CIRCULANT PRECONDITIONERS FOR SECOND-ORDER HYPERBOLIC EQUATIONS

Xiao-Qing Jin and Raymond H. Chan

Department of Mathematics

University of Hong Kong

Hong Kong

September 1991

Revised December 1991

Abstract. Linear systems arising from implicit time discretizations and finite difference space discretizations of second-order hyperbolic equations in two-dimension are considered. We propose and analyze the use of circulant preconditioners for the solution of linear systems via preconditioned iterative methods such as the conjugate gradient method. Our motivation is to exploit the fast inversion of circulant systems with Fast Fourier Transform (FFT). For the second-order hyperbolic equations with initial and Dirichlet boundary conditions, we prove that the condition number of the preconditioned system is of $O(\alpha)$ or $O(m)$, where α is the quotient between the time and space steps and m is the number of interior gridpoints in each direction. The results are extended to parabolic equations. Numerical experiments also indicate that the preconditioned systems exhibit favorable clustering of eigenvalues that leads to a fast convergence rate.

Abbreviated Title. Circulant Preconditioners for Hyperbolic Equations

Key Words. Hyperbolic equation, circulant matrix, condition number, preconditioned conjugate gradient method

AMS(MOS) Subject Classifications. 65F10, 65N22

Typeset by $\mathcal{A}\mathcal{M}\mathcal{S}$-TEX

§1 Introduction.

In this paper, we are concerned with the numerical solution of initial and Dirichlet boundary value problems of second-order hyperbolic equations by iterative methods. After discretization by using an implicit time-marching method, such problems reduce to the solution of linear systems of the form $Ax = b$ in each time step. We shall only consider the case where A is symmetric and positive definite.

The problems that we want to discuss are the second-order hyperbolic equations of the form

$$\frac{\partial^2 z}{\partial t^2} = \frac{\partial}{\partial x_1}(a\frac{\partial z}{\partial x_1}) + \frac{\partial}{\partial x_2}(b\frac{\partial z}{\partial x_2}) + g$$

with given initial and Dirichlet boundary conditions. An application of such hyperbolic equations is the vibration problem of light homogeneous membrane, see [15]. A number of common methods for solving such kind of problems are explicit finite difference schemes, see [14]. For explicit methods, however, the maximal time step $k_{\max}$ is limited by the CFL-criterion, which in some situations may be unrealistically strict. An example is when the time-dependence of the problem is much weaker than the space-dependence and hence large time step could be used.

An alternative is to use implicit schemes. Usually, the numerical solution of two-dimensional second-order hyperbolic equation on a uniform grid, using an implicit time-marching scheme, involves the solution of block tridiagonal systems of equations in each time step. The important properties for the matrices of such systems are their sparsity and bandwidth. If the grid has m interior gridpoints in each direction, then the block tridiagonal matrix is m^2-by-m^2 and contains only about $5m^2$ nonzero entries. The bandwidth of the matrix is $(2m+1)$. It is desirable to retain the sparsity in solving procedure, and therefore interest has been shifted to iterative methods.

A popular iterative method for solving symmetric positive definite systems is the preconditioned conjugate gradient method, see [1] and [10]. A successful type of preconditioners is the modified incomplete LU (MILU) factorizations, see for instance, [11]. We note that though the conjugate gradient method is highly parallelizable, see [16], both the computation and the application of the MILU preconditioner have limited degree of parallelism because of the inherently sequential way in which the grid points are ordered.

The purpose of this paper is to propose another class of preconditioners, one that is based on averaging the coefficients of A to form a circulant approximation to A. Recent research on circulant preconditioners for Toeplitz systems shows that the preconditioned systems often have clustering of eigenvalues which is favorable to the convergence rate, see [2], [3], [4], [5], [6] and [8]. Particularly, Holmgren and Otto [12] have used the circulant preconditioners for implicit systems arising from first-order hyperbolic equations where the coefficient matrix A is highly nonsymmetric and non-diagonally dominant. Hence many classical preconditioning techniques are not effective and sometimes not well-defined. For these problems, the circulant preconditioners are often the only ones that work. Circulant preconditioners have also been used by R. Chan and T. Chan [7] for the solution of linear systems arising from elliptic problems.

In this paper, we will extend the idea explored in [7] to construct our preconditioners for the hyperbolic and parabolic cases. We note that both the computation (based on averaging of the coefficients of A) and the inversion (using FFT's) of our circulant preconditioner are highly parallelizable, see [17]. Our main results in this paper is that for the second-order hyperbolic or parabolic equations defined on unit square with initial and Dirichlet boundary conditions, the condition number of the preconditioned system is of $O(\alpha)$ or $O(m)$, where α is the quotient between the time and space steps and m is the number of interior gridpoints in each direction.

The outline of the paper is as follows. We define the circulant preconditioner in §2 and analyze a model problem in §3. The results are extended to variable coefficient case in §4 and to parabolic equations in §5. Numerical experiments are presented in §6 to verify these theoretical results and to illustrate the effect of clustering of the spectrum and the effect of variation of coefficients.

§2 Circulant Approximation to Discretized System.

In this section, we derive the discretized system of second-order hyperbolic equation in two-dimensional case. The preconditioner for solving this linear system is also constructed.

We consider the following second-order hyperbolic equation

$$\frac{\partial^2 z}{\partial t^2} = \frac{\partial}{\partial x_1}(a\frac{\partial z}{\partial x_1}) + \frac{\partial}{\partial x_2}(b\frac{\partial z}{\partial x_2}) + g \ , \tag{1}$$

where $0 < x_1 < 1$, $0 < x_2 < 1$, $t > 0$ and $a = a(x_1, x_2)$, $b = b(x_1, x_2)$, $g = g(t, x_1, x_2)$ are given functions with

$$0 < c_{\min} \leq a(x_1, x_2),\ b(x_1, x_2) \leq c_{\max} \tag{2}$$

for some constants $c_{\min}$ and $c_{\max}$. The initial conditions are given as follows:

$$z(0, x_1, x_2) = f_0(x_1, x_2) \ , \quad \text{and} \quad z_t(0, x_1, x_2) = f_1(x_1, x_2) \ ,$$

and the boundary conditions are given by

$$\begin{aligned} z(t, 0, x_2) &= z_0(t, x_2) \ , \quad z(t, 1, x_2) = z_1(t, x_2) \ , \\ z(t, x_1, 0) &= z_2(t, x_1) \ , \quad z(t, x_1, 1) = z_3(t, x_1) \ . \end{aligned} \tag{3}$$

In this way, we obtain a mixed initial and boundary value problem.

Let

$$u = z_t \ , \quad w = a z_{x_1} \ , \quad v = b z_{x_2} \ ,$$

then we have the following first-order system

$$\begin{cases} \frac{\partial u}{\partial t} = \frac{\partial w}{\partial x_1} + \frac{\partial v}{\partial x_2} + g \\ \frac{\partial w}{\partial t} = a \frac{\partial u}{\partial x_1} \\ \frac{\partial v}{\partial t} = b \frac{\partial u}{\partial x_2} \ . \end{cases} \tag{4}$$

The grid is uniform in the computational domain with $(m+2) \times (m+2)$ gridpoints, where $m \geq 2$. Let $u_{i,j}, v_{i,j}, w_{i,j}$ denote the calculated approximate solutions of u, v, w at point $(x_{1,i}, x_{2,j})$ and $a_{i,j}, b_{i,j}, g_{i,j}$ denote the values of a, b, g at point $(x_{1,i}, x_{2,j})$ respectively, where

$$\begin{cases} x_{1,i} = ih \ , \ i = 0, \cdots, m+1 \ , \\ x_{2,j} = jh \ , \ j = 0, \cdots, m+1 \ , \end{cases}$$

and h is the space step. By using the trapezoidal rule with time step k to do the time-discretization of equations (4), and then followed by using central-differencing schemes to approximate the spatial derivatives, we then have

$$\begin{cases} \frac{u_{i,j}^{n+1} - u_{i,j}^n}{k} - \frac{w_{i+\frac{1}{2},j}^{n+1} - w_{i-\frac{1}{2},j}^{n+1} + w_{i+\frac{1}{2},j}^n - w_{i-\frac{1}{2},j}^n}{2h} \\ - \frac{v_{i,j+\frac{1}{2}}^{n+1} - v_{i,j-\frac{1}{2}}^{n+1} + v_{i,j+\frac{1}{2}}^n - v_{i,j-\frac{1}{2}}^n}{2h} = \frac{1}{2}(g_{i,j}^{n+1} + g_{i,j}^n) \\ \frac{w_{i-\frac{1}{2},j}^{n+1} - w_{i-\frac{1}{2},j}^n}{k} - a_{i-\frac{1}{2},j} \frac{u_{i,j}^{n+1} - u_{i-1,j}^{n+1} + u_{i,j}^n - u_{i-1,j}^n}{2h} = 0 \\ \frac{v_{i,j-\frac{1}{2}}^{n+1} - v_{i,j-\frac{1}{2}}^n}{k} - b_{i,j-\frac{1}{2}} \frac{u_{i,j}^{n+1} - u_{i,j-1}^{n+1} + u_{i,j}^n - u_{i,j-1}^n}{2h} = 0 \ . \end{cases} \tag{5}$$

Let $\alpha = \frac{k}{h}$ and substitute the last two equations in (5) to the first equation, we have

$$\begin{aligned} &(\frac{4}{\alpha^2} + a_{i+\frac{1}{2},j} + a_{i-\frac{1}{2},j} + b_{i,j+\frac{1}{2}} + b_{i,j-\frac{1}{2}})u_{i,j}^{n+1} - a_{i+\frac{1}{2},j}u_{i+1,j}^{n+1} - a_{i-\frac{1}{2},j}u_{i-1,j}^{n+1} \\ &- b_{i,j+\frac{1}{2}}u_{i,j+1}^{n+1} - b_{i,j-\frac{1}{2}}u_{i,j-1}^{n+1} = \frac{d_{i,j}^{n+1}}{\alpha^2} \ , \end{aligned}$$

where $d_{i,j}^{n+1}$ are known quantities. Observe that $u_{i,0}^{n+1}$, $u_{i,m+1}^{n+1}$, $u_{0,j}^{n+1}$ and $u_{m+1,j}^{n+1}$ for $i, j = 0, \cdots, m+1$ are given directly by the boundary conditions (3). This implies that we have to solve for m^2 unknowns in each time step.

Let

$$u^{n+1} = (u_{1,1}^{n+1}, u_{2,1}^{n+1}, \cdots, u_{m,1}^{n+1}, u_{2,1}^{n+1}, \cdots, u_{m,m}^{n+1})$$

and

$$d^{n+1} = \frac{1}{\alpha^2}(d_{1,1}^{n+1}, d_{2,1}^{n+1}, \cdots, d_{m,1}^{n+1}, d_{2,1}^{n+1}, \cdots, d_{m,m}^{n+1}) \ ,$$

then we finally obtain the following linear system

$$Au^{n+1} = d^{n+1} \ . \tag{6}$$

Here the matrix A is an m^2-by-m^2 block tridiagonal matrices where the diagonal blocks are tridiagonal matrices and the off-diagonal blocks are diagonal matrices. Once we get u^{n+1}, we could obtain $z_{i,j}^{n+1}$, the approximation of $z(t_{n+1}, x_{1,i}, x_{2,j})$, by the following difference scheme

$$u_{i,j}^{n+1} = \frac{z_{i,j}^{n+1} - z_{i,j}^{n}}{k} \ ,$$

i.e., $z_{i,j}^{n+1} = ku_{i,j}^{n+1} + z_{i,j}^{n}$. Hence, we only need to discuss the solution of the system (6) in the remainder of this paper.

For any given m-by-m matrix B, the optimal circulant approximation T, first proposed in [8], is defined as the minimizer of $\|B - C\|_F$ over all circulant matrices C. Here $\|\cdot\|_F$ denotes the Frobenius norm. Let the elements of B be denoted by $b_{i,j}$ and the first column of T be denoted by $(t_0, t_1, \cdots, t_{m-1})^T$. We then have following formula,

$$t_j = \frac{1}{m} \sum_{p-q \equiv j \pmod{m}} b_{pq}, \qquad j = 0, \cdots, m-1.$$

Now consider applying this result to solve system (6). We introduce the following circulant preconditioner which preserves the block structure of A. The preconditioner C is defined as follows:

$$C = I \otimes C^a + C^b \otimes I. \tag{7}$$

Here I is an identity matrix of order m and C^a, C^b are m-by-m circulant matrices with their first columns defined by:

$$\begin{aligned}
c_0^a &= 2\bar{a} + \frac{2\beta}{\alpha^2} + \frac{1}{m^2}(1 + \frac{1}{\alpha^2}) \ , \\
c_1^a &= c_{m-1}^a = -\bar{a} \ , \\
c_i^a &= 0 \ , \ i = 2, \cdots, m-2 \ ; \\
c_0^b &= 2\bar{b} + \frac{2\beta}{\alpha^2} + \frac{1}{m^2}(1 + \frac{1}{\alpha^2}) \ , \\
c_1^b &= c_{m-1}^b = -\bar{b} \ , \\
c_i^b &= 0 \ , \ i = 2, \cdots, m-2 \ ,
\end{aligned}$$

where

$$\bar{a} = \frac{1}{m^2} \sum_{j=1}^{m} \sum_{i=1}^{m-1} a_{i+\frac{1}{2},j} \ , \quad \bar{b} = \frac{1}{m^2} \sum_{i=1}^{m} \sum_{j=1}^{m-1} b_{i,j+\frac{1}{2}} \ , \quad \beta = \frac{m-1}{m} \quad \text{and} \quad \alpha = \frac{k}{h} \ .$$

The shift $\frac{1}{m^2}(1+\frac{1}{\alpha^2})$ can guarantee the reduction of the condition number for the preconditioned system. We will illustrate this in §3.

The question we are facing now is how good this preconditioner is in the sense of minimizing $\kappa(C^{-1}A)$, where $\kappa(\cdot)$ denotes the condition number. We will show that

1. For any α, when m is sufficiently large ($m \gg \alpha$), $\kappa(C^{-1}A) \leq O(\alpha)$, while for the original matrix, $\kappa(A) \leq O(\alpha^2)$.
2. For any m, when α is sufficiently large ($\alpha \gg m$), $\kappa(C^{-1}A) \leq O(m)$, while for the original matrix, $\kappa(A) \leq O(m^2)$.

We first prove the claims above for a model problem in case of $a(x_1,x_2) = b(x_1,x_2) = 1$ in §3 and then extend the results to general variable-coefficient case in §4.

§3 Analysis for Model Problem.

In the constant-coefficient case of $a(x_1,x_2) = b(x_1,x_2) = 1$, A is an m^2-by-m^2 matrix of the following form

$$A = A_0 \otimes I + I \otimes A_0 \ , \tag{8}$$

where A_0 is an m-by-m matrix given by

$$A_0 = \begin{pmatrix} 2+\frac{2}{\alpha^2} & -1 & & \underset{\sim}{0} \\ -1 & \ddots & \ddots & \\ & \ddots & \ddots & -1 \\ \underset{\sim}{0} & & -1 & 2+\frac{2}{\alpha^2} \end{pmatrix} .$$

In this case, $\bar{a} = \bar{b} = \beta = \frac{m-1}{m}$. In particular, the circulant preconditioner C is given by

$$C = C_0 \otimes I + I \otimes C_0 \ , \tag{9}$$

where C_0 is an m-by-m matrix given by

$$C_0 = \beta \begin{pmatrix} 2+\frac{2}{\alpha^2} & -1 & & & -1 \\ -1 & \ddots & \ddots & \underset{\sim}{0} & \\ & \ddots & \ddots & \ddots & \\ & \underset{\sim}{0} & \ddots & \ddots & -1 \\ -1 & & & -1 & 2+\frac{2}{\alpha^2} \end{pmatrix} + \frac{1}{m^2}(1+\frac{1}{\alpha^2})I \ .$$

Hence C_0 is a positive definite circulant matrix. For the eigenvalues of A_0 and C_0, we have the following Lemma.

Lemma 1. *The eigenvalues of A_0 and C_0 are given as follows:*

$$\lambda_j(A_0) = \frac{2}{\alpha^2} + 4\sin^2\frac{\pi(j+1)}{2m+2} \tag{10}$$

$$\lambda_j(C_0) = \frac{2\beta}{\alpha^2} + \frac{1}{m^2}(1+\frac{1}{\alpha^2}) + 4\beta\sin^2\frac{\pi j}{m}\ , \tag{11}$$

for $j = 0, \cdots, m-1$.

Proof. For (10), one can refer to [14]. For (11), since C_0 is a circulant matrix, we have $C_0 = F\Lambda F^*$, where

$$F = [\frac{1}{\sqrt{m}}e^{2\pi ijk/m}]_{0\le j\le m-1, 0\le k\le m-1},$$

is the Fourier matrix, F^* is the complex conjugate transpose of F and Λ is a diagonal matrix containing the eigenvalues of C_0, see Davis [9]. By using this spectral decomposition, one can easily obtain (11). $\square$

By (8) and (10), we know that the eigenvalues of A are given by

$$\lambda_{i,j}(A) = \frac{4}{\alpha^2} + 4\sin^2\frac{\pi(i+1)}{2m+2} + 4\sin^2\frac{\pi(j+1)}{2m+2}, \tag{12}$$

for $0 \le i, j \le m-1$. From (12), we know that when α is sufficiently large ($\alpha \gg m$), then the smallest eigenvalue of A decreases to zero like $O(\frac{1}{m^2})$. Since $\lambda_{i,j}(A) \le 9$ for $\alpha \ge 4$ and $0 \le i, j \le m-1$, as a consequence, we have

$$\kappa(A) \le O(m^2)\ .$$

If m is sufficiently large ($m \gg \alpha$), then the smallest eigenvalue of A decreases to zero like $O(\frac{1}{\alpha^2})$. As a consequence, we have

$$\kappa(A) \le O(\alpha^2)\ .$$

For the condition number of $C_0^{-1}A_0$, we have the following two Lemmas.

Lemma 2. *Let $\lambda(C_0^{-1}A_0)$ denote any eigenvalue of $C_0^{-1}A_0$. For any α, when m is sufficiently large ($m \gg \alpha$), we have*

$$\frac{1}{2} \leq \lambda(C_0^{-1}A_0) \leq O(\alpha) \ .$$

As a consequence, we have

$$\kappa(C_0^{-1}A_0) \leq O(\alpha) \ , \qquad \textit{when } m \gg \alpha \ .$$

Proof. Let e_j be the j-th unit m-vector. Since

$$C_0 = \beta(A_0 - e_1 e_m^* - e_m e_1^*) + \frac{1}{m^2}(1 + \frac{1}{\alpha^2})I \ ,$$

we note that for all m-vectors x,

$$\begin{aligned} x^* C_0 x =& \beta x^* A_0 x + \beta x^*(e_1 e_1^* + e_m e_m^*)x - \beta x^*(e_1 + e_m)(e_1 + e_m)^* x + \frac{1}{m^2}(1 + \frac{1}{\alpha^2})x^* x \\ =& 2\beta x^* A_0 x - \beta x^*[A_0 - (e_1 e_1^* + e_m e_m^*) - \frac{1}{\beta m^2}(1 + \frac{1}{\alpha^2})I]x \\ & - \beta x^*(e_1 + e_m)(e_1 + e_m)^* x \ . \end{aligned}$$

We note that the matrix $(e_1 + e_m)(e_1 + e_m)^*$ is positive semi-definite and the matrix

$$A_0 - (e_1 e_1^* + e_m e_m^*) - \frac{1}{\beta m^2}(1 + \frac{1}{\alpha^2})I$$

is also positive semi-definite when $m \geq \sqrt{\alpha^2 + 1}$. We then have for any α, when m is sufficiently large,

$$x^* C_0 x \leq 2\beta x^* A_0 x \ .$$

Thus

$$\frac{1}{2} \leq \frac{1}{2\beta} \leq \min_{\|x\| \neq 0} \frac{x^* A_0 x}{x^* C_0 x} \leq \lambda(C_0^{-1}A_0) \ .$$

On the other hand, we note that for all m-vectors x,

$$\begin{aligned} \beta x^* A_0 x =& x^* C_0 x + \frac{\beta}{2} x^*(e_1 + e_m)(e_1 + e_m)^* x - \frac{\beta}{2} x^*(e_1 - e_m)(e_1 - e_m)^* x \\ & - \frac{1}{m^2}(1 + \frac{1}{\alpha^2})x^* x \ , \end{aligned}$$

where the last two terms on the right hand side is always non-positive. Thus

$$\beta x^* A_0 x \le x^* C_0 x + \frac{\beta}{2} x^* e e^* x \ ,$$

where $e = e_1 + e_m$, i.e.,

$$\frac{x^* A_0 x}{x^* C_0 x} \le \frac{1}{\beta} + \frac{1}{2} \frac{x^* e e^* x}{x^* C_0 x} \ . \tag{13}$$

We note that for all nonzero m-vectors x,

$$\frac{x^* e e^* x}{x^* C_0 x} \le \|C_0^{-1/2} e e^* C_0^{-1/2}\|_2 = e^* C_0^{-1} e \ . \tag{14}$$

From the proof of Lemma 1, we know that $C_0 = F \Lambda F^*$ where by (11), the entries of Λ are given by

$$[\Lambda]_{j,j} = \lambda_j(C_0) = \frac{2\beta}{\alpha^2} + \frac{1}{m^2}(1 + \frac{1}{\alpha^2}) + 4\beta \sin^2 \theta_j \ ,$$

where $\theta_j = \pi j/m$, $0 \le j \le m-1$. Hence

$$\begin{aligned} e^* C_0^{-1} e =& e^* F \Lambda^{-1} F^* e = \frac{4}{m} \sum_{j=0}^{m-1} \frac{\cos^2 \theta_j}{\frac{2\beta}{\alpha^2} + \frac{1}{m^2}(1 + \frac{1}{\alpha^2}) + 4\beta \sin^2 \theta_j} \\ =& \frac{4m\alpha^2}{2m^2 - 2m + \alpha^2 + 1} + \frac{8}{m} \sum_{j=1}^{m/2-1} \frac{\cos^2 \theta_j}{\frac{2\beta}{\alpha^2} + \frac{1}{m^2}(1 + \frac{1}{\alpha^2}) + 4\beta \sin^2 \theta_j} \\ \le& \frac{4m\alpha^2}{2m^2 - 2m + \alpha^2 + 1} + \frac{4}{\pi} \int_{\frac{\pi}{m}}^{\frac{\pi}{2}} \frac{\cos^2 \theta d\theta}{\frac{1}{2\alpha^2} + \sin^2 \theta} \ . \end{aligned} \tag{15}$$

For any α, when m is sufficiently large ($m \gg \alpha$), we have by (15)

$$\begin{aligned} e^* C_0^{-1} e &\le 1 + \frac{4}{\pi} \int_0^{\frac{\pi}{2}} \frac{\cos^2 \theta d\theta}{\frac{1}{2\alpha^2} + \sin^2 \theta} \le 1 + \frac{4 \cos^2 \hat{\theta}}{\pi} \int_0^{\frac{\pi}{2}} \frac{d\theta}{\frac{1}{2\alpha^2} + \sin^2 \theta} \\ &= 1 + \frac{4 \cos^2 \hat{\theta}}{\pi} \frac{\pi}{2\sqrt{\frac{1}{2\alpha^2}(\frac{1}{2\alpha^2} + 1)}} \ , \end{aligned} \tag{16}$$

where $0 < \hat{\theta} < \frac{\pi}{2}$. By (14) and (16), we then have

$$\frac{x^* e e^* x}{x^* C_0 x} \le e^* C_0^{-1} e \le 1 + c\alpha = O(\alpha) \ , \tag{17}$$

where c is a constant. By (13) and (17), we have

$$\lambda(C_0^{-1} A_0) \le \max_{\|x\| \ne 0} \frac{x^* A_0 x}{x^* C_0 x} \le O(\alpha) \ . \qquad \square$$

Lemma 3. *For any m, when α is sufficiently large ($\alpha \gg m$), we have*

$$O(1) \leq \lambda(C_0^{-1}A_0) \leq O(m) \ .$$

As a consequence, we have

$$\kappa(C_0^{-1}A_0) \leq O(m) \ , \qquad \text{when } \alpha \gg m \ .$$

Proof. We note that

$$\begin{aligned} x^*C_0x =& 2\beta x^*A_0x - \beta x^*[A_0 - (e_1e_1^* + e_me_m^*)]x \\ & - \beta x^*(e_1 + e_m)(e_1 + e_m)^*x + \frac{1}{m^2}(1 + \frac{1}{\alpha^2})x^*x \ . \end{aligned}$$

Since the matrices $(e_1 + e_m)(e_1 + e_m)^*$ and $A_0 - (e_1e_1^* + e_me_m^*)$ are positive semi-definite, we have

$$x^*C_0x \leq 2\beta x^*A_0x + \frac{1}{m^2}(1 + \frac{1}{\alpha^2})x^*x \ . \tag{18}$$

When α is sufficiently large ($\alpha \gg m$), we know that $x^*x \leq O(m^2)x^*A_0x$. Using this fact, we see from (18) that $(2\beta + O(1))^{-1} \leq \lambda(C_0^{-1}A_0)$, i.e.,

$$O(1) \leq \lambda(C_0^{-1}A_0) \ .$$

On the other hand, for any m, when α is sufficiently large ($\alpha \gg m$), we have by (14) and (15)

$$\begin{aligned} \frac{x^*ee^*x}{x^*C_0x} &\leq e^*C_0^{-1}e \leq O(m) + \frac{4}{\pi}\int_{\frac{\pi}{m}}^{\frac{\pi}{2}} \frac{\cos^2\theta d\theta}{\sin^2\theta} \\ &\leq O(m) + \frac{4}{\pi}\int_{\frac{\pi}{m}}^{\frac{\pi}{2}} \frac{d\theta}{\sin^2\theta} = O(m) + \frac{4}{\pi}\cot(\frac{\pi}{m}) = O(m) \ . \end{aligned} \tag{19}$$

By (13) and (19), we thus have

$$\lambda(C_0^{-1}A_0) \leq O(m) \ . \qquad \square$$

By using Lemmas 2 and 3, we then have

Theorem 1. *For the circulant preconditioned systems of the model problem, we have*

(i) For any α, when m is sufficiently large ($m \gg \alpha$), we have

$$\frac{1}{2} \leq \lambda(C^{-1}A) \leq O(\alpha) \ .$$

(ii) For any m, when α is sufficiently large ($\alpha \gg m$), we have

$$O(1) \leq \lambda(C^{-1}A) \leq O(m) \ .$$

As a consequence, we have

$$\kappa(C^{-1}A) \leq O(\alpha) \ , \qquad \text{when } m \gg \alpha \ ;$$

and

$$\kappa(C^{-1}A) \leq O(m) \ , \qquad \text{when } \alpha \gg m \ .$$

Proof. For (i), we note that for any m-vector x, when $m \geq \sqrt{\alpha^2+1}$, by Lemma 2,

$$\frac{1}{2}x^*C_0x \leq x^*A_0x \leq O(\alpha)x^*C_0x \ .$$

Hence, for any m^2-vector x, one can easily prove that

$$\frac{1}{2}x^*(C_0 \otimes I)x \leq x^*(A_0 \otimes I)x \leq O(\alpha)x^*(C_0 \otimes I)x$$

and

$$\frac{1}{2}x^*(I \otimes C_0)x \leq x^*(I \otimes A_0)x \leq O(\alpha)x^*(I \otimes C_0)x \ .$$

Combining these two inequalities togather, we have (i). Similarly, we can prove (ii) by using Lemma 3. $\square$

For conjugate gradient method, it is important that the spectrum of $C^{-1}A$ has highly multiple eigenvalues or the eigenvalues are clustered in a small interval (a, b) which keeps a clear gap between a and 0. For $m = 4,\ 8$ and 16, the following tables show the distributions of the eigenvalues for increasing α. In these tables, the eigenvalues are ordered as

$$\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_{m^2-1} \leq \lambda_{m^2} \ .$$

	A			$C^{-1}A$		
α	λ_1	λ_{m^2-1}	λ_{m^2}	λ_1	λ_{m^2-1}	λ_{m^2}
10	0.80393	6.2761	7.2761	0.80923	1.8355	7.0293
100	0.76433	6.2365	7.2365	0.80529	1.8460	8.3609
1000	0.76394	6.2361	7.2361	0.80525	1.8462	8.3775

Table 1. Eigenvalues Distribution for $m = 4$

	A			$C^{-1}A$		
α	λ_1	λ_{m^2-1}	λ_{m^2}	λ_1	λ_{m^2-1}	λ_{m^2}
10	0.28123	7.4515	7.7988	0.64169	2.4046	9.1196
100	0.24373	7.4140	7.7613	0.63434	2.4791	16.896
1000	0.24123	7.4115	7.7588	0.63427	2.4798	17.040

Table 2. Eigenvalues Distribution for $m = 8$

	A			$C^{-1}A$		
α	λ_1	λ_{m^2-1}	λ_{m^2}	λ_1	λ_{m^2-1}	λ_{m^2}
10	0.10811	7.8709	7.9719	0.57667	3.5551	8.5180
100	0.07061	7.8334	7.9344	0.56344	4.0515	32.934
1000	0.068112	7.8309	7.9319	0.56329	4.0577	34.342

Table 3. Eigenvalues Distribution for $m = 16$

We observe that the eigenvalues of $C^{-1}A$ are all located in a relatively small interval (c, d) except one outlying eigenvalue which increases like $O(m)$ for α large enough, just as our Theorem 1 predicts. Here d is increased slowly with α and m increasing. Since the spectrum of the preconditioned matrix $C^{-1}A$ is clustered, which is favorable to the conjugate gradient method, we can expect fast convergence. This fact is confirmed numerically in §6.

§4 Analysis for Variable Coefficient Problem.

In this section, we extend the results in the last section to variable coefficient case. We consider the second-order hyperbolic equations of the form given by (1). Let $\tilde{A}$ be the m^2-by-m^2 matrix given by (6). Define $A_{\max} = c_{\max} \cdot A$ and $A_{\min} = c_{\min} \cdot A$, where $c_{\max}$, $c_{\min}$ are given in (2) and A is given by (8). Without loss of generality, we assume $c_{\min} \leq 1$ and $c_{\max} \geq 1$. Let $\tilde{C}$, $C_{\max}$ and $C_{\min}$ be the the circulant approximations of $\tilde{A}$,

$A_{\max}$ and $A_{\min}$ respectively. Clearly, $C_{\max} = c_{\max} \cdot C$ and $C_{\min} = c_{\min} \cdot C$, where C is given by (9). We then have the following Lemma. The proof of the Lemma can be found in [7], we therefore omit it.

Lemma 4. *All the matrices $A_{\max} - \tilde{A}$, $\tilde{A} - A_{\min}$, $C_{\max} - \tilde{C}$ and $\tilde{C} - C_{\min}$ are positive semi-definite.*

By Lemma 4, for all nonzero vectors x, we have

$$0 < x^* A_{\min} x \leq x^* \tilde{A} x \leq x^* A_{\max} x \tag{20}$$

and

$$0 < x^* C_{\min} x \leq x^* \tilde{C} x \leq x^* C_{\max} x. \tag{21}$$

Combining (20) with (21), we get

$$0 < \frac{c_{\min}}{c_{\max}} \frac{x^* A x}{x^* C x} = \frac{x^* A_{\min} x}{x^* C_{\max} x} \leq \frac{x^* \tilde{A} x}{x^* \tilde{C} x} \leq \frac{x^* A_{\max} x}{x^* C_{\min} x} = \frac{c_{\max}}{c_{\min}} \frac{x^* A x}{x^* C x}.$$

Recalling the results from Theorem 1, we then have our main results.

Theorem 2. *Let $\tilde{A}$ be the discretization matrix of (1) defined by (6) under the condition (2) and $\tilde{C}$ be the circulant preconditioner as defined in (7). We have*

(i) For any α, when m is sufficiently large ($m \gg \alpha$), we have

$$O(1) \leq \lambda(\tilde{C}^{-1}\tilde{A}) \leq O(\alpha) \ .$$

(ii) For any m, when α is sufficiently large ($\alpha \gg m$), we have

$$O(1) \leq \lambda(\tilde{C}^{-1}\tilde{A}) \leq O(m) \ .$$

As a consequence, we have

$$\kappa(\tilde{C}^{-1}\tilde{A}) \leq O(\alpha) \ , \qquad \text{when } m \gg \alpha \ ;$$

and

$$\kappa(\tilde{C}^{-1}\tilde{A}) \leq O(m) \ , \qquad \text{when } \alpha \gg m \ .$$

§5 Extension to Parabolic Equation.

In this section, we extend our results to parabolic equations. We consider the following parabolic equation

$$\frac{\partial z}{\partial t} = \frac{\partial}{\partial x_1}(a\frac{\partial z}{\partial x_1}) + \frac{\partial}{\partial x_2}(b\frac{\partial z}{\partial x_2}) + g \ , \tag{22}$$

where

$$0 < x_1 < 1 \ , \quad 0 < x_2 < 1 \ , \quad t > 0 \ ,$$

and

$$a = a(x_1, x_2), \quad b = b(x_1, x_2), \quad g = g(t, x_1, x_2)$$

are given functions with

$$0 < c_{\min} \le a(x_1, x_2),\ b(x_1, x_2) \le c_{\max}$$

for some constants $c_{\min}$ and $c_{\max}$. The initial condition is given by

$$z(0, x_1, x_2) = g_0(x_1, x_2) \ ,$$

and the boundary conditions are given by

$$z(t, 0, x_2) = z_0(t, x_2) \ , \quad z(t, 1, x_2) = z_1(t, x_2) \ ,$$

$$z(t, x_1, 0) = z_2(t, x_1) \ , \quad z(t, x_1, 1) = z_3(t, x_1) \ .$$

By using the uniform grid and notations introduced in §2, for any function $f(x_1, x_2)$ defined on unit square, we define the following forward and backward differences as

$$\triangle_i \, f_{i,j} = f_{i+1,j} - f_{i,j} \ , \quad \nabla_i f_{i,j} = f_{i,j} - f_{i-1,j} \ ,$$

$$\triangle^j \, f_{i,j} = f_{i,j+1} - f_{i,j} \ , \quad \nabla^j f_{i,j} = f_{i,j} - f_{i,j-1} \ .$$

Then by applying Crank–Nicholson scheme to (22), see [13], we have

$$\begin{aligned} &\frac{z_{i,j}^{n+1} - z_{i,j}^{n}}{k} - \frac{1}{2h^2}[\triangle_i(a_{i-\frac{1}{2},j} \nabla_i \, z_{i,j}^{n+1}) + \triangle^j(b_{i,j-\frac{1}{2}} \nabla^j \, z_{i,j}^{n+1})] \\ &- \frac{1}{2h^2}[\triangle_i(a_{i-\frac{1}{2},j} \nabla_i \, z_{i,j}^{n}) + \triangle^j(b_{i,j-\frac{1}{2}} \nabla^j \, z_{i,j}^{n})] = g_{i,j}^{n+\frac{1}{2}} \ . \end{aligned} \tag{23}$$

Thus, we have from (23)

$$
\begin{aligned}
&(\frac{2}{\alpha^2} + a_{i+\frac{1}{2},j} + a_{i-\frac{1}{2},j} + b_{i,j+\frac{1}{2}} + b_{i,j-\frac{1}{2}})z_{i,j}^{n+1} - a_{i+\frac{1}{2},j}z_{i+1,j}^{n+1} - a_{i-\frac{1}{2},j}z_{i-1,j}^{n+1} \\
&- b_{i,j+\frac{1}{2}}z_{i,j+1}^{n+1} - b_{i,j-\frac{1}{2}}z_{i,j-1}^{n+1} = \frac{d_{i,j}^{n+1}}{\alpha^2} \ ,
\end{aligned}
$$

where $\alpha^2 = \frac{k}{h^2}$ and $d_{i,j}^{n+1}$ are known quantities. Finally, we obtain the following linear system

$$
Az^{n+1} = d^{n+1}
$$

with a block tridiagonal matrix A. Thus, we can define a preconditioner C as we did in §2. By using the same trick as we introduced in §3 and §4, we can obtain the same results as in Theorem 2.

§6 Numerical Results.

In this section, we compare the performance of our method to the MILU preconditioner, see [11]. In these tests, we mainly compare the number of iterations. The equation we used is

$$
\frac{\partial^2 z}{\partial t^2} = \frac{\partial}{\partial x_1}\left[\left(1 + \epsilon e^{x_1 x_2}\right)\frac{\partial z}{\partial x_1}\right] + \frac{\partial}{\partial x_2}\left[\left(1 + \frac{\epsilon}{2}\cos(\pi(x_1 + x_2))\right)\frac{\partial z}{\partial x_2}\right] + g(t, x_1, x_2), \quad (24)
$$

defined on the unit square. The ϵ here is a parameter. When $\epsilon = 0$, (24) is the model problem discussed in §3. We discretize the equations by using the schemes we introduced in §2. The right hand side and the initial guess are chosen to be random vectors and are the same for different methods. Computations are done with double precision on a VAX 6420 and the iterations are stopped when $\frac{\|r_k\|_2}{\|r_0\|_2} < 10^{-7}$. Here r_k is the residual vector at the k-th iteration. The circulant preconditioner we used is defined by (7).

Since the circulant preconditioners are based on averaging of these coefficients over the grid points, their performance will deteriorate as the variations in the coefficficients increase. We therefore first symmetrically scale A by its diagonal before applying the circulant preconditioners. In our tests, we apply diagonal scaling to all methods.

We note that the application of the circulant preconditioners require $O(m^2 \log m)$ flops, which is slightly more expensive than the $O(m^2)$ flops required by the MILU preconditioners. The FFTs, however, can be computed in $O(\log m)$ parallel steps with $O(m^2)$ processors, see [17], whereas the MILU preconditioners require at least $O(m)$ steps regardless of how many processors could be used.

The following tables show the number of iterations required for convergence for different choices of ϵ and α. In the tables, I, C and M represent the systems with no preconditioning, circulant preconditioner and MILU preconditioner respectively. We see that for small values of ϵ (e.g. $\epsilon \leq 0.01$) and large values of α (e.g. $\alpha \geq 100$), the number of iteration of the our preconditioners is less than that of MILU. We also note that the MILU method is less sensitive to the changes in ϵ but more sensitive to the changes in α. In contrast, the circulant preconditioner is less sensitive to the changes in α when α is large. In all cases, the number of iterations grows slower than as predicted by Theorems 1 and 2.

ε	0.0			0.01			0.1			1.0		
m	I	C	M	I	C	M	I	C	M	I	C	M
8	24	12	10	24	14	10	28	14	10	29	15	10
16	44	16	13	47	18	13	50	18	13	53	20	13
32	72	19	15	72	22	15	78	22	15	89	25	16
64	94	26	15	94	30	15	103	30	15	120	33	17
128	107	37	15	107	43	15	113	44	15	139	47	17

Table 4. Number of iterations for different systems with $\alpha = 10$

ε	0.0			0.01			0.1			1.0		
m	I	C	M	I	C	M	I	C	M	I	C	M
8	24	12	11	25	13	11	29	14	10	29	15	10
16	47	16	15	53	18	15	54	18	15	57	20	14
32	89	19	21	102	22	21	103	23	21	109	26	20
64	171	25	30	198	29	30	201	30	30	213	33	29
128	326	32	40	351	38	40	356	40	40	416	45	40

Table 5. Number of iterations for different systems with $\alpha = 100$

ε	0.0			0.01			0.1			1.0		
m	I	C	M	I	C	M	I	C	M	I	C	M
8	24	12	11	25	13	11	29	14	10	29	15	10
16	47	16	15	53	18	15	54	18	15	57	20	14
32	89	19	21	103	22	21	103	23	21	110	26	20
64	173	25	31	201	29	31	202	30	31	215	34	30
128	336	32	45	367	38	45	403	40	45	430	46	43

Table 6. Number of iterations for different systems with $\alpha = 1000$

References

[1] O. Axelsson and V. Barker, *Finite Element Solution of Boundary Value Problems: Theory and Computation*, Academic Press, Orlando, Fl. (1983).

[2] R. Chan and G. Strang, *Toeplitz Equations by Conjugate Gradients with Circulant Preconditioner*, SIAM J. Sci. Statist. Comput., 10 (1989), pp. 104–119.

[3] R. Chan, *Circulant Preconditioners for Hermitian Teoplitz Systems*, SIAM J. Matrix Anal. Appl., 10 (1989), pp. 542–550.

[4] R. Chan, X. Jin and M. Yeung, *The Circulant Operator in the Banach Algebra of Matrices*, Lin. Alg. Applic. 149 (1991), pp. 41–53.

[5] R. Chan and X. Jin, *Circulant and Skew-Circulant Preconditioners for Skew-Hermitian Type Toeplitz Systems*, BIT, to appear.

[6] R. Chan and X. Jin, *A Family of Block Preconditioners for Block Systems*, SIAM J. Sci. Stat. Comp., to appear.

[7] R. Chan and T. Chan, *Circulant Preconditioners for Elliptic Problems*, J. of Numerical Linear Algebra with Applications, to appear.

[8] T. Chan, *An Optimal Circulant Preconditioner for Toeplitz Systems*, SIAM J. Sci. Stat. Comp., 9 (1988), pp. 766-771.

[9] P. Davis, *Circulant Matrices*, John Wiley and Sons, Inc., New York (1979).

[10] G. Golub and C. Van Loan, *Matrix Computations*, 2nd edition, Johns Hopkins Univ. Press, Baltimore (1989).

[11] I. Gustafsson, *A Class of First Order Factorization Methods*, BIT 18 (1978), pp. 142-156.

[12] S. Holmgren and K. Otto, *Iterative Solution Methods and Preconditioners for Non-symmetric, Non-diagonally Dominant Block-Tridiagonal Systems of Equations*, Uppsala University, Dept. of Scientific Computing, Report No. 123 (1989).

[13] H. Keller, *A New Difference Scheme for Parabolic Problems*, Numerical Solution of Partial Differential Equations-II, Academic Press (1971), pp. 327-350.

[14] L. Lapidus and G. Pinder, *Numerical Solution of Partial Differential Equations in Science and Engineering*, Wiley, New York (1982).

[15] R. Sakamoto, *Hyperbolic Boundary Value Problems*, Cambridge University Press, London (1978).

[16] M. Seager, *Parallelizing Conjugate Gradient for the CRAY X-MP*, Parallel Comuting 3 (1986), pp. 35-47.

[17] P. Swarztrauber, *Multiprocessor FFTs*, Parallel Comuting 5 (1987), pp. 197-210.