Toeplitz Preconditioners for Hermitian Toeplitz Systems

Raymond H. Chan* and Kwok-Po Ng
Department of Mathematics
University of Hong Kong
Hong Kong
People's Republic of China

Submitted by Robert J. Plemmons

ABSTRACT

We propose a new type of preconditioners for Hermitian positive definite Toeplitz systems $A_n x = b$ where A_n are assumed to be generated by functions f that are positive and 2π-periodic. Our approach is to precondition A_n by the Toeplitz matrix $\tilde{A}_n$ generated by $1/f$. We prove that the resulting preconditioned matrix $\tilde{A}_n A_n$ will have clustered spectrum. When $\tilde{A}_n$ cannot be formed efficiently, we use quadrature rules and convolution products to construct nearby approximations to $\tilde{A}_n$. We show that the resulting approximations are Toeplitz matrices which can be written as sums of $\{\omega\}$-circulant matrices. As a side result, we prove that any Toeplitz matrix can be written as a sum of $\{\omega\}$-circulant matrices. We then show that our Toeplitz preconditioners T_n are generalizations of circulant preconditioners and the way they are constructed is similar to the approach used in the additive Schwarz method for elliptic problems. We finally prove that the preconditioned systems $T_n A_n$ will have clustered spectra around 1.

1. INTRODUCTION

Toeplitz systems arise in a variety of practical applications in mathematics and engineering. For instance, in signal processing, solutions of Toeplitz systems are required in order to obtain the filter coefficients in the design of recursive digital filters; see Chui and A. Chan [10]. Time-series analysis also

*Research supported in part by ONR contract N00014-90-J-1695 and DOE grant DE-FG03-87ER25037.

655 Avenue of the Americas, New York, NY 10010 0024-3795/93/$6.00

involves solutions of Toeplitz systems for the unknown parameters of stationary autoregressive models; see King et al. [17, pp. 368–379].

There are a number of specialized fast direct methods for solving Toeplitz systems; see for instance Trench [23]. For an n-by-n Toeplitz system $A_n x = b$, these algorithms require $O(n^2)$ operations to solve it. Around 1980, superfast direct solvers of complexity $O(n \log^2 n)$ were developed; see for instance Brent, Gustavson, and Yun [3]. However, recent research on using a preconditioned conjugate-gradient method as an iterative method for solving Toeplitz systems has received much attention. The most important result of this methodology is that the complexity of solving a large class of Toeplitz systems can be reduced to $O(n \log n)$.

The iterative approach is to use a preconditioned conjugate-gradient method with circulant matrices as preconditioners for the solution of Toeplitz systems; see Strang [21]. Several successful circulant preconditioners have been proposed and analyzed; see for instance Chan [4], T. Chan [8], Huckle [15], Ku and Kuo [18], Tismenetsky [22], and Tyrtyshnikov [24]. In these papers, the Toeplitz matrix A_n is assumed to be generated by a generating function f, i.e., the diagonals of A_n are given by the Fourier coefficients of f. It has been shown that if f is a positive function in the Wiener class, then these circulant preconditioned systems converge superlinearly.

A unifying approach of constructing circulant preconditioners is given in Chan and Yeung [7], where it is shown that many of the abovementioned circulant preconditioners can be derived by using the convolution products of f with some well-known kernels. For example, Strang's and T. Chan's circulant preconditioners are generated by using the Dirichlet and Fejér kernels respectively. We remark that the convolution products of f with these kernels are just smooth approximations of f. Chan and Yeung [7] proved that if the convolution product converges to f uniformly, i.e. if the convolution product is a good approximation of f, then the circulant preconditioned systems will converge fast.

As alternatives to circulant preconditioners, band-Toeplitz matrices have also been proposed as preconditioners for Toeplitz systems when the generating function f is not positive, but only nonnegative with countable zeros. In this case, most of the circulant preconditioners will fail, whereas the spectra of band-Toeplitz preconditioned matrices are still uniformly bounded by constants independent of n; see Chan [5]. The motivation behind using band-Toeplitz matrices is to approximate f by trigonometric polynomials of fixed degree rather than by convolution products of f with some kernels. The advantage here is that trigonometric polynomials can be chosen to match the zeros of f, so that the method still works when f has zeros. By using Remez's algorithm to search for the best trigonometric approximation of f, band-Toeplitz preconditioned systems can be made to converge at about the same rate as those circulant preconditioned systems even when f is positive; see Chan and Tang [6].

In this paper, we propose a new type of preconditioners for Hermitian positive definite Toeplitz systems. Our approach is to use the Toeplitz matrix $\tilde{A}_n$ generated by $1/f$ to approximate the inverse of A_n, i.e., the preconditioned matrix will be $\tilde{A}_n A_n$. We remark that the inverse of A_n is non-Toeplitz in general, but it is closely related to Toeplitz matrices; see Friedlander et al. [13]. Since $\tilde{A}_n$ is a Toeplitz matrix, the matrix-vector product $\tilde{A}_n y$, which is required in every iteration of the preconditioned conjugate-gradient method, can be performed in $O(n \log n)$ operations by using fast Fourier transforms (FFTs); see Strang [21]. Hence the cost per iteration is $O(n \log n)$.

As for the convergence rate, it is well known that it depends on the spectrum of the preconditioned matrix $\tilde{A}_n A_n$: the more clustered it is, the faster the convergence rate will be; see Axelsson and Barker [2, p. 26]. Presumably, we want $\tilde{A}_n A_n = I_n + L_n + U_n$, where I_n is the n-by-n identity matrix, L_n is a low-rank matrix, and U_n is a small-norm matrix. We will first show that if f is a finite trigonometric series, then the rank of $\tilde{A}_n A_n - I_n$ is fixed independent of n. Then in the general case when f is a 2π-periodic continuous function, we show that $\tilde{A}_n A_n - I_n$ is indeed equal to a low-rank matrix plus a small-norm matrix. Hence we can then conclude that the spectrum of the preconditioned matrix is clustered around 1, and therefore, if the preconditioned conjugate-gradient method is applied to the preconditioned system, we expect fast convergence.

We note however that in general it may be difficult to compute the Fourier coefficients of $1/f$ explicitly, and hence $\tilde{A}_n$ cannot be formed efficiently. In these cases, we derive families of Toeplitz preconditioners $T_n^{(s)}$ by using different kernel functions and different levels of approximation in approximating the Fourier coefficients of $1/f$. We will show that for the first level of approximation, $s = 1$, our Toeplitz preconditioners $T_n^{(1)}$ reduce to the well-known circulant preconditioners mentioned above, depending on the kernel function we used. As an example, if the kernel function is the Fejér function, then $T_n^{(1)}$ is just the inverse of T. Chan's circulant preconditioner proposed in [8].

For integers $s > 1$, we will show that the Toeplitz preconditioner $T_n^{(s)}$ thus constructed can be written as a sum of so-called $\{\omega\}$-circulant matrices (see Davis [11, p. 84] or Section 4 for definition). More precisely, we have

$$T_n^{(s)} = \frac{1}{s} \sum_{t=0}^{s-1} V_t,$$

where V_t are $\{\omega_t\}$-circulant matrices with $\omega_t = e^{-2\pi i t/s}$. As a side result, we will see that given any Toeplitz matrix A_n and integer $s > 1$, we have

$$A_n = \frac{1}{s}\sum_{t=0}^{s-1} W_t,$$

where W_t are also $\{\omega_t\}$-circulant matrices. We note that for $s = 2$, this formula was first discovered by Pustylnikov [19]. We further show that for any $0 \leqslant t < s$ we have $W_t^{-1} = V_t$, provided that the Dirichlet kernel is used and W_t is invertible. In particular, if all W_t are invertible, we have

$$T_n^{(s)} = \frac{1}{s}\sum_{t=0}^{s-1} W_t^{-1}.$$

In this aspect, our Toeplitz preconditioner is closely related to the additive Schwarz-type preconditioners proposed by Dryja and Widlund [12].

For the convergence rate, we will prove that the preconditioned system $T_n^{(s)}A_n$ has clustered spectrum around 1 and converges at the same rate as other well-known circulant preconditioned systems. Numerical results show that our methods converge faster than those preconditioned by circulant preconditioners or best band-Toeplitz preconditioners.

The outline of the paper is as follows. In Section 2, we study Toeplitz preconditioners generated by $1/f$ and prove some of their clustering properties. The preconditioners serve as motivation for the general Toeplitz preconditioners $T_n^{(s)}$ we construct in Section 3. Two ways of constructing $T_n^{(s)}$ are given. In Section 4, we show that $T_n^{(s)}$, and in fact any Toeplitz matrix, can be written as a sum of $\{\omega\}$-circulant matrices. In Section 5, we prove that Toeplitz preconditioners have clustering and superlinear convergence properties. Finally, numerical examples and concluding remarks are given in Sections 6 and 7.

2. TOEPLITZ PRECONDITIONER GENERATED BY $1/f$

Let $\mathscr{C}_{2\pi}$ be the set of all 2π-periodic continuous real-valued functions. For all $f \in \mathscr{C}_{2\pi}$, let

$$a_k = \frac{1}{2\pi}\int_0^{2\pi} f(\theta)e^{-ik\theta}\,d\theta, \qquad k = 0, \pm 1, \pm 2, \dots,$$

be the Fourier coefficients of f, $\mathscr{T}[f]$ be the semiinfinite Toeplitz matrix with the (j,k)th entry given by a_{j-k}, and $\mathscr{T}_n[f]$ be the n-by-n principal submatrix of $\mathscr{T}[f]$. Since f is real-valued, we have

$$a_{-k} = \bar{a}_k, \qquad k = 0, \pm 1, \pm 2, \dots .$$

It follows that $\mathscr{T}[f]$ and $\mathscr{T}_n[f]$ are Hermitian. We note that the spectrum $\sigma(\mathscr{T}_n[f])$ of $\mathscr{T}_n[f]$ satisfies

$$\sigma(\mathscr{T}_n[f]) \subseteq [f_{\min}, f_{\max}] \qquad \forall n \geqslant 1, \tag{1}$$

where $f_{\min}$ and $f_{\max}$ are the minimum and maximum of f respectively; see, for instance, Grenander and Szegö [14, pp. 63–65]. In particular, if f is positive, then $\mathscr{T}_n[f]$ is positive definite for all n.

For the Toeplitz systems $A_n x = b$ considered in this paper, we will assume that $A_n = \mathscr{T}_n[f]$ for some functions f in $\mathscr{C}_{2\pi}$. The systems will be solved by using a preconditioned conjugate-gradient method; see Axelsson and Barker [2, p. 26]. Thus instead of solving the original system, we solve $P_n A_n x = P_n b$. In order to have fast convergence, the preconditioner P_n should be chosen such that the spectrum of $P_n A_n$ is clustered. Specifically, we want $P_n A_n$ to be of the form $I_n + L_n + U_n$ where I_n is an n-by-n identity matrix, L_n is a matrix of low rank, and U_n is a matrix of small l_2 norm.

In this section, we will consider using the Toeplitz matrix $\mathscr{T}_n[1/f]$ generated by $1/f$ as preconditioner for $\mathscr{T}_n[f]$. Our motivation for choosing $\mathscr{T}_n[1/f]$ as preconditioner is given by the following lemma due to Widom [26, p. 192]. We first note that a function f (not necessarily real-valued) is said to be of *analytic type* (or respectively *coanalytic type*) if $a_k = 0$ for $k < 0$ (or respectively $a_k = 0$ for $k > 0$).

LEMMA 1. *Let f be of analytic type (or respectively coanalytic type) and $a_0 \neq 0$. Then $\mathscr{T}[f]$ is invertible if and only if $1/f$ is bounded and of analytic type (or respectively coanalytic type). In either case, we have $\mathscr{T}[1/f]\mathscr{T}[f] = \mathscr{T}[f]\mathscr{T}[1/f] = I$, where I is the identity operator.*

As an immediate corollary, we have $\mathscr{T}_n[1/f]\mathscr{T}_n[f] = I_n$ for all $n \geqslant 1$, i.e., if $\mathscr{T}_n[f]$ is an upper or lower triangular Toeplitz matrix, then its inverse is the Toeplitz matrix $\mathscr{T}_n[1/f]$ generated by $1/f$. In the remainder of this section, we assume that the Fourier coefficients of $1/f$ are given explicitly or easily found and hence $\mathscr{T}_n[1/f]$ is readily available.

LEMMA 2. *Let f be a positive trigonometric polynomial of degree K in $\mathscr{C}_{2\pi}$, i.e.*

$$f(\theta) = \sum_{k=-K}^{K} a_k e^{ik\theta}.$$

Then for $n > 2K$, $\operatorname{rank}(\mathscr{T}_n[1/f]\mathscr{T}_n[f] - I_n) \leqslant 2K$.

Proof. Let

$$\frac{1}{f(\theta)} = \sum_{k=-\infty}^{\infty} \rho_k e^{ik\theta}.$$

We see that

$$\sum_{k=-K}^{K} a_k \rho_{m-k} = \begin{cases} 1 & \text{if} \quad m = 0, \\ 0 & \text{otherwise.} \end{cases}$$

Hence for $n > 2K$, the entries of the matrix $\mathscr{T}_n[1/f]\mathscr{T}_n[f] - I_n$ are all zeros except possibly entries in its first and last K columns. ■

As an example, consider the Kac-Murdock-Szegö matrices [16], whose generating function is given by

$$f(\theta) = \frac{1 + \alpha^2 - \alpha e^{i\theta} - \alpha e^{-i\theta}}{1 - \alpha^2}$$

for $|\alpha| < 1$. Hence $\mathscr{T}_n[f]$ is a tridiagonal Toeplitz matrix. Since

$$\frac{1}{f(\theta)} = \sum_{k=-\infty}^{\infty} \alpha^{|k|} e^{ik\theta} = \frac{1 - \alpha^2}{(1 - \alpha e^{i\theta})(1 - \alpha e^{-i\theta})},$$

$\mathscr{T}_n[1/f]$ is a dense Toeplitz matrix. However, by Lemma 2, the rank of the matrix $\mathscr{T}_n[1/f]\mathscr{T}_n[f] - I_n$ is at most two, and therefore the conjugate gradient method will converge in at most three steps; see Axelsson and Barker [2, p. 14].

We end this section by considering general f in $\mathscr{C}_{2\pi}$.

LEMMA 3. *Let $f \in \mathscr{C}_{2\pi}$ be positive. Then for all $\varepsilon > 0$, there exist positive integers M and N such that for all $n > N$,*

$$\mathscr{T}_n[1/f]\mathscr{T}_n[f] = I_n + L_n + U_n, \tag{2}$$

where rank $L_n \leqslant M$ *and* $\|U_n\|_2 < \varepsilon$.

Proof. By the Weierstrass theorem (see Cheney [9, p. 144]), there exists a positive trigonometric polynomial

$$p_K(\theta) = \sum_{k=-K}^{K} \rho_k e^{ik\theta}$$

with $\rho_{-k} = \bar{\rho}_k$ such that $p_K(\theta)$ satisfies the following conditions:

$$\tfrac{1}{2} f_{\min} \leqslant p_K(\theta) \leqslant 2 f_{\max} \qquad \forall \theta \in [0, 2\pi], \tag{3}$$

and

$$\max_{\theta \in [0, 2\pi]} |f(\theta) - p_K(\theta)| \leqslant \frac{f_{\min}}{2}\left(-1 + \sqrt{1+\varepsilon}\right) \min\left\{\frac{f_{\min}}{2 f_{\max}}, 1\right\}. \tag{4}$$

Since f is positive, it follows from (1) and (3) that the matrices $\mathscr{T}_n[1/f]$, $\mathscr{T}_n[p_K]$, and $\mathscr{T}_n[1/p_K]$ are all positive definite for all n. Write

$$\begin{aligned} \mathscr{T}_n[1/f]\mathscr{T}_n[f] &= \mathscr{T}_n[1/f]\mathscr{T}_n^{-1}[1/p_K]\mathscr{T}_n[1/p_K]\mathscr{T}_n[p_K]\mathscr{T}_n^{-1}[p_K]\mathscr{T}_n[f] \\ &= (I_n + V_n)(\mathscr{T}_n[1/p_K]\mathscr{T}_n[p_K])(I_n + W_n), \end{aligned} \tag{5}$$

where

$$V_n = (\mathscr{T}_n[1/f] - \mathscr{T}_n[1/p_K])\mathscr{T}_n^{-1}[1/p_K]$$

and

$$W_n = \mathscr{T}_n^{-1}[p_K](\mathscr{T}_n[f] - \mathscr{T}_n[p_K]).$$

Note that by (1), (3), and (4), we have

$$\|\mathcal{T}_n^{-1}[p_K]\|_2 \leqslant \frac{2}{f_{\min}}, \tag{6}$$

$$\|\mathcal{T}_n^{-1}[1/p_K]\|_2 \leqslant 2 f_{\max}, \tag{7}$$

$$\|\mathcal{T}_n[f] - \mathcal{T}_n[p_K]\|_2 \leqslant \frac{(-1 + \sqrt{1+\varepsilon}) f_{\min}}{2}, \tag{8}$$

and

$$\begin{aligned} \|\mathcal{T}_n[1/f] - \mathcal{T}_n[1/p_K]\|_2 &\leqslant \max_{\theta \in [0, 2\pi]} \left| \frac{1}{f(\theta)} - \frac{1}{p_K(\theta)} \right| \\ &\leqslant \frac{2}{f_{\min}^2} \max_{\theta \in [0, 2\pi]} |f(\theta) - p_K(\theta)| \qquad (9) \\ &\leqslant \frac{-1 + \sqrt{1+\varepsilon}}{2 f_{\max}}. \end{aligned}$$

From Lemma 2 we have, when $n > 2K$,

$$\mathcal{T}_n[1/p_K] \mathcal{T}_n[p_K] = I_n + \tilde{L}_n$$

with rank $\tilde{L}_n \leqslant 2K$. Therefore, (5) becomes

$$\mathcal{T}_n[1/f] \mathcal{T}_n[f] = (I_n + V_n)(I_n + \tilde{L}_n)(I_n + W_n) \equiv I_n + L_n + U_n, \tag{10}$$

where

$$U_n = V_n + W_n + V_n W_n$$

and

$$L_n = \tilde{L}_n (I_n + W_n) + V_n \tilde{L}_n (I_n + W_n).$$

It is clear that rank $L_n \leqslant 4K$, and from (6), (7), (8), and (9) we see that $\|U_n\|_2 \leqslant \varepsilon$. ∎

We now show that the spectrum of $\mathcal{T}_n[1/f]\mathcal{T}_n[f]$ is clustered around 1.

THEOREM 1. *Let $f \in \mathcal{C}_{2\pi}$ be positive. Then for all $\varepsilon > 0$, there exist positive integers M and $N > 0$ such that for all $n > N$, at most M eigenvalues of $\mathcal{T}_n[1/f]\mathcal{T}_n[f] - I_n$ have absolute values greater than ε.*

Proof. First we note that since f is positive, it follows from (1) that $\mathcal{T}_n[1/f]$ is a Hermitian positive definite matrix. Hence its square root $\mathcal{T}_n^{1/2}[1/f]$ is well defined and is also a Hermitian positive definite matrix. Moreover, the norms $\|\mathcal{T}_n^{-1/2}[1/f]\|_2$ and $\|\mathcal{T}_n^{1/2}[1/f]\|_2$ are uniformly bounded independent of n. Next we note that the non-Hermitian matrix $\mathcal{T}_n[1/f]\mathcal{T}_n[f]$ is similar to the Hermitian positive definite matrix

$$X_n \equiv \mathcal{T}_n^{1/2}[1/f]\mathcal{T}_n[f]\mathcal{T}_n^{1/2}[1/f];$$

therefore the eigenvalues of $\mathcal{T}_n[1/f]\mathcal{T}_n[f]$ are the same as the singular values of X_n. In the following, we will show that the singular values of X_n are clustered.

By (2), we have

$$X_n = I_n + \mathcal{T}_n^{-1/2}[1/f]L_n\mathcal{T}_n^{1/2}[1/f] + \mathcal{T}_n^{-1/2}[1/f]U_n\mathcal{T}_n^{1/2}[1/f].$$

Using the properties of L_n and U_n as stated in Lemma 3 and the uniform boundedness of $\|\mathcal{T}_n^{-1/2}[1/f]\|_2$ and $\|\mathcal{T}_n^{1/2}[1/f]\|_2$, we see that the matrices $\mathcal{T}_n^{-1/2}[1/f]L_n\mathcal{T}_n^{1/2}[1/f]$ and $\mathcal{T}_n^{-1/2}[1/f]U_n\mathcal{T}_n^{1/2}[1/f]$ are matrices of low rank and small l_2 norm respectively. Therefore, we have

$$X_n^* X_n = I_n + \hat{L}_n + \hat{U}_n,$$

where $\hat{L}_n$ is of low rank, $\hat{U}_n$ is of small l_2 norm, and both matrices are Hermitian. Using Cauchy's interlacing theorem (see, for instance, Wilkinson [27, p. 103]), we see that the singular values of X_n are clustered around 1. ∎

Using Theorem 1, we can easily prove that if the conjugate-gradient method is used to solve the preconditioned system

$$\mathcal{T}_n[1/f]\mathcal{T}_n[f]x = \mathcal{T}_n[1/f]b,$$

the method will converge superlinearly; see Chan [4]. Thus, we see that $\mathcal{T}_n[1/f]$ is a good choice of preconditioner for $\mathcal{T}_n[f]$. However, we remark that in order to construct $\mathcal{T}_n[1/f]$, the first n Fourier coefficients of $1/f$ should be easily available, and this may not be true in general.

3. CONSTRUCTION OF GENERAL TOEPLITZ PRECONDITIONERS

In this section, we construct our Toeplitz preconditioners for cases where the Fourier coefficients of $1/f$, i.e.

$$\frac{1}{2\pi}\int_0^{2\pi}\frac{1}{f(\theta)}e^{-ik\theta}\,d\theta, \tag{11}$$

cannot be evaluated efficiently. There are three different cases where this can happen:

(a) f is given explicitly, but the evaluation of the definite integral (11) cannot be done efficiently.

(b) f is given, but the evaluations of $1/f$ are costly, e.g., f is given in series form.

(c) f is not given explicitly, e.g., only the Toeplitz matrix A_n is given rather than f.

Our approach is to approximate the integral by the rectangular rule and f by the convolution product of f with some kernel functions.

Let us begin with case (a). We subdivide the interval $[0, 2\pi]$ into $sn - 1$ subintervals of equal length. Here s is a positive integer independent of n. Then we approximate (11) by

$$z_k^{(s)} = \frac{1}{sn}\sum_{j=0}^{sn-1}\frac{1}{f(2\pi j/sn)}e^{-2\pi ijk/sn}, \qquad k = 0, \pm 1, \dots, \pm (n-1). \tag{12}$$

Our preconditioner is then defined to be the Toeplitz matrix $\mathcal{T}_n[g_n^{(s)}]$ generated by

$$g_n^{(s)}(\theta) \equiv \sum_{k=-(n-1)}^{n-1} z_k^{(s)} e^{ik\theta} \qquad \forall\theta \in [0, 2\pi]. \tag{13}$$

We remark that we have defined a family of Toeplitz preconditioners indexed by s. Notice that the first column of the Toeplitz matrix $\mathcal{T}_n[g_n^{(s)}]$ is given by the numbers $\{z_k^{(s)}\}_{k=0}^{n-1}$.

In case (b), we further approximate f in (12) by using its $(n-1)$th partial sum, i.e., we replace f in (12) by

$$f_{n-1}(\theta) = \sum_{k=-(n-1)}^{n-1} a_k e^{ik\theta} \qquad \forall\theta \in [0, 2\pi], \tag{14}$$

and the numbers $\{z_k^{(s)}\}_{k=0}^{n-1}$ so obtained will again give the first column of the Toeplitz preconditioner $\mathcal{T}_n[g_n^{(s)}]$. In case (c), we associate the entries of the first column of A_n with a generating function $f_{n-1}(\theta)$ given by (14). Then the numbers $\{z_k^{(s)}\}_{k=0}^{n-1}$ can be obtained as in case (b).

We remark that we can unify the notations employed above by using convolution products. Given a kernel function $\mathcal{K}$ and a positive integer s, we define our approximation to the Fourier coefficients in (11) to be

$$z_k^{(s)} = \frac{1}{sn} \sum_{j=0}^{sn-1} \frac{1}{(\mathcal{K} * f)(2\pi j/sn)} e^{-2\pi ijk/sn}, \quad k = 0, \pm 1, \dots, \pm (n-1). \tag{15}$$

Here $\mathcal{K} * f$ is the convolution product of $\mathcal{K}$ and f; see Walker [25, p. 86]. In the first case (12) above, we are just using the Dirac delta kernel $\mathcal{K} = \delta$, and in the second case (14), $\mathcal{K} = \mathcal{D}_{n-1}$, the Dirichlet kernel; see Walker [25, pp. 87, 45] respectively. We note that there are other kernels that one can use, such as the Fejér kernel $\mathcal{F}_n$; see Walker [25, p. 76]. We remark that in (15), we are assuming that the values of $\mathcal{K} * f$ at the sampled points $\{2\pi j/sn\}_{j=0}^{sn-1}$ are nonzero.

In all cases, the Toeplitz preconditioner $\mathcal{T}_n[g_n^{(s)}]$ is the Toeplitz matrix with the first column given by $z_k^{(s)}$ in (15). The cost of obtaining the numbers $z_k^{(s)}$ depends on the kernel we use. For the Dirichlet and the Fejér kernels, or more generally, for kernels that can be written as

$$(\mathscr{K}*f)(\theta) = \sum_{k=-(n-1)}^{n-1} b_{n,k}e^{ik\theta} \qquad \forall\theta \in [0, 2\pi], \tag{16}$$

the values $\{(\mathscr{K}*f)(2\pi j/sn)\}_{j=0}^{sn-1}$ can be obtained in $O(sn \log sn)$ operations by using an sn-dimensional FFT. After getting the values, the numbers $\{z_k^{(s)}\}_{k=0}^{n-1}$ in (15) can then be obtained by using another sn-dimensional FFT in $O(sn \log sn)$ operations. For a list of kernels that satisfy (16) and their corresponding $b_{n,k}$, see Chan and Yeung [7].

We note that another way of constructing the Toeplitz preconditioners is by embedding. In fact, by (16), we have

$$(\mathscr{K}*f)\left(\frac{2\pi j}{sn}\right) = \sum_{k=-(n-1)}^{n-1} b_{n,k}e^{2\pi ijk/sn} = \sum_{k=0}^{sn-1} \hat{b}_{n,k}e^{2\pi ijk/sn},$$

where for $s = 1$,

$$\hat{b}_{n,k} \equiv b_{n,k} + b_{n,k-n}, \qquad k = 1, \dots, n-1,$$

and for $s > 1$,

$$\hat{b}_{n,k} \equiv \begin{cases} b_{n,k} & 0 < k < n, \\ 0 & n \leqslant k \leqslant sn - n, \\ b_{n,k-sn} & sn - n < k < sn. \end{cases}$$

Thus $(\mathscr{K}*f)(2\pi j/sn)$, $j = 0, \dots, sn-1$, are eigenvalues of an sn-by-sn circulant matrix with the first column given by $\{\hat{b}_{n,k}\}_{k=0}^{sn-1}$; see Davis [11, p. 74]. Let us denote this circulant matrix by C_{sn}. Clearly, the eigenvalues of C_{sn}^{-1} are given by $1/[(\mathscr{K}*f)(2\pi j/sn)]$. Therefore, the first column of the circulant matrix C_{sn}^{-1} will be given by

$$[C_{sn}^{-1}]_{0,k} = \frac{1}{sn}\sum_{j=0}^{sn-1} \frac{1}{(\mathscr{K}*f)(2\pi j/sn)}e^{-2\pi ijk/sn}, \qquad 0 \leqslant k < sn;$$

see also Davis [11, p. 74]. By comparing this formula with (15), we see that our Toeplitz matrix $\mathscr{T}_n[g_n^{(s)}]$ is just the n-by-n principal submatrix of C_{sn}^{-1}.

Notice that if $b_{n,k}$ are known, then the second method requires only one sn-dimensional FFT and we don't need to generate the values $\{(\mathscr{K} * f)(2\pi j/sn)\}_{j=0}^{sn-1}$ explicitly. For example, if the Dirichlet kernel $\mathscr{D}_{n-1}$ is used, then $b_{n,k} = a_k$ for all n and k. Hence in this case, we just embed A_n into an sn-by-sn circulant matrix C_{sn} as defined by $\hat{b}_{n,k}$ above, and our Toeplitz preconditioner is given by the n-by-n principal submatrix of C_{sn}^{-1}.

Let us end the section by considering the cost per iteration in applying the preconditioned conjugate-gradient method to the preconditioned system

$$\mathscr{T}_n[g_n^{(s)}]A_n x = \mathscr{T}_n[g_n^{(s)}]b.$$

We first recall that the multiplication of an n-vector to an n-by-n circulant matrix requires only two n-dimensional FFTs. Since both matrices $\mathscr{T}_n[g_n^{(s)}]$ and A_n are Toeplitz, products of the form $\mathscr{T}_n[g_n^{(s)}]v$ and $A_n v$ can be obtained by first embedding the matrices into $2n$-by-$2n$ circulant matrices and using $2n$-dimensional FFTs; see Strang [21]. Thus the cost per iteration is about the same as the cost of applying four $2n$-dimensional FFTs. For circulant preconditioned systems, we still have to compute a product of the form $A_n v$ in each iteration, but the product $\mathscr{T}_n[g_n^{(s)}]v$ will be replaced by a circulant matrix-vector multiplication which can be done by two n-dimensional FFTs. Thus the actual cost per iteration of our method is roughly $\frac{4}{3}$ times higher than that required by circulant preconditioned systems on sequential machines. On parallel computers using a single-instruction-stream, multiple-data-stream (SIMD) architecture (see for instance Aki [1, p. 5]), because the real time required by a $2n$-dimensional FFT is of $O(\log 2n)$ (see Aki [1, p. 238]), which is about the same as the cost of an n-dimensional FFT, there will be no significant time difference per iteration between our method and those that use circulant preconditioners.

4. PROPERTIES OF TOEPLITZ PRECONDITIONERS

In this section, we give some interesting properties of the Toeplitz preconditioners which will be useful in proving the convergence rate of the Toeplitz preconditioners in the next section. We first show below that the Toeplitz preconditioner can always be written as a sum of so-called $\{\omega\}$-circulant matrices, which are defined as follows (see also Davis [11, p. 84] for an equivalent definition):

DEFINITION. Let $\omega = e^{i\theta_0}$ with $\theta_0 \in [0, 2\pi)$. A matrix W_n is said to be an $\{\omega\}$*-circulant matrix* if it has the spectral decomposition

$$W_n = D_n F_n \Lambda_n F_n^* D_n^*. \tag{17}$$

Here F_n is the Fourier matrix with entries

$$[F_n]_{k,j} = \frac{1}{\sqrt{n}} e^{-2\pi ijk/n}, \tag{18}$$

D_n is given by

$$D_n = \operatorname{diag}[1, \omega^{1/n}, \dots, \omega^{(n-1)/n}],$$

and Λ_n is a diagonal matrix holding the eigenvalues of W_n.

Notice that $\{\omega\}$-circulant matrices are Toeplitz matrices with the first entry of each row obtained by multiplying the last entry of the preceding row by ω. In particular, $\{1\}$-circulant matrices are circulant matrices, while $\{-1\}$-circulant matrices are skew-circulant matrices. Also, from the spectral decomposition in (17) we see that the entries in the first column of W_n and the eigenvalues $\lambda_j(W_n)$ of W_n are related by the following formula:

$$[W_n]_{k,0} = \frac{\omega^{k/n}}{n} \sum_{j=0}^{n-1} \lambda_j(W_n) e^{-2\pi ijk/n}, \qquad k = 0, \dots, n-1. \tag{19}$$

THEOREM 2. *Let* $(\mathscr{K} * f)(2\pi j/sn) \neq 0$ *for* $0 \leqslant j < sn$. *Then the Toeplitz preconditioner* $\mathscr{T}_n[g_n^{(s)}]$ *can be expressed as*

$$\mathscr{T}_n[g_n^{(s)}] = \frac{1}{s} \sum_{t=0}^{s-1} \mathscr{T}_n[g_n^{(s,t)}], \tag{20}$$

where $\mathscr{T}_n[g_n^{(s,t)}]$, $0 \leqslant t < s$, *are* $\{\omega_t\}$*-circulant matrices with* $\omega_t = e^{-2\pi it/s}$

and eigenvalues given by

$$\lambda_j\left(\mathscr{T}_n\left[g_n^{(s,t)}\right]\right) = \frac{1}{(\mathscr{K} * f)\left(\dfrac{2\pi j}{n} + \dfrac{2t\pi}{sn}\right)}, \qquad 0 \leqslant j < n, \quad 0 \leqslant t < s. \tag{21}$$

In particular, if $(\mathscr{K} * f)(2\pi j/sn) > 0$ *for* $0 \leqslant j < sn$, *the Toeplitz preconditioner* $\mathscr{T}_n[g_n^{(s)}]$ *is positive definite.*

Proof. We replace the index j in (15) by $sj + t$ where $0 \leqslant t < s$ and $0 \leqslant j < n$. Then we have

$$z_k^{(s)} = \frac{1}{s}\sum_{t=0}^{s-1}\left\{\frac{e^{-2\pi i t k/sn}}{n}\sum_{j=0}^{n-1}\frac{1}{(\mathscr{K} * f)\left(\dfrac{2\pi j}{n} + \dfrac{2t\pi}{sn}\right)}e^{-2\pi i jk/n}\right\}$$

$$\equiv \frac{1}{s}\sum_{t=0}^{s-1} z_k^{(s,t)},$$

for $k = 0, \pm 1, \dots, \pm (n-1)$. Here

$$z_k^{(s,t)} = \frac{\omega_t^{k/n}}{n}\sum_{j=0}^{n-1}\left\{\frac{1}{(\mathscr{K} * f)\left(\dfrac{2\pi j}{n} + \dfrac{2t\pi}{sn}\right)}e^{-2\pi i jk/n}\right\}, \tag{22}$$

for $0 \leqslant t < s$, $0 \leqslant j < n$. Correspondingly, we define

$$g_n^{(s,t)}(\theta) \equiv \sum_{k=-(n-1)}^{n-1} z_k^{(s,t)} e^{ik\theta}, \quad 0 \leqslant t < s, \qquad \forall\theta \in [0, 2\pi],$$

and rewrite (13) as

$$g_n^{(s)}(\theta) = \frac{1}{s}\sum_{t=0}^{s-1} g_n^{(s,t)}(\theta) = \frac{1}{s}\sum_{t=0}^{s-1}\sum_{k=-(n-1)}^{n-1} z_k^{(s,t)} e^{ik\theta}, \quad s \geqslant 1,$$

$$\forall\theta \in [0, 2\pi].$$

By the linearity of the operator $\mathcal{T}_n[\cdot]$, we see that (20) holds. Moreover, since $\mathcal{T}_n[g_n^{(s,t)}]$ are Toeplitz matrices with their first columns given by $\{z_k^{(s,t)}\}_{k=0}^{n-1}$, by comparing (19) with (22), we see that $\mathcal{T}_n[g_n^{(s,t)}]$ are $\{\omega_t\}$-circulant matrices with eigenvalues given by (21). If $(\mathcal{K} * f)(2\pi j/sn) > 0$ for $0 \leqslant j < sn$, then $\mathcal{T}_n[g_n^{(s,t)}]$ will be positive definite for $0 \leqslant t < s$. Hence $\mathcal{T}_n[g_n^{(s)}]$ is positive definite. ■

As an application, we note that our Toeplitz preconditioners are generalization of circulant preconditioners. Indeed, when $s = 1$, then by Theorem 2, $\mathcal{T}_n[g_n^{(1)}]$ is a circulant matrix. This can also be seen simply from (15), as

$$z_{n-k}^{(1)} = \bar{z}_k^{(1)}, \qquad k = 1, \ldots, n-1.$$

Using the characterization of circulant preconditioners in Chan and Yeung [7], we can further show that if in (15) we choose the kernel $\mathcal{K}$ to be $\mathcal{D}_{\lfloor n/2 \rfloor}$, $\mathcal{D}_{n-1}$, or $\mathcal{F}_n$, then the inverse of $\mathcal{T}_n[g_n^{(1)}]$ equals the Strang, Chan, or T. Chan circulant preconditioner respectively; see Chan and Yeung [7].

We next show that indeed any Toeplitz matrix can be written as a sum of $\{\omega_t\}$-circulant matrices. We first note that from the definition of $\{\omega_t\}$-circulant matrix, the inverse $\mathcal{T}_n^{-1}[g_n^{(s,t)}]$ of $\mathcal{T}_n[g_n^{(s,t)}]$ is still an $\{\omega_t\}$-circulant matrix. Moreover, by (21), its eigenvalues are given by

$$\lambda_j\left(\mathcal{T}_n^{-1}\left[g_n^{(s,t)}\right]\right) = (\mathcal{K} * f)\left(\frac{2\pi j}{n} + \frac{2t\pi}{sn}\right), \qquad 0 \leqslant j < n, \quad 0 \leqslant t < s.$$

Therefore by (19), we see that

$$\mathcal{T}_n^{-1}\left[g_n^{(s,t)}\right] = \mathcal{T}_n\left[h_n^{(s,t)}\right],$$

where

$$h_n^{(s,t)}(\theta) \equiv \sum_{k=-(n-1)}^{n-1} y_k^{(s,t)} e^{ik\theta}, \quad 0 \leqslant t < s \qquad \forall 0 \in [0, 2\pi],$$

with

$$y_k^{(s,t)} \equiv \frac{\omega_t^{k/n}}{n} \sum_{j=0}^{n-1} \left\{(\mathcal{K} * f)\left(\frac{2\pi j}{n} + \frac{2t\pi}{sn}\right)\right\} e^{-2\pi ijk/n} \tag{23}$$

for $0 \leqslant t < s$, $0 \leqslant j < n$. Clearly, we also have

$$\mathscr{T}_n^{-1}\left[h_n^{(s,t)}\right] = \mathscr{T}_n\left[g_n^{(s,t)}\right] \tag{24}$$

and

$$\lambda_j\left(\mathscr{T}_n\left[h_n^{(s,t)}\right]\right) = (\mathscr{K}*f)\left(\frac{2\pi j}{n} + \frac{2t\pi}{sn}\right), \qquad 0 \leqslant j < n, \quad 0 \leqslant t < s. \tag{25}$$

Now let us add the matrices $\mathscr{T}_n[h_n^{(s,t)}]$ together. More precisely, let

$$h_n^{(s)}(\theta) = \frac{1}{s}\sum_{t=0}^{s-1} h_n^{(s,t)}(\theta) = \sum_{k=-(n-1)}^{n-1}\left\{\frac{1}{s}\sum_{t=0}^{s-1} y_k^{(s,t)}\right\}e^{ik\theta}. \tag{26}$$

We now show that for most kernels $\mathscr{K}$, $h_n^{(s)}$ does give us back $\mathscr{K}*f$ exactly.

LEMMA 4. *Let $\mathscr{K}$ be a kernel of the form given by* (16). *Then for all* $s > 1$,

$$h_n^{(s)}(\theta) = (\mathscr{K}*f)(\theta) \qquad \forall\theta \in [0, 2\pi].$$

Proof. By comparing (16) and (26), it suffices to show that

$$b_{n,k} = \frac{1}{s}\sum_{t=0}^{s-1} y_k^{(s,t)}, \qquad k = 0, \pm 1, \dots \pm (n-1). \tag{27}$$

However, by (23),

$$\frac{1}{s}\sum_{t=0}^{s-1} y_k^{(s,t)} = \frac{1}{s}\sum_{t=0}^{s-1}\frac{e^{-2\pi itk/sn}}{n}\sum_{j=0}^{n-1}\left\{(\mathscr{K}*f)\left(\frac{2\pi j}{n} + \frac{2t\pi}{sn}\right)\right\}e^{-2\pi ijk/n}$$

$$= \frac{1}{sn}\sum_{l=0}^{sn-1}(\mathscr{K}*f)\left(\frac{2\pi l}{sn}\right)e^{-2\pi ilk/sn},$$

$$k = 0, \pm 1, \dots, \pm (n-1),$$

where the last equality is obtained by setting the index $sj + t$ equal to l.

Using (16) again, we have, for any $s \geqslant 1$ and $k = 0, \pm 1, \dots, \pm (n-1)$,

$$\frac{1}{s}\sum_{t=0}^{s-1} y_k^{(s,t)} = \frac{1}{sn}\sum_{l=0}^{sn-1}\left\{\sum_{j=-(n-1)}^{n-1} b_{n,j} e^{2\pi ijl/sn}\right\} e^{-2\pi ilk/sn}$$

$$= \sum_{j=-(n-1)}^{n-1} b_{n,j}\left\{\frac{1}{sn}\sum_{l=0}^{sn-1} e^{2\pi il(j-k)/sn}\right\}.$$

Since

$$\frac{1}{sn}\sum_{l=0}^{sn-1} e^{2\pi il(j-k)/sn} = \begin{cases} 1 & j = k, k \pm sn, k \pm 2sn, \dots, \\ 0 & \text{otherwise}, \end{cases}$$

(27) follows by noting that $s > 1$. ■

We can now show that any Toeplitz matrix can be written as the sum of $\{\omega_t\}$-circulant matrices where $0 \leqslant t < s$, $s > 1$.

THEOREM 3. *Given any Toeplitz matrix A_n and $s > 1$, we have*

$$A_n = \frac{1}{s}\sum_{t=0}^{s-1} W_n^{(s,t)},$$

where $W_n^{(s,t)}$ are $\{\omega_t\}$-circulant matrices with $\omega_t = e^{-2\pi it/s}$. Moreover, if all $W_n^{(s,t)}$ are invertible, then the Toeplitz preconditioner $\mathcal{T}_n[g_n^{(s)}]$ corresponding to the Dirichlet kernel $\mathcal{D}_{n-1}$ is given by

$$\mathcal{T}_n[g_n^{(s)}] = \frac{1}{s}\sum_{t=0}^{s-1}\left(W_n^{(s,t)}\right)^{-1}.$$

Proof. Given A_n with the first column entries $\{a_k\}_{k=0}^{n-1}$, we can write it as $A_n = \mathcal{T}_n[f_{n-1}]$ where

$$f_{n-1}(\theta) = \sum_{k=-(n-1)}^{n-1} a_k e^{ik\theta}.$$

Since

$$(\mathscr{D}_{n-1} * f_{n-1})(\theta) = \sum_{k=-(n-1)}^{n-1} a_k e^{ik\theta} = f_{n-1}(\theta),$$

we have, by Lemma 4 and (25),

$$A_n = \mathscr{T}_n[f_{n-1}] = \mathscr{T}_n[\mathscr{D}_{n-1} * f_{n-1}] = \mathscr{T}_n[h_n^{(s)}] = \frac{1}{s}\sum_{t=0}^{s-1} \mathscr{T}_n[h_n^{(s,t)}],$$

where $\mathscr{T}_n[h_n^{(s,t)}]$ $\{\omega_t\}$-circulant matrices corresponding to the Dirichlet kernel $\mathscr{D}_{n-1}$. Moreover, by (20) and (24),

$$\mathscr{T}_n[g_n^{(s)}] = \frac{1}{s}\sum_{t=0}^{s-1} \mathscr{T}_n[g_n^{(s,t)}] = \frac{1}{s}\sum_{t=0}^{s-1} \mathscr{T}_n^{-1}[h_n^{(s,t)}],$$

provided that $\mathscr{T}_n[h_n^{(s,t)}]$ are invertible. ∎

When $s = 2$, the theorem gives

$$A_n = \tfrac{1}{2}\left(W_n^{(2,0)} + W_n^{(2,1)}\right),$$

where $W_n^{(2,0)}$ is a circulant matrix and $W_n^{(2,1)}$ is a skew-circulant matrix. We remark that this formula was first discovered by Pustylnikov [19]. Also from the theorem, we see that any Toeplitz matrix can be decomposed as a sum of $\{\omega_t\}$-circulant matrices and that our Toeplitz preconditioner is just the sum of the inverses of these $\{\omega_t\}$-circulant matrices.

We recall that in the additive Schwarz method, a matrix A is first decomposed into a sum of individual matrices,

$$A = A^{(1)} + A^{(2)} + \cdots + A^{(s)},$$

and then the generalized inverses of these individual matrices are added back together to form a preconditioner P of the original matrix A, i.e.

$$P = A^{(1)+} + A^{(2)+} + \cdots + A^{(s)+};$$

see Dryja and Widlund [12]. Thus, the construction of our Toeplitz preconditioner is very similar to the approach used in the additive Schwarz method.

5. ANALYSIS OF CONVERGENCE RATE

In this section, we discuss the convergence rate of the preconditioned systems $\mathscr{T}_n[g_n^{(s)}]A_n$. Before we start, we recall the following two lemmas which are useful in the following analysis. The proofs can be found in Chan [5] and Chan and Yeung [7] respectively.

LEMMA 5. *Let $f \in \mathscr{C}_{2\pi}$ and $\tilde{f}(\theta) = f(\theta + \theta_0)$, where $\theta_0 \in [0, 2\pi)$. Then for all $n > 0$,*

$$\mathscr{T}_n[\tilde{f}] = D_n^* \mathscr{T}_n[f] D_n,$$

where

$$D_n = \operatorname{diag}(1, e^{i\theta_0}, e^{i2\theta_0}, \dots, e^{i(n-1)\theta_0}).$$

LEMMA 6. *Let $f \in \mathscr{C}_{2\pi}$, and $\mathscr{K}$ be a kernel such that $\mathscr{K} * f$ converges to f uniformly on $[0, 2\pi]$. Define Λ_n to be the diagonal matrix with diagonal entries*

$$[\Lambda_n]_{j,j} = (\mathscr{K} * f)\left(\frac{2\pi j}{n}\right), \qquad 0 \leqslant j < n.$$

Then for all $\varepsilon > 0$, there exist positive integers N and M such that for all $n > N$, at most M eigenvalues of $\mathscr{T}_n[f] - F_n \Lambda_n F_n^*$ have absolute value greater than ε.

We note that the matrix F_n in Lemma 6 is the Fourier matrix defined in (18), and hence $F_n \Lambda_n F_n^*$ is an n-by-n circulant matrix, and by (25), it is equal to $\mathscr{T}_n[h_n^{(s,0)}]$. The lemma thus states that the matrix $\mathscr{T}_n[f] - \mathscr{T}_n[h_n^{(s,0)}]$ has clustered spectrum around zero. Using Lemmas 5 and 6, we now show that the spectrum of $\mathscr{T}_n[f] - \mathscr{T}_n[h_n^{(s,t)}]$ is also clustered around zero for $0 \leqslant t < s$.

THEOREM 4. *Let $f \in \mathscr{C}_{2\pi}$ and $s \geqslant 1$. Let $\mathscr{K}$ be a kernel such that $\mathscr{K} * f$ converges to f uniformly on $[0, 2\pi]$, and*

$$W_n^{(s,t)} = D_n F_n \Lambda_n^{(s,t)} F_n^* D_n^*$$

be $\{\omega_t\}$-circulant matrices with $\omega_t = e^{-2\pi i t/s}$ and

$$[\Lambda_n^{(s,t)}]_{jj} = (\mathscr{K} * f)\left(\frac{2\pi j}{n} + \frac{2t\pi}{sn}\right), \qquad 0 \leqslant j < n, \quad 0 \leqslant t < s. \quad (28)$$

Then for all $\varepsilon > 0$, there exist positive integers N and M such that for all $n > N$, at most M eigenvalues of $\mathcal{T}_n[f] - W_n^{(s,t)}$ have absolute value greater than ε.

Proof. For all $0 \leqslant t < s$, define

$$\tilde{f}_t(\theta) = f\left(\theta + \frac{2\pi t}{sn}\right).$$

Then we have

$$\left[\Lambda_n^{(s,t)}\right]_{jj} = (\mathscr{K} * f)\left(\frac{2\pi j}{n} + \frac{2\pi t}{sn}\right) = \left(\mathscr{K} * \tilde{f}_t\right)\left(\frac{2\pi j}{n}\right),$$

$$j = 0, 1, \ldots, n-1.$$

Since by Lemma 5 we have

$$D_n^* \mathcal{T}_n[f] D_n = \mathcal{T}_n\left[\tilde{f}_t\right],$$

it follows that

$$\mathcal{T}_n[f] - W_n^{(s,t)} = D_n\left(\mathcal{T}_n\left[\tilde{f}_t\right] - F_n \Lambda_n^{(s,t)} F_n^*\right) D_n^*.$$

As $\mathscr{K} * f$ converges uniformly to f on $[0, 2\pi]$, $\mathscr{K} * \tilde{f}_t$ also converges to $\tilde{f}_t$ uniformly on $[0, 2\pi]$ for all $0 \leqslant t < s$. Hence the theorem follows by applying Lemma 6 and noting that $\|D_n\|_2 = 1$. ∎

As an immediate corollary, we can show that each $\mathcal{T}_n[g_n^{(s,t)}]$, $0 \leqslant t < s$, is already a good approximation to $\mathcal{T}_n[f]$.

LEMMA 7. *Let $f \in \mathscr{C}_{2\pi}$ be positive and $s \geqslant 1$. Let $\mathscr{K}$ be a kernel such that $\mathscr{K} * f$ converges to f uniformly on $[0, 2\pi]$. Then for all $\varepsilon > 0$ and $0 \leqslant t < s$, there exist positive integers N and M such that for all $n > N$, at most M eigenvalues of $I_n - \mathcal{T}_n[g_n^{(s,t)}]\mathcal{T}_n[f]$ have absolute value greater than ε.*

Proof. For any fixed $0 \leqslant t < s$, by comparing (25) and (28) and recalling $\mathcal{T}_n[h_n^{(s,t)}]$ are $\{\omega_t\}$-circulant matrices, we see that the spectrum of $\mathcal{T}_n[f] -$

$\mathscr{T}_n[h_n^{(s,t)}]$ is clustered around zero. Since $\mathscr{K} * f$ converges to f uniformly and $f_{\min} > 0$, it follows that for sufficiently large n, $\mathscr{K} * f$ will be positive. Therefore, by (25) and (21), $\mathscr{T}_n[h_n^{(s,t)}]$ and its inverse $\mathscr{T}_n[g_n^{(s,t)}]$ are positive definite and uniformly invertible for large n. The lemma then follows by noting that

$$\begin{aligned} I_n - \mathscr{T}_n[g_n^{(s,t)}]\mathscr{T}_n[f] &= I_n - \mathscr{T}_n^{-1}[h_n^{(s,t)}]\mathscr{T}_n[f] \\ &= \mathscr{T}_n^{-1}[h_n^{(s,t)}]\left(\mathscr{T}_n[h_n^{(s,t)}] - \mathscr{T}_n[f]\right). \end{aligned}$$ ∎

Now we can prove the main theorem of this section, namely that the spectrum of the preconditioned system $\mathscr{T}_n[g_n^{(s)}]\mathscr{T}_n[f]$ is clustered around 1.

THEOREM 5. *Let $f \in \mathscr{C}_{2\pi}$ be positive and $s \geqslant 1$. Let $\mathscr{K}$ be a kernel such that $\mathscr{K} * f$ converges to f uniformly on $[0, 2\pi]$, and $\mathscr{T}_n[g_n^{(s)}]$ be the Toeplitz preconditioner defined in* (20). *Then for all $\varepsilon > 0$, there exist positive integers N and M such that for all $n > N$, at most M eigenvalues of $I_n - \mathscr{T}_n[g_n^{(s)}]\mathscr{T}_n[f]$ have absolute value greater than ε.*

Proof. Since the spectrum of $\mathscr{T}_n[g_n^{(s,t)}]\mathscr{T}_n[f]$ is clustered around 1 for $0 \leqslant t < s$, we have

$$\mathscr{T}_n[g_n^{(s,t)}]\mathscr{T}_n[f] = I_n + L_n^{(s,t)} + U_n^{(s,t)},$$

where $L_n^{(s,t)}$ is a matrix with rank independent of n, and $U_n^{(s,t)}$ is a matrix with l_2 norm less than ε. We note that by (20)

$$\begin{aligned} \mathscr{T}_n[g_n^{(s)}]\mathscr{T}_n[f] &= \frac{1}{s}\sum_{t=0}^{s-1}\left(\mathscr{T}_n[g_n^{(s,t)}]\mathscr{T}_n[f]\right) \\ &= \frac{1}{s}\sum_{t=0}^{s-1}\left(I_n + L_n^{(s,t)} + U_n^{(s,t)}\right) = I_n + L_n^{(s)} + U_n^{(s)} \end{aligned}$$

where $L_n^{(s)} = (1/s)\sum_{t=0}^{s-1} L_n^{(s,t)}$ and $U_n^{(s)} = (1/s)\sum_{t=0}^{s-1} U_n^{(s,t)}$. As s is independent of n, the rank of $L_n^{(s)}$ is also independent of n and $\|U_n^{(s)}\|_2 < \varepsilon$. The remaining part of the proof is similar to that in Theorem 1. ∎

It follows easily by Theorem 5 that the conjugate-gradient method, when applied to the preconditioned system $\mathscr{T}_n[g_n^{(s)}]A_n$, converges superlinearly.

Recall that in each iteration, the work is $O(n \log n)$; therefore, the work of solving the equation $A_n x = b$ to a given accuracy is also $O(n \log n)$.

6. NUMERICAL EXAMPLES

In this section, we compare our Toeplitz preconditioners with band-Toeplitz preconditioners and circulant preconditioners. We test their performance on six continuous functions defined on $[-\pi, \pi]$. They are (i) $\theta^4 + 1$, (ii) $\sum_{k=-\infty}^{\infty}(1 + |k|)^{-1.1}e^{ik\theta}$, (iii) $(1 - 0.1e^{i\theta})/(1 - 0.8e^{i\theta}) + (1 - 0.1e^{-i\theta})/(1 - 0.8e^{-i\theta})$, (iv) $1 + (\theta + \pi)^2$, (v) θ^4, and (vi) $(\theta - 1)^2(\theta + 1)^2$. We note that the first two functions are 2π-periodic continuous, the third one is a positive rational function and can be written as

$$\frac{2.16 - 1.8\cos\theta}{1.64 - 1.6\cos\theta},$$

the fourth one has a jump at $\theta = \pm\pi$, and the last two are functions with zeros. The matrices A_n are formed by evaluating the Fourier coefficients of the test functions.

In the test, we used the vector of all ones as the right-hand-side vector and the zero vector as the initial guess. The stopping criterion is $\|r_q\|_2/\|r_0\|_2 \leqslant 10^{-7}$, where r_q is the residual vector after q iterations. All computations were done on a Vax 6420 with double-precision arithmetic. Tables 1–6 show the numbers of iterations required for convergence with different choices of preconditioners. In the tables, I denotes that no preconditioner was used, and $T_\delta^{(s)}$, $T_{\mathscr{D}}^{(s)}$, and $T_{\mathscr{F}}^{(s)}$ are the Toeplitz preconditioners based on the Dirac delta function, the Dirichlet kernel $\mathscr{D}_{n-1}$, and the Fejér kernel $\mathscr{F}_n$

TABLE 1
NUMBERS OF ITERATIONS FOR $f(\theta) = \theta^4 + 1$

		Number of Iterations										
n	I	$T_\delta^{(1)}$	$T_\delta^{(2)}$	$T_\delta^{(4)}$	$T_{\mathscr{D}}^{(1)}$	$T_{\mathscr{D}}^{(2)}$	$T_{\mathscr{D}}^{(4)}$	$T_{\mathscr{F}}^{(1)}$	$T_{\mathscr{F}}^{(2)}$	$T_{\mathscr{F}}^{(4)}$	C_S	B_5
16	8	5	4	4	6	5	4	8	8	8	6	7
32	19	5	4	4	5	4	4	7	8	8	5	7
64	36	5	4	4	5	4	4	7	7	7	5	7
128	54	5	4	4	5	4	4	6	6	6	5	7
256	66	5	4	4	5	4	4	6	5	5	5	7
512	70	5	4	4	5	4	4	6	5	5	5	7

respectively. For comparison, we also used Strang's circulant preconditioner C_S (see Strang [21]) and the best band-Toeplitz preconditioner B_5 with half bandwidth 5 (see Chan and Tang [6]). We emphasize that for the circulant and band-Toeplitz preconditioners, the inverse of the matrix is used as the preconditioner. In particular, $T_{\mathscr{D}}^{(1)}$ and $T_{\mathscr{F}}^{(1)}$ are the inverses of the circulant preconditioners proposed by Chan [4] and T. Chan [8] respectively, whereas C_S is the inverse of the Toeplitz preconditioner corresponding to the Dirichlet kernel $\mathscr{D}_{\lfloor n/2 \rfloor}$ with $s = 1$.

In Table 2, since the generating function is not known explicitly, B_5 and $T_{\delta}^{(s)}$ are not available. In Table 4, Remez's algorithm fails to give the best trigonometric approximation to the discontinuous generating function. Hence B_5 is also not available in that case. In Tables 5 and 6, since f has zeros, the kernel functions may be zero at some of the mesh points $2\pi j/sn$, and hence some of the matrices $\mathscr{T}_n[g_n^{(s,t)}]$ are undefined; see (21). In that case, we just replace those eigenvalues of $\mathscr{T}_n[g_n^{(s,t)}]$ by zeros. We note that although

TABLE 2
NUMBERS OF ITERATIONS FOR $f(\theta) = \sum_{k=-\infty}^{\infty}(1 + |k|)^{-1.1}e^{ik\theta}$

		Numbers of Iterations										
n	I	$T_{\delta}^{(1)}$	$T_{\delta}^{(2)}$	$T_{\delta}^{(4)}$	$T_{\mathscr{D}}^{(1)}$	$T_{\mathscr{D}}^{(2)}$	$T_{\mathscr{D}}^{(4)}$	$T_{\mathscr{F}}^{(1)}$	$T_{\mathscr{F}}^{(2)}$	$T_{\mathscr{F}}^{(4)}$	C_S	B_5
16	8	—	—	—	5	3	4	4	4	4	5	—
32	11	—	—	—	5	3	3	5	3	3	5	—
64	14	—	—	—	4	3	4	5	4	4	5	—
128	17	—	—	—	5	4	4	5	4	4	5	—
256	20	—	—	—	5	4	4	5	4	4	5	—
512	22	—	—	—	5	4	4	5	4	4	5	—

TABLE 3
NUMBERS OF ITERATIONS FOR $f(\theta) = (1 - 0.1e^{i\theta})/(1 - 0.8e^{i\theta}) + (1 - 0.1e^{-i\theta})/(1 - 0.8e^{-i\theta})$

		Number of Iterations										
n	I	$T_{\delta}^{(1)}$	$T_{\delta}^{(2)}$	$T_{\delta}^{(4)}$	$T_{\mathscr{D}}^{(1)}$	$T_{\mathscr{D}}^{(2)}$	$T_{\mathscr{D}}^{(4)}$	$T_{\mathscr{F}}^{(1)}$	$T_{\mathscr{F}}^{(2)}$	$T_{\mathscr{F}}^{(4)}$	C_S	B_5
16	6	2	2	2	5	4	4	3	3	3	5	7
32	9	2	2	2	5	4	4	3	2	2	5	8
64	11	2	2	2	5	5	5	2	2	2	3	9
128	15	2	2	2	5	4	4	2	2	2	2	9
256	18	2	2	2	4	4	4	2	2	2	2	9
512	18	2	2	2	4	4	4	2	2	2	2	9

TABLE 4
NUMBERS OF ITERATIONS FOR $f(\theta) = (\theta + \pi)^2 + 1$

		Number of Iterations										
n	I	$T_{\delta}^{(1)}$	$T_{\delta}^{(2)}$	$T_{\delta}^{(4)}$	$T_{\mathcal{D}}^{(1)}$	$T_{\mathcal{D}}^{(2)}$	$T_{\mathcal{D}}^{(4)}$	$T_{\mathcal{F}}^{(1)}$	$T_{\mathcal{F}}^{(2)}$	$T_{\mathcal{F}}^{(4)}$	C_S	B_5
16	16	7	6	6	8	7	7	9	9	9	10	—
32	33	7	6	6	9	8	8	10	10	10	14	—
64	45	8	7	7	9	10	9	11	11	11	17	—
128	49	10	7	7	10	10	10	12	12	12	19	—
256	50	10	8	8	10	11	10	12	12	12	20	—
512	51	11	8	8	11	11	10	12	13	13	21	—

TABLE 5
NUMBERS OF ITERATIONS FOR $f(\theta) = \theta^4$

		Number of Iterations										
n	I	$T_{\delta}^{(1)}$	$T_{\delta}^{(2)}$	$T_{\delta}^{(4)}$	$T_{\mathcal{D}}^{(1)}$	$T_{\mathcal{D}}^{(2)}$	$T_{\mathcal{D}}^{(4)}$	$T_{\mathcal{F}}^{(1)}$	$T_{\mathcal{F}}^{(2)}$	$T_{\mathcal{F}}^{(4)}$	C_S	B_5
16	9	—	6	7	8	8	6	8	8	8	9	8
32	28	—	6	7	11	11	10	16	17	17	10	11
64	103	—	7	7	16	18	14	25	25	25	13	11
128	420	—	13	10	27	30	20	38	40	40	16	12
256	> 1000	—	13	12	45	70	30	109	102	102	19	12
512	> 1000	—	14	13	119	179	66	340	305	305	27	13

TABLE 6
NUMBERS OF ITERATIONS FOR $f(\theta) = (\theta - 1)^2(\theta + 1)^2$

		Number of Iterations										
n	I	$T_{\delta}^{(1)}$	$T_{\delta}^{(2)}$	$T_{\delta}^{(4)}$	$T_{\mathcal{D}}^{(1)}$	$T_{\mathcal{D}}^{(2)}$	$T_{\mathcal{D}}^{(4)}$	$T_{\mathcal{F}}^{(1)}$	$T_{\mathcal{F}}^{(2)}$	$T_{\mathcal{F}}^{(4)}$	C_S	B_5
16	8	5	5	4	12	9	10	8	8	8	7	7
32	24	5	5	4	8	7	6	14	13	13	8	8
64	67	5	5	4	8	8	6	17	18	18	9	8
128	185	6	6	4	10	8	7	22	21	21	6	8
256	450	8	4	6	10	7	7	27	28	28	8	8
512	> 1000	8	6	6	10	9	10	36	35	35	8	8

$\mathscr{T}_n[g_n^{(s,t)}]$ may be singular, the preconditioners

$$\mathscr{T}_n[g_n^{(s)}] = \frac{1}{s}\sum_{t=0}^{s-1}\mathscr{T}_n[g_n^{(s,t)}]$$

are nonsingular in all the cases we tested, except in the Table 4, $T_8^{(1)}$ is singular, as $f(0) = 0$.

From the numerical results, we see that in all tests, the Toeplitz preconditioner $T_8^{(4)}$ performs better than the other preconditioners and the differences are more profound when f is either discontinuous or nonnegative. For other Toeplitz preconditioners, the number of iterations in most cases decreases as s is increased. We note that the larger s is, the better the rectangular rule (12) will be in approximating the definite integral (11).

7. CONCLUDING REMARKS

In this paper, we have proposed and analyzed new types of preconditioners for Hermitian positive definite Toeplitz systems. The preconditioners are Toeplitz matrices and can be considered as generalizations of circulant preconditioners proposed previously by other authors. In this preliminary report, we have only considered using the rectangular rule to approximate the definite integral (11). We note that other Newton-Cotes formulas can also be employed (see Stoer and Bulirsch [20, pp. 119–120]). The definite integral (11) will then be approximated by

$$z_k^{(s)} = \frac{1}{sn}\sum_{j=0}^{sn-1}\frac{\beta_j}{f(2\pi j/sn)}e^{-2\pi ijk/sn}, \qquad k = 0, \pm 1, \dots, \pm (n-1),$$

where β_j are the weights used in the approximating formula. For example, for Simpson's rule, (11) will be approximated by

$$z_k^{(s)} = \frac{1}{3sn}\left\{\frac{1}{f(0)} + \frac{4}{f(2\pi/sn)}e^{-2\pi ik/sn} + \frac{2}{f(4\pi/sn)}e^{-4\pi ik/sn} + \cdots + \frac{1}{f(2(sn-1)\pi/sn)}e^{-2\pi i(sn-1)k/sn}\right\},$$

for $k = 0, \pm 1, \dots, \pm (n-1)$. Presumably, such higher-order quadrature rules will yield better preconditioners.

We would like to thank Professors M. Pourahmadi and G. Ammar of Northern Illinois University for their valuable discussions.

REFERENCES

1 S. Aki, *The Design and Analysis of Parallel Algorithms*, Prentice-Hall International, London, 1989.

2 O. Axelsson and V. Barker, *Finite Element Solution of Boundary Value Problems, Theory and Computation*, Academic, New York, 1984.

3 R. Brent, F. Gustavson, and D. Yun, Fast solution of Toeplitz systems of equations and computations of Padé approximations, *J. Algorithms* 1:259–295 (1980).

4 R. Chan, Circulant preconditioners for Hermitian Toeplitz systems, *SIAM J. Matrix Anal. Appl.* 10:542–550 (1989).

5 R. Chan, Toeplitz preconditioners for Toeplitz systems with nonnegative generating functions, *IMA J. Numer. Anal.* 11:333–345 (1991).

6 R. Chan and P. Tang, Fast band-Toeplitz preconditioners for Toeplitz systems, *SIAM J. Sci. Statist. Comput.*, to appear.

7 R. Chan and M. Yeung, Circulant preconditioners constructed from kernels, *SIAM J. Numer. Anal.* 29:1093–1103 (1992).

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

9 E. Cheney, *Introduction to Approximation Theory*, McGraw-Hill, New York, 1966.

10 C. Chui and A. Chan, Application of approximation theory methods to recursive digital filter design, *IEEE Trans. Acoust. Speech Signal Process.* 30:18–24 (1982).

11 P. Davis, *Circulant Matrices*, Wiley, New York, 1979.

12 M. Dryja and O. Widlund, Some domain decomposition algorithms for elliptic problems, in *Proceedings of the Conference on Iterative Methods for Large Linear Systems*, Austin, Texas, October, 1989, Academic.

13 B. Friedlander, M. Morf, T. Kailath, and L. Ljung, New inversion formulas for matrices classified in terms of their distance from Toeplitz matrices, *Linear Algebra Appl.* 27:31–60 (1979).

14 U. Grenander and G. Szegö, *Toeplitz Forms and Their Applications*, 2nd ed., Chelsea, New York, 1984.

15 T. Huckle, Circulant and skew-circulant matrices for solving Toeplitz matrix problems, *SIAM J. Matrix Appl.*, to appear.

16 M. Kac, W. Murdock, and G. Szegö, On the eigenvalues of certain Hermitian forms, *Arch. Rational Mech. Anal.* 2:767–800 (1953).

17 R. King, M. Ahmadi, R. Gorgui-Naguib, A. Kwabwe, and M. Azimi-Sadjadi, *Digital Filtering in One and Two Dimensions: Design and Applications*, Plenum, New York, 1989.

18 T. Ku and C. Kuo, Design and analysis of Toeplitz preconditioners, *IEEE Trans. Acoust. Speech Signal Process.* 40:129–141 (1991).

19 L. Pustylnikov, The algebraic structure of spaces of Toeplitz and Hankel matrices (in Russian), *Dokl. Akad. Nauk USSR* 250(3):556–559 (1980).

20 J. Stoer and R. Bulirsch, *Introduction to Numerical Analysis*, Springer-Verlag, New York, 1980.

21 G. Strang, A proposal for Toeplitz matrix calculations, *Stud. Appl. Math.* 74:171–176 (1986).

22 M. Tismenetsky, A decomposition of Toeplitz matrices and optimal circulant preconditioning, *Linear Algebra Appl.* 154:105–121 (1991).

23 W. Trench, An algorithm for the inversion of finite Toeplitz matrices, *SIAM J. Appl. Math.* 12:515–522 (1964).

24 E. Tyrtyshnikov, Optimal and super-optimal circulant preconditioners, *SIAM J. Matrix Anal. Appl.* 13:459–473 (1992).

25 J. Walker, *Fourier Analysis*, Oxford U.P., New York, 1988.

26 H. Widom, Toeplitz matrices, in *Studies in Real and Complex Analysis*, Math. Assoc. Amer., Vol. 3, 1965, pp. 179–209.

27 J. Wilkinson, *The Algebraic Eigenvalue Problem*, Clarendon, Oxford, 1965.

Received 28 June 1992; final manuscript accepted 14 September 1992