulants) is invertible for a wider choice of differential problems with respect to the one based on the natural circulants. To this end, we can observe that, if the eigenvalues μ_r, $r=1,\ldots,m$ of the matrix J of the IVP (2) have non-positive real part, then the matrix P as (9) is non-singular for all step size $h>0$ for all schemes and circulants $\breve{A}$, $\breve{B}$ approximating A, B such that

$$\mathrm{Re}\left(\frac{\phi_j}{\psi_j}\right) \tag{20}$$

is positive for all $j=0,\ldots,s$, where ϕ_j, ψ_j, $j=0,\ldots,s$ are the eigenvalues of $\breve{A}$, $\breve{B}$.

When (20) is positive, it can be observed that we can guarantee the invertibility of the preconditioner also in presence of increasing modes (i.e. $\mathrm{Re}(\mu_r)>0$ for some r) by choosing the stepsize h_j such as to satisfy the non singularity of both the matrices (1) and (9).

Proposition 4.2. *If the Jacobian matrix J of* (2) *has eigenvalues* μ_r *such that* $\mathrm{Re}(\mu_r)\leqslant 0$, $r=1,\ldots,m$, *then the preconditioner* (9) *based on MS-circulants is invertible for A-stable* ($A_{v,k-v}$-*stable*) *linear multistep formulae.*

Proof

Follows from the above arguments by observing that, when using MS-circulants, the expression (20) cannot vanish for finite values of s. Indeed, using Theorem 4.7.2 in Reference [5], which states mainly that

$$\mathrm{Re}\left(\frac{\rho(\mathrm{e}^{\mathrm{i}\theta})}{\sigma(\mathrm{e}^{\mathrm{i}\theta})}\right)\geqslant 0, \quad \theta\in\mathbb{R}$$

for all A-stable ($A_{v,k-v}$-stable if $k>v$) LMFs, $\rho(z)$, $\sigma(z)$ characteristic polynomials as (5), it is straightforward to observe that

$$\mathrm{Re}\left(\frac{\rho(\mathrm{e}^{\mathrm{i}\theta})}{\sigma(\mathrm{e}^{\mathrm{i}\theta})}\right)=\mathrm{Re}\left(\frac{\mathrm{e}^{-\mathrm{i}v\theta}\rho(\mathrm{e}^{\mathrm{i}\theta})}{\mathrm{e}^{-\mathrm{i}v\theta}\sigma(\mathrm{e}^{\mathrm{i}\theta})}\right)=\mathrm{Re}\left(\frac{\phi_j}{\psi_j}\right)>0, \quad \theta=\frac{2\pi j}{s+1},\ j=1,\ldots,s \tag{21}$$

and ϕ_j, ψ_j, $j=0,\ldots,s$ are the eigenvalues of the natural circulants $\breve{A}_{\mathrm{NAT}}$, $\breve{B}_{\mathrm{NAT}}$, respectively, while, for $j=0$, the ratio (20) is equal to $\hat{\phi}_0/\psi_0=\chi/\sigma(1)=\chi$, where $\mathrm{Re}(\chi)>0$ by the definition of MS-circulant approximation and for the scaling condition on $\sigma(z)$ (see Section 2). The eigenvalues of P as (9) are

$$\lambda_{j,r}(P)=\phi_j-h\psi_j\mu_r,\ j=0,\ldots,s,\ r=1,\ldots,m \tag{22}$$

Then, since (21) holds true, for the consistency conditions (5) and the above arguments, if $\mathrm{Re}(\mu_r)\leqslant 0$, $\lambda_{j,r}(P)$ cannot be zero. ∎

For the P-circulants, the strictly positivity of expression (20) has been verified directly for the classes of A-stable ($A_{v,k-v}$-stable if $k>v$) methods considered in References [6, 10]. For those methods, the same results of the proposition 4.2 hold true.

4.3. Convergence of preconditioned iterations

In Reference [6] we have observed that, if P^{-1} is well defined, the spectrum of the P-circulant and of the natural circulant preconditioned matrix $P^{-1}M$, M as (6), is clustered around

$(1,0)\in\mathbb{C}$. More precisely, in the notation of References [20, 21], the (complex) eigenvalues of $P^{-1}M$ have a proper cluster in that point, i.e. the number of the outliers do not depend on s, e.g., if $\mathrm{Re}(\mu_r)<0,\ r=1,\ldots,m,\ \mu_r$ eigenvalues of the Jacobian matrix. Otherwise, we can have simply a cluster in $(1,0)\in\mathbb{C}$, see Reference [6, Section 4.2] for more details. To this end, we write $P^{-1}M$ as the sum of the identity, a low rank and a small $l2$-norm matrix.

Theorem 4.3. *Let M be the matrix of the linear system* (1), *P its block P-circulant preconditioner as* (9). *Then, for fixed $\delta>0$, there exists $C_\delta\geqslant 0$, $s_\delta\geqslant k$ such that, for all $s\geqslant s_\delta$ ($s+1$ is the size of A,B),*

$$P^{-1}M=I+M_\delta^{(1)}+M_\delta^{(2)} \tag{23}$$

where $\mathrm{rank}(M_\delta^{(2)})\leqslant m[2(k+1)+C_\delta]$ *and* $\|M_\delta^{(1)}\|\leqslant\delta c$, *where c does not depend on s.*

Proof
See Reference [6, Theorem 4.1]. ∎

As a corollary of the above theorem, we have:

Proposition 4.4. *Let M be the matrix of system* (6), *P its MS-circulant* (*natural circulant*) *block preconditioner* (9). *Then, if the block preconditioner P is invertible,*

$$P^{-1}M=I+\tilde{M}^{(2)} \tag{24}$$

where $\mathrm{rank}(\tilde{M}^{(2)})\leqslant 2m(k+2)$ $(\mathrm{rank}(\tilde{M}^{(2)})\leqslant 2m(k+1)$ *for the natural circulant*).

Proof
Let us consider the MS-circulant approximations $\breve{A}_{\mathrm{MS}}$, $\breve{B}_{\mathrm{MS}}$ for A, B. Then,

$$E_A^{(2)}=A-\breve{A}_{\mathrm{MS}},\quad E_B^{(2)}=B-\breve{B}_{\mathrm{MS}}$$

have at most rank $k+2$ and $k+1$ for natural circulants, since $\breve{A}_{\mathrm{MS}}$ is a one-rank correction of $\breve{A}_{\mathrm{NAT}}$, see (19). Indeed, it is easy to check from the definition of the Strang's approximation, (7) and (8) that $E_A^{(2)}$, $E_B^{(2)}$ are $(s+1)\times(s+1)$ matrices whose non-zero entries $e_{i,j}^{(2)}$ have indices

$$\begin{aligned}
&i=1,\ldots,v,\quad j=1,\ldots,k+1;\quad i=1,\ldots,v,\quad j=s+2-v,\ldots,s+1;\\
&i=s+2-k+v,\ldots,s+1,\quad j=1,\ldots,k-v;\quad i=s+2-k+v,\ldots,s+1,\\
&j=s+1-k,\ldots,s+1.
\end{aligned}$$

Thus, if P is invertible, we have

$$P^{-1}M=I+P^{-1}(M-P)=I+P^{-1}(E_A^{(2)}\otimes I_m-hE_B^{(2)}\otimes J)=I+\tilde{M}^{(2)} \tag{25}$$

Then, $\tilde{M}^{(2)}$ has at most rank $2m(k+1)$ and $2m(k+2)$ if MS-circulants are considered. ∎

Notice that, under the same assumptions of the above theorems, also the singular values of the preconditioned matrix have a (proper) cluster at 1. The proof follows by using [20, Lemma 2.2].

From the results stated above, we expect that preconditioned iterations will be much faster than the unpreconditioned ones. To this end, we can use similar arguments as in Reference [26] for the rate of convergence of MS-circulant preconditioned CG-iterations. We can experience a fast convergence also if the preconditioner is used for the GMRES method, see, e.g. Reference [27].

Notice that, given a block preconditioner $\tilde{P}$ as (9) using a certain approximation for A, B as in (1), if $\tilde{P}^{-1}M$ has clustered spectrum but some singular values lie much closer to the origin with respect to 1, a delay of a possibly very high number of iterations for CG, GMRES and other Krylov methods with respect to the other preconditioners can be expected, see e.g. Reference [28, Section 5]. This is the case of the preconditioners using Strang's approximations, see example 1 in the next section and [6, Section 5].

A detailed analysis of the convergence behaviour of the GMRES iterations for the class of block circulant preconditioners as (9) will be presented in a forthcoming paper.

5. NUMERICAL RESULTS

To compare the effectiveness of our preconditioner (9) with the various circulant approximations we have considered some linear test problems. In this section we will compare the number of iterations needed to converge for BiCGStab [8] and (unrestarted) GMRES [7] to solve the linear systems as (6).

More numerical tests can be found in References [6, 9, 26], some implementation details in Reference [6].

The initial guess for those iterative solvers is zero. The stopping criterion is $\|r_j\|_2 < 10^{-6}\|b\|_2$, r_j updated true residual after j iterations. All experiments are performed in MATLAB.

Example 1. Wave equation of first order.

$$\begin{aligned} u_t - u_x &= 0, \\ u(x,0) &= g(x), \quad x \in [0,\pi] \\ u(\pi,t) &= 0, \quad t \in [0,2\pi] \end{aligned} \tag{26}$$

We discretize the partial derivative $\partial/\partial x$ with the first-order forward difference and step size $\Delta x = \pi/m$, $x_j = j\Delta x$ (upwind discretization). We obtain a family of systems of m ODEs

$$\begin{aligned} y'(t) &= L_m y(t), t \in [0,2\pi] \\ y(0) &= \eta, \eta = (g(x_0) \ \cdots \ g(x_{m-1}))^{\mathrm{T}} \\ L_m &= \frac{1}{\Delta x} \begin{pmatrix} -1 & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & -1 \end{pmatrix} \end{aligned} \tag{27}$$

The generalized Adams method with $k = 4$ (order 5, see Reference [5] for the coefficients) is used to solve (27) and $g(x) = \sin(x)$.

Table I. Number of matrix–vector multiplications required for convergence of iterative methods in example 1.

		GMRES					BiCGstab				
m	s	I	C	S	P	MS	I	C	S	P	MS
20	8	38	9	9	9	9	100	9	10	10	10
	16	29	8	8	8	8	54	9	8	9	9
	64	68	7	6	7	6	76	7	7	7	7
50	8	100	11	11	11	11	—	12	13	13	13
	16	87	9	9	9	9	—	9	9	10	9
	64	64	7	7	7	7	95	7	7	7	8
100	8	235	12	12	12	12	—	12	14	14	13
	16	217	9	9	9	9	—	9	9	10	10
	50	145	8	7	8	7	—	8	8	8	8

Example 2. Heat equation with a variable diffusion coefficient.

$$\begin{aligned} \frac{\partial u}{\partial t} - \frac{\partial}{\partial x}\left(a(x)\frac{\partial u}{\partial x}\right) &= 0, \quad a(x) > 0, \\ u(0,t) &= u(x_{\max}, t) = 0, \quad t \in [0, 2\pi] \\ u(x,0) &= g(x), \quad x \in [0,\pi] \end{aligned} \tag{28}$$

If we discretize the operator $\partial/\partial x$ with centered differences and stepsize $\delta x = \pi/(m+1)$, we obtain a system of m ODEs such as (27) whose $m \times m$ Jacobian matrix L_m is symmetric and tridiagonal

$$L_m = \frac{1}{2\Delta x}\begin{pmatrix} -a(-1/2) - a(1/2) & a(1/2) & & & \\ a(1/2) & -a(-1/2) - a(3/2) & a(3/2) & & \\ & \ddots & \ddots & & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & \end{pmatrix} \tag{29}$$

but Toeplitz if and only if $a(x)$ is constant. We note that the above Jacobian matrices have real and strictly negative eigenvalues. The generalized Adams method with $k = 4$ (order 5, see Reference [5] for the coefficients) is used to solve (27). In the numerical examples, we will consider $a(x) = \exp(-x^r)$, $r = 0$ (constant diffusion coefficient) and $r = 3$, while two different functions for $g(x)$ (see below).

In Tables I–III we can see the number of matrix–vector multiplications $P^{-1}M$ required for the convergence of GMRES and BICGStab (BiCGstab(2) when the convergence behaviour of BICGStab is too much erratic) for examples 1 and 2. For the GMRES, one matrix–vector multiplication is required at each iteration, while two are needed for BiCGStab.

The columns labeled I, C, S, P, MS give the number of matrix–vector products for unpreconditioned, T. Chan's, Strang's, P-circulant and MS-circulant preconditioned iterations respectively. A "–" means that convergence was not attained after 500 matrix–vector multiplications.

Table II. Number of matrix–vector multiplications required for convergence of iterative methods in example 2, $a(x) = \exp(-x^r)$, $r = 0, 3$, $g(x) = \sin(x)$.

			GMRES					BiCGstab				
r	m	s	I	C	S	P	MS	I	C	S	P	MS
0	24	12	39	6	3	6	3	148	10	6	8	6
		48	128	6	3	6	3	—	10	4	10	5
0	96	12	233	6	3	6	4	—	10	6	8	6
		48	—	6	3	6	4	—	10	4	10	5
3	24	12	119	14	16	14	14	318	20	198	20	20
		48	233	15	15	15	14	—	20	168	20	20
3	96	12	—	15	39	15	15	—	20	—	20	20
		48	—	16	47	16	14	—	22	—	20	20

Table III. Number of matrix–vector multiplications required for convergence of iterative methods in example 2, $a(x) = \exp(-x^r)$, $r = 0, 3$, $g(x) = \sin(x)$.

			GMRES					BiCGstab				
r	m	s	I	C	S	P	MS	I	C	S	P	MS
0	24	12	132	9	8	8	8	270	12	10	12	12
		48	161	9	8	9	8	320	12	12	12	10
0	96	12	–	8	6	7	7	–	12	12	10	10
		48	–	9	7	9	7	–	12	12	12	12
3	24	12	117	13	10	14	14	310	18	192	18	18
		48	232	14	24	14	14	–	20	178	20	20
3	96	12	–	15	48	14	13	–	20	–	20	20
		48	–	16	134	15	14	–	22	–	20	20

The computational cost of the block preconditioner was analyzed in Reference [6], here we remark only that each iteration costs $O(ms \log s + ms)$ flops. Indeed, the Jacobian matrices considered contain only a few non-zero diagonals. We stress that the above examples are linear and therefore they need the solution of one linear system such as (6) for their entire interval of integration. Moreover, the amount of the work for a matrix–vector multiplication is the same for all circulant approximations considered.

We have observed that the block preconditioner based on the Strang's circulant, when it is well conditioned, converge usually in almost the same number of iterations as its more flexible and reliable counterpart, the MS-circulant-based one.

In Figures 1 and 2 we can see the spectrum of the eigenvalues of M as (6) and of $P^{-1}M$ for example 1 and the distribution of the singular values for the example 2. The Strang's circulant can be ill conditioned even for small k and for problems whose Jacobian matrix is non-singular such as example 2. More precisely, notice that already for small positive values of r in the diffusion coefficient $a(x) = \exp(-x^r)$, the Jacobian matrix (29) has some of its negative eigenvalues that can be small in modulus. As a consequence, the preconditioner (9)

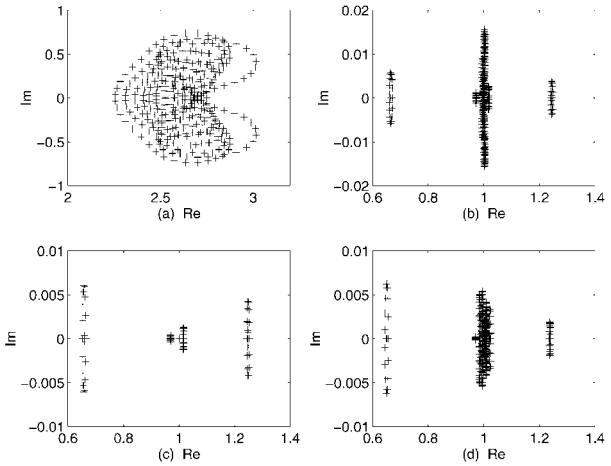

Figure 1. Example 1. Spectrum of the eigenvalues of the matrix M before (a) and after (b) optimal, (c) Strang's/MS-circulant, (d) P-circulant preconditioning.

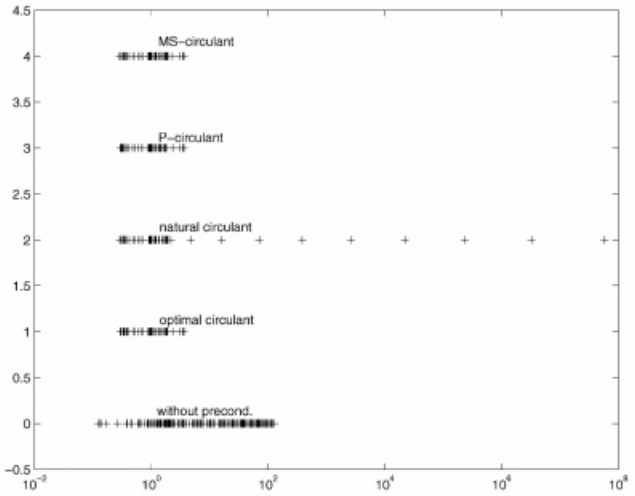

Figure 2. Example 2. Distribution of the singular values of the matrix M before and after the T. Chan's optimal, Strang's, P-circulant and MS-circulant preconditioning for $r = 3$, $m = 24$, $s = 12$.

based on the Strang's approximation can be severely ill conditioned, see Tables II and III and Figure 2. Notice that we have considered two initial values for problem 2, both occurring in practice. The first, given by $g(x) = \sin(x)$ in (28), is used also in the first test problem in Reference [22], is an eigenvector of the Jacobian matrix L_m when the diffusion coefficient $a(x)$ is constant (e.g. $r = 0$ in our case). As a consequence, we can see that the Strang's natural (see also [22]) and the MS-circulant based block preconditioner converge sensibly faster than the others, when $a(x)$ is constant and similarly if $g(x) = c \sin(nx)$, n integer, c constant. On

the contrary, as is the case of $a(x) = \exp(-x^3)$, the Strang's approximation can be unsuitable, see Table II. Let us consider the other initial value, i.e. $g(x) = x$. We can observe that, except for the natural circulant, all the considered approximation are more or less equivalent, perhaps the MS-circulant is slightly better.

6. CONCLUDING REMARKS

In this paper, a block preconditioner for the linear systems arising in certain numerical integrators based on linear multistep formulae was considered. Some properties of the preconditioner were discussed in relation to the various circulant approximations it can be based on such as the T. Chan's, the Strang's, the P-circulant and the MS-circulant. The latest is a one-rank correction of the Strang's circulant that we have introduced here to keep non-singular the approximations used in the underlying block preconditioner and to retain fast convergence. Indeed, the MS-circulant-based block preconditioner has shown to be fast and reliable with respect to the above mentioned in the numerical integration of certain differential problems such as some evolutionary partial differential equations.

REFERENCES

1. Chan TF, Jackson KR. The use of iterative linear-equation solvers in codes for large systems of stiff IVPs for ODEs. *SIAM Journal on Scientific and Statistical Computing* 1986; **7**:378–417.
2. Gear CW, Saad Y. Iterative solution of linear equations in ODE codes. *SIAM Journal on Scientific and Statistical Computing* 1983; **4**:583–601.
3. Hairer E. Wanner G. *Solving Ordinary Differential Equations II. Stiff and Differential-Algebraic Problems*. Springer: Berlin, 1991.
4. Lambert JD. *Numerical Methods for Ordinary Differential Systems*. Wiley: Chichester, 1991.
5. Brugnano L, Trigiante D. *Solving ODE by Linear Multistep Methods*: *Initial and Boundary Value Methods*. Gordon & Breach: Reading, UK, 1998.
6. Bertaccini D. A circulant preconditioner for the systems of LMF-based ODE codes. *SIAM Journal on Scientific Computing* 2000; **22**(3):767–786.
7. Saad Y, Schultz MH, GMRES: a generalized minimal residual algorithm for solving nonsymmetric linear systems. *SIAM Journal on Scientific and Statistical Computing* 1986; **7**:856–869.
8. van der Vorst HA. BICGSTAB: a fast and smoothly converging variant of Bi-CG for the solution of nonsymmetric linear systems. *SIAM Journal on Scientific and Statistical Computing* 1992; **10**:631–644.
9. Bertaccini D. P-circulant preconditioners and the systems of ODE codes. In *Iterative Methods in Scientific Computation* IV, Kincaid DR, Elster AC (eds), IMACS Series in Computational and Applied Mathematics. IMACS: NJ, 1999; 179–193.
10. Bertaccini D. An analysis of circulant preconditioners for linear multistep integrators. submitted.
11. Davis PJ. *Circulant Matrices*. Wiley: New York, 1979.
12. Chan RH, Ng MK. Conjugate gradient methods for Toeplitz systems. *SIAM Review* 1996; **38**(3):427–482.
13. Chan RH. Circulant preconditioners for Hermitian Toeplitz systems. *SIAM Journal on Matrix Analysis and Applications* 1989; **10**:542–550.
14. Chan TF. An optimal circulant preconditioner for Toeplitz systems. *SIAM Journal on Scientific and Statistical Computing* 1988; **9**:766–771.
15. Fiorentino G, Serra S. Fast parallel solvers for elliptic problems. *Computers & Mathematics with Applications* 1996; **32**(2):61–68.
16. Serra S. Toeplitz preconditioners constructed from linear approximation processes. *SIAM Journal on Matrix Analysis and Applications* 1998; **20**:446–465.
17. Strang G. A proposal for Toeplitz matrix calculations. *Studies in Applied Mathematics* 1986; **74**:171–176.
18. Strela VV, Tyrtyshnikov EE. Which circulant preconditioner is better? *Mathematics of Computations* 1996; **65**:137–150.
19. Tyrtyshnikov EE. Optimal and superoptimal circulant preconditioners. *SIAM Journal on Matrix Analysis and Applications* 1992; **13**:459–473.

20. Tyrtyshnikov EE. Circulant preconditioners with unbounded inverses. *Linear Algebra and its Applications* 1995; **216**:1–23.
21. Tyrtyshnikov EE. A unifying approach to some old and new theorems on distribution and clustering. *Linear Algebra and its Applications* 1996; **232**:1–43.
22. Chan RH, Ng MK, Jin XQ. Circulant preconditioners for solving ordinary differential equations. submitted .
23. Axelsson AOH, Verwer JG, Boundary value techniques for initial value problems in ordinary differential equations. *Mathematics of Computations* 1985; **45**(171):153–171.
24. Hanke M. Nagy J. Plemmons R. Preconditioned iterative regularization for Ill-posed problems. In *Numerical Linear Algebra*, Reichel L, Ruttan A, Varga R (eds). de Gruyter: Berlin, 1993
25. Chan RH, Ng M, Yip A. The best circulant preconditioners for Hermitian Toeplitz matrices. *SIAM Journal on Numerical Analysis* 2000; **38**:876–896.
26. Bertaccini D, Ng MK. The Convergence rate of Block Preconditioned Systems Arising from LMF-based ODE codes. BIT-To appear.
27. Greenbaum A. *Iterative Methods for Solving Linear Systems*. SIAM: Philadelphia, 1997.
28. Driscoll TA, Toh KC, Trefethen LN. From potential theory to matrix iterations in six steps. *SIAM Review* 1998; **40**(3): 547–578.
29. Saad Y. *Iterative Methods for Sparse Linear Systems*. PWS Publishing company: Boston, MA, 1995.

