

## CIRCULANT PRECONDITIONERS FOR TOEPLITZ MATRICES WITH PIECEWISE CONTINUOUS GENERATING FUNCTIONS

MAN-CHUNG YEUNG AND RAYMOND H. CHAN

**ABSTRACT.** We consider the solution of  $n$ -by- $n$  Toeplitz systems  $T_n x = b$  by preconditioned conjugate gradient methods. The preconditioner  $C_n$  is the T. Chan circulant preconditioner, which is defined to be the circulant matrix that minimizes  $\|B_n - T_n\|_F$  over all circulant matrices  $B_n$ . For Toeplitz matrices generated by positive  $2\pi$ -periodic continuous functions, we have shown earlier that the spectrum of the preconditioned system  $C_n^{-1}T_n$  is clustered around 1 and hence the convergence rate of the preconditioned system is superlinear. However, in this paper, we show that if instead the generating function is only piecewise continuous, then for all  $\epsilon$  sufficiently small, there are  $O(\log n)$  eigenvalues of  $C_n^{-1}T_n$  that lie outside the interval  $(1 - \epsilon, 1 + \epsilon)$ . In particular, the spectrum of  $C_n^{-1}T_n$  cannot be clustered around 1. Numerical examples are given to verify that the convergence rate of the method is no longer superlinear in general.

### 1. INTRODUCTION

An  $n$ -by- $n$  matrix  $T_n$  is said to be *Toeplitz* if it has constant diagonals, i.e.,  $[T_n]_{j,k} = t_{j-k}$  for all  $0 \leq j, k < n$ . It is said to be *circulant* if we further have  $[T_n]_{j,n-1} = [T_n]_{j+1,0}$  for all  $0 \leq j < n-1$ . In this paper, we consider the convergence rate of the preconditioned conjugate gradient method for solving Toeplitz systems  $T_n x = b$  with circulant matrices as preconditioners. Strang in [13] showed that for such a method the cost per iteration is of  $O(n \log n)$  operations. In contrast, superfast direct Toeplitz solvers require  $O(n \log^2 n)$  operations (see, for instance, Ammar and Gragg [1]). Thus, one has to analyze the convergence rate of the iterative method in order to compare it with direct methods.

To analyze the convergence rate, which is a function of the matrix size  $n$ , we assume that the given Toeplitz matrix  $T_n$  is the  $n$ -by- $n$  principal submatrix of a semi-infinite Toeplitz matrix  $T$ . The function  $f$  which has the diagonals  $\{t_j\}_{j=-\infty}^{\infty}$  of  $T$  as Fourier coefficients is called the *generating function* of the sequence of Toeplitz matrices  $\{T_n\}_{n=1}^{\infty}$ . Chan and Strang [5] proved that if the Strang preconditioner  $S_n$  [13] is used, the method will converge superlinearly

---

Received by the editor August 5, 1992.

1991 *Mathematics Subject Classification.* Primary 65F10, 65F15.

*Key words and phrases.* Toeplitz matrix, circulant matrix, generating function, preconditioned conjugate gradient method, superlinear convergence rate.

Research for the second author supported in part by ONR contract no. N00014-90-J-1695 and DOE grant no. DE-FG03-87ER25037.

whenever  $f$  is a positive function in the Wiener class, i.e., when the sequence  $\{t_j\}_{j=-\infty}^{\infty}$  is absolutely summable. The superlinear result is established by first showing that the spectra of the preconditioned matrices  $S_n^{-1}T_n$  are clustered around 1.

Since then, several other circulant preconditioners have been proposed and analyzed under the same assumption that the  $T_n$  are generated by a fixed function  $f$  (see T. Chan [9], Huckle [10], Ku and Kuo [11], Tismenetsky [14], Trefethen [15], and Tyrtyshnikov [16]). The most notable one is the T. Chan preconditioner  $C_n$  [9], which is defined to be the minimizer  $\|B_n - T_n\|_F$  over all circulant matrices  $B_n$ . Here,  $\|\cdot\|_F$  denotes the Frobenius norm. The preconditioner  $C_n$  has a distinct advantage over  $S_n$  in that  $C_n$  is always positive definite whenever  $T_n$  is (see Tyrtyshnikov [16]). Chan [2] proved that under the Wiener class assumption, the spectra of  $C_n^{-1}T_n$  and  $S_n^{-1}T_n$  will be the same as  $n$  tends to infinity, and hence for sufficiently large  $n$ , the preconditioned system  $C_n^{-1}T_n$  converges at the same rate as the system  $S_n^{-1}T_n$  provided that  $f$  is in the Wiener class.

However, in our recent papers, we have shown that the two preconditioners are fundamentally different. By using Weierstrass's theorem, we showed in [6] that if the underlying generating function  $f$  is a positive  $2\pi$ -periodic *continuous* function, then the T. Chan preconditioned systems  $C_n^{-1}T_n$  have clustered spectrum around 1, and hence the systems converge superlinearly if the conjugate gradient method is employed. But the proof used there does not work for Strang's preconditioner. In [7], we resorted to a stronger form of Weierstrass's theorem, namely the Jackson theorem in approximation theory, and we were able to show that the Strang preconditioned systems  $S_n^{-1}T_n$  have clustered spectrum around 1, and hence converge superlinearly whenever  $f$  is a positive  $2\pi$ -periodic *Lipschitz continuous* function. One explanation of this fundamental difference, though not a formal mathematical proof, is that we can associate the Strang preconditioner  $S_n$  with the Dirichlet kernel whereas the T. Chan preconditioner  $C_n$  can be associated with the Fejér kernel (see Chan and Yeung [8]). It is well known in Fourier analysis that if  $f$  is  $2\pi$ -periodic continuous (or respectively Lipschitz continuous), then the convolution product of  $f$  with the Fejér kernel (or respectively the Dirichlet kernel) will converge to  $f$  uniformly (see, for instance, Walker [18, pp. 59, 79]).

In this paper, we will consider functions  $f$  that are not positive  $2\pi$ -periodic continuous, but only nonnegative *piecewise continuous*. We will show that for these generating functions, the spectra of  $C_n^{-1}T_n$  will no longer cluster around 1. More precisely, we show that for all sufficiently small  $\epsilon > 0$ , the number of eigenvalues of  $C_n^{-1}T_n$  that lie outside  $(1 - \epsilon, 1 + \epsilon)$  will be at least  $O(\log n)$ . If moreover  $f$  is strictly positive, then we can show further that the number of outlying eigenvalues is exactly  $O(\log n)$ . Numerical examples are then given to demonstrate that for the preconditioned systems the number of iterations required for convergence does increase like  $O(\log n)$ , and hence the convergence rate of the method cannot be superlinear in general. In view of the explanation made in the preceding paragraph, it is interesting to note that for piecewise continuous  $f$  its convolution product with the Fejér kernel will no longer converge to  $f$  uniformly.

The outline of the paper is as follows. In §2, we list some of the useful lemmas

that will be used in later sections. In §3, we show that for piecewise continuous generating functions  $f$  the number of outlying eigenvalues of the matrix  $T_n - C_n$  is at least  $O(\log n)$ , and hence the spectra of  $T_n - C_n$  cannot be clustered around zero. Using this result, we prove in §4 that the spectra of  $C_n^{-1}T_n$  cannot be clustered around 1 for any nonnegative piecewise continuous function  $f$ . We then prove in §5 that if  $f$  is strictly positive, then the number of outlying eigenvalues of  $C_n^{-1}T_n$  is exactly  $O(\log n)$ . Numerical results are given in §6 to illustrate how the discontinuities in  $f$  affect the rate of convergence. They show that the convergence rate is no longer superlinear, and in general the number of iterations required for convergence increases at least like  $O(\log n)$  when  $n$  increases. Concluding remarks are given in §7.

2. PRELIMINARY LEMMAS

Let  $\mathcal{L}_{2\pi}$  be the space of all  $2\pi$ -periodic Lebesgue integrable real-valued functions defined on the real line  $\mathbb{R}$ . For  $f \in \mathcal{L}_{2\pi}$ , its Fourier coefficients are defined as

$$t_k[f] = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\theta)e^{-ik\theta} d\theta, \quad k = 0, \pm 1, \pm 2, \dots$$

Let  $\mathcal{T}_n[f]$  be the  $n$ -by- $n$  Toeplitz matrix with the  $(j, k)$ th entry given by  $t_{j-k}[f]$ ,  $0 \leq j, k < n$ , and  $\mathcal{E}_n[f]$  be the  $n$ -by- $n$  circulant matrix that minimizes  $\|C_n - \mathcal{T}_n[f]\|_F$  over all  $n$ -by- $n$  circulant matrices  $C_n$ . The matrix  $\mathcal{E}_n[f]$  is called the *T. Chan circulant preconditioner* and its  $(j, l)$ th entry is given by  $c_{j-l}[f]$ , where

$$c_k[f] = \begin{cases} \frac{(n-k)t_k[f] + kt_{k-n}[f]}{n}, & 0 \leq k < n, \\ c_{n+k}[f], & 0 < -k < n \end{cases}$$

(see T. Chan [9]). In this paper, we will consider the spectrum of  $\mathcal{E}_n^{-1}[f]\mathcal{T}_n[f]$  as  $n$  goes to infinity for piecewise continuous functions  $f \in \mathcal{L}_{2\pi}$ . Since  $f$  is real-valued,  $t_{-k}[f] = \bar{t}_k[f]$  and hence  $\mathcal{T}_n[f]$  and  $\mathcal{E}_n[f]$  are Hermitian matrices for all  $n$ . For  $f \in \mathcal{L}_{2\pi}$ , let  $f_{\max}$  and  $f_{\min}$  be its essential supremum and infimum, respectively.

**Lemma 1.** *Let  $f \in \mathcal{L}_{2\pi}$  with  $f_{\max} \neq f_{\min}$ . Then for all  $n \geq 1$ ,*

$$f_{\min} < \lambda_{\min}(\mathcal{T}_n[f]) \leq \lambda_{\min}(\mathcal{E}_n[f]) \leq \lambda_{\max}(\mathcal{E}_n[f]) \leq \lambda_{\max}(\mathcal{T}_n[f]) < f_{\max},$$

where  $\lambda_{\max}$  and  $\lambda_{\min}$  denote the maximum and minimum eigenvalues, respectively.

*Proof.* For the two strict inequalities, see Chan [3, Lemma 1]. For the other inner inequalities, see Tyrtyshnikov [16, Theorem 3.1].  $\square$

Notice that if  $f_{\max} = f_{\min}$ , then  $\mathcal{T}_n[f] = \mathcal{E}_n[f] = f_{\min} \cdot I_n$ , where  $I_n$  is the  $n$ -by- $n$  identity matrix. Thus, in the following, we assume for simplicity that  $f$  is nonconstant.

Given a Hermitian matrix  $A$ , we use  $N(\epsilon; A)$  to denote the number of eigenvalues of  $A$  with absolute values exceeding  $\epsilon$ . A sequence of Hermitian matrices  $\{A_n\}_{n=1,2,\dots}$  is said to have *clustered spectra around  $\alpha$*  if for any  $\epsilon > 0$ , there exists a  $c > 0$  such that for all  $n \geq 1$ ,  $N(\epsilon; A_n - \alpha I_n) \leq c$ . If  $\alpha = 0$ , we simply say  $\{A_n\}_{n=1,2,\dots}$  has *clustered spectra*.

**Lemma 2.** Let  $A_n$  and  $B_n$  be  $n$ -by- $n$  Hermitian matrices and  $\lambda$  and  $\mu$  be any positive numbers. Then

- (i)  $N(\lambda; \pm\mu A_n) = N(\lambda/\mu; A_n)$ ,  
(ii)  $N(\lambda + \mu, A_n + B_n) \leq N(\lambda, A_n) + N(\mu, B_n)$ .

*Proof.* (i) is trivial and (ii) can be proved by Cauchy's interlacing theorem (see Wilkinson [20, p. 103] or Widom [19, p. 11]).  $\square$

It follows immediately from Lemma 2 that if  $\{A_n\}$  and  $\{B_n\}$  are two sequences of Hermitian matrices with clustered spectra, then  $\{\alpha A_n + \beta B_n\}$  also has clustered spectra for any real numbers  $\alpha$  and  $\beta$ .

**Lemma 3.** Let  $\{A_n\}_{n=1,2,\dots}$  be a sequence of Hermitian matrices. If  $\sup_n \|A_n\|_F < \infty$ , then  $\{A_n\}$  has clustered spectra.

*Proof.* Since  $\|A_n\|_F^2$  is equal to the sum of the square of the eigenvalues of  $A_n$ , it follows that for any given  $\epsilon > 0$ ,  $N(\epsilon; A_n) \leq \sup_n \|A_n\|_F^2 / \epsilon^2$ .  $\square$

**Lemma 4** (Chan and Yeung [6, Theorem 1]). Let  $f \in \mathcal{L}_{2\pi}$  be continuous. Then the sequence of matrices

$$\Delta_n[f] \equiv \mathcal{F}_n[f] - \mathcal{E}_n[f], \quad n = 1, 2, \dots,$$

has clustered spectra.

**Lemma 5** (Widom [19, p. 30]). Let  $H_n$  be the  $n$ -by- $n$  Hilbert matrix

$$H_n = \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{3} & \cdots & \frac{1}{n} \\ \frac{1}{2} & \frac{1}{3} & \cdots & \frac{1}{n+1} \\ \frac{1}{3} & \cdots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ \frac{1}{n} & \frac{1}{n+1} & \cdots & \cdots & \frac{1}{2n-1} \end{bmatrix}.$$

Then for any  $0 < \epsilon < \pi$ , we have

$$N(\epsilon, H_n) = \frac{2}{\pi} \log n \cdot \operatorname{sech}^{-1} \frac{\epsilon}{\pi} \cdot (1 + o(1)),$$

where  $o(1)$  tends to zero as  $n$  increases.

**Lemma 6.** Let  $f \in \mathcal{L}_{2\pi}$  be bounded. Define  $\mathcal{H}_n[f]$  to be the  $n$ -by- $n$  Hankel matrix with entries given by

$$[\mathcal{H}_n[f]]_{j,k} = (t_{j+k}[f]), \quad j, k = 0, 1, \dots, n-1.$$

Then  $\|\mathcal{H}_n[f]\|_2 \leq \|f\|_\infty$ .

*Proof.* By Nehari's theorem [12, Theorem 1], the infinite Hankel matrix  $\mathcal{H}[f]$  satisfies

$$\|\mathcal{H}[f]\|_2^2 = \max_{\|x\|_2=1} \{x^* \mathcal{H}[f]^* \mathcal{H}[f] x\} \leq \|f\|_\infty^2.$$

Hence, for any  $n$ -vector  $y$  with  $\|y\|_2 = 1$ , we have

$$\|f\|_\infty^2 \geq (y^*, 0) \mathcal{H}[f]^* \mathcal{H}[f] \begin{pmatrix} y \\ 0 \end{pmatrix} \geq y^* \mathcal{H}_n^*[f] \mathcal{H}_n[f] y.$$

In particular,  $\|\mathcal{H}_n[f]\|_2 \leq \|f\|_\infty$ .  $\square$

### 3. SPECTRA OF $\Delta_n[f]$

In this section, we prove that if  $f \in \mathcal{L}_{2\pi}$  is piecewise continuous, the spectrum of  $\Delta_n[f] = \mathcal{T}_n[f] - \mathcal{E}_n[f]$  cannot be clustered around zero. More precisely, we show that  $N(\epsilon; \Delta_n[f]) \geq O(\log n)$ . For simplicity, we will present the proof for the case  $n = 2m$ . When  $n$  is odd, the proof can be modified accordingly.

Before we start, let us give a brief motivation of our proof. Suppose we have an  $f \in \mathcal{L}_{2\pi}$  which has only one jump discontinuity at  $\xi \in (-\pi, \pi]$ . Then by adding multiples of the function  $g(\theta)$  defined in Lemma 8 below, the sum of the functions will be a  $2\pi$ -periodic continuous function. In view of Lemmas 2 and 4, we then only have to consider  $N(\epsilon; \Delta_n[g])$ . In Lemma 8, we will show that the spectrum of  $\Delta_n[g]$  is basically the same as the spectrum of the Hilbert matrix  $H_n$  with only small norm perturbation. Hence, by Lemma 5, we get the result. The proof below, however, will be more complicated because we need to show further that if  $f$  has multiple jumps, then the outlying eigenvalues derived from one jump will not be canceled by the outlying eigenvalues from the other jumps and thus leave us with a clustered spectrum.

Let  $f \in \mathcal{L}_{2\pi}$  be a piecewise continuous function with points of discontinuity in  $(-\pi, \pi]$  at  $-\pi < \theta_1 < \dots < \theta_\nu \leq \pi$  and jumps

$$\alpha_k = \lim_{\theta \rightarrow \theta_k^+} f(\theta) - \lim_{\theta \rightarrow \theta_k^-} f(\theta), \quad k = 1, \dots, \nu.$$

Let the biggest jump be at  $\theta_{k_0}$ , i.e.,

$$|\alpha_{k_0}| = \max_{1 \leq k \leq \nu} |\alpha_k|.$$

Insert  $\nu$  arbitrary points  $\phi_1, \phi_2, \dots, \phi_\nu$  into  $\{\theta_1, \theta_2, \dots, \theta_\nu\}$  such that

$$-\pi < \phi_1 < \theta_1 < \phi_2 < \theta_2 < \dots < \phi_\nu < \theta_\nu \leq \pi.$$

Define the functions

$$g_0(\theta) = \begin{cases} \theta + \pi - \theta_{k_0}, & -\pi < \theta \leq \theta_{k_0}, \\ \theta - \pi - \theta_{k_0}, & \theta_{k_0} < \theta \leq \pi, \end{cases}$$

$$g_k(\theta) = \begin{cases} 0, & -\pi < \theta \leq \phi_k, \\ \frac{\theta - \phi_k}{2(\theta_k - \phi_k)}, & \phi_k < \theta \leq \theta_k, \\ \frac{\theta - \phi_{k+1}}{2(\phi_{k+1} - \theta_k)}, & \theta_k < \theta \leq \phi_{k+1}, \\ 0, & \phi_{k+1} < \theta \leq \pi, \end{cases}$$

for  $k = 1, 2, \dots, \nu - 1$ , and

$$g_\nu(\theta) = \begin{cases} 0, & -\pi < \theta \leq \phi_\nu, \\ \frac{\theta - \phi_\nu}{2(\theta_\nu - \phi_\nu)}, & \phi_\nu < \theta \leq \theta_\nu, \\ \frac{\theta - \pi}{2(\pi - \theta_\nu)}, & \theta_\nu < \theta \leq \pi, \end{cases}$$

if  $\theta_\nu < \pi$  or

$$g_\nu(\theta) = \begin{cases} \frac{\theta - \phi_1}{2(\phi_1 + \pi)}, & -\pi < \theta \leq \phi_1, \\ 0, & \phi_1 < \theta \leq \phi_\nu, \\ \frac{\theta - \phi_\nu}{2(\pi - \phi_\nu)}, & \phi_\nu < \theta \leq \pi, \end{cases}$$

if  $\theta_\nu = \pi$ . All functions  $g_k(\theta)$ ,  $k = 0, 1, \dots, \nu$ , are to be extended into functions in  $\mathcal{L}_{2\pi}$ .

Now we write  $f$  as

$$f = f + \frac{\alpha_{k_0}}{\pi} g_0 + \sum_{k=1}^{\nu} \alpha_k \delta_k g_k - \frac{\alpha_{k_0}}{\pi} g_0 - \sum_{k=1}^{\nu} \alpha_k \delta_k g_k,$$

where

$$\delta_k = \begin{cases} -1, & k = k_0, \\ 1, & k \neq k_0. \end{cases}$$

Then we have

$$(1) \Delta_{2m}[f] = \Delta_{2m} \left[ f + \frac{\alpha_{k_0}}{\pi} g_0 + \sum_{k=1}^{\nu} \alpha_k \delta_k g_k \right] - \frac{\alpha_{k_0}}{\pi} \Delta_{2m}[g_0] - \Delta_{2m} \left[ \sum_{k=1}^{\nu} \alpha_k \delta_k g_k \right].$$

In the next three lemmas, we consider the limiting behavior, as  $m$  tends to infinity, of the eigenvalues of the three terms in the right-hand side of (1), respectively.

**Lemma 7.** *The sequence of matrices*

$$\left\{ \Delta_{2m} \left[ f + \frac{\alpha_{k_0}}{\pi} g_0 + \sum_{k=1}^{\nu} \alpha_k \delta_k g_k \right] \right\}_{m=1, 2, \dots}$$

*has clustered spectra.*

*Proof.* By Lemma 4, it suffices to show that the function

$$f + \frac{\alpha_{k_0}}{\pi} g_0 + \sum_{k=1}^{\nu} \alpha_k \delta_k g_k$$

is a  $2\pi$ -periodic continuous function. However, from the definitions of  $g_k$ ,  $k = 0, 1, \dots, \nu$ , it is clear that the function is already  $2\pi$ -periodic and that the points  $\theta_j$ ,  $j = 1, 2, \dots, \nu$ , are its only possible points of discontinuity in

$(-\pi, \pi]$ . However, for  $\theta_j \neq \theta_{k_0}$ , we have

$$\begin{aligned} & \lim_{\theta \rightarrow \theta_j^-} \left[ f(\theta) + \frac{\alpha_{k_0}}{\pi} g_0(\theta) + \sum_{k=1}^{\nu} \alpha_k \delta_k g_k(\theta) \right] \\ &= \lim_{\theta \rightarrow \theta_j^-} \left[ f(\theta) + \frac{\alpha_{k_0}}{\pi} g_0(\theta) + \sum_{\substack{k=1 \\ k \neq j}}^{\nu} \alpha_k \delta_k g_k(\theta) + \alpha_j \delta_j g_j(\theta) \right] \\ &= \lim_{\theta \rightarrow \theta_j^-} f(\theta) + \frac{\alpha_{k_0}}{\pi} g_0(\theta_j) + \sum_{\substack{k=1 \\ k \neq j}}^{\nu} \alpha_k \delta_k \cdot 0 + \frac{1}{2} \alpha_j \delta_j \\ &= \lim_{\theta \rightarrow \theta_j^+} f(\theta) + \frac{\alpha_{k_0}}{\pi} g_0(\theta_j) + \sum_{\substack{k=1 \\ k \neq j}}^{\nu} \alpha_k \delta_k \cdot 0 - \frac{1}{2} \alpha_j \delta_j \\ &= \lim_{\theta \rightarrow \theta_j^+} \left[ f(\theta) + \frac{\alpha_{k_0}}{\pi} g_0(\theta) + \sum_{\substack{k=1 \\ k \neq j}}^{\nu} \alpha_k \delta_k g_k(\theta) + \alpha_j \delta_j g_j(\theta) \right] \\ &= \lim_{\theta \rightarrow \theta_j^+} \left[ f(\theta) + \frac{\alpha_{k_0}}{\pi} g_0(\theta) + \sum_{k=1}^{\nu} \alpha_k \delta_k g_k(\theta) \right]. \end{aligned}$$

At  $\theta_{k_0}$ , we have

$$\begin{aligned} & \lim_{\theta \rightarrow \theta_{k_0}^-} \left[ f(\theta) + \frac{\alpha_{k_0}}{\pi} g_0(\theta) + \sum_{k=1}^{\nu} \alpha_k \delta_k g_k(\theta) \right] \\ &= \lim_{\theta \rightarrow \theta_{k_0}^-} \left[ f(\theta) + \frac{\alpha_{k_0}}{\pi} g_0(\theta) - \alpha_{k_0} g_{k_0}(\theta) + \sum_{\substack{k=1 \\ k \neq k_0}}^{\nu} \alpha_k \delta_k g_k(\theta) \right] \\ &= \lim_{\theta \rightarrow \theta_{k_0}^-} f(\theta) + \frac{\alpha_{k_0}}{\pi} \cdot \pi - \frac{1}{2} \alpha_{k_0} + \sum_{\substack{k=1 \\ k \neq k_0}}^{\nu} \alpha_k \delta_k \cdot 0 \\ &= \lim_{\theta \rightarrow \theta_{k_0}^+} f(\theta) + \frac{\alpha_{k_0}}{\pi} (-\pi) - \left(-\frac{1}{2}\right) \alpha_{k_0} + \sum_{\substack{k=1 \\ k \neq k_0}}^{\nu} \alpha_k \delta_k \cdot 0 \\ &= \lim_{\theta \rightarrow \theta_{k_0}^+} \left[ f(\theta) + \frac{\alpha_{k_0}}{\pi} g_0(\theta) - \alpha_{k_0} g_{k_0}(\theta) + \sum_{\substack{k=1 \\ k \neq k_0}}^{\nu} \alpha_k \delta_k g_k(\theta) \right] \\ &= \lim_{\theta \rightarrow \theta_{k_0}^+} \left[ f(\theta) + \frac{\alpha_{k_0}}{\pi} g_0(\theta) + \sum_{k=1}^{\nu} \alpha_k \delta_k g_k(\theta) \right]. \quad \square \end{aligned}$$

**Lemma 8.** Let  $\xi$  be an arbitrary point in  $(-\pi, \pi]$ . Let  $g \in \mathcal{L}_{2\pi}$  be defined by

$$g(\theta) = \begin{cases} \theta + \pi - \xi, & -\pi < \theta \leq \xi, \\ \theta - \pi - \xi, & \xi < \theta \leq \pi. \end{cases}$$

Then

$$(2) \quad \Delta_{2m}[g] = A_{2m} + B_{2m},$$

where  $A_{2m}$  and  $B_{2m}$  are both Hermitian matrices with

$$(3) \quad N(\epsilon; A_{2m}) = \frac{4}{\pi} \log m \cdot \operatorname{sech}^{-1} \frac{\epsilon}{\pi} \cdot (1 + o(1))$$

for any  $0 < \epsilon < \pi$ , and

$$(4) \quad \sup_m \|B_{2m}\|_F < 2 + 2\sqrt{\ln 2} < \infty.$$

*Proof.* The Fourier coefficients  $t_k[g]$  of  $g$  are given by

$$t_k[g] = \frac{1}{2\pi} \int_{-\pi}^{\pi} g(\theta) e^{-ik\theta} d\theta = \begin{cases} 0, & k = 0, \\ \frac{i}{k} e^{-ik\xi}, & k = \pm 1, \pm 2, \dots \end{cases}$$

Thus, the first row of  $\Delta_{2m}[g]$  is given by

$$\begin{aligned} & \left( 0, \frac{1}{2m}(t_{-1}[g] - t_{2m-1}[g]), \dots, \frac{j}{2m}(t_{-j}[g] - t_{2m-j}[g]), \right. \\ & \qquad \qquad \qquad \left. \dots, \frac{2m-1}{2m}(t_{-2m+1}[g] - t_1[g]) \right) \\ & = \left( 0, -\frac{ie^{-i(2m-1)\xi}}{2m-1}, \dots, -\frac{ie^{-i(2m-j)\xi}}{2m-j}, \dots, -ie^{-i\xi} \right) \\ & \quad + \left( 0, \frac{i(e^{-i(2m-1)\xi} - e^{i\xi})}{2m}, \dots, \frac{i(e^{-i(2m-j)\xi} - e^{ij\xi})}{2m}, \dots, \frac{i(e^{-i\xi} - e^{i(2m-1)\xi})}{2m} \right). \end{aligned}$$

Let  $\tilde{A}_{2m}$  and  $\tilde{B}_{2m}$  be the  $2m$ -by- $2m$  Hermitian Toeplitz matrices with their first rows given by

$$(5) \quad \left( 0, -\frac{ie^{-i(2m-1)\xi}}{2m-1}, \dots, -\frac{ie^{-i(2m-j)\xi}}{2m-j}, \dots, -ie^{-i\xi} \right)$$

and

$$(6) \quad \left( 0, \frac{i(e^{-i(2m-1)\xi} - e^{i\xi})}{2m}, \dots, \frac{i(e^{-i(2m-j)\xi} - e^{ij\xi})}{2m}, \dots, \frac{i(e^{-i\xi} - e^{i(2m-1)\xi})}{2m} \right),$$

respectively. Then we have  $\Delta_{2m}[g] = \tilde{A}_{2m} + \tilde{B}_{2m}$ . From (6), we have

$$(7) \quad \begin{aligned} \|\tilde{B}_{2m}\|_F^2 &= 2 \sum_{j=1}^{2m-1} (2m-j) \left| \frac{i}{2m} (e^{-i(2m-j)\xi} - e^{ij\xi}) \right|^2 \\ &\leq 2 \sum_{j=1}^{2m-1} \frac{2m-j}{m^2} = \frac{2(2m-1)}{m} < 4. \end{aligned}$$

We next partition  $\tilde{A}_{2m}$  as

$$\tilde{A}_{2m} = \begin{bmatrix} 0 & U_m \\ U_m^* & 0 \end{bmatrix} + \begin{bmatrix} V_m & 0 \\ 0 & V_m \end{bmatrix}.$$

By (5), we see that  $V_m$  is a Hermitian Toeplitz matrix with its first row given by

$$\left( 0, -\frac{ie^{-i(2m-1)\xi}}{2m-1}, \dots, -\frac{ie^{-i(2m-j)\xi}}{2m-j}, \dots, -\frac{ie^{-i(m+1)\xi}}{m+1} \right).$$

Hence,

$$\begin{aligned} \|V_m\|_F^2 &= \sum_{j=1}^{m-1} (m-j) \left[ \left| -\frac{ie^{-i(2m-j)\xi}}{2m-j} \right|^2 + \left| \frac{ie^{i(2m-j)\xi}}{2m-j} \right|^2 \right] \\ (8) \quad &= 2 \sum_{j=1}^{m-1} \frac{m-j}{(2m-j)^2} \leq 2 \sum_{j=1}^{m-1} \frac{1}{2m-j} < 2 \int_m^{2m-1} \frac{1}{x} dx < 2 \ln 2. \end{aligned}$$

Thus, if we define

$$B_{2m} = \tilde{B}_{2m} + \begin{bmatrix} V_m & 0 \\ 0 & V_m \end{bmatrix},$$

then, by (7) and (8), we have

$$\|B_{2m}\|_F \leq \|\tilde{B}_{2m}\|_F + \sqrt{2}\|V_m\|_F < 2 + 2\sqrt{\ln 2}.$$

It remains to show that the matrix

$$A_{2m} \equiv \Delta_{2m}[g] - B_{2m} = \tilde{A}_{2m} + \tilde{B}_{2m} - B_{2m} = \tilde{A}_{2m} - \begin{bmatrix} V_m & 0 \\ 0 & V_m \end{bmatrix} = \begin{bmatrix} 0 & U_m \\ U_m^* & 0 \end{bmatrix}$$

satisfies (3). To prove that, we first define

$$J_m \equiv \begin{bmatrix} 0 & & & 1 \\ & \ddots & & \\ & & 1 & \\ 1 & & & 0 \end{bmatrix},$$

$$P_m \equiv \text{diag}(1, e^{i\xi}, \dots, e^{i(m-2)\xi}, e^{i(m-1)\xi}),$$

and

$$Q_m \equiv \text{diag}(-ie^{-im\xi}, -ie^{-i(m-1)\xi}, \dots, -ie^{-2i\xi}, -ie^{-i\xi}).$$

It is straightforward to check that  $U_m = P_m^* H_m J_m Q_m$ , where  $H_m$  is the Hilbert matrix defined in Lemma 5. Hence

$$\begin{aligned} A_{2m} &= \begin{bmatrix} 0 & U_m \\ U_m^* & 0 \end{bmatrix} = \begin{bmatrix} P_m^* & 0 \\ 0 & Q_m^* \end{bmatrix} \begin{bmatrix} 0 & H_m J_m \\ J_m H_m & 0 \end{bmatrix} \begin{bmatrix} P_m & 0 \\ 0 & Q_m \end{bmatrix} \\ &= \frac{1}{2} \begin{bmatrix} P_m^* & 0 \\ 0 & Q_m^* \end{bmatrix} \begin{bmatrix} I_m & I_m \\ J_m & -J_m \end{bmatrix} \begin{bmatrix} H_m & 0 \\ 0 & -H_m \end{bmatrix} \begin{bmatrix} I_m & J_m \\ I_m & -J_m \end{bmatrix} \begin{bmatrix} P_m & 0 \\ 0 & Q_m \end{bmatrix} \\ &= \frac{1}{2} \begin{bmatrix} P_m^* & P_m^* J_m \\ Q_m^* J_m & -Q_m^* J_m \end{bmatrix} \begin{bmatrix} H_m & 0 \\ 0 & -H_m \end{bmatrix} \begin{bmatrix} P_m & J_m Q_m \\ P_m & -J_m Q_m \end{bmatrix}. \end{aligned}$$

Since

$$\frac{1}{\sqrt{2}} \begin{bmatrix} P_m & J_m Q_m \\ P_m & -J_m Q_m \end{bmatrix}$$

is an orthogonal matrix for all  $\xi$ , we have that  $A_{2m}$  is orthogonally similar to

$$\begin{bmatrix} H_m & 0 \\ 0 & -H_m \end{bmatrix}.$$

By Lemma 5, assertion (3) follows.  $\square$

**Lemma 9.** *The matrix  $\Delta_{2m}[\sum_{k=1}^{\nu} \alpha_k \delta_k g_k]$  can be written as*

$$(9) \quad \Delta_{2m} \left[ \sum_{k=1}^{\nu} \alpha_k \delta_k g_k \right] = D_{2m} + E_{2m},$$

where  $D_{2m}$  and  $E_{2m}$  are Hermitian matrices with

$$(10) \quad N\left(\frac{|\alpha_{k_0}|}{2}; D_{2m}\right) = 0$$

and

$$(11) \quad \sup_m \|E_{2m}\|_F \leq c < \infty$$

for some  $c$  independent of  $m$ .

*Proof.* For simplicity, let us write

$$h = \sum_{k=1}^{\nu} \alpha_k \delta_k g_k.$$

Define  $W_m$  to be the  $m$ -by- $m$  Toeplitz matrix

$$(12) \quad W_m = \begin{bmatrix} t_m[h] & t_{m-1}[h] & \dots & t_1[h] \\ t_{m+1}[h] & \ddots & & t_2[h] \\ \vdots & & & \\ t_{2m-1}[h] & t_{2m-2}[h] & & t_m[h] \end{bmatrix}.$$

It is clear that the entries of the Hankel matrix  $W_m J_m$  are just Fourier coefficients of the function  $h(\theta)e^{-i\theta}$ . Therefore, by Lemma 6, we have

$$\|W_m J_m\|_2 \leq \sup_{\theta} |h(\theta)e^{-i\theta}| = \|h\|_{\infty},$$

where by the definitions of  $h$  and  $g_k$ ,  $1 \leq k \leq \nu$ ,

$$\|h\|_{\infty} = |h(\theta_{k_0})| = \frac{|\alpha_{k_0}|}{2}.$$

Hence, if we let

$$D_{2m} = \begin{bmatrix} 0 & -W_m \\ -W_m^* & 0 \end{bmatrix},$$

then we have

$$\|D_{2m}\|_2 = \|W_m\|_2 = \|W_m J_m J_m\|_2 \leq \|W_m J_m\|_2 \|J_m\|_2 \leq \frac{|\alpha_{k_0}|}{2}.$$

Thus,  $N(|\alpha_{k_0}|/2; D_{2m}) = 0$ .

It remains to show that  $E_{2m} \equiv \Delta_{2m}[h] - D_{2m}$  satisfies (11). To estimate  $\|E_{2m}\|_F$ , we partition the Hermitian Toeplitz matrix  $\Delta_{2m}[h]$  as

$$\Delta_{2m}[h] = \begin{bmatrix} X_m & Y_m \\ Y_m^* & X_m \end{bmatrix}.$$

Clearly,  $X_m$  is an  $m$ -by- $m$  Hermitian Toeplitz matrix with its first row given by

$$(13) \quad \left( 0, \frac{1}{2m}(t_{-1}[h] - t_{2m-1}[h]), \frac{2}{2m}(t_{-2}[h] - t_{2m-2}[h]), \dots, \frac{m-1}{2m}(t_{-m+1}[h] - t_{m+1}[h]) \right)$$

and  $Y_m$  is given by the  $m$ -by- $m$  Toeplitz matrix

$$(14) \quad \begin{bmatrix} \frac{m}{2m}(t_{-m}[h] - t_m[h]) & \frac{m+1}{2m}(t_{-m-1}[h] - t_{m-1}[h]) & \dots & \frac{2m-1}{2m}(t_{-2m+1}[h] - t_1[h]) \\ \frac{m-1}{2m}(t_{-m+1}[h] - t_{m+1}[h]) & \ddots & & \vdots \\ \vdots & & & \\ \frac{1}{2m}(t_{-1}[h] - t_{2m-1}[h]) & \frac{2}{2m}(t_{-2}[h] - t_{2m-2}[h]) & \dots & \frac{m}{2m}(t_{-m}[h] - t_m[h]) \end{bmatrix}.$$

Therefore,

$$E_{2m} = \begin{bmatrix} X_m & W_m + Y_m \\ W_m^* + Y_m^* & X_m \end{bmatrix},$$

and hence

$$(15) \quad \|E_{2m}\|_F^2 = 2\|X_m\|_F^2 + 2\|W_m + Y_m\|_F^2.$$

By direct computation, the Fourier coefficients  $t_j[h]$  of  $h$  are given by

$$t_j[h] = \frac{1}{4\pi j} \sum_{k=1}^{\nu-1} \alpha_k \delta_k \left\{ 2ie^{-ij\theta_k} + \frac{e^{-ij\theta_k} - e^{-ij\phi_k}}{(\theta_k - \phi_k)j} + \frac{e^{-ij\phi_{k+1}} - e^{-ij\theta_k}}{(\phi_{k+1} - \theta_k)j} \right\} + \frac{\alpha_\nu \delta_\nu}{4\pi j} \left\{ 2ie^{-ij\theta_\nu} + \frac{e^{-ij\theta_\nu} - e^{-ij\phi_\nu}}{(\theta_\nu - \phi_\nu)j} + \frac{e^{-ij\pi} - e^{-ij\theta_\nu}}{(\pi - \theta_\nu)j} \right\},$$

$j = \pm 1, \pm 2, \dots$ , when  $\theta_\nu < \pi$ . If  $\theta_\nu = \pi$ , then

$$t_j[h] = \frac{1}{4\pi j} \sum_{k=1}^{\nu-1} \alpha_k \delta_k \left\{ 2ie^{-ij\theta_k} + \frac{e^{-ij\theta_k} - e^{-ij\phi_k}}{(\theta_k - \phi_k)j} + \frac{e^{-ij\phi_{k+1}} - e^{-ij\theta_k}}{(\phi_{k+1} - \theta_k)j} \right\} + \frac{\alpha_\nu \delta_\nu}{4\pi j} \left\{ ie^{ij\pi} + ie^{-ij\pi} + \frac{e^{-ij\pi} - e^{-ij\phi_\nu}}{(\pi - \phi_\nu)j} + \frac{e^{-ij\phi_1} - e^{ij\pi}}{(\phi_1 + \pi)j} \right\},$$

$j = \pm 1, \pm 2, \dots$ . In either case, there exists a constant  $c$  such that

$$|t_j[h]| \leq \frac{c}{|j|}, \quad j = \pm 1, \pm 2, \dots$$

Hence, by (13),

$$(16) \quad \begin{aligned} \|X_m\|_F^2 &= 2 \sum_{j=1}^{m-1} (m-j) \left| \frac{j}{2m}(t_{-j}[h] - t_{2m-j}[h]) \right|^2 \\ &\leq \frac{1}{2m^2} \sum_{j=1}^{m-1} (m-j) j^2 \left( \frac{c}{j} + \frac{c}{2m-j} \right)^2 \\ &= 2c^2 \sum_{j=1}^{m-1} \frac{m-j}{(2m-j)^2} = 2c^2 \sum_{j=1}^{m-1} \frac{j}{(m+j)^2} \leq \frac{2c^2}{m^2} \sum_{j=1}^{m-1} j < c^2. \end{aligned}$$

Moreover, by (12) and (14),

$$\begin{aligned}
\|W_m + Y_m\|_F^2 &= \sum_{j=0}^{m-1} (m-j) \left| t_{m-j}[h] + \frac{m+j}{2m} (t_{-m-j}[h] - t_{m-j}[h]) \right|^2 \\
&\quad + \sum_{j=1}^{m-1} (m-j) \left| t_{m+j}[h] + \frac{m-j}{2m} (t_{-m+j}[h] - t_{m+j}[h]) \right|^2 \\
&= \sum_{j=0}^{m-1} (m-j) \left| \frac{m+j}{2m} t_{-m-j}[h] + \frac{m-j}{2m} t_{m-j}[h] \right|^2 \\
&\quad + \sum_{j=1}^{m-1} (m-j) \left| \frac{m-j}{2m} t_{-m+j}[h] + \frac{m+j}{2m} t_{m+j}[h] \right|^2 \\
&\leq \frac{1}{4m^2} \sum_{j=0}^{m-1} (m-j) \{ (m+j) |t_{-m-j}[h]| + (m-j) |t_{m-j}[h]| \}^2 \\
&\quad + \frac{1}{4m^2} \sum_{j=1}^{m-1} (m-j) \{ (m-j) |t_{-m+j}[h]| + (m+j) |t_{m+j}[h]| \}^2 \\
&\leq \frac{c^2}{m^2} \sum_{j=0}^{m-1} (m-j) + \frac{c^2}{m^2} \sum_{j=1}^{m-1} (m-j) = c^2.
\end{aligned}$$

Putting this and (16) back into (15), we have  $\|E_{2m}\|_F < 2c$ .  $\square$

We now combine Lemmas 7–9 to show that the spectra of  $\Delta_{2m}[f]$  cannot be clustered.

**Theorem 1.** *Let  $f \in \mathcal{L}_{2\pi}$  be piecewise continuous with points of discontinuity in  $(-\pi, \pi]$  at  $-\pi < \theta_1 < \dots < \theta_\nu \leq \pi$  and jumps*

$$\alpha_k \equiv \lim_{\theta \rightarrow \theta_k^+} f(\theta) - \lim_{\theta \rightarrow \theta_k^-} f(\theta), \quad k = 1, \dots, \nu.$$

*Define  $|\alpha_{k_0}| = \max_{1 \leq k \leq \nu} |\alpha_k|$ . Then for any  $0 < \epsilon < |\alpha_{k_0}|/4$ , there exists a constant  $b$ , independent of  $m$ , such that*

$$N(\epsilon; \Delta_{2m}[f]) \geq \frac{4}{\pi} (1 + o(1)) \log m \cdot \operatorname{sech}^{-1} \left( \frac{1}{2} + \frac{2\epsilon}{|\alpha_{k_0}|} \right) - b,$$

where  $o(1)$  tends to zero as  $m$  increases.

*Proof.* Putting (2) and (9) into (1), we find

$$\begin{aligned}
\frac{\alpha_{k_0}}{\pi} A_{2m} &= \left\{ \Delta_{2m} \left[ f + \frac{\alpha_{k_0}}{\pi} g_0 + \sum_{k=1}^{\nu} \alpha_k \delta_k g_k \right] - \frac{\alpha_{k_0}}{\pi} B_{2m} - E_{2m} \right\} - D_{2m} - \Delta_{2m}[f] \\
&= G_{2m} - D_{2m} - \Delta_{2m}[f],
\end{aligned}$$

where

$$G_{2m} \equiv \Delta_{2m} \left[ f + \frac{\alpha_{k_0}}{\pi} g_0 + \sum_{k=1}^{\nu} \alpha_k \delta_k g_k \right] - \frac{\alpha_{k_0}}{\pi} B_{2m} - E_{2m}.$$

We note that by (4), (11), and Lemmas 7 and 3, the sequence of matrices  $\{G_{2m}\}$  has clustered spectra. Moreover, by Lemma 2 and (10),

$$\begin{aligned} N\left(\frac{|\alpha_{k_0}|}{2} + 2\epsilon; \frac{\alpha_{k_0}}{\pi} A_{2m}\right) &\leq N(\epsilon; G_{2m}) + N\left(\frac{|\alpha_{k_0}|}{2}; -D_{2m}\right) + N(\epsilon; -\Delta_{2m}[f]) \\ &= N(\epsilon; G_{2m}) + N\left(\frac{|\alpha_{k_0}|}{2}; D_{2m}\right) + N(\epsilon; \Delta_{2m}[f]) \\ &= N(\epsilon; G_{2m}) + N(\epsilon; \Delta_{2m}[f]) \end{aligned}$$

for all  $\epsilon > 0$ . Thus, by (3),

$$\begin{aligned} N(\epsilon; \Delta_{2m}[f]) &\geq N\left(\frac{|\alpha_{k_0}|}{2} + 2\epsilon; \frac{\alpha_{k_0}}{\pi} A_{2m}\right) - N(\epsilon; G_{2m}) \\ &= N\left(\frac{|\alpha_{k_0}| + 4\epsilon}{2|\alpha_{k_0}|} \pi; A_{2m}\right) - N(\epsilon; G_{2m}) \\ &= \frac{4}{\pi}(1 + o(1)) \log m \cdot \operatorname{sech}^{-1}\left(\frac{1}{2} + \frac{2\epsilon}{|\alpha_{k_0}|}\right) - N(\epsilon; G_{2m}) \end{aligned}$$

for all  $0 < \epsilon < |\alpha_{k_0}|/4$ . Finally, since  $\{G_{2m}\}$  has clustered spectra, it follows that for any  $0 < \epsilon < |\alpha_{k_0}|/4$ , there exists a constant  $b$  such that  $N(\epsilon; G_{2m}) \leq b$  for all  $m$ . Hence, the theorem is proved.  $\square$

#### 4. SPECTRUM OF THE PRECONDITIONED SYSTEMS

In this section, we consider the spectrum of the preconditioned matrices  $\mathcal{E}_n^{-1}[f]\mathcal{T}_n[f]$ . We note that by Lemma 1,  $f$  should be nonnegative to guarantee that  $\mathcal{T}_n[f]$  and  $\mathcal{E}_n[f]$  are positive definite. When  $\mathcal{E}_n[f]$  is positive definite,  $\mathcal{E}_n^{-1/2}[f]$  is well defined and  $\mathcal{E}_n^{-1}[f]\mathcal{T}_n[f]$  is similar to the Hermitian matrix  $\mathcal{E}_n^{-1/2}[f]\mathcal{T}_n[f]\mathcal{E}_n^{-1/2}[f]$ . The following theorem shows that the spectrum of

$$\mathcal{E}_n[f]^{-1/2}\Delta_n[f]\mathcal{E}_n[f]^{-1/2} = \mathcal{E}_n^{-1/2}[f]\mathcal{T}_n[f]\mathcal{E}_n^{-1/2}[f] - I_n$$

cannot be clustered around zero.

**Theorem 2.** *Let  $f \in \mathcal{L}_{2\pi}$  be nonnegative and piecewise continuous. Let its points of discontinuity in  $(-\pi, \pi]$  be at  $-\pi < \theta_1 < \dots < \theta_\nu \leq \pi$  with jumps*

$$\alpha_k = \lim_{\theta \rightarrow \theta_k^+} f(\theta) - \lim_{\theta \rightarrow \theta_k^-} f(\theta), \quad k = 1, \dots, \nu,$$

and  $|\alpha_{k_0}| = \max_{1 \leq k \leq \nu} |\alpha_k|$ . Then to any  $\epsilon$  with  $0 < \epsilon < |\alpha_{k_0}|/4\|f\|_\infty$ , there corresponds a constant  $b$  such that

$$N(\epsilon; \mathcal{E}_n[f]^{-1/2}\Delta_n[f]\mathcal{E}_n[f]^{-1/2}) \geq \frac{4}{\pi}(1 + o(1)) \log \frac{n}{2} \cdot \operatorname{sech}^{-1}\left(\frac{1}{2} + \frac{2\epsilon\|f\|_\infty}{|\alpha_{k_0}|}\right) - b,$$

where  $o(1)$  tends to zero as  $n$  increases.

*Proof.* For simplicity, we write  $\Delta_n[f]$  and  $\mathcal{E}_n[f]$  as  $\Delta_n$  and  $\mathcal{E}_n$ , respectively. For any nonzero vector  $x$ , let  $y = \Delta_n \mathcal{E}_n^{-1/2} x$ . Then

$$\frac{x^* \mathcal{E}_n^{-1/2} \Delta_n \mathcal{E}_n^{-1} \Delta_n \mathcal{E}_n^{-1/2} x}{x^* x} = \frac{y^* \mathcal{E}_n^{-1} y}{x^* x}.$$

If  $y \neq 0$ , then by Lemma 1,

$$(17) \quad y^* \mathcal{E}_n^{-1} y = \frac{y^* \mathcal{E}_n^{-1} y}{y^* y} \cdot y^* y \geq \lambda_{\min}(\mathcal{E}_n^{-1}) \cdot y^* y \geq \frac{1}{f_{\max}} \cdot y^* y.$$

Since this is trivially true also for  $y = 0$ , we always have

$$\frac{y^* \mathcal{E}_n^{-1} y}{x^* x} \geq \frac{1}{f_{\max}} \frac{y^* y}{x^* x} = \frac{1}{f_{\max}} \frac{x^* \mathcal{E}_n^{-1/2} \Delta_n^2 \mathcal{E}_n^{-1/2} x}{x^* x}.$$

Let  $z = \mathcal{E}_n^{-1/2} x$ . Notice that  $z \neq 0$  since  $x \neq 0$ . Therefore, we have

$$\frac{y^* \mathcal{E}_n^{-1} y}{x^* x} \geq \frac{1}{f_{\max}} \frac{z^* \Delta_n^2 z}{z^* z} = \frac{1}{f_{\max}} \cdot \frac{z^* \Delta_n^2 z}{z^* z} \cdot \frac{z^* z}{x^* x}.$$

Since again by Lemma 1,

$$\frac{z^* z}{x^* x} = \frac{x^* \mathcal{E}_n^{-1} x}{x^* x} \geq \lambda_{\min}(\mathcal{E}_n^{-1}) \geq \frac{1}{f_{\max}},$$

it follows that

$$\frac{x^* \mathcal{E}_n^{-1/2} \Delta_n \mathcal{E}_n^{-1} \Delta_n \mathcal{E}_n^{-1/2} x}{x^* x} = \frac{y^* \mathcal{E}_n^{-1} y}{x^* x} \geq \frac{1}{f_{\max}^2} \cdot \frac{z^* \Delta_n^2 z}{z^* z}.$$

Hence, by the Courant and Fischer theorem (see Wilkinson [20, p. 101]), we have, for any nonzero vectors  $\{v_k\}_{k=1}^{j-1}$  in  $\mathbb{C}^n$ ,

$$\begin{aligned} \max_{\substack{x \neq 0 \\ x \in \langle v_1, \dots, v_{j-1} \rangle^\perp}} \frac{x^* \mathcal{E}_n^{-1/2} \Delta_n \mathcal{E}_n^{-1} \Delta_n \mathcal{E}_n^{-1/2} x}{x^* x} &\geq \max_{\substack{x \neq 0 \\ x \in \langle v_1, \dots, v_{j-1} \rangle^\perp}} \frac{1}{f_{\max}^2} \cdot \frac{z^* \Delta_n^2 z}{z^* z} \\ &= \max_{\substack{z \neq 0 \\ z \in \langle \mathcal{E}_n^{1/2} v_1, \dots, \mathcal{E}_n^{1/2} v_{j-1} \rangle^\perp}} \frac{1}{f_{\max}^2} \cdot \frac{z^* \Delta_n^2 z}{z^* z} \geq \frac{1}{f_{\max}^2} \cdot \lambda_j(\Delta_n^2), \end{aligned}$$

where the eigenvalues  $\lambda_j$  are ordered as  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$ . Since  $\{v_k\}_{k=1}^{j-1}$  are arbitrary, again by the Courant-Fischer theorem, we have

$$\lambda_j(\mathcal{E}_n^{-1/2} \Delta_n \mathcal{E}_n^{-1} \Delta_n \mathcal{E}_n^{-1/2}) \geq \frac{1}{f_{\max}^2} \cdot \lambda_j(\Delta_n^2).$$

Therefore, for  $0 < \epsilon < |\alpha_{k_0}|/4f_{\max}$ ,

$$\begin{aligned} N(\epsilon; \mathcal{E}_n^{-1/2} \Delta_n \mathcal{E}_n^{-1/2}) &= N(\epsilon^2; (\mathcal{E}_n^{-1/2} \Delta_n \mathcal{E}_n^{-1/2})^* (\mathcal{E}_n^{-1/2} \Delta_n \mathcal{E}_n^{-1/2})) \\ &= N(\epsilon^2; \mathcal{E}_n^{-1/2} \Delta_n \mathcal{E}_n^{-1} \Delta_n \mathcal{E}_n^{-1/2}) \\ &\geq N\left(\epsilon^2; \frac{1}{f_{\max}^2} \Delta_n^2\right) = N(f_{\max}^2 \epsilon^2; \Delta_n^2) = N(f_{\max} \epsilon; \Delta_n). \end{aligned}$$

Hence, by Theorem 1, we have

$$N(\epsilon; \mathcal{E}_n^{-1/2} \Delta_n \mathcal{E}_n^{-1/2}) \geq \frac{4}{\pi} (1 + o(1)) \log \frac{n}{2} \cdot \operatorname{sech}^{-1} \left( \frac{1}{2} + \frac{2\epsilon f_{\max}}{|\alpha_{k_0}|} \right) - b. \quad \square$$

### 5. BOUNDS ON THE NUMBER OF OUTLYING EIGENVALUES

In this section, we show that if  $f$  is strictly positive, then the number of outlying eigenvalues of  $\mathcal{E}_n^{-1}[f]\mathcal{I}_n[f]$  cannot be more than  $O(\log n)$ . We begin with the following lemma.

**Lemma 10.** *Let  $f \in \mathcal{L}_{2\pi}$  be piecewise continuous with points of discontinuity in  $(-\pi, \pi]$  at  $-\pi < \theta_1 < \dots < \theta_\nu \leq \pi$  and jumps  $\alpha_k, k = 1, \dots, \nu$ . Then for all sufficiently small  $\epsilon > 0$ , there exist positive constants  $c_1$  and  $c_2$ , independent of  $m$ , such that*

$$c_1 \log m \leq N(\epsilon; \Delta_{2m}[f]) \leq c_2 \log m .$$

*Proof.* For  $k = 1, 2, \dots, \nu$ , we define

$$\tilde{g}_k(\theta) = \begin{cases} \theta + \pi - \theta_k, & -\pi < \theta \leq \theta_k, \\ \theta - \pi - \theta_k, & \theta_k < \theta \leq \pi, \end{cases}$$

and write  $f$  as

$$f = \left\{ f + \sum_{k=1}^{\nu} \frac{\alpha_k}{2\pi} \tilde{g}_k \right\} - \sum_{k=1}^{\nu} \frac{\alpha_k}{2\pi} \tilde{g}_k .$$

It is easy to check that the first term on the right side is continuous. By Lemma 8,

$$\begin{aligned} \Delta_{2m}[f] &= \Delta_{2m} \left[ f + \sum_{k=1}^{\nu} \frac{\alpha_k}{2\pi} \tilde{g}_k \right] - \sum_{k=1}^{\nu} \frac{\alpha_k}{2\pi} \Delta_{2m}[\tilde{g}_k] \\ &= \Delta_{2m} \left[ f + \sum_{k=1}^{\nu} \frac{\alpha_k}{2\pi} \tilde{g}_k \right] - \sum_{k=1}^{\nu} \frac{\alpha_k}{2\pi} (A_{k,2m} + B_{k,2m}) , \end{aligned}$$

where  $A_{k,2m}$  and  $B_{k,2m}$  satisfy the properties in (3) and (4), respectively. Hence, by Lemmas 2–4, for any  $0 < \epsilon < \frac{\nu+1}{2} |\alpha_{k_0}|$  (where as before  $|\alpha_{k_0}| = \max_{1 \leq k \leq \nu} |\alpha_k|$ ), there exists a positive constant  $c$  such that

$$\begin{aligned} N(\epsilon; \Delta_{2m}[f]) &\leq N \left( \frac{\epsilon}{\nu+1}; \Delta_{2m} \left[ f + \sum_{k=1}^{\nu} \frac{\alpha_k}{2\pi} \tilde{g}_k \right] - \sum_{k=1}^{\nu} \frac{\alpha_k}{2\pi} B_{k,2m} \right) \\ &\quad + N \left( \frac{\nu\epsilon}{\nu+1}; - \sum_{k=1}^{\nu} \frac{\alpha_k}{2\pi} A_{k,2m} \right) \\ &\leq c + \sum_{k=1}^{\nu} N \left( \frac{\epsilon}{\nu+1}; \frac{\alpha_k}{2\pi} A_{k,2m} \right) \\ &= c + \sum_{k=1}^{\nu} N \left( \frac{2\pi\epsilon}{|\alpha_k|(\nu+1)}; A_{k,2m} \right) \\ &= c + \sum_{k=1}^{\nu} \frac{4}{\pi} \log m \cdot \operatorname{sech}^{-1} \left( \frac{2\epsilon}{|\alpha_k|(\nu+1)} \right) \cdot (1 + o(1)) \\ &\leq c + \frac{4}{\pi} \log m \sum_{k=1}^{\nu} (1 + o(1)) \cdot \operatorname{sech}^{-1} \left( \frac{2\epsilon}{|\alpha_{k_0}|(\nu+1)} \right) \\ &= c + \frac{4\nu}{\pi} \log m \cdot (1 + o(1)) \operatorname{sech}^{-1} \left( \frac{2\epsilon}{|\alpha_{k_0}|(\nu+1)} \right) . \end{aligned}$$

By combining this result with Theorem 1, the lemma follows.  $\square$

As a corollary, we can show that the matrix  $\mathcal{E}_n[f]^{-1} \mathcal{F}_n[f] - I_n$  will have at most  $O(\log n)$  outlying eigenvalues provided that  $f_{\min} > 0$ .

**Theorem 3.** Let  $f \in \mathcal{L}_{2\pi}$  be piecewise continuous with  $f_{\min} > 0$ . Then for all sufficiently small  $\epsilon > 0$ , there exist positive constants  $c_3$  and  $c_4$  such that

$$c_3 \log \frac{n}{2} \leq N(\epsilon; \mathcal{E}_n[f]^{-1/2} \Delta_n[f] \mathcal{E}_n[f]^{-1/2}) \leq c_4 \log \frac{n}{2}.$$

*Proof.* The proof is similar to the one of Theorem 2 with (17) replaced by

$$y^* \mathcal{E}_n^{-1} y = \frac{y^* \mathcal{E}_n^{-1} y}{y^* y} \cdot y^* y \leq \lambda_{\max}(\mathcal{E}_n^{-1}) \cdot y^* y \leq \frac{1}{f_{\min}} \cdot y^* y,$$

where the last inequality above follows from Lemma 1.  $\square$

## 6. NUMERICAL RESULTS

In this section, we illustrate by numerical examples how the discontinuities in the generating function  $f$  affect the convergence rate of the method. In the examples, test functions  $f$  defined on  $(-\pi, \pi]$  are used to generate Toeplitz matrices  $\mathcal{T}_n[f]$  and the systems  $\mathcal{T}_n[f]x = b$ , where  $b = \frac{1}{\sqrt{n}}(1, 1, \dots, 1, 1)^*$ , are then solved by the preconditioned conjugate gradient method with or without the preconditioner  $\mathcal{E}_n[f]$ . All computations are done by Matlab on a Sparc II workstation at UCLA. The zero vector is used as the initial guess and the stopping criterion is  $\|r_q\|_2 / \|r_0\|_2 \leq 10^{-7}$ , where  $r_q$  is the residual vector after  $q$  iterations. Table 1 shows the number of iterations required for convergence. In the table, the first row gives the generating functions and the second row indicates the preconditioner used. The function  $f_{\beta, \gamma}$ ,  $0 \leq \gamma < \beta$ , is a piecewise linear function defined by

$$f_{\{\beta, \gamma\}}(\theta) = \begin{cases} \frac{\beta - \gamma}{\pi} \theta + \beta, & -\pi < \theta \leq 0, \\ \frac{\beta - \gamma}{\pi} \theta + \gamma, & 0 < \theta \leq \pi, \end{cases} \quad 0 \leq \gamma < \beta,$$

where  $\beta$  and  $\gamma$  are the maximum and minimum values of  $f_{\{\beta, \gamma\}}$  on  $(-\pi, \pi]$ , respectively.

TABLE 1. Number of iterations for different generating functions

$n$	$\theta^4 + 1$		$(\theta + \pi)^2 + 1$		$f_{\{10, 0.1\}}$		$(\theta + \pi)^2$		$f_{\{10, 0\}}$	
	None	$\mathcal{E}_n[f]$	None	$\mathcal{E}_n[f]$	None	$\mathcal{E}_n[f]$	None	$\mathcal{E}_n[f]$	None	$\mathcal{E}_n[f]$
16	8	8	16	9	8	7	16	10	8	7
32	20	7	33	10	16	8	35	14	16	8
64	37	7	45	11	26	9	77	19	27	9
128	56	6	49	11	36	10	167	27	39	10
256	67	6	50	13	47	11	356	41	55	12
512	70	6	51	13	59	13	743	65	76	15
1024	71	5	51	13	68	14	1486	107	106	18

We note that the first generating function  $\theta^4 + 1$  is a  $2\pi$ -periodic function and the convergence rate obtained here is typical for such a class of functions (see Chan [2]). The other four functions are all piecewise continuous. Note that the second and the third functions are strictly positive. Therefore,  $\mathcal{T}_n[f]$ ,  $\mathcal{E}_n[f]$ , and hence  $\mathcal{E}_n^{-1}[f] \mathcal{T}_n[f]$  are all well conditioned in view of Lemma 1. In particular, the corresponding systems will converge linearly, i.e., the method

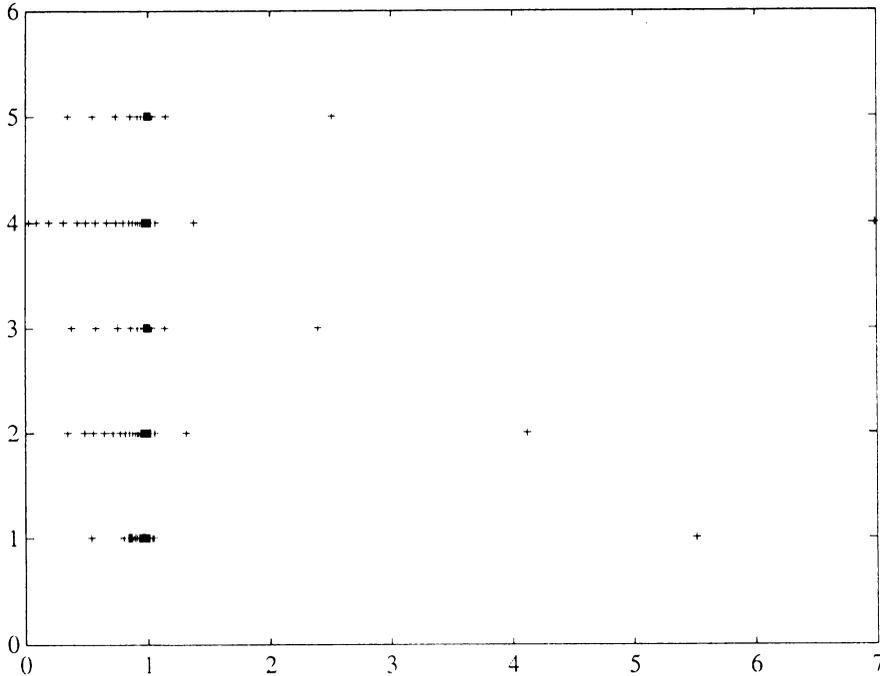


FIGURE 1. Spectra of preconditioned systems for  $n = 64$

will converge in a finite number of steps independent of the matrix size. So, although the  $O(\log n)$  effect can be seen for the preconditioned systems for small  $n$ , it will level off when  $n$  gets larger. For the last two functions, since their  $f_{\min} = 0$ , the matrices  $\mathcal{T}_n[f]$  will no longer be well conditioned. In fact, we see that for the nonpreconditioned systems, the numbers of iterations required for convergence increase like  $O(n)$  and  $O(\sqrt{n})$ , respectively (cf. Chan [3, p.338]). In these cases, the number of iterations for the preconditioned systems grows even faster than  $O(\log n)$ .

For comparison, the spectra of the preconditioned systems for  $n = 64$  were computed and shown in Figure 1 with the first test function  $\theta^4 + 1$  at the bottom (i.e.,  $y = 1$  in the figure) to the fifth one  $f_{\{10,0\}}$  at the top. For the last four functions, we can see that their corresponding spectra are less clustered than the first one.

## 7. CONCLUDING REMARKS

We have proved in this paper that when the T. Chan circulant preconditioner is used to precondition Toeplitz matrices that are generated by nonnegative piecewise continuous functions, the resulting matrices cannot have spectrum clustered around 1 and the number of outlying eigenvalues grows at least like  $O(\log n)$ . We then show by numerical examples that these outlying eigenvalues do affect the convergence rate of the method, and in general the convergence rate is no longer superlinear and the number of iterations required for convergence increases at least like  $O(\log n)$ . For such systems, it is better to use band Toeplitz preconditioners instead of circulant preconditioners, for they guarantee linear convergence rate whenever  $f$  is nonnegative piecewise continuous (see

Chan and Ng [4, Theorem 1]). We finally remark that, recently, Tyrtyshnikov [17] has established a generalized Szegő theorem and used it to prove that if  $f$  is in  $L_2$  with  $f_{\min} > 0$ , then the number of outlying eigenvalues grows no more than  $o(n)$ . Theorem 3 in this paper can be viewed as a stronger form of his result under a stronger assumption.

#### ACKNOWLEDGMENT

We would like to thank Dr. K. M. Tsang for helpful discussions.

#### BIBLIOGRAPHY

1. G. Ammar and W. Gragg, *Superfast solution of real positive definite Toeplitz systems*, SIAM J. Matrix Anal. Appl. **9** (1988), 61–76.
2. R. Chan, *The spectrum of a family of circulant preconditioned Toeplitz systems*, SIAM J. Numer. Anal. **26** (1989), 503–506.
3. —, *Toeplitz preconditioners for Toeplitz systems with nonnegative generating functions*, IMA J. Numer. Anal. **11** (1991), 333–345.
4. R. Chan and K. Ng, *Fast iterative solvers for Toeplitz-plus-band systems*, SIAM J. Sci. Statist. Comput. (to appear).
5. R. Chan and G. Strang, *Toeplitz equations by conjugate gradients with circulant preconditioner*, SIAM J. Sci. Statist. Comput. **10** (1989), 104–119.
6. R. Chan and M. Yeung, *Circulant preconditioners for Toeplitz matrices with positive continuous generating functions*, Math. Comp. **58** (1992), 233–240.
7. —, *Jackson's Theorem and circulant preconditioned Toeplitz systems*, J. Approx. Theory **70** (1992), 191–205.
8. —, *Circulant preconditioners constructed from kernels*, SIAM J. Numer. Anal. **29** (1992), 1093–1103.
9. T. Chan, *An optimal circulant preconditioner for Toeplitz systems*, SIAM J. Sci. Statist. Comput. **9** (1988), 766–771.
10. T. Huckle, *Circulant and skew-circulant matrices for solving Toeplitz matrix problems*, SIAM J. Matrix Anal. Appl. (to appear).
11. T. Ku and C. Kuo, *Design and analysis of Toeplitz preconditioners*, IEEE Trans. Signal Process. **40** (1991), 129–141.
12. Z. Nehari, *On bounded bilinear forms*, Ann. of Math. (2) **65** (1957), 153–162.
13. G. Strang, *A proposal for Toeplitz matrix calculations*, Stud. Appl. Math. **74** (1986), 171–176.
14. M. Tismenetsky, *A decomposition of Toeplitz matrices and optimal circulant preconditioning*, Linear Algebra Appl. **154** (1991), 105–121.
15. L. Trefethen, *Approximation theory and numerical linear algebra*, Algorithms for Approximation II (M. Cox and J. Mason, eds.), Chapman and Hall, London, 1990.
16. E. Tyrtyshnikov, *Optimal and super-optimal circulant preconditioners*, SIAM J. Matrix Anal. Appl. **13** (1992), 459–473.
17. —, *A unifying approach to some old and new theorems on distribution and clustering*, Linear Algebra Appl. (to appear).
18. J. Walker, *Fourier analysis*, Oxford Univ. Press, New York, 1988.
19. H. Widom, *Hankel matrices*, Trans. Amer. Math. Soc. **121** (1966), 1–35.
20. J. Wilkinson, *The algebraic eigenvalue problem*, Clarendon Press, Oxford, 1965.

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF HONG KONG, POKFULAM ROAD, HONG KONG  
 Current address, M.-C. Yeung: Department of Mathematics, University of California, Los Angeles, 405 Hilgard Avenue, Los Angeles, California 90024  
 E-mail address, M.-C. Yeung: myeung@math.ucla.edu  
 E-mail address, R. H. Chan: na.rchan@na-net.ornl.gov