Electron. J. Math. Phys. Sci., **2002**, *1, 1*, 14–46

Electronic Journal of
Mathematical
and Physical Sciences
EJMAPS
ISSN: 1538-263X
www.ejmaps.org

STRANG-TYPE PRECONDITIONERS FOR SOLVING SYSTEMS OF ODES BY BOUNDARY VALUE METHODS

Raymond H. Chan[1*], Xiao-Qing Jin[2] and Yue-Hung Tam[1]

[1]Department of Mathematics, The Chinese University of Hong Kong, Shatin, Hong Kong.

[2]Faculty of Science and Technology, The University of Macau, Macau.

[*]Corresponding Author. E-mail: rchan@math.cuhk.edu.hk

Received: 2 April 2002 / Accepted: 2 May 2002 / Published: 22 August 2002

Abstract: In this paper, we survey some of the latest developments in using boundary value methods for solving systems of ordinary differential equations with initial values. These methods require the solutions of one or more nonsymmetric, large and sparse linear systems. The GMRES method with the Strang-type preconditioner is proposed for solving these linear systems. One of the main results is that if an A_{ν_1,ν_2}-stable boundary value method is used for an m-by-m system of ODEs, then the preconditioner is invertible and the preconditioned matrix can be decomposed as $I+L$ where I is the identity matrix and the rank of L is at most $2m(\nu_1+\nu_2)$. It follows that when the GMRES method is applied to solving the preconditioned systems, the method will converge in at most $2m(\nu_1+\nu_2)+1$ iterations. Applications to differential algebraic equations and delay differential equations are also given.

Keywords: boundary value method, GMRES, ordinary differential equation.

AMS Mathematical Subject Classification Codes: 65L05, 65L06, 65N10.

---

1 Introduction

Consider the initial value problem

$$\begin{cases} \dfrac{\mathrm{d}\mathbf{y}(t)}{\mathrm{d}t} = J_m\mathbf{y}(t) + \mathbf{g}(t), \quad t \in (t_0, T], \\ \mathbf{y}(t_0) = \mathbf{z}, \end{cases} \tag{1}$$

where $\mathbf{y}(t)$, $\mathbf{g}(t) : \mathbb{R} \to \mathbb{R}^m$, $\mathbf{z} \in \mathbb{R}^m$, and J_m is the stiffness matrix in $\mathbb{R}^{m\times m}$. The initial value methods (IVMs), such as the Runge-Kutta methods and the waveform relaxation methods are well-known methods for solving (1), see [26]. This paper however discusses another class of methods called the boundary value methods (BVMs), see [5].

Using BVMs to discretize (1), we get a linear system

$$M\mathbf{y} \equiv (A \otimes I_m - hB \otimes J_m)\mathbf{y} = \mathbf{e}_1 \otimes \mathbf{z} + h(B \otimes I_m)\mathbf{g},$$

where $\mathbf{y}$, $\mathbf{e_1}$, $\mathbf{z}$, and $\mathbf{g}$ are vectors, I_m is the m-by-m identity matrix, and A and B are matrices depending on the multistep rule we used to discretize the time-derivative. The advantage of using BVMs is that the methods are more stable and the resulting linear system is hence more well-conditioned. However, the system is in general large and sparse (with band-structure), and solving it is a major problem in the application of the BVMs.

In this paper, we consider the use the GMRES method, which is one of Krylov subspace methods, for solving the discrete system. In order to speed up the convergence of the GMRES iterations, we use the Strang-type preconditioners to precondition the discrete system. The Strang-type block-circulant preconditioner of M is of the form

$$S = s(A) \otimes I_m - h \cdot s(B) \otimes J_m.$$

where $s(A)$ and $s(B)$ are the Strang-type circulant preconditioners [9, 12] for A and B, which will be discussed in §3.

The advantage of the Strang-type preconditioner is that if an A_{ν_1,ν_2}-stable BVM is used in (1), then S is invertible and the preconditioned matrix can be decomposed as

$$S^{-1}M = I_{m(s+1)} + L,$$

where the rank of L is at most $2m(\nu_1+\nu_2)$ which is independent of the integration step size h. It follows that the GMRES method applied to the preconditioned system will converge in at most $2m(\nu_1+\nu_2)+1$ iterations in exact arithmetic.

The outline of the paper is as follows. In §2, we will give some background knowledge about the linear multistep formulae and the preconditioned GMRES method. Then, we will investigate the properties of the Strang-type block-circulant preconditioner in §3. The convergence and cost analysis of the method will also be given with a numerical example. In §4, we use the block-circulant preconditioner with circulant blocks (BCCB preconditioner) to speed up the convergence rate. A modified version of the Strang-type BCCB preconditioner will be proposed in this section to handle the problem for singular or nearly singular stiffness matrix J_m. In §5, we will combine the idea of waveform relaxation method and the BVM to solve (1). We will use the well-known circulant and skew-circulant decomposition for splitting of the stiffness matrix J_m. Comparisons between this method and the classical splitting methods are also given. We will briefly discuss the applications of the Strang-type preconditioner with BVMs for solving the differential algebraic equations (DAEs) and delay differential equations (DDEs) in §6.

2 Background

Before discussing how to solve (1), we first give some background information on linear multistep formulae and the preconditioned GMRES method in this section.

2.1 Linear Multistep Formulae

Consider an initial value problem

$$\begin{cases} y' = f(t,y), & t \in (t_0, T], \\ y(t_0) = y_0. \end{cases}$$

The μ-step *linear multistep formula* (LMF) over a uniform mesh with stepsize h is defined as follows:

$$\sum_{j=0}^{\mu} \alpha_j y_{n+j} - h \sum_{j=0}^{\mu} \beta_j f_{n+j} = 0 \tag{2}$$

where y_n is the discrete approximation to $y(t_n)$, f_n denotes $f(t_n, y_n)$, and the coefficients satisfies

$$\sum_{j=0}^{\mu} \alpha_j = 1, \quad \sum_{j=0}^{\mu} \beta_j = 1.$$

To get the solution by (2), we need μ initial conditions $y_0, y_1, \cdots, y_{\mu-1}$. Since only y_0 is provided from the original problem, we have to find additional conditions for the remaining values $y_1, y_2, \cdots,, y_{\mu-1}$. The method in (2) with the $(\mu-1)$ additional conditions is called *Initial Value Methods* (IVMs). An IVM is called *implicit* if $\beta_\mu \neq 0$ and *explicit* if $\beta_\mu = 0$. If an IVM is applied to an initial value problem on the interval $[t_0, t_{N+\mu-1}]$, we have the following discrete problem

$$A_N \mathbf{y} = h B_N \mathbf{f} + \begin{pmatrix} \sum_{i=0}^{\mu-1} (\alpha_i y_i - h\beta_i f_i) \\ \vdots \\ \alpha_0 y_{\mu-1} - h\beta_0 f_{\mu-1} \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \tag{3}$$

where $\mathbf{y} = (y_\mu, y_{\mu+1}, \cdots, y_{N+\mu-1})^T$ and $\mathbf{f} = (f_\mu, f_{\mu+1}, \cdots, f_{N+\mu-1})^T$,

$$A_N = \begin{pmatrix} \alpha_\mu & & & & \\ \vdots & \ddots & & & \\ \alpha_0 & & \ddots & & \\ & \ddots & & \ddots & \\ & & \alpha_0 & \cdots & \alpha_\mu \end{pmatrix}_{N\times N}$$

and

$$B_N = \begin{pmatrix} \beta_\mu & & & & \\ \vdots & \ddots & & & \\ \beta_0 & & \ddots & & \\ & \ddots & & \ddots & \\ & & \beta_0 & \cdots & \beta_\mu \end{pmatrix}_{N\times N}.$$

Note that the matrices A_N and B_N are lower triangular Toeplitz matrices. We recall that a matrix is said to be Toeplitz if its entries are constant along its diagonals. Moreover, the linear system (3) can be solved easily by forward recursion. A classical example of IVM is the second order backward differentiation formulae (BDF),

$$3y_{n+2} - 4y_{n+1} + y_n = 2hf_{n+2},$$

which is a two-step method with $\alpha_0 = 1$, $\alpha_1 = -4$, $\alpha_2 = 3$ and $\beta_2 = 2$.

Instead of using μ initial conditions for solving (1) by (2), we can also use the so-called *Boundary Value Methods* (BVMs). Given ν_1, $\nu_2 \geq 0$ such that $\nu_1 + \nu_2 = \mu$, then the corresponding BVM requires ν_1 initial addition conditions $y_0, y_1, \cdots, y_{\nu_1-1}$ and ν_2 final addition conditions $y_N, y_{N+1}, \cdots, y_{N+\nu_2-1}$, which are called (ν_1, ν_2)-boundary conditions. Note that the class of BVMs contains the class of IVMs (i.e. $\nu_1 = \mu$, $\nu_2 = 0$).

The discrete problem generated by a μ-step BVM with (ν_1, ν_2)-boundary conditions can be written in the following matrix form

$$A\mathbf{y} = hB\mathbf{f} + \begin{pmatrix} \sum_{i=0}^{\nu_1-1}(\alpha_i y_i - h\beta_i f_i) \\ \vdots \\ \alpha_0 y_{\nu_1-1} - h\beta_0 f_{\nu_1-1} \\ 0 \\ \vdots \\ 0 \\ \alpha_\mu y_N - h\beta_\mu f_N \\ \vdots \\ \sum_{i=1}^{\nu_2}(\alpha_{\nu_1+i} y_{N-1+i} - h\beta_{\nu_1+i} f_{N-1+i}) \end{pmatrix},$$

where $\mathbf{y} = (y_{\nu_1}, y_{\nu_1+1}, \cdots, y_{N-1})^T$, $\mathbf{f} = (f_{\nu_1}, f_{\nu_1+1}, \cdots, f_{N-1})^T$, A and $B \in \mathbb{R}^{(N-\nu_1)\times(N-\nu_1)}$ are defined as follows,

$$A = \begin{pmatrix} \alpha_{\nu_1} & \cdots & \alpha_\mu & & \\ \vdots & \ddots & \ddots & \ddots & \\ \alpha_0 & \ddots & \ddots & \ddots & \alpha_\mu \\ & \ddots & & \ddots & \vdots \\ & & \alpha_0 & \cdots & \alpha_{\nu_1} \end{pmatrix}, \quad B = \begin{pmatrix} \beta_{\nu_1} & \cdots & \beta_\mu & & \\ \vdots & \ddots & \ddots & \ddots & \\ \beta_0 & \ddots & \ddots & \ddots & \beta_\mu \\ & \ddots & & \ddots & \vdots \\ & & \beta_0 & \cdots & \beta_{\nu_1} \end{pmatrix}.$$

Note that the coefficient matrices are Toeplitz with lower bandwidth ν_1 and upper bandwidth ν_2. An example of BVMs is the third order generalized backward differentiation formulae (GBDF),

$$2y_{n+1} + 3y_n - 6y_{n-1} + y_{n-2} = 6hf_n,$$

which is a three-step method with (2, 1)-boundary conditions where $\alpha_0 = 1$, $\alpha_1 = -6$, $\alpha_2 = 3$, $\alpha_4 = 2$ and $\beta_2 = 6$.

Although IVMs are more efficient than BVMs (which cannot be solved by forward recursion), the advantage in using BVMs over IVMs comes from their stability properties. For example, the usual BDF are not A-stable for $\mu > 2$ but the GBDF are $A_{\nu,\mu-\nu}$-stable for any $\mu \geq 1$, see for instances [1, 4] and [5, p. 79 and Figures 5.1–5.3].

2.2 Preconditioned GMRES Method

The generalized minimal residual (GMRES) method was proposed in 1986 as a Krylov subspace method for solving nonsymmetric linear systems $A\mathbf{x} = \mathbf{b}$. Unlike the normalized conjugate gradient method, the GMRES method does not require computation of the action of A^T on a vector. The k-th iteration of the GMRES method is the solution to the least squares problem

$$\min_{\mathbf{x}\in\mathbf{x}_0+\mathcal{K}_k} \|\mathbf{b} - A\mathbf{x}\|_2$$

where $\mathbf{x}_0$ is the initial iterate, $\mathcal{K}_k = \text{span}\{\mathbf{r}_0, A\mathbf{r}_0, \cdots, A^{k-1}\mathbf{r}_0\}$ is the k-th Krylov subspace and $\mathbf{r}_0 = \mathbf{b} - A\mathbf{x}_0$. If $A = I + L$, where I is the identity matrix, then the GMRES method will converge in at most $\text{rank}(L) + 1$ iterations in exact arithmetic, see [12, 23] for details.

Also, it is well-known that for any circulant matrix C_n, it can be diagonalized by the discrete Fourier matrix F_n, i.e.,

$$C_n = F_n^* \Lambda_n F_n, \tag{4}$$

where the entries of F_n are given by

$$(F_n)_{j,k} = \frac{1}{\sqrt{n}} e^{2\pi \mathbf{i} jk/n}, \quad 0 \le j, k \le n-1, \tag{5}$$

and Λ_n is a diagonal matrix holding the eigenvalues of C_n. We note that Λ_n can be obtained in $O(n \log n)$ operations by taking the fast Fourier transform (FFT) of the first column of C_n. Once Λ_n is obtained, the products $C_n\mathbf{y}$ and $C_n^{-1}\mathbf{y}$ for any vector $\mathbf{y}$ can be computed by FFTs in $O(n \log n)$ operations. Thus, if we use a circulant matrix to precondition the Toeplitz system, then in each GMRES iteration, we need to solve the preconditioned system

$$C_n^{-1} A\mathbf{x} = C_n^{-1}\mathbf{b},$$

which can be done in $O(n \log n)$ operations by using Strang's embedding algorithm with FFTs, see [9]. For more details about the circulant preconditioners for Toeplitz systems, we refer to [9, 12].

3 Strang-Type Preconditioners with BVMs

In this section, we construct the Strang-type block-circulant preconditioner for solving the systems discretized by BVMs. The main advantage of the Strang-type preconditioners is that the preconditioned systems are invertible and the operation cost for each iteration is smaller than that of the direct solvers.

3.1 Block-BVMs and Their Matrix Forms

By using the LMF stated in (2), the μ-step block-BVM over a uniform mesh h for (1) is defined as follows:

$$\sum_{i=-\nu}^{\mu-\nu} \alpha_{i+\nu}\mathbf{y}_{s+i} = h\sum_{i=-\nu}^{\mu-\nu} \beta_{i+\nu}\mathbf{f}_{s+i}, \quad n = \nu, \ldots, s-\mu+\nu. \tag{6}$$

Here, $\mathbf{y}_n$ is the discrete approximation to $\mathbf{y}(t_n)$, $\mathbf{f}_n = J_m\mathbf{y}_n + \mathbf{g}_n$ and $\mathbf{g}_n = \mathbf{g}(t_n)$. Also, (6) requires ν initial conditions and $\mu-\nu$ final conditions which are provided by the following $(\mu-1)$ additional equations:

$$\sum_{i=0}^{\mu} \alpha_i^{(j)}\mathbf{y}_i = h\sum_{i=0}^{\mu} \beta_i^{(j)}\mathbf{f}_i, \quad j = 1, \ldots, \nu-1, \tag{7}$$

and

$$\sum_{i=0}^{\mu} \alpha_{\mu-i}^{(j)}\mathbf{y}_{s-i} = h\sum_{i=0}^{\mu} \beta_{\mu-i}^{(j)}\mathbf{f}_{s-i}, \quad j = s-\mu+\nu+1, \ldots, s. \tag{8}$$

The coefficients $\alpha^{(j)}$ and $\beta^{(j)}$ in (7) and (8) should be chosen such that the truncation errors for these initial and final conditions are of the same order as that in (6). By combining (6), (7) and (8), the discrete system of (1) is given by the following block form

$$M\mathbf{y} \equiv (A\otimes I_m - hB\otimes J_m)\mathbf{y} = \mathbf{e}_1\otimes\mathbf{z} + h(B\otimes I_m)\mathbf{g}. \tag{9}$$

Here $\mathbf{e}_1 = (1, 0, \cdots, 0)^T \in \mathbb{R}^{(s+1)}$, $\mathbf{y} = (\mathbf{y}_0, \cdots, \mathbf{y}_s)^T \in \mathbb{R}^{(s+1)m}$, $\mathbf{g} = (\mathbf{g}_0, \cdots, \mathbf{g}_s)^T \in$

$\mathbb{R}^{(s+1)m}$, and A and B are $(s+1)$-by-$(s+1)$ matrices given by:

$$
A = \begin{pmatrix}
1 & \cdots & 0 & & & & \\
\alpha_0^{(1)} & \cdots & \alpha_\mu^{(1)} & & & & \\
\vdots & \vdots & \vdots & & & 0 & \\
\alpha_0^{(\nu-1)} & \cdots & \alpha_\mu^{(\nu-1)} & & & & \\
\alpha_0 & \cdots & \alpha_\mu & & & & \\
 & \alpha_0 & \cdots & \alpha_\mu & & & \\
 & & \ddots & \ddots & \ddots & & \\
 & & & \ddots & \ddots & \ddots & \\
 & & & & \alpha_0 & \cdots & \alpha_\mu \\
 & & 0 & & \alpha_0^{(s-\mu+\nu+1)} & \cdots & \alpha_\mu^{(s-\mu+\nu+1)} \\
 & & & & \vdots & \vdots & \vdots \\
 & & & & \alpha_0^{(s)} & \cdots & \alpha_\mu^{(s)}
\end{pmatrix}
$$

and

$$
B = \begin{pmatrix}
0 & \cdots & 0 & & & & \\
\beta_0^{(1)} & \cdots & \beta_\mu^{(1)} & & & & \\
\vdots & \vdots & \vdots & & & & \\
\beta_0^{(\nu-1)} & \cdots & \beta_\mu^{(\nu-1)} & & & 0 & \\
\beta_0 & \cdots & \beta_\mu & & & & \\
 & \beta_0 & \cdots & \beta_\mu & & & \\
 & & \ddots & \ddots & \ddots & & \\
 & & & \ddots & \ddots & \ddots & \\
 & & & & \beta_0 & \cdots & \beta_\mu \\
 & & 0 & & \beta_0^{(s-\mu+\nu+1)} & \cdots & \beta_\mu^{(s-\mu+\nu+1)} \\
 & & & & \vdots & \vdots & \vdots \\
 & & & & \beta_0^{(s)} & \cdots & \beta_\mu^{(s)}
\end{pmatrix}.
$$

3.2 Construction of the Strang-type Preconditioner

In [10], we proposed the following preconditioner for (9):

$$S = s(A) \otimes I_m - hs(B) \otimes J_m, \tag{10}$$

where

$$s(A) = \begin{pmatrix} \alpha_\nu & \cdots & \alpha_\mu & & & \alpha_0 & \cdots & \alpha_{\nu-1} \\ \vdots & \ddots & & \ddots & & & \ddots & \vdots \\ \alpha_0 & & \ddots & & \ddots & & & \alpha_0 \\ & \ddots & & \ddots & & \ddots & 0 & \\ & & \ddots & & \ddots & & \ddots & \\ & 0 & & \ddots & & \ddots & & \ddots \\ \alpha_\mu & & & & \ddots & & \ddots & \alpha_\mu \\ \vdots & \ddots & & & & \ddots & \ddots & \vdots \\ \alpha_{\nu+1} & \cdots & \alpha_\mu & & & \alpha_0 & \cdots & \alpha_\nu \end{pmatrix}$$

and $s(B)$ is defined similarly by using $\{\beta_i\}_{i=0}^{\mu}$ instead of $\{\alpha_i\}_{i=0}^{\mu}$ in $s(A)$. The $\{\alpha_i\}_{i=0}^{\mu}$ and $\{\beta_i\}_{i=0}^{\mu}$ here are the coefficients in (6). We note that $s(A)$ and $s(B)$ are the generalized Strang-type circulant preconditioners of A and B respectively, see [9].

We will show that the preconditioner S is invertible provided that the given BVM is stable and the eigenvalues of J_m are in the negative half of the complex plane $\mathbb{C}$. The stability of a BVM is closely related to two characteristic polynomials of degree μ, defined as follows:

$$\rho(z) \equiv z^\nu \sum_{j=-\nu}^{\mu-\nu} \alpha_{j+\nu} z^j \quad \text{and} \quad \sigma(z) \equiv z^\nu \sum_{j=-\nu}^{\mu-\nu} \beta_{j+\nu} z^j. \tag{11}$$

Definition 1 *[5, p.101] Consider a BVM with the characteristic polynomials $\rho(z)$ and $\sigma(z)$ given by (11). The region*

$$\mathcal{D}_{\nu,\mu-\nu} = \{q \in \mathbb{C} : \rho(z) - q\sigma(z) \text{ has } \nu \text{ zeros inside } |z| = 1$$
$$\text{and } \mu - \nu \text{ zeros outside } |z| = 1\}$$

is called the region of $A_{\nu,\mu-\nu}$-stability *of the given BVM. Moreover, the BVM is said to be* $A_{\nu,\mu-\nu}$-stable *if*

$$\mathbb{C}^- \equiv \{q \in \mathbb{C} : Re(q) < 0\} \subseteq \mathcal{D}_{\nu,\mu-\nu}.$$

Theorem 2 *If the BVM for (2) is $A_{\nu,\mu-\nu}$-stable and $h\lambda_k(J_m) \in \mathcal{D}_{\nu,\mu-\nu}$ where $\lambda_k(J_m)$ ($k = 1, \cdots, m$) are the eigenvalues of J_m, then the preconditioner S in (10) is invertible.*

Proof. Since $s(A)$ and $s(B)$ are circulant matrices, their eigenvalues are given by

$$g_A(z) \equiv \alpha_\mu z^{\mu-\nu} + \cdots + \alpha_\nu + \alpha_{\nu-1}\frac{1}{z} + \cdots + \alpha_0\frac{1}{z^\nu} = \frac{\rho(z)}{z^\nu}$$

and

$$g_B(z) \equiv \beta_\mu z^{\mu-\nu} + \cdots + \beta_\nu + \beta_{\nu-1}\frac{1}{z} + \cdots + \beta_0\frac{1}{z^\nu} = \frac{\sigma(z)}{z^\nu},$$

evaluated at $\omega_j = e^{2\pi \mathbf{i} j/(s+1)}$ for $j = 0, \ldots, s$, see [9]. The eigenvalues $\lambda_{jk}(S)$ of S are therefore given by

$$\lambda_{jk}(S) = g_A(\omega_j) - h\lambda_k(J_m) g_B(\omega_j), \quad j = 0, \cdots, s, \quad k = 1, \cdots, m.$$

Since the BVM is $A_{\nu,\mu-\nu}$-stable, if $h\lambda_k(J_m) \in \mathcal{D}_{\nu,\mu-\nu}$, the μ-degree polynomial $\rho(z) - h\lambda_k(J_m)\sigma(z)$ will have no roots on the unit circle $|z| = 1$. Thus for all $k = 1, \cdots, m$,

$$g_A(z) - h\lambda_k(J_m) g_B(z) = \frac{1}{z^\nu}\{\rho(z) - h\lambda_k(J_m)\sigma(z)\} \neq 0, \quad \forall |z| = 1.$$

It follows that $\lambda_{jk}(S) \neq 0$ for all $j = 0, \cdots, s$, and $k = 1, \cdots, m$. Thus S is invertible. ■

In particular, we have

Corollary 3 *If the BVM is $A_{\nu,\mu-\nu}$-stable and $\lambda_k(J_m) \in \mathbb{C}^-$, then the preconditioner S is invertible.*

3.3 Convergence Rate and Operation Cost

As we have stated in §2.2, we have the following theorem for the convergence rate.

Theorem 4 *We have*

$$S^{-1}M = I + L$$

where I is the identity matrix and the rank of L is at most $2m\mu$. Therefore, when the GMRES method is applied to solving $S^{-1}M\mathbf{y} = S^{-1}\mathbf{b}$, the method will converge in at most $2m\mu + 1$ iterations in exact arithmetic.

Proof. Let $E = M - S$, we have by (9) and (10),

$$E = (A - s(A)) \otimes I_m - h(B - s(B)) \otimes J_m = L_A \otimes I_m - hL_B \otimes J_m.$$

It is easy to check that L_A and L_B are $(s+1)$-by-$(s+1)$ matrices with nonzero entries only in the following four corners: a ν-by-$(\mu+1)$ block in the upper left; a ν-by-ν block in the upper right; a $(\mu-\nu)$-by-$(\mu+1)$ block in the lower right; and a $(\mu-\nu)$-by-$(\mu-\nu)$ block in the lower left.

Since $\mu > \nu$, rank $(L_A) \leq \mu$ and rank $(L_B) \leq \mu$. Thus, we have

$$\text{rank}(L_A \otimes I_m) = \text{rank}(L_A) \cdot m \leq m\mu$$

and

$$\text{rank}(L_B \otimes J_m) = \text{rank}(L_B) \cdot m \leq m\mu.$$

Therefore

$$S^{-1}M = I_{m(s+1)} + S^{-1}E = I_{m(s+1)} + L,$$

where the rank of L is at most $2m\mu$. ■

Regarding the cost per iteration, the main work in each iteration for the GMRES method is the matrix-vector multiplication

$$S^{-1}M\mathbf{z} = (s(A) \otimes I_m - hs(B) \otimes J_m)^{-1}(A \otimes I_m - hB \otimes J_m)\mathbf{z} \tag{12}$$

see for instance Saad [23]. Since A and B are band matrices and J_m is assumed to be sparse, the matrix-vector multiplication $M\mathbf{z} = (A \otimes I_m - hB \otimes J_m)\mathbf{z}$ can be done very fast.

To compute $S^{-1}(M\mathbf{z})$, since $s(A)$ and $s(B)$ are circulant matrices, we have the following decompositions by (4),

$$s(A) = F\Lambda_A F^* \quad \text{and} \quad s(B) = F\Lambda_B F^*$$

where Λ_A and Λ_B are diagonal matrices containing the eigenvalues of $s(A)$ and $s(B)$ respectively and F is the Fourier matrix defined in (5). It follows that

$$S^{-1}(M\mathbf{z}) = (F^* \otimes I_m)(\Lambda_A \otimes I_m - h\Lambda_B \otimes J_m)^{-1}(F \otimes I_m)(M\mathbf{z}).$$

This product can be obtained by using FFTs and solving s linear systems of order m. Since J_m is sparse, the matrix

$$\Lambda_A \otimes I_m - h\Lambda_B \otimes J_m$$

will also be sparse. Thus $S^{-1}(M\mathbf{z})$ can be obtained by solving s sparse linear systems of order m. It follows that the total number of operations per iteration is $\gamma_1 ms\log s + \gamma_2 smn$, where n is the number of nonzeros of J_m, and γ_1 and γ_2 are some positive constants. For comparing the computational cost of the method with direct solvers for the linear system (9), we refer to [10].

3.4 Numerical Result

Now we give an example to illustrate the efficiency of the preconditioner by solving the test problems given in [2]. The experiments were performed in MATLAB. We used the MATLAB-provided M-file "gmres" (see MATLAB on-line documentation) to solve the preconditioned systems. In our tests, the zero vector is the initial guess and the stopping criterion is $||\mathbf{r}_q||_2/||\mathbf{r}_0||_2 < 10^{-6}$, where $\mathbf{r}_q$ is the residual after q iterations. In the example, the BVM we used is the third order generalized Adam's method (GAM) which has $\mu = 2$. Its formulae and the additional initial and final conditions can be found in [5, p.153].

Example 1. Heat equation:

$$\begin{cases} \dfrac{\partial u}{\partial t} = \dfrac{\partial^2 u}{\partial x^2}, & \\ u(0,t) = \dfrac{\partial u}{\partial x}(\pi, t) = 0, & t \in [0, 2\pi], \\ u(x,0) = x, & x \in [0, \pi]. \end{cases}$$

We discretize the partial differential operator $\partial^2/\partial x^2$ with central differences and step size equals to $\pi/(m+1)$. The system of ODEs obtained is:

$$\begin{cases} \mathbf{y}'(t) = J_m \mathbf{y}(t), & t \in [0, 2\pi] \\ \mathbf{y}(0) = (x_1, x_2, \cdots, x_m)^T, & \end{cases}$$

where J_m is a scaled discrete Laplacian matrix

$$J_m = \frac{(m+1)^2}{\pi^2} \begin{pmatrix} -2 & 1 & & & \\ 1 & \ddots & \ddots & & \\ & \ddots & \ddots & \ddots & \\ & & 1 & -2 & 1 \\ & & & 1 & -1 \end{pmatrix}. \tag{13}$$

Table 1 lists the numbers of iterations required for convergence of the GMRES method for different m and s. In the table, I means no preconditioner is used and S denotes the Strang-type block-circulant preconditioner which is defined in (10). We see that the number of iterations required for convergence, when a circulant preconditioner is used, is always less than that when no preconditioner is used. As expected from Theorem 2, the numbers under column S stay constant for increasing s and m.

4 Strang-Type BCCB Preconditioner

The Strang-type preconditioner proposed in §3 is a block-circulant preconditioner. In this section, Strang-type block-circulant preconditioner with circulant blocks (BCCB preconditioner) is proposed for solving (1) when the stiffness matrix J_m is a Toeplitz matrix in the Wiener class. The main advantage of the use of BCCB preconditioner is that the operation cost for each GMRES iteration can be reduced.

m	s	I	S	m	s	I	S
24	6	19	4	48	6	47	4
	12	70	4		12	167	4
	24	152	4		24	359	4
	48	227	3		48	>400	3
	96	314	3		96	>400	3

Table 1: Number of iterations for convergence.

4.1 Construction of BCCB Preconditioners

Instead of using the block-circulant preconditioner, the Strang-type BCCB preconditioner can also be constructed for solving (9):

$$S^{(2)} \equiv s(A) \otimes I_m - hs(B) \otimes s(J_m) \tag{14}$$

for J_m being a full Toeplitz matrix. The advantage of BCCB preconditioners is that the operation cost in each iteration of Krylov subspace methods for the preconditioned system is much less than that required by using any block-circulant preconditioners.

Similar to Theorem 1 in §3, we can show that if the BVM for (1) is $A_{\nu,\mu-\nu}$-stable and the eigenvalues of $s(J_m)$ satisfy

$$\lambda_k(s(J_m)) \in \mathbb{C}^-$$

for $k = 1, \cdots, m$, then the preconditioner $S^{(2)}$ is invertible.

For some linear evolutionary partial differential equations, the matrix J_m is usually Toeplitz and $s(J_m)$ is singular. Note that the eigenvalues of $S^{(2)}$ are given by

$$\lambda_{jk}(S^{(2)}) = \phi_j - h\psi_j \lambda_k(s(J_m)), \quad j = 0, \cdots, s, \quad k = 1, \cdots, m, \tag{15}$$

where ϕ_j and ψ_j are eigenvalues of $s(A)$ and $s(B)$ respectively. When some eigenvalues of $s(J_m)$ are zero, then some eigenvalues of $S^{(2)}$ is the same as the eigenvalues of the matrix $s(A)$. It is well-known that the eigenvalues of the circulant matrix $s(A)$ can be expressed as the following sum, see [11],

$$\phi_j = \sum_{r=-\nu}^{\mu-\nu} \alpha_{r+\nu}\omega^{rj}, \quad \omega = e^{2\pi \mathbf{i}/(s+1)}, \quad j = 0, \cdots, s,$$

where $\alpha_{r+\nu}$ are given by (6).

From the characteristic polynomials defined in (11), the coefficients must satisfy the consistent conditions,

$$\rho(1) = 0 \quad \text{and} \quad \rho'(1) = \sigma(1).$$

Thus, we have

$$\phi_0 = \rho(1) = 0$$

for any consistent BVM. From (15), we know that $S^{(2)}$ is singular when some eigenvalues of $s(J_m)$ are zero. In this case, we move the zero eigenvalue of $s(A)$ to a nonzero value. More precisely, we change the matrix $s(A) = F\mathrm{diag}(\phi_0, \cdots, \phi_s)\ F^*$ to

$$\widetilde{s}(A) \equiv F\mathrm{diag}(\widetilde{\phi}_0, \cdots, \phi_s)F^*,$$

where $\widetilde{\phi}_0 \equiv \mathrm{Re}(\phi_s)$ and F is the Fourier matrix. Define

$$\widetilde{S}^{(2)} \equiv \widetilde{s}(A) \otimes I_m - hs(B) \otimes s(J_m), \tag{16}$$

we can also prove that $\widetilde{S}^{(2)}$ is invertible, see [16] for a detail.

4.2 Convergence Rate and Operation Cost

Let

$$E \equiv M - \widetilde{S}^{(2)}, \qquad E_1 \equiv M - S, \qquad E_2 \equiv S - \widetilde{S}^{(2)}$$

where S is defined by (10). Then $E = E_1 + E_2$. From Theorem 2 in §3, we know that

$$\mathrm{rank}(E_1) \leq 2m\mu = O(m)$$

where μ is given by the BVM used for (1). For the matrix E_2, by (10) and (16), we have

$$\begin{aligned} E_2 &= (s(A) - \widetilde{s}(A)) \otimes I_m - hs(B) \otimes (J_m - s(J_m)) \\ &= L_A \otimes I_m - hs(B) \otimes L_J \end{aligned}$$

where $L_A \equiv s(A) - \widetilde{s}(A)$ and $L_J \equiv J_m - s(J_m)$. Since

$$L_A = F\mathrm{diag}(\widetilde{\phi}_0, 0, \cdots, 0)F^*$$

is a matrix of rank one, we have

$$\mathrm{rank}(L_A \otimes I_m) \leq 1 \cdot m = m = O(m). \tag{17}$$

For $s(B)\otimes L_J$ in E_2, let J_m be a Toeplitz matrix in the Wiener class (i.e., $\sum_{k=-\infty}^{\infty} |a_k| < \infty$ where a_k is the k-th diagonals of J_m as $m \to \infty$). The matrix L_J can be expressed as a sum of a matrix with low rank and a matrix with small norm, see [8, 12]. More precisely, for any given $\epsilon > 0$, there exists a constant $C(\epsilon)$ such that

$$L_J = U + V \quad \text{with} \quad \mathrm{rank}(U) \leq C(\epsilon) \quad \text{and} \quad \|V\|_2 \leq \epsilon, \tag{18}$$

when m is sufficiently large. Then we have

$$s(B) \otimes L_J = s(B) \otimes U + s(B) \otimes V \tag{19}$$

with

$$\mathrm{rank}(s(B) \otimes U) \leq s \cdot C(\epsilon) = O(s). \tag{20}$$

For $\|s(B) \otimes V\|_2$, we note that

$$\|s(B)\|_1 = M_1 < \infty, \quad \|s(B)\|_\infty = M_2 < \infty,$$

where M_1 and M_2 are two constants independent of the size of the matrices. Therefore,

$$\|s(B)\|_2 \leq (\|s(B)\|_1 \|s(B)\|_\infty)^{1/2} = (M_1 M_2)^{1/2} = M_3 < \infty. \tag{21}$$

Furthermore, by (18) and (21), we have

$$\|s(B) \otimes V\|_2 = \|s(B)\|_2 \|V\|_2 \leq \epsilon M_3. \tag{22}$$

By using (17), (19), (20) and (22), we know that for any $\epsilon > 0$, the matrix E_2 can be decomposed as

$$E_2 = L_{O(m)} + hL_{O(s)} + hW \tag{23}$$

with

$$\text{rank}(L_{O(m)}) = O(m), \quad \text{rank}(L_{O(s)}) = O(s), \quad \|W\|_2 \leq \epsilon.$$

Thus, if J_m is a Toeplitz matrix in the Wiener class, then the spectrum of $(\widetilde{S}^{(2)})^{-1}M$ is clustered around $(1, 0) \in \mathbb{C}$. As a consequence, when the GMRES method is applied to solving the preconditioned system

$$(\widetilde{S}^{(2)})^{-1}M\mathbf{y} = (\widetilde{S}^{(2)})^{-1}\mathbf{b},$$

we can expect a fast convergence rate.

For simplicity, we assume that $s + 1 = m$ in the following analysis of the operation cost. Regarding the cost in each iteration of the GMRES method, the main work is the matrix-vector multiplication

$$(\widetilde{S}^{(2)})^{-1}M\mathbf{v} \equiv (\tilde{s}(A) \otimes I_m - hs(B) \otimes s(J_m))^{-1}M\mathbf{v},$$

where $\mathbf{v}$ is a vector. Since $(\widetilde{S}^{(2)})^{-1}$ can be diagonalized by the 2-dimensional Fourier matrix, the matrix-vector multiplication can be obtained within $\mathcal{O}(m^2 \log m)$ operations by using FFTs. For the Strang-type block-circulant preconditioner S stated in §3, in each iteration, there are m Toeplitz systems of order m needed to be solved. Thus, the complexity in each iteration of our method is much lower.

4.3 Numerical Result

We give two examples to compare the Strang-type BCCB preconditioners $S^{(2)}$ and $\widetilde{S}^{(2)}$ with the Strang-type block-circulant preconditioner S. In the examples, the BVM we used is the fifth order GAM which has $\mu = 4$.

Example 2. Consider the wave equation:

$$\begin{cases} u_t - u_x = 0, & \\ u(x,0) = \sin(x), & x \in [0,\pi], \\ u(\pi,t) = 0, & t \in [0,2\pi]. \end{cases}$$

We discretize the partial differential operator $\partial/\partial x$ with the first order forward differences and step size $\Delta x = \pi/m$, $x_i = i\Delta x$. The system of ODEs is obtained as follows:

$$\begin{cases} \mathbf{y}'(t) = J_m \mathbf{y}(t), & t \in [0,2\pi], \\ \mathbf{y}(0) = (\sin(x_1), \sin(x_2), \cdots, \sin(x_m))^T, & \end{cases}$$

where

$$J_m = \frac{1}{\Delta x} \begin{pmatrix} -1 & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & -1 \end{pmatrix}.$$

Example 3. Consider

$$\begin{cases} \mathbf{y}'(t) = J_m \mathbf{y}(t), & t \in [0,1], \\ \mathbf{y}(0) = (1,2,3,\cdots,m)^T, & \end{cases}$$

where

$$J_m = \begin{pmatrix} -6 & 2 & -1 & & & \\ 2 & -6 & 2 & -1 & & \\ -1 & \ddots & \ddots & \ddots & \ddots & \\ & \ddots & \ddots & \ddots & \ddots & -1 \\ & & \ddots & \ddots & \ddots & 2 \\ & & & -1 & 2 & -6 \end{pmatrix}.$$

Table 2 lists the number of iterations required for convergence of the GMRES method with different preconditioners. From the table, we see that the numbers of iterations of $S^{(2)}$ and $\widetilde{S}^{(2)}$ are slightly larger than those of S. But we should emphasize that the operation costs per iteration of $S^{(2)}$ and $\widetilde{S}^{(2)}$ are less than those of S. We remark that for Example 2, $S^{(2)}$ is singular.

m	s	I	S	$\widetilde{S}^{(2)}$
20	16	29	8	14
	32	38	7	13
	64	70	6	13
	128	133	5	13
40	16	68	9	16
	32	53	8	15
	64	66	7	15
	128	120	6	15
80	16	157	10	19
	32	126	8	18
	64	98	7	18
	128	116	6	17

m	s	I	S	$S^{(2)}$	$\widetilde{S}^{(2)}$
20	16	22	5	9	10
	32	38	5	9	9
	64	70	4	9	9
	128	134	4	9	9
40	16	22	5	9	9
	32	37	5	9	9
	64	70	4	9	9
	128	134	4	9	9
80	16	21	5	9	9
	32	37	5	9	9
	64	69	4	9	9
	128	133	4	9	9

Table 2: Number of iterations for convergence in Examples 2 (left) and 3 (right).

5 Preconditioned Waveform Relaxation

Waveform relaxation (WR) is a classical method to solve (1) by splitting the stiffness matrix J_m into $Q - P$. Typical examples of waveform relaxation are the so-called Jacobi WR and Gauss-Seidel WR. In this section, we use the circulant and skew-circulant decomposition for splitting the matrix J_m. We call it the $C + S$ version of WR. We will see that the convergence rate of the $C+S$ version of WR is faster than that of the Jacobi and Gauss-Seidel WR.

5.1 Waveform Relaxation

By splitting the matrix J_m as

$$J_m = Q - P, \tag{24}$$

we can construct an iteration of the form for (1)

$$\begin{cases} \dfrac{\mathrm{d}\mathbf{y}^{(k+1)}(t)}{\mathrm{d}t} + P\mathbf{y}^{(k+1)}(t) = Q\mathbf{y}^{(k)}(t) + \mathbf{g}(t), \quad t \in (t_0, T], \\ \mathbf{y}^{(k+1)}(t_0) = \mathbf{z}, \end{cases} \tag{25}$$

where $k = 0, 1, \cdots$, and $\mathbf{y}^{(0)}$ is a given initial guess usually given by $\mathbf{y}^{(0)}(t) = \mathbf{z}$ for $t \in [t_0, T]$. The iteration (25) is called the waveform relaxation method or dynamic iteration, see [6]. This method originated from electrical network simulation, see [17]. It differs from standard iterative techniques in that it is a continuous-in-time analogue of stationary method by iterating with functions.

The Jacobi and Gauss-Seidel versions of the WR technique are classical methods. To be more precise, the matrix J_m is first decomposed as $J_m = L + D + U$, where D is a diagonal matrix, L is a strictly lower triangular matrix and U is a strictly upper triangular matrix. The splittings

$$P = D, \quad Q = L + U,$$

and

$$P = L + D, \quad Q = U,$$

define, respectively, the Jacobi and Gauss-Seidel WR iterations.

If J_m in (24) is Toeplitz, by using the well-known circulant and skew-circulant decomposition of Toeplitz matrix, we decompose the matrix J_m as $J_m = Q - P$, where P is a circulant matrix and Q is a skew-circulant matrix, see [8]. More precisely, for a Toeplitz matrix

$$T_m = (t_{i-j})_{i,j=1}^{m} = (t_k)_{k=0}^{m-1},$$

we can decompose the matrix T_m as

$$T_m = C_m + S_m \tag{26}$$

where

$$C_m = \begin{pmatrix} c_0 & c_1 & \cdots & c_{m-2} & c_{m-1} \\ c_{m-1} & c_0 & \ddots & \ddots & c_{m-2} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ c_2 & \ddots & \ddots & c_0 & c_1 \\ c_1 & c_2 & \cdots & c_{m-1} & c_0 \end{pmatrix},$$

$$S_m = \begin{pmatrix} s_0 & s_1 & \cdots & s_{m-2} & s_{m-1} \\ -s_{m-1} & s_0 & \ddots & \ddots & s_{m-2} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ -s_2 & \ddots & \ddots & s_0 & s_1 \\ -s_1 & -s_2 & \cdots & -s_{m-1} & s_0 \end{pmatrix}$$

with $c_0 + s_0 = t_0$, $c_k = \frac{1}{2}(t_k + t_{-m+k})$ and $s_k = \frac{1}{2}(t_k - t_{-m+k})$, where $k = 1, 2, \cdots, m-1$. The WR method with this new scheme is called the $C + S$ version.

The convergence behavior of the WR methods has been studied extensively in a series of papers in [19, 21, 22] where the authors formulated the convergence characteristics of the method in terms of the spectral radius of the corresponding waveform relaxation operator. To accelerate the WR iterations, the multigrid technique was studied in [25] and the preconditioning technique was discussed in [6].

5.2 Invertibility of the Strang-type preconditioners

We know that if the BVM for (25) is $A_{\nu,\mu-\nu}$-stable and the eigenvalues of P satisfy

$$\mathrm{Re}(\lambda_k(P)) \in \mathbb{C}^-,$$

for $k = 1, \cdots, m$, then the Strang-type block-circulant preconditioner

$$S = s(A) \otimes I_m + hs(B) \otimes P, \tag{27}$$

is invertible, see Theorem 1 in §3.

In some cases, there is a $\lambda_l(P)$ which does not satisfy the condition in Theorem 1, say, $\mathrm{Re}(\lambda_l(P)) \notin \mathbb{C}^-$. If P is diagonalizable by a unitary matrix, we can "move" $\lambda_l(P)$ into $\mathbb{C}^-$ by subtracting $\lambda_{\max} + \varepsilon$ from the main diagonal of the matrix P, where

$$\lambda_{\max} = \max_l \{\mathrm{Re}(\lambda_l(P)) \notin \mathbb{C}^-\}$$

and ε is a positive real number. After such a modification, a new matrix $\widetilde{P}$ can be written as

$$\widetilde{P} = P - (\lambda_{\max} + \varepsilon) I_m.$$

It yields a new decomposition of the matrix J_m:

$$J_m = \widetilde{Q} - \widetilde{P}$$

where

$$\widetilde{Q} = Q + (\lambda_{\max} + \varepsilon) I_m.$$

Obviously, all the eigenvalues of $\widetilde{P}$ are now in $\mathbb{C}^-$ and therefore Theorem 1 is still applicable.

5.3 Convergence rate and operation cost

Let $M = A \otimes I_m - hB \otimes P$ and $G = -h(B \otimes Q)$. For the convergence of the WR method, we require that $\rho(M^{-1}G) < 1$ where $\rho(\cdot)$ is the spectral radius. For the convergence rate of the GMRES method, as shown in §3, by Theorem 2, we have $S^{-1}M = I + L$ where $\mathrm{rank}(L) \leq 2m\mu$. Thus, the GMRES will converge in at most $2m\mu + 1$ iterations in exact arithmetic. Now, we compare the operation cost with different WR splittings for Toeplitz matrix

$$J_m = (q_{i-j})_{i,j=1}^m = (q_k)_{k=0}^{m-1}.$$

(i) In the Jacobi WR iterations, since M is a block-Toeplitz matrix with Toeplitz blocks (plus a small rank perturbation), by using Strang's embedding algorithm with FFTs, see [9, 12], $M\mathbf{v}$ can be computed within $O(ms \log ms)$ operations. Meanwhile,

$$S = s(A) \otimes I_m + q_0 hs(B) \otimes I_m,$$

therefore, S^{-1} can be calculated within $O(ms \log s)$ operations. Thus, computing $S^{-1}(M\mathbf{v})$ requires $O(ms \log ms)$ operations.

(ii) In the Guass-Seidel WR iterations, we note that

$$\Lambda_A \otimes I_m + h\Lambda_B \otimes P$$

is a block diagonal matrix with lower triangular Toeplitz blocks. Therefore, we have to solve $s+1$ lower triangular Toeplitz systems of size m-by-m. By using the super-fast direct Toeplitz solver, see [9], it requires $O(sm \log^2 m)$ operations to calculate $S^{-1}\mathbf{w}$ for some vector $\mathbf{w}$. As in (i), $M\mathbf{v}$ can also be computed within $O(ms \log ms)$ operations. Therefore, it requires $O(ms \log ms + ms \log^2 m)$ operations to compute $S^{-1}(M\mathbf{v})$.

(iii) In the $C+S$ version of WR iterations, since the matrix P is a circulant matrix, we have

$$S^{-1}(M\mathbf{v}) = (F_{s+1} \otimes F_m)(\Lambda_A \otimes I_m + h\Lambda_B \otimes \Lambda_P)^{-1}(F^*_{s+1} \otimes F^*_m)(M\mathbf{v})$$

By using the FFT, S^{-1} can be calculated within $O(ms \log ms)$ operations. As in (i), $M\mathbf{v}$ can also be computed within $O(ms \log ms)$ operations. Therefore, it requires $O(ms \log ms)$ operations to compute $S^{-1}(M\mathbf{v})$.

Consequently, by Theorem 2, the total complexity of each WR iteration is bounded by $O(m^2 s \log ms)$ operations by using the $C+S$ version and the Jacobi version, while is bounded by $O(m^2 s \log ms + m^2 s \log^2 m)$ operations by using the Gauss-Seidel version.

5.4 Numerical Result

So far, we have introduced our method which combines the WR iterations, the BVM and the GMRES method together with the Strang-type preconditioner for solving (1). From the preceding analysis, we choose diagonal dominant Toeplitz matrices as our stiffness matrices in order to guarantee the convergence of the WR iterations. The BVM we used in the experiments is the fifth order GAM, see [5]. The stopping criterion of the WR iterations is

$$\frac{\|\mathbf{y}^{(k+1)} - \mathbf{y}^{(k)}\|_2}{\|\mathbf{y}^{(k)}\|_2} \leq 10^{-6}$$

where $\mathbf{y}^{(k)}$ is the solution after the k-th WR iteration.

Example 4 Consider

$$\begin{cases} \mathbf{y}'(t) &= J_m \mathbf{y}(t), \quad t \in (0,1]\,, \\ \mathbf{y}(0) &= (1, 2, \cdots, m)^T, \end{cases}$$

where

$$J_m = \begin{pmatrix} -6 & 2 & -1 & & & \\ 2 & -6 & 2 & -1 & & \\ -1 & 2 & -6 & \ddots & \ddots & \\ & -1 & \ddots & \ddots & \ddots & -1 \\ & & \ddots & \ddots & \ddots & 2 \\ & & & -1 & 2 & -6 \end{pmatrix}.$$

Table 3 shows the number of WR iterations and total CPU time (on a 886MHz PC) required for convergence with different combinations of matrix sizes m and s. As expected, the number of iterations required for convergence remains almost constant for increasing m and s.

6 Application to DAEs and DDEs

In electrical engineering, control theory and biomathematics [7, 14], many problems are formulated by the so-called differential algebraic equations (DAEs) and delay differential equations (DDEs). In §6.1, we introduce the DAE solver to solve a system of linear DAEs, which transforms the system of DAEs into two systems of differential equations. We will

m	s	$C+S$	Jacobi	GS
20	16	11	18	11
	32	11	18	11
	64	11	18	11
	128	11	17	11
40	16	11	18	11
	32	11	19	11
	64	11	17	11
	128	10	17	11
60	16	11	18	11
	32	11	17	11
	64	11	17	11
	128	10	17	10

m	s	$C+S$	Jacobi	GS
20	16	2.64	3.63	2.47
	32	3.02	4.61	2.91
	64	4.39	6.54	4.17
	128	7.63	11.04	7.74
40	16	2.97	4.72	2.96
	32	4.12	6.92	4.23
	64	7.52	11.48	7.75
	128	27.24	44.54	30.65
60	16	3.63	5.71	3.74
	32	5.44	8.95	5.66
	64	16.58	25.98	18.90
	128	59.75	106.01	64.49

Table 3: Number of WR iterations (left) and total CPU time (sec) (right) for convergence.

show that one of them is a system of ODEs. Thus, we can apply the method discussed in the previous sections to get the solutions of the discrete system. In §6.2, we introduce the DDE solver to solve a system of linear DDEs.

6.1 Introduction to DAE Solver

Consider a system of linear DAEs

$$\begin{cases} A\dfrac{\mathrm{d}\mathbf{x}(t)}{\mathrm{d}t} + B\mathbf{x}(t) = \mathbf{f}(t), \qquad t \in (t_0, T], \\ \mathbf{x}(t_0) = \mathbf{z}, \end{cases} \tag{28}$$

where A, B are n-by-n matrices and A is singular. A matrix pencil is defined by $\lambda A + B$ with $\lambda \in \mathbb{C}$. A pencil is said to be regular if $\det(\lambda A + B)$ is not identically zero. When $\lambda A + B$ is regular, then (28) is solvable and there exists two invertible matrices P and Q

such that

$$PAQ = \begin{pmatrix} I & 0 \\ 0 & N \end{pmatrix}_{n\times n}, \quad PBQ = \begin{pmatrix} G & 0 \\ 0 & I \end{pmatrix}_{n\times n}.$$

Here the sum of the matrix sizes of N and G is n and N is a nilpotent matrix, i.e., there exists a positive integer ν such that $N^\nu = 0$ and $N^{\nu-1} \neq 0$, see [3]. To compute the matrix P and Q, we can follow a constructive approach given in [24].

Applying the coordinate changes P and Q to the DAEs in (28), we have

$$\begin{cases} \mathbf{y}_1' + G\mathbf{y}_1 = \mathbf{g}_1(t), \\ N\mathbf{y}_2' + \mathbf{y}_2 = \mathbf{g}_2(t), \end{cases} \tag{29}$$

where $Q^{-1}\mathbf{x} = (\mathbf{y}_1^T, \mathbf{y}_2^T)^T$ and $P\mathbf{f} = (\mathbf{g}_1^T, \mathbf{g}_2^T)^T$. The first equation in (29) is a system of ODEs and a solution exists for any initial value of $\mathbf{y}_1$. The second equation has only one solution

$$\mathbf{y}_2(t) = \sum_{i=0}^{\nu-1} (-1)^i N^i \mathbf{g}_2^{(i)}(t)$$

where $\mathbf{g}_2^{(i)}(t)$ denotes the i-th order derivative of $\mathbf{g}_2(t)$ with respect to t. Thus, we can apply the Strang-type preconditioner with BVM discussed in the previous sections to solve the first equation in (29) with a given initial condition. For readers interested in the numerical implementation of our algorithm, we refer to [15].

6.2 Introduction to DDE Solver

We consider the solution of differential equation with multi-delays:

$$\begin{cases} \dfrac{\mathrm{d}\mathbf{y}(t)}{\mathrm{d}t} = J_n\mathbf{y}(t) + D_n^{(1)}\mathbf{y}(t-\tau_1) + \cdots + D_n^{(s)}\mathbf{y}(t-\tau_s) + \mathbf{f}(t), & t \geq t_0, \\ \mathbf{y}(t) = \boldsymbol{\phi}(t), & t \leq t_0, \end{cases} \tag{30}$$

where $\mathbf{y}(t)$, $\mathbf{f}(t)$, $\boldsymbol{\phi}(t) : \mathbb{R} \to \mathbb{R}^n$; J_n, $D_n^{(1)}, \cdots, D_n^{(s)} \in \mathbb{R}^{n\times n}$ and $\tau_1, \cdots, \tau_s > 0$ are some rational numbers.

For (30), in order to find a reasonable numerical solution, we require that the solution of (30) is asymptotically stable. We have the following lemma, see [20, 27].

Lemma 5 *For any $s \geq 1$, if $\eta(J_n) \equiv \frac{1}{2}\lambda_{\max}(J_n + J_n^T) < 0$ and*

$$\eta(J_n) + \sum_{j=1}^{s} ||D_n^{(j)}||_2 < 0, \tag{31}$$

then solution of (30) is asymptotically stable.

In the following, for simplicity, we only consider the case of $s = 2$ in (30). The generalization to arbitrary s is straightforward. Let

$$h = \tau_1/m_1 = \tau_2/m_2$$

be the step size where m_1 and m_2 are positive integers with $m_2 > m_1$ ($\tau_2 > \tau_1$). By using a BVM with (ν_1, ν_2)-boundary conditions, we have

$$\sum_{i=0}^{\mu} \alpha_i \mathbf{y}_{p+i-\nu_1} = h \sum_{i=0}^{\mu} \beta_i (J_n \mathbf{y}_{p+i-\nu_1} + D_n^{(1)} \mathbf{y}_{p+i-\nu_1-m_1} + D_n^{(2)} \mathbf{y}_{p+i-\nu_1-m_2} + \mathbf{f}_{p+i-\nu_1}), \tag{32}$$

for $p = \nu_1, \ldots, N-1$, where $\mu = \nu_1 + \nu_2$, and $\{\alpha_i\}_{i=0}^{\mu}$, $\{\beta_i\}_{i=0}^{\mu}$ are coefficients of the given BVM. By providing the values

$$y_{-m_2}, \ldots, y_{-m_1}, \ldots, y_0, \quad y_1, \ldots, y_{\nu_1-1}, \quad y_N, \ldots, y_{N+\nu_2-1}, \tag{33}$$

the equation (32) can be written in a matrix form as

$$M\mathbf{y} = \mathbf{b} \tag{34}$$

where

$$M = A \otimes I_n - hB \otimes J_n - hC^{(1)} \otimes D_n^{(1)} - hC^{(2)} \otimes D_n^{(2)}, \tag{35}$$

$$\mathbf{y}^T = (\mathbf{y}_{\nu_1}^T, \mathbf{y}_{\nu_1+1}^T, \ldots, \mathbf{y}_{N-1}^T) \in \mathbb{R}^{n(N-\nu_1)},$$

and $\mathbf{b} \in \mathbb{R}^{n(N-\nu_1)}$ depends on $\mathbf{f}$, the boundary values, and the coefficients of the method. The matrices A, B, $C^{(1)}$, $C^{(2)} \in \mathbb{R}^{(N-\nu_1)\times(N-\nu_1)}$ in (35) are defined as follows,

$$A = \begin{pmatrix} \alpha_{\nu_1} & \cdots & \alpha_\mu & & \\ \vdots & \ddots & \ddots & \ddots & \\ \alpha_0 & \ddots & \ddots & \ddots & \alpha_\mu \\ & \ddots & & \ddots & \vdots \\ & & \alpha_0 & \cdots & \alpha_{\nu_1} \end{pmatrix}, \quad B = \begin{pmatrix} \beta_{\nu_1} & \cdots & \beta_\mu & & \\ \vdots & \ddots & \ddots & \ddots & \\ \beta_0 & \ddots & \ddots & \ddots & \beta_\mu \\ & \ddots & & \ddots & \vdots \\ & & \beta_0 & \cdots & \beta_{\nu_1} \end{pmatrix},$$

$$C^{(1)} = \begin{pmatrix} \mathbf{0} & & & & & \\ \beta_\mu & \ddots & & & & \\ \vdots & \ddots & \ddots & & & \\ \beta_0 & \cdots & \beta_\mu & \ddots & & \\ & \ddots & & \ddots & \ddots & \\ & & \beta_0 & \cdots & \beta_\mu & \mathbf{0} \end{pmatrix}, \qquad C^{(2)} = \begin{pmatrix} \mathbf{0} & & & & & \\ \beta_\mu & \ddots & & & & \\ \vdots & \ddots & \ddots & & & \\ \beta_0 & \cdots & \beta_\mu & \ddots & & \\ & \ddots & & \ddots & \ddots & \\ & & \beta_0 & \cdots & \beta_\mu & \mathbf{0} \end{pmatrix},$$

see [5]. We remark that the first column of $C^{(1)}$ is given by

$$(\underbrace{0,\ldots,0}_{m_1+\nu_1-\mu}, \beta_\mu, \ldots, \beta_0, \underbrace{0,\ldots,0}_{N-m_1-2\nu_1-1})^T$$

and the first column of $C^{(2)}$ is given by

$$(\underbrace{0,\ldots,0}_{m_2+\nu_1-\mu}, \beta_\mu, \ldots, \beta_0, \underbrace{0,\ldots,0}_{N-m_2-2\nu_1-1})^T.$$

The Strang-type block-circulant preconditioner for (35) is defined as follows:

$$S = s(A) \otimes I_n - hs(B) \otimes J_n - hs(C^{(1)}) \otimes D_n^{(1)} - hs(C^{(2)}) \otimes D_n^{(2)} \tag{36}$$

where $s(E)$ is Strang's circulant preconditioner of matrix E, for $E = A$, B, $C^{(1)}$ and $C^{(2)}$. We have the following theorem for the invertability of our preconditioner. The proof of the theorem is similar to that of Theorem 1.

Theorem 6 *[13] If the BVM for (30) is A_{ν_1,ν_2}-stable and (31) holds, the Strang-type block-circulant preconditioner S defined in (36) is invertible.*

For the convergence rate, we have

Theorem 7 *[13] When Krylov subspace methods are applied to solving the preconditioned system*

$$S^{-1}M\mathbf{y} = S^{-1}\mathbf{b},$$

the methods will converge in at most $(2\mu + m_1 + m_2 + 2\nu_1 + 2)n = O(n)$ iterations in exact arithmetic.

For the operation cost of our algorithm, we refer to [13, 18].

6.3 Numerical Result

We illustrate the efficiency of our preconditioner by solving the following problem. In the example, the BVM we used is the third order GBDF for $t \in [0, 4]$.

Example 5. Consider

$$\begin{cases} \mathbf{y}'(t) = J_n \mathbf{y}(t) + D_n^{(1)} \mathbf{y}(t - 0.5) + D_n^{(2)} \mathbf{y}(t-1), & t \geq 0, \\ \mathbf{y}(t) = (\sin t, 1, \ldots, 1)^T, & t \leq 0, \end{cases}$$

where

$$J_n = \begin{pmatrix} -10 & 2 & & & \\ 2 & \ddots & \ddots & & \\ 1 & \ddots & \ddots & \ddots & \\ & \ddots & \ddots & \ddots & 2 \\ & & 1 & 2 & -10 \end{pmatrix}, \qquad D_n^{(1)} = \frac{1}{n} \begin{pmatrix} 2 & -1 & & \\ -1 & \ddots & \ddots & \\ & \ddots & \ddots & -1 \\ & & -1 & 2 \end{pmatrix},$$

and

$$D_n^{(2)} = \frac{1}{n} \begin{pmatrix} 2 & 1 & & \\ 1 & \ddots & \ddots & \\ & \ddots & \ddots & 1 \\ & & 1 & 2 \end{pmatrix}.$$

Table 4 shows the number of iterations required for convergence of the GMRES method with different combinations of matrix sizes n and s. We see that the numbers of iterations required for convergence increase slowly for increasing n and s under the column S.

Acknowledgment. The research was partially supported by the Hong Kong Research Grant Council grant CUHK4243/01P and CUHK DAG 2060220 (R. H. Chan), and by the research grant RG024/01-02S/JXQ/FST from the University of Macau (X. Q. Jin).

References

1. P. Amodio, F. Mazzia and D. Trigiante, *Stability of Some Boundary Value Methods for the Solution of Initial Value Problems*, BIT, **1993**, *33*, 3.

n	s	I	S	n	s	I	S
24	10	52	9	48	10	53	12
	20	97	11		20	98	14
	40	185	15		40	189	14
	80	367	19		80	378	17

Table 4: Number of iterations for convergence.

2. D. Bertaccini, *A Circulant Preconditioner for the Systems of LMF-Based ODE Codes*, SIAM J. Sci. Comput., **2000**, *22* , 3.

3. K. Brenan, S. Campbell and L. Petzold, *Numerical Solution of Initial-Value Problems in Differential-Algebraic Equations*, SIAM Press, Philadelphia, **1996**.

4. L. Brugnano and D. Trigiante, *Stability Properties of Some BVM Methods*, Appl. Numer. Math., **1993**, *13*, 4.

5. L. Brugnano and D. Trigiante, *Solving Differential Problems by Multistep Initial and Boundary Value Methods*, Gordon and Berach Science Publishers, Amsterdam, **1998**.

6. K. Burrage, Z. Jackiewicz, S. Nøresett and R. Renaut, *Preconditioning Waveform Relaxation Iterations for Differential Systems*, BIT, **1996**, *36*, 1.

7. S. Campbell, *Singular Systems of Differential Equations*, Pitman, **1980**.

8. R. Chan, *Circulant Preconditioners for Hermitian Toeplitz Systems*, SIAM J. Matrix Anal. Appl., **1989**, *10*, 4.

9. R. Chan and M. Ng, *Conjugate Gradient Methods for Toeplitz Systems*, SIAM Review, **1996**, *38*, 3.

10. R. Chan, M. Ng and X. Jin, *Strang-Type Preconditioners for Systems of LMF-Based ODE Codes*, IMA J. Numer. Anal., **2001**, *21*, 2.

11. P. Davis, *Circulant Matrices*, John Wiley & Sons Inc., New York, **1979**.

12. X. Jin, *Developments and Applications of Block Toeplitz Iterative Solvers*, Kluwer Academic Publishers & Science Press, **2002**.

13. X. Jin, S. Lei and Y. Wei, *Circulant Preconditioners for Solving Differential Equations with Multi-delays*, submitted.

14. Y. Kuang, *Delay Differential Equations with Applications in Population Dynamics*, Academic Press, INC., **1993**.

15. S. Lei and X. Jin, *Strang-Type Preconditioners for Solving Differential-Algebraic Equations*, In *NAA 2000, LNCS*; L. Vulkov, J. Wasniewski and P. Yalamov, Eds.; Springer, New York, 2001; Vol. 1988, p 505–512.

16. S. Lei and X. Jin, *BCCB Preconditioners for Systems of BVM-Based Numerical Integrators*, Numer. Linear Algebra, to appear.

17. E. Lelarasmee, A. Ruehli and A. Sangiovanni-Vincentelli, *The Waveform Relaxation Method for Time-domain Analysis of Large Scale Integrated Circuits*, IEEE Trans. CAD IC Sys., **1982**, *1*, 3.

18. F. Lin, X. Jin and S. Lei, *Strang-type Preconditioners for Solving Linear Systems form Delay Differential Equations*, BIT, to appear.

19. U. Miekkala and O. Nevanlinna, *Convergence of Dynamic Iteration Methods for Initial Value Problems*, SIAM J. Sci. Stat. Comput., **1987**, *8*, 4.

20. T. Mori, N. Fukuma and M. Kuwahara, *Simple Stability Criteria for Single and Composite Linear Systems with Time Delays*, Int. J. Control, **1981**, *34*, 6.

21. O. Nevanlinna, *Remarks on Picard-Lindelöf Iteration*, Part I., BIT, **1989**, *29*, 2.

22. O. Nevanlinna, *Remarks on Picard-Lindelöf Iteration*, Part II, BIT, **1989**, *29*, 3.

23. Y. Saad and M. Schultz, *GMRES: a Generalized Minimal Residual Algorithm for Solving Non-Symmetric Linear Systems*, SIAM J. Sci. Stat. Comput., **1986**, *7*, 3.

24. M. Shirvani and J. So, *Solutions of Linear Differential Algebraic Equations*, SIAM Review, **1998**, *40*, 2.

25. S. Vandewalle, *Parallel Multigrid Waveform Relaxation for Parabolic Problems*, B.G. Teubner, Stuttgart, **1993**.

26. S. Vandewalle and R. Piessens, *On Dynamic Iteration Methods for Solving Time-Periodic Differential Equations*, SIAM J. Num. Anal., **1993**, *30*, 1.

27. S. Wang, *Further Results on Stability of* $\dot{X}(t) = AX(t) + BX(t-\tau)$, Syst. Cont. Letters, **1992**, *19*, 2.