On the Asymptotic Equivalence of Circulant and Toeplitz Matrices

Zhihui Zhu and Michael B. Wakin

***Abstract*—Any sequence of uniformly bounded $N \times N$ Hermitian Toeplitz matrices $\{\boldsymbol{H}_N\}$ is asymptotically equivalent to a certain sequence of $N \times N$ circulant matrices $\{\boldsymbol{C}_N\}$ derived from the Toeplitz matrices in the sense that $\|\boldsymbol{H}_N - \boldsymbol{C}_N\|_F = o(\sqrt{N})$ as $N \to \infty$. This implies that certain collective behaviors of the eigenvalues of each Toeplitz matrix are reflected in those of the corresponding circulant matrix and supports the utilization of the computationally efficient fast Fourier transform (instead of the Karhunen-Loève transform) in applications like coding and filtering. In this paper, we study the asymptotic performance of the *individual eigenvalue estimates*. We show that the asymptotic equivalence of the circulant and Toeplitz matrices implies the individual asymptotic convergence of the eigenvalues for certain types of Toeplitz matrices. We also show that these estimates asymptotically approximate the largest and smallest eigenvalues for more general classes of Toeplitz matrices.**

***Keywords*—Szegő's theorem, Toeplitz matrices, circulant matrices, asymptotic equivalence, Fourier analysis, eigenvalue estimates**

I. Introduction

A. Szegő's Theorem

Toeplitz matrices are of considerable interest in statistical signal processing and information theory [1–5]. An $N \times N$ Toeplitz matrix $\boldsymbol{H}_N$ has the form[1]

$$\boldsymbol{H}_N = \begin{bmatrix} h[0] & h[-1] & h[-2] & \cdots & h[-(N-1)] \\ h[1] & h[0] & h[-1] & & \\ h[2] & h[1] & h[0] & & \vdots \\ \vdots & & & \ddots & \\ h[N-1] & & \cdots & & h[0] \end{bmatrix}$$

or $\boldsymbol{H}_N[m,n] = h[m-n]; m,n \in [N] := \{0,1,\dots,N-1\}$. The covariance matrix of a discrete-time wide-sense stationary (WSS) random process is an example of such a matrix.

Throughout this paper, we consider $\boldsymbol{H}_N$ that is Hermitian, i.e., $\boldsymbol{H}_N^H = \boldsymbol{H}_N$, and we suppose that the eigenvalues of $\boldsymbol{H}_N$ are denoted and arranged as $\lambda_0(\boldsymbol{H}_N) \geq \cdots \geq \lambda_{N-1}(\boldsymbol{H}_N)$. Here the Hermitian transpose of a matrix $\boldsymbol{A}$ is denoted by $\boldsymbol{A}^H$.

Szegő's theorem [1] describes the collective asymptotic behavior (as $N \to \infty$) of the eigenvalues of a sequence of Hermitian Toeplitz matrices $\{\boldsymbol{H}_N\}$ by defining a function $\widetilde{h}(f) \in L^2([0,1])$ with Fourier series[2]

$$h[k] = \int_0^1 \widetilde{h}(f) e^{-j2\pi kf} df, \quad k \in \mathbb{Z},$$
$$\widetilde{h}(f) = \sum_{k=-\infty}^{\infty} h[k] e^{j2\pi kf}, \quad f \in [0,1].$$

Usually $\widetilde{h}(f)$ is referred to as the symbol or generating function for the $N \times N$ Toeplitz matrices $\{\boldsymbol{H}_N\}$.

Suppose $\widetilde{h} \in L^\infty([0,1])$. Szegő's theorem [1] states that

$$\lim_{N\to\infty} \frac{1}{N} \sum_{l=0}^{N-1} \vartheta(\lambda_l(\boldsymbol{H}_N)) = \int_0^1 \vartheta(\widetilde{h}(f)) df, \tag{1}$$

where ϑ is any function continuous on the range of $\widetilde{h}$. As one example, choosing $\vartheta(x) = x$ yields

$$\lim_{N\to\infty} \frac{1}{N} \sum_{l=0}^{N-1} \lambda_l(\boldsymbol{H}_N) = \int_0^1 \widetilde{h}(f) df.$$

In words, this says that as $N \to \infty$, the average eigenvalue of $\boldsymbol{H}_N$ converges to the average value of the symbol $\widetilde{h}(f)$ that generates $\boldsymbol{H}_N$. As a second example, suppose $\widetilde{h}(f) > 0$ (and thus $\lambda_l(\boldsymbol{H}_N) > 0$ for all $l \in [N]$ and $N \in \mathbb{N}$) and let ϑ be the log function. Then Szegő's theorem indicates that

$$\lim_{N\to\infty} \frac{1}{N} \log\left(\det\left(\boldsymbol{H}_N\right)\right) = \int_0^1 \log\left(\widetilde{h}(f)\right) df.$$

This relates the determinant of the Toeplitz matrix to its symbol.

Szegő's theorem has been widely used in the areas of signal processing, communications, and information theory. A paper and review by Gray [2, 7] serve as a remarkable elementary introduction in the engineering literature and offer a simplified proof of Szegő's original theorem. The result has also been extended in several ways. For example, the Avram-Parter theorem [8, 9], a generalization of Szegő's theorem, relates the collective asymptotic behavior of the singular values of a general (non-Hermitian) Toeplitz matrix to the absolute value of its symbol, i.e., $|\widetilde{h}(f)|$. Tyrtyshnikov [10] proved that Szegő's theorem holds if $\widetilde{h}(f) \in \mathbb{R}$ and $\widetilde{h}(f) \in L^2([0,1])$, and Zamarashkin and Tyrtyshnikov [11] further extended Szegő's

This work was supported by NSF grant CCF-1409261.

Z. Zhu and M. B. Wakin are with the Department of Electrical Engineering and Computer Science, Colorado School of Mines, Golden, CO 80401 USA. e-mail: {zzhu,mwakin}@mines.edu.

[1]Through the paper, finite-dimensional vectors and matrices are indicated by bold characters and we index such vectors and matrices beginning at 0.

[2]This can also be interpreted using the discrete-time Fourier transform (DTFT). That is, we can define $\widehat{h}(f) = \sum_{k=-\infty}^{\infty} h[k] e^{-j2\pi kf} = \widetilde{h}(1-f)$. However, it is more common to view h as the Fourier series of the symbol $\widetilde{h}$; see [1, 6].

theorem to the case when $\widetilde{h}(f) \in \mathbb{R}$ and $\widetilde{h}(f) \in L^1([0,1])$. Sakrison [12] extended Szegő's theorem to high dimensions. Gazzah et al. [13] and Gutiérrez-Gutiérrez and Crespo [14] extended Gray's results on Toeplitz and circulant matrices to block Toeplitz and block circulant matrices and derived Szegő's theorem for block Toeplitz matrices.

Most relevant to our work, Bogoya et al. [15] studied the *individual* asymptotic behavior of the eigenvalues of Toeplitz matrices by interpreting Szegő's theorem in probabilistic language. In the case that the range of $\widetilde{h}$ is connected, Bogoya et al. related the eigenvalues to the values obtained by sampling the symbol $\widetilde{h}(f)$ uniformly in frequency on $[0,1]$.

B. Motivation

Despite the power of Szegő's theorem, in many scenarios (such as certain coding and filtering applications [2, 3]), one may only have access to $\boldsymbol{H}_N$ and not $\widetilde{h}$. In such cases, it is still desirable to have practical and efficiently computable estimates of individual eigenvalues of $\boldsymbol{H}_N$. We elaborate on two example applications below.

i. **Estimating the condition number of a positive-definite Toeplitz matrix.** The linear system $\boldsymbol{H}_N \boldsymbol{y} = \boldsymbol{b}$ arises naturally in many signal processing and estimation problems such as linear prediction [4, 5]. The condition number $\kappa(\boldsymbol{H}_N)$ of the Toeplitz matrix $\boldsymbol{H}_N$ is important when solving such systems. For example, the speed of solving such linear systems via the widely used conjugate gradient method is determined by the condition number: the larger $\kappa(\boldsymbol{H}_N)$, the slower convergence of the algorithm. In case of large $\kappa(\boldsymbol{H}_N)$, preconditioning can be applied to ensure fast convergence. Thus estimating the smallest and largest eigenvalues of a symmetric positive-definite Toeplitz matrix (such as the covariance matrix of a stationary random process) is of considerable interest [16, 17].

ii. **Spectrum sensing algorithm for cognitive radio.** Spectrum sensing is a fundamental task in cognitive ratio, which aims to best utilize the available spectrum by identifying unoccupied bands [18–20]. Zeng and Ling [20] have proposed spectrum sensing methods for cognitive radio based on the eigenvalues of a Toeplitz covariance matrix. These eigenvalue-based algorithms overcome the noise uncertainty problem which exists in alternative methods based on energy detection.

Aside from the above applications, approximate and efficiently computable eigenvalue estimates can also be used as the starting point for numerical algorithms that iteratively compute eigenvalues with high precision.

C. Contributions

In this paper, we consider estimates for the eigenvalues of a Toeplitz matrix that are obtained through a two-step process:

1) Transform the Toeplitz matrix into a circulant matrix using a certain procedure described below.
2) Compute the eigenvalues of the circulant matrix.

Both of these steps can be performed efficiently; in particular, the eigenvalues of an $N \times N$ circulant matrix can be computed in $O(N \log N)$ time[3] using the fast Fourier transform (FFT). The individual eigenvalues of the circulant matrix approximate those of the Toeplitz matrix. We study the quality of this approximation.

An $N \times N$ circulant matrix $\boldsymbol{C}_N$ is a special Toeplitz matrix of the form

$$\boldsymbol{C}_N = \begin{bmatrix} c[0] & c[1] & c[2] & \cdots & c[N-1] \\ c[N-1] & c[0] & c[1] & & \\ c[N-2] & c[N-1] & c[0] & & \vdots \\ \vdots & & & \ddots & \\ c[1] & & \cdots & & c[0] \end{bmatrix}.$$

Circulant matrices arise naturally in applications involving the discrete Fourier transform (DFT) [3]; in particular, any circulant matrix can be unitarily diagonalized using the DFT matrix. Circulant matrices offer a nontrivial but simple set of objects that can be used for problems involving Toeplitz matrices. For example, the product $\boldsymbol{H}_N \boldsymbol{x}$ can be computed in $O(N \log N)$ time by embedding $\boldsymbol{H}_N$ into a $(2N-1)\times(2N-1)$ circulant matrix and using the FFT to perform matrix-vector multiplication. Also Gray [2, 7] showed that Toeplitz and circulant matrices are asymptotically equivalent in a certain sense; this implies that their eigenvalues have similar *collective behavior*. See Section II for formal definitions. Finally, we note that circulant matrices have been used as preconditioners [21, 22] of Toeplitz matrices in iterative methods for solving linear systems of the form $\boldsymbol{H}_N \boldsymbol{y} = \boldsymbol{b}$.

We consider estimates for the eigenvalues of a Toeplitz matrix obtained from a well-constructed circulant matrix. The eigenvalues of the circulant matrix can be computed efficiently without constructing the whole matrix; one merely applies the FFT to the first row of the matrix. We do *not* provide new circulant approximations to Toeplitz matrices in this paper; rather we sharpen the analysis on the asymptotic equivalence of Toeplitz and certain circulant matrices [2, 3, 7] by establishing results in terms of *individual eigenvalues* rather than collective behavior. To the best of our knowledge, this is the first work that provides guarantees for asymptotic equivalence in terms of individual eigenvalues.

D. Circulant Approximations to $\boldsymbol{H}_N$

We consider the following circulant approximations that have been widely used in information theory and applied mathematics.

1) $\widetilde{\boldsymbol{C}}_N$*:* Bogoya et al. [15] proved that the samples of the symbol $\widetilde{h}$ are the main asymptotic terms of the eigenvalues of the Toeplitz matrix $\boldsymbol{H}_N$. Given only $\boldsymbol{H}_N$, one practical strategy for estimating the eigenvalues is to first approximate $\widetilde{h}$ by the $(N-1)^{\text{th}}$ partial Fourier sum $S_{N-1}(f) = \sum_{k=-(N-1)}^{N-1} h[k] e^{j2\pi f k}$. Then construct a circulant matrix whose eigenvalues are samples of $S_{N-1}(f)$, i.e., $S_{N-1}(\frac{l}{N})$. We let $\widetilde{\boldsymbol{C}}_N$ denote the corresponding circulant matrix, whose

[3] We say $g_1(N) = O(g_2(N))$ if and only if there exist a positive real number t and $M \in \mathbb{N}$ such that $g_1(N) \le t g_2(N)$ for all $N \ge M$.

top row $(\widetilde{c}[0], \widetilde{c}[1], \ldots, \widetilde{c}[N-1])$ can be obtained as

$$\begin{aligned}\widetilde{c}[k] &= \frac{1}{N}\sum_{n=0}^{N-1} S_{N-1}(\frac{2\pi n}{N}) e^{j2\pi kn/N} \\ &= \frac{1}{N}\sum_{n=0}^{N-1}\sum_{k'=-(N-1)}^{N-1} h[k'] e^{j2\pi(k+k')n/N} \\ &= \sum_{k'=-(N-1)}^{N-1} h[k'] \left(\sum_{n=0}^{N-1} \frac{1}{N} e^{j2\pi(k+k')n/N}\right) \\ &= \begin{cases} h[0], & k=0, \\ h[-k]+h[N-k], & k=1,2,\ldots,N-1, \end{cases}\end{aligned}$$

where the last line utilizes the fact

$$\sum_{n=0}^{N-1} \frac{1}{N} e^{j2\pi(k+k')n/N} = \begin{cases} 1, & \text{mod}(k+k', N)=0, \\ 0, & \text{otherwise.} \end{cases}$$

2) $\widehat{C}_N$: Following the same strategy, we first compute the $\left(\lfloor \frac{N-1}{2} \rfloor\right)^{\text{th}}$ partial Fourier sum

$$S_{\lfloor \frac{N-1}{2} \rfloor}(f) = \sum_{k=-\lfloor \frac{N-1}{2} \rfloor}^{\lfloor \frac{N-1}{2} \rfloor} h[k] e^{j2\pi fk}.$$

Let $\widehat{C}_N$ denote the $N \times N$ circulant matrix whose eigenvalues are samples of $S_{\lfloor \frac{N-1}{2} \rfloor}(f)$, i.e, $S_{\lfloor \frac{N-1}{2} \rfloor}(\frac{l}{N})$. With simple manipulations, the top row $(\widehat{c}[0], \widehat{c}[1], \ldots, \widehat{c}[N-1])$ of $\widehat{C}_N$ is given by

$$\widehat{c}[k] = \begin{cases} h[-k], & 0 \le k \le \lfloor \frac{N-1}{2} \rfloor, \\ h[N-k], & \lceil \frac{N+1}{2} \rceil \le k < N, \\ 0, & k = N/2, \end{cases}$$

when N is even, and

$$\widehat{c}[k] = \begin{cases} h[-k], & 0 \le k \le \lfloor \frac{N-1}{2} \rfloor, \\ h[N-k], & \lceil \frac{N+1}{2} \rceil \le k < N, \end{cases}$$

when N is odd.

Strang [21] first employed such circulant matrices as preconditioners to speed up the convergence of iterative methods for solving Toeplitz linear systems. This approach is quite simple. The underlying idea is that the sequence $h[k]$ usually decays quickly as k grows large, and thus we keep the largest part of the Toeplitz matrix and fill in the remaining part to form a circulant approximation.

3) $\overline{C}_N$: In the Fourier analysis literature, it is known that Cesàro sum has rather better convergence than the partial Fourier sum [23]. The N^{th} Cesàro sum is defined as

$$\sigma_N(f) = \frac{\sum_{n=0}^{N-1} S_n(f)}{N}.$$

We use $\overline{C}_N$ to denote the $N \times N$ circulant matrix whose eigenvalues are samples of $\sigma_N(f)$, i.e., $\sigma_N(\frac{l}{N})$. The top row $(\overline{c}[0], \overline{c}[1], \ldots, \overline{c}[N-1])$ of $\overline{C}_N$ can be obtained as follows

$$\begin{aligned}\overline{c}[k] &= \frac{1}{N}\sum_{l=0}^{N-1} \sigma_N(\frac{l}{N}) e^{j2\pi kl/N} \\ &= \frac{1}{N}\sum_{l=0}^{N-1} \frac{1}{N}\sum_{n=0}^{N-1}\sum_{k'=-n}^{n} h[k'] e^{j2\pi l(k+k')/N} \\ &= \frac{1}{N}\sum_{n=0}^{N-1}\sum_{k'=-n}^{n} \left(h[k'] \sum_{l=0}^{N-1} \frac{1}{N} e^{j2\pi l(k+k')/N}\right) \\ &= \frac{1}{N}\left((N-k)h[-k] + kh[N-k]\right).\end{aligned}$$

Pearl [3] first analyzed such a circulant approximation and its applications in coding and filtering. The same circulant approximation (referred to as an optimal preconditioner) was also proposed by Chan [22]. The optimal preconditioner is the solution to the following optimization problem

$$\text{minimize } \|C_N - H_N\|_F$$

over all $N \times N$ circulant matrices. One can verify that $\overline{C}_N$ is the solution to the above problem.

E. Main Results

Let $\{\lambda_l(C_N)\}_{l\in[N]}$ denote the eigenvalues of the circulant matrix C_N for all $C_N \in \left\{\widetilde{C}_N, \widehat{C}_N, \overline{C}_N\right\}$. Let $\lambda_l(C_N)$ be permuted that such that $\lambda_{\rho(0)}(C_N) \ge \lambda_{\rho(1)}(C_N) \ge \cdots \ge \lambda_{\rho(N-1)}(C_N)$. In this paper, we establish the following results.

Theorem I.1. *Suppose that the sequence $h[k]$ is absolutely summable. Then*

$$\lim_{N\to\infty} \max_{l\in[N]} \left|\lambda_l(H_N) - \lambda_{\rho(l)}(C_N)\right| = 0, \tag{2}$$

for all $C_N \in \left\{\widetilde{C}_N, \widehat{C}_N, \overline{C}_N\right\}$.

Theorem I.1 states that the individual asymptotic convergence of the eigenvalues between the Toeplitz matrices H_N and circulant matrices $C_N \in \left\{\widetilde{C}_N, \widehat{C}_N, \overline{C}_N\right\}$ holds as long as $h[k]$ is absolutely summable. Its proof involves the uniform convergence of a Fourier series and the fact that the equal distribution of two sequences implies individual asymptotic equivalence of two sequences in a certain sense. By utilizing the Sturmian separation theorem [24], we also provide the convergence rate for band Toeplitz matrices as follows.

Theorem I.2. *Suppose that $h[k] = 0$ for all $k > r$, i.e., H_N is a band Toeplitz matrix when $N > r$. Then*

$$\max_{l\in[N]} \left|\lambda_l(H_N) - \lambda_{\rho(l)}(C_N)\right| = O(\frac{1}{N}) \tag{3}$$

as $N \to \infty$ for all $C_N \in \left\{\widetilde{C}_N, \widehat{C}_N, \overline{C}_N\right\}$.

Utilizing the fact that the Cesàro sum has rather better convergence than the partial Fourier sum, the following result establishes a weaker condition on $h[k]$ for the individual asymptotic convergence of the eigenvalues between $\boldsymbol{H}_N$ and $\overline{\boldsymbol{C}}_N$.

Theorem I.3. *Suppose that $h[k]$ is square summable and $\widetilde{h} \in L^{\infty}([0,1])$ is Riemann integrable and the essential range of $\widetilde{h}$ is $\left[\operatorname{ess\,inf} \widetilde{h}, \operatorname{ess\,sup} \widetilde{h}\right]$, i.e., the essential range of $\widetilde{h}$ is connected. Then*

$$\lim_{N\to\infty} \max_{l\in[N]} \left|\lambda_l(\boldsymbol{H}_N) - \lambda_{\rho(l)}(\overline{\boldsymbol{C}}_N)\right| = 0. \quad (4)$$

Note that the sequence $h[k]$ being absolutely summable implies that $h[k]$ is square summable, that $\widetilde{h} \in L^{\infty}([0,1])$ is Riemann integrable, and that its range is connected. However, the converse of this statement does not hold. We provide an example in Section IV-B.

Finally, the following result concerns the convergence of the largest and smallest eigenvalues for more general classes of Toeplitz matrices.

Theorem I.4. *Suppose that $\widetilde{h} \in L^{\infty}([0,1])$ is Riemann integrable. Then*

$$\lim_{N\to\infty} \lambda_0(\boldsymbol{H}_N) = \lim_{N\to\infty} \lambda_{\rho(0)}\left(\overline{\boldsymbol{C}}_N\right) = \operatorname{ess\,sup} \widetilde{h},$$
$$\lim_{N\to\infty} \lambda_{N-1}(\boldsymbol{H}_N) = \lim_{N\to\infty} \lambda_{\rho(N-1)}\left(\overline{\boldsymbol{C}}_N\right) = \operatorname{ess\,inf} \widetilde{h}.$$

Laudadio et al. [17] summarized several algorithms to estimate the smallest eigenvalue of a symmetric positive-definite Toeplitz matrix. These algorithms need $O(N^2)$ flops. Computing $\lambda_{\rho(N-1)}\left(\overline{\boldsymbol{C}}_N\right)$ via the FFT requires $O(N \log N)$ flops, and at the same time, we are guaranteed that $\lambda_{\rho(N-1)}(\overline{\boldsymbol{C}}_N)$ is asymptotically equivalent to $\lambda_{N-1}(\boldsymbol{H}_N)$ by Theorem I.4.

The above results—characterizing the individual asymptotic convergence of the eigenvalues between Toeplitz and circulant matrices—serve as complements to the literature on asymptotic equivalence that has focused on the collective behavior of the eigenvalues. Before moving on, we briefly review said literature. In [2, 7], Gray showed the asymptotic equivalence[4] of $\{\boldsymbol{H}_N\}$ and $\{\widetilde{\boldsymbol{C}}_N\}$ when the sequence $h[k]$ is absolutely summable. Pearl showed the asymptotic equivalence of $\{\boldsymbol{H}_N\}$ and $\{\overline{\boldsymbol{C}}_N\}$ when the sequence $h[k]$ is square summable and $\boldsymbol{H}_N$ and $\overline{\boldsymbol{C}}_N$ have bounded eigenvalues for all $N \in \mathbb{N}$. The spectrum of the preconditioned matrix $\boldsymbol{C}_N^{-1}\boldsymbol{H}_N$ asymptotically clustering around one was investigated in [10, 25–27]. Finally, as noted previously, Bogoya et al. [15] studied the *individual* asymptotic behavior of the eigenvalues of Toeplitz matrices by interpreting Szegő's theorem in probabilistic language. Our estimates for the eigenvalues of a Toeplitz matrix differ from [15] in that they are only dependent on the entries of $\boldsymbol{H}_N$ (instead of the symbol $\widetilde{h}(f)$). For our proof, we utilize the same approach of interpreting Szegő's theorem in probabilistic language. However, [15] requires the sequences of the eigenvalues to be strictly inside the range of $\widetilde{h}$, while our work covers more general cases where the sequences of the eigenvalues can be outside of the range of $\widetilde{h}$ as illustrated in Theorem III.1. See also our remark at the end of Section III-C.

[4] We define asymptotically equivalent sequences of matrices in Section II.

The rest of the paper is organized as follows. Section II states preliminary results on the asymptotic equivalence of Toeplitz and circulant matrices. We prove our main results in Section III. Section IV presents examples to illustrate our results, and Section V concludes the paper.

II. Preliminaries

A. Asymptotically Equivalent Matrices

We begin with the notion of equal distribution of two real sequences, using a definition attributed to Weyl [1].

Definition II.1. [1] Assume that the sequences $\{\{u_{N,l}\}_{l\in[N]}\}_{N=1}^{\infty}$ and $\{\{v_{N,l}\}_{l\in[N]}\}_{N=1}^{\infty}$ are absolutely bounded, i.e., there exist a, b such that $a \le u_{N,l} \le b$ and $a \le v_{N,l} \le b$ for all $l \in [N]$ and $N \in \mathbb{N}$. Then $\{\{u_{N,l}\}_{l\in[N]}\}_{N=1}^{\infty}$ and $\{\{v_{N,l}\}_{l\in[N]}\}_{N=1}^{\infty}$ are *equally distributed* if

$$\lim_{N\to\infty} \frac{1}{N} \sum_{l=0}^{N-1} \left(\vartheta\left(u_{N,l}\right) - \vartheta\left(v_{N,l}\right)\right) = 0.$$

for every continuous function ϑ on $[a, b]$.

The asymptotic equivalence of two sequences of matrices is defined as follows.

Definition II.2. [2, 7] Two sequences of $N \times N$ matrices $\{\boldsymbol{A}_N\}$ and $\{\boldsymbol{B}_N\}$ (where $\boldsymbol{A}_N$ and $\boldsymbol{B}_N$ denote $N \times N$ matrices) are said to be *asymptotically equivalent* if

$$\lim_{N\to\infty} \frac{\|\boldsymbol{A}_N - \boldsymbol{B}_N\|_F}{\sqrt{N}} = 0$$

and there exists a constant $M < \infty$ such that

$$\|\boldsymbol{A}_N\|_2, \|\boldsymbol{B}_N\|_2 \le M, \quad \forall N \in \mathbb{N}.$$

Following the convention in Gray's monograph [7], we write $\boldsymbol{A}_N \sim \boldsymbol{B}_N$ if $\{\boldsymbol{A}_N\}$ and $\{\boldsymbol{B}_N\}$ are asymptotically equivalent. This kind of asymptotic equivalence is transitive, i.e., if $\boldsymbol{A}_N \sim \boldsymbol{B}_N$ and $\boldsymbol{B}_N \sim \boldsymbol{C}_N$, then $\boldsymbol{A}_N \sim \boldsymbol{C}_N$. Additional properties of $\sim$ can be found in [7]. The following result concerns the asymptotic eigenvalue behavior of asymptotically equivalent Hermitian matrices.

Theorem II.3. *[7, Theorem 2.4] Let $\{\boldsymbol{A}_N\}$ and $\{\boldsymbol{B}_N\}$ be asymptotically equivalent sequences of Hermitian matrices with eigenvalues $\{\{\lambda_l(\boldsymbol{A}_N)\}_{l\in[N]}\}_{N=1}^{\infty}$ and $\{\{\lambda_l(\boldsymbol{B}_N)\}_{l\in[N]}\}_{N=1}^{\infty}$. Then there exist constants a and b such that*

$$a \le \lambda_l(\boldsymbol{A}_N), \lambda_l(\boldsymbol{B}_N) \le b, \ \forall \ l \in [N], N \in \mathbb{N}.$$

Let ϑ be any function continuous on $[a, b]$. We have

$$\lim_{N\to\infty} \frac{1}{N} \sum_{l=0}^{N-1} \left(\vartheta\left(\lambda_l(\boldsymbol{A}_N)\right) - \vartheta\left(\lambda_l(\boldsymbol{B}_N)\right)\right) = 0.$$

In light of this theorem, Definition II.2 can be viewed as the matrix equivalent of Definition II.1.

B. Asymptotic Equivalence of Circulant and Toeplitz Matrices

Any circulant matrix $\boldsymbol{C}_N$ is characterized by its top row. Let

$$\boldsymbol{e}_f := \begin{bmatrix} e^{j2\pi f0} \\ e^{j2\pi f1} \\ \vdots \\ e^{j2\pi f(N-1)} \end{bmatrix} \in \mathbb{C}^N, \ f \in [0,1]$$

denote a length-N vector of samples from a discrete-time complex exponential signal with digital frequency f. Note that

$$\begin{aligned}\left(\boldsymbol{C}_N \boldsymbol{e}_{(N-l)/N}\right)[k] &= \sum_{n=0}^{N-1} c[n] e^{j2\pi(N-l)(k+n)/N} \\ &= e^{j2\pi(N-l)k/N}\left(\sum_{n=0}^{N-1} c[n] e^{-j2\pi l n/N}\right),\end{aligned}$$

which implies that

$$\boldsymbol{C}_N \boldsymbol{e}_{(N-l)/N} = \left(\sum_{n=0}^{N-1} c[n] e^{-j2\pi l n/N}\right) \boldsymbol{e}_{(N-l)/N}.$$

Thus the normalized DFT basis vectors $\left\{\frac{1}{\sqrt{N}}\boldsymbol{e}_{l/N}\right\}_{l\in[N]}$ are the eigenvectors of any circulant matrix $\boldsymbol{C}_N$, and the corresponding eigenvalues are obtained by taking the DFT of the first row of $\boldsymbol{C}_N$. Specifically,

$$\lambda_l\left(\boldsymbol{C}_N\right) = \sum_{n=0}^{N-1} c[n] e^{-j2\pi l n/N},$$

which can be computed efficiently via the FFT. We note that $\{\lambda_l\left(\boldsymbol{C}_N\right)\}_{l\in[N]}$ are not necessarily arranged in any particular order; namely, they do not necessarily decrease with l.

For a sequence of Toeplitz matrices $\{\boldsymbol{H}_N\}$ and their respective circulant approximations discussed in Section I-D, the following result establishes asymptotic equivalence in terms of the collective behaviors of the eigenvalues. As a reminder, we assume throughout this paper that each $\boldsymbol{H}_N$ is Hermitian; this ensures that all $\boldsymbol{C}_N \in \left\{\widetilde{\boldsymbol{C}}_N, \widehat{\boldsymbol{C}}_N, \overline{\boldsymbol{C}}_N\right\}$ are Hermitian as well.

Lemma II.4. *Suppose that the sequence $h[k]$ is square summable and $\boldsymbol{H}_N, \widetilde{\boldsymbol{C}}_N, \widehat{\boldsymbol{C}}_N, \overline{\boldsymbol{C}}_N$ are absolutely bounded*[5] *for all $N \in \mathbb{N}$. Then*

$$\boldsymbol{H}_N \sim \widehat{\boldsymbol{C}}_N \sim \widetilde{\boldsymbol{C}}_N \sim \overline{\boldsymbol{C}}_N,$$

and

$$\lim_{N\to\infty} \frac{1}{N} \sum_{l=0}^{N-1} \left(\vartheta(\lambda_l(\boldsymbol{H}_N)) - \vartheta(\lambda_l(\boldsymbol{C}_N))\right) = 0,$$

where ϑ is any continuous function on $[a,b]$ and $\boldsymbol{C}_N \in \left\{\widetilde{\boldsymbol{C}}_N, \widehat{\boldsymbol{C}}_N, \overline{\boldsymbol{C}}_N\right\}$. Here $[a,b]$ is the smallest interval that covers all the eigenvalues of $\boldsymbol{H}_N, \widetilde{\boldsymbol{C}}_N, \widehat{\boldsymbol{C}}_N$, and $\overline{\boldsymbol{C}}_N$.

[5] We say a matrix $\boldsymbol{A}$ is absolutely bounded if its spectral norm (or largest singular value) $\|\boldsymbol{A}\|_2$ is bounded.

Proof. See Appendix A.

A stronger result follows simply from the elementary view of Weyl's theory of equal distribution [28], which is presented in Lemma B.1. As a reminder, we do assume that the eigenvalues of each Toeplitz matrix are ordered such that $\lambda_0(\boldsymbol{H}_N) \geq \cdots \geq \lambda_{N-1}(\boldsymbol{H}_N)$.

Lemma II.5. *Suppose that the sequence $h[k]$ is square summable and $\boldsymbol{H}_N, \widetilde{\boldsymbol{C}}_N, \widehat{\boldsymbol{C}}_N, \overline{\boldsymbol{C}}_N$ are absolutely bounded. Let $\lambda_l(\boldsymbol{C}_N)$ be permuted that such that $\lambda_{\rho(0)}(\boldsymbol{C}_N) \geq \lambda_{\rho(1)}(\boldsymbol{C}_N) \geq \cdots \geq \lambda_{\rho(N-1)}(\boldsymbol{C}_N)$. Then*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{l=0}^{N-1} \left|\vartheta(\lambda_l(\boldsymbol{H}_N)) - \vartheta(\lambda_{\rho(l)}(\boldsymbol{C}_N))\right| = 0$$

for every function ϑ that is continuous on $[a,b]$ and $\boldsymbol{C}_N \in \left\{\widetilde{\boldsymbol{C}}_N, \widehat{\boldsymbol{C}}_N, \overline{\boldsymbol{C}}_N\right\}$. Here $[a,b]$ is the smallest interval that covers all the eigenvalues of $\boldsymbol{H}_N, \widetilde{\boldsymbol{C}}_N, \widehat{\boldsymbol{C}}_N$, and $\overline{\boldsymbol{C}}_N$.

Proof. This result follows simply from Lemmas II.4 and B.1. ∎

III. Proofs of Main Theorems

Let $\operatorname{ess}\mathcal{R}(\cdot)$ be the essential range of a function. For any $\Omega \subset \mathbb{R}$, let $\operatorname{int}(\Omega)$ be the interior of the set Ω.

A. Proof of Theorem I.1

We first provide a strong condition under which the equal distribution of two sequences is equivalent to individual asymptotic equivalence.

Theorem III.1. *Assume that the sequences $\{\{u_{N,l}\}_{l\in[N]}\}_{N=1}^{\infty}$ and $\{\{v_{N,l}\}_{l\in[N]}\}_{N=1}^{\infty}$ are absolutely bounded , i.e., there exist a', b' such that $b' \geq u_{N,0} \geq u_{N,1} \geq \cdots \geq u_{N,N-1} \geq a'$ and $b' \geq v_{N,0} \geq v_{N,1} \geq \cdots \geq v_{N,N-1} \geq a'$ for all $N \in \mathbb{N}$. Furthermore, suppose there exists a non-constant continuous function $g(x) : [c,d] \to \mathbb{R}$ such that*

$$\lim_{N\to\infty} u_{N,0} = \lim_{N\to\infty} v_{N,0} = \max_{x\in[c,d]} g(x),$$
$$\lim_{N\to\infty} u_{N,N-1} = \lim_{N\to\infty} v_{N,N-1} = \min_{x\in[c,d]} g(x),$$

and

$$\lim_{N\to\infty} \frac{1}{N} \sum_{l=0}^{N-1} \vartheta(u_{N,l}) = \frac{1}{d-c}\int_c^d \vartheta(g(x))dx < \infty$$

for every function ϑ that is continuous on $[a,b]$, where [a,b] is the smallest interval that covers $[a',b']$ and the range of $g(x)$. Then the following are equivalent:

$$\lim_{N\to\infty} \frac{1}{N} \sum_{l=0}^{N-1} \left(\vartheta(u_{N,l}) - \vartheta(v_{N,l})\right) = 0; \tag{5}$$

$$\lim_{N\to\infty} \max_{l\in[N]} |u_{N,l} - v_{N,l}| = 0. \tag{6}$$

Proof (of Theorem III.1). See Appendix B. ∎

If $\widetilde{h}(f) \equiv C$ is a constant function, then $\boldsymbol{H}_N$ and $\boldsymbol{C}_N$ are diagonal matrices with all diagonals being C for all $\boldsymbol{C}_N \in \left\{\widetilde{\boldsymbol{C}}_N, \widehat{\boldsymbol{C}}_N, \overline{\boldsymbol{C}}_N\right\}$ and $N \in \mathbb{N}$. Thus $\lambda_l\left(\boldsymbol{H}_N\right) = \lambda_l\left(\boldsymbol{C}_N\right) = C$ for all $l \in [N]$ and $\boldsymbol{C}_N \in \left\{\widetilde{\boldsymbol{C}}_N, \widehat{\boldsymbol{C}}_N, \overline{\boldsymbol{C}}_N\right\}$. The following result establishes the range of the eigenvalues of $\overline{\boldsymbol{C}}_N$ and $\boldsymbol{H}_N$ for the case when $\widetilde{h}(f)$ is not a constant function.

Lemma III.2. *Suppose that $\widetilde{h} \in L^\infty([0,1])$ and $\widetilde{h}$ is not a constant function. Also let $\lambda_l(\overline{\boldsymbol{C}}_N)$ be permuted such that $\lambda_{\rho(0)}(\overline{\boldsymbol{C}}_N) \geq \lambda_{\rho(1)}(\overline{\boldsymbol{C}}_N) \geq \cdots \geq \lambda_{\rho(N-1)}(\overline{\boldsymbol{C}}_N)$. Then*

$$\operatorname{ess\,inf} \widetilde{h} < \lambda_{N-1}(\boldsymbol{H}_N) < \lambda_{\rho(N-1)}(\overline{\boldsymbol{C}}_N)$$

and

$$\lambda_{\rho(0)}(\overline{\boldsymbol{C}}_N) \leq \lambda_0(\boldsymbol{H}_N) < \operatorname{ess\,sup} \widetilde{h}.$$

Proof (of Lemma III.2). We first rewrite $\lambda_l\left(\overline{\boldsymbol{C}}_N\right)$ as

$$\begin{aligned}
\lambda_l\left(\overline{\boldsymbol{C}}_N\right) &= \sum_{n=0}^{N-1} \overline{c}[n] e^{\frac{-j2\pi l n}{N}} \\
&= \sum_{n=0}^{N-1} \frac{1}{N}\left((N-n)h[-n] + nh[N-n]\right) e^{\frac{-j2\pi l n}{N}} \\
&= \left\langle \boldsymbol{H}_N \frac{1}{\sqrt{N}} \boldsymbol{e}_{l/N}, \frac{1}{\sqrt{N}} \boldsymbol{e}_{l/N} \right\rangle.
\end{aligned}$$

By definition, $\lambda_0(\boldsymbol{H}_N) = \max_{\|\boldsymbol{v}\|_2=1} \langle \boldsymbol{H}_N \boldsymbol{v}, \boldsymbol{v}\rangle$ and $\lambda_{N-1}(\boldsymbol{H}_N) = \min_{\|\boldsymbol{v}\|_2=1} \langle \boldsymbol{H}_N \boldsymbol{v}, \boldsymbol{v}\rangle$, we obtain

$$\lambda_{N-1}(\boldsymbol{H}_N) \leq \lambda_l(\overline{\boldsymbol{C}}_N) \leq \lambda_0(\boldsymbol{H}_N), \ \forall\ l.$$

For arbitrary $\boldsymbol{v} \in \mathbb{C}^N$, $\|\boldsymbol{v}\|_2 = 1$, we extend $\boldsymbol{v}$ to an infinite sequence $v[n], n \in \mathbb{Z}$ by zero-padding. Then

$$\begin{aligned}
\langle \boldsymbol{H}_N \boldsymbol{v}, \boldsymbol{v}\rangle &= \sum_{m=0}^{N-1} \boldsymbol{v}^*[m] \sum_{n=0}^{N-1} h[m-n] \boldsymbol{v}[n] \\
&= \sum_{m=-\infty}^{\infty} v^*[m] \sum_{n=-\infty}^{\infty} h[m-n] v[n] \\
&= \int_0^1 \left|\widetilde{\boldsymbol{v}}(f)\right|^2 \widetilde{h}(f) df
\end{aligned}$$

where $\widetilde{\boldsymbol{v}}(f) = \sum_{n=0}^{N-1} \boldsymbol{v}[n] e^{j2\pi f n}$. If $\widetilde{h}(f)$ is not a constant function of $[0,1]$, we conclude

$$\begin{aligned}
\operatorname{ess\,inf} \widetilde{h} &= \int_0^1 \left|\widetilde{\boldsymbol{v}}(f)\right|^2 df \cdot \operatorname{ess\,inf} \widetilde{h} < \langle \boldsymbol{H}_N \boldsymbol{v}, \boldsymbol{v}\rangle \\
&< \int_0^1 \left|\widetilde{\boldsymbol{v}}(f)\right|^2 df \cdot \operatorname{ess\,sup} \widetilde{h} = \operatorname{ess\,sup} \widetilde{h}.
\end{aligned}$$

■

Theorem I.1 holds trivially when $\widetilde{h}(f)$ is a constant function since for this case $\boldsymbol{H}_N$ and $\boldsymbol{C}_N$ have the same eigenvalues for all $\boldsymbol{C}_N \in \left\{\widetilde{\boldsymbol{C}}_N, \widehat{\boldsymbol{C}}_N, \overline{\boldsymbol{C}}_N\right\}$ and $N \in \mathbb{N}$. In what follows, we suppose $\widetilde{h}(f)$ is not a constant function. The assumption of absolute summability of the sequence $h[k]$ indicates that its DTFT $\widetilde{h}(f)$ is continuous on $[0,1]$, and moreover, its partial Fourier sum $S_N(f)$ converges uniformly to $\widetilde{h}(f)$ on $[0,1]$ as $N \to \infty$ [23]. Thus, given $\epsilon > 0$, there exists $N_0 \in \mathbb{N}$ such that

$$\left|\widetilde{h}(f) - S_{N-1}(f)\right| \leq \epsilon$$

for all $f \in [0,1]$ and $N \geq N_0$. The Cesàro sum $\sigma_N(f)$ also converges to $\widetilde{h}(f)$ uniformly on $[0,1]$ as $N \to \infty$.

Since the eigenvalues of $\widetilde{\boldsymbol{C}}_N$ and $\widehat{\boldsymbol{C}}_N$ are, respectively, the samples of $S_{N-1}(f)$ and $S_{\lfloor \frac{N-1}{2} \rfloor}(f)$, we conclude that $\widetilde{\boldsymbol{C}}_N$ and $\widehat{\boldsymbol{C}}_N$ are absolutely bounded. Lemma III.2 implies that $\overline{\boldsymbol{C}}_N$ and $\boldsymbol{H}_N$ are also absolutely bounded.

We next show $\lim_{N\to\infty} \max_l \lambda_l\left(\boldsymbol{C}_N\right) = \max_{f\in[0,1]} \widetilde{h}(f)$ for all $\boldsymbol{C}_N \in \left\{\widetilde{\boldsymbol{C}}_N, \widehat{\boldsymbol{C}}_N, \overline{\boldsymbol{C}}_N\right\}$. The extreme value theorem states that $\widetilde{h}(f)$ must attain a maximum and a minimum each at least once since $\widetilde{h}(f) \in \mathbb{R}$ is continuous on $[0,1]$. Let

$$\widehat{f} := \arg\max_f \widetilde{h}(f)$$

denote any point at which $\widetilde{h}$ achieves its maximum value. Also let

$$\widehat{l}_N := \arg\min_{l\in[N]} \left|\widehat{f} - \frac{l}{N}\right|$$

denote any closest on-grid point to $\widehat{f}$. For arbitrary $\epsilon > 0$, by uniform convergence, there exists N_0 such that

$$\left|\lambda_{\widehat{l}_N}\left(\boldsymbol{C}_N\right) - \widetilde{h}\left(\frac{\widehat{l}_N}{N}\right)\right| \leq \epsilon$$

for all $N \geq N_0$. Noting that $\left|\frac{\widehat{l}_N}{N} - \widehat{f}\right| \leq \frac{1}{2N}$ and $\widetilde{h}$ is continuous on $[0,1]$, there exists $N_1 \in \mathbb{N}$ so that

$$\left|\widetilde{h}\left(\widehat{f}\right) - \widetilde{h}\left(\frac{\widehat{l}_N}{N}\right)\right| \leq \epsilon$$

when $N \geq N_1$. Thus we conclude

$$\left|\lambda_{\widehat{l}_N}\left(\boldsymbol{C}_N\right) - \widetilde{h}\left(\widehat{f}\right)\right| \leq 2\epsilon$$

for all $N \geq \max\{N_0, N_1\}$. Since ϵ is arbitrary,

$$\lim_{N\to\infty} \max_l \lambda_l\left(\boldsymbol{C}_N\right) = \max_{f\in[0,1]} \widetilde{h}(f)$$

for all $\boldsymbol{C}_N \in \left\{\widetilde{\boldsymbol{C}}_N, \widehat{\boldsymbol{C}}_N, \overline{\boldsymbol{C}}_N\right\}$. Noting that $\lambda_l\left(\overline{\boldsymbol{C}}_N\right) \leq \lambda_0(\boldsymbol{H}_N) \leq \max_{f\in[0,1]} \widetilde{h}(f)$, we obtain

$$\lim_{N\to\infty} \max_l \lambda_l\left(\boldsymbol{H}_N\right) = \max_{f\in[0,1]} \widetilde{h}(f).$$

The asymptotic argument for the smallest eigenvalues can be obtained with a similar approach. It follows from Lemma II.4 that $\boldsymbol{H}_N \sim \widehat{\boldsymbol{C}}_N \sim \widetilde{\boldsymbol{C}}_N \sim \overline{\boldsymbol{C}}_N$ and from Szegő's theorem (1) that

$$\lim_{N\to\infty} \frac{1}{N} \sum_{l=0}^{N-1} \vartheta(\lambda_l(\boldsymbol{H}_N)) = \int_0^1 \vartheta(\widetilde{h}(f))df.$$

Finally, the proof of Theorem I.1 is completed by applying Theorem III.1 with $g = \widetilde{h}$. ∎

B. Proof of Theorem I.2

The proof is given in Appendix C. We outline the main idea here. Let $[\boldsymbol{H}_N]_{N-r}$ be the $(N-r)\times(N-r)$ matrix obtained by deleting the last r columns and the last r rows of $\boldsymbol{H}_N$. Similar notation holds for $[\widetilde{\boldsymbol{C}}_N]_{N-r}$. Note that $[\boldsymbol{H}]_{N-r}$ and $[\widetilde{\boldsymbol{C}}_N]_{N-r}$ have the same eigenvalues when $N > 2r$ since $[\boldsymbol{H}]_{N-r}$ is exactly the same as $[\widetilde{\boldsymbol{C}}_N]_{N-r}$. Also $\widehat{C}_N$ is equivalent to $\widetilde{C}_N$ when $N > 2r$. We first apply the Sturmian separation theorem for the Toeplitz and circulant matrices to obtain a bound on the distance between $\lambda_l(\boldsymbol{H}_N)$ and $\lambda_{\rho(l)}(\widetilde{\boldsymbol{C}}_N)$. We then utilize the fact that $\widetilde{h}(f)$ is Lipschitz continuous to guarantee the closeness between $\lambda_l(\widetilde{\boldsymbol{C}}_N)$ and $\lambda_{l+r}(\widetilde{\boldsymbol{C}}_N)$. Finally, we show $\lambda_l(\widetilde{\boldsymbol{C}}_N)$ is close to $\lambda_l(\overline{\boldsymbol{C}}_N)$ since the Cesàro sum and partial Fourier sum converge to the same function in this case. ∎

C. Proof of Theorem I.3

We first provide another condition (which, informally speaking, is weaker than that in Theorem III.1) under which the equal distribution of two sequences implies individual asymptotic equivalence.

Theorem III.3. *Assume that* $b \geq u_{N,0} \geq u_{N,1} \geq \cdots \geq u_{N,N-1} \geq a$ *and* $b \geq v_{N,0} \geq v_{N,1} \geq \cdots \geq v_{N,N-1} \geq a$. *Furthermore, suppose there is a Riemann integrable function* $g(x) : [c,d] \to [a,b]$ *such that*

$$u_{N,l}, v_{N,l} \in \text{int}\left(\text{ess}\,\mathcal{R}(g)\right), \ \ \forall \ l \in [N], N \in \mathbb{N},$$

and

$$\lim_{N\to\infty} \frac{1}{N} \sum_{l=0}^{N-1} \vartheta(u_{N,l}) = \frac{1}{d-c}\int_c^d \vartheta(g(x))dx < \infty$$

for all ϑ *that are continuous on* $[a,b]$. *Then the following are equivalent:*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{l=0}^{N-1} \left(\vartheta(u_{N,l}) - \vartheta(v_{N,l})\right) = 0; \tag{7}$$

$$\lim_{N\to\infty} \max_l |u_{N,l} - v_{N,l}| = 0. \tag{8}$$

Proof (of Theorem III.3). See Appendix D. ∎

If $\widetilde{h}(f) \equiv C$ is a constant function, then $\lambda_l\left(\boldsymbol{H}_N\right) = \lambda_l\left(\overline{\boldsymbol{C}}_N\right) = C$ for all $l \in [N]$. Thus Theorem I.3 holds trivially. On the other hand, suppose that $\widetilde{h} \in L^\infty([0,1])$ is not a constant function and the essential range of $\widetilde{h}$ is $\left[\text{ess}\inf \widetilde{h}, \text{ess}\sup \widetilde{h}\right]$. It follows from Lemma III.2 that $\lambda_l\left(\boldsymbol{H}_N\right), \lambda_l\left(\overline{\boldsymbol{C}}_N\right) \in \text{int}\left(\mathcal{R}\left(\widetilde{h}\right)\right)$ for all $l \in [N]$ and $N \in \mathbb{N}$. Using Lemma II.4 and Szegő's theorem (see (1)), the fact that $h[k]$ is square summable together with the fact that $\boldsymbol{H}_N, \overline{\boldsymbol{C}}_N$ are absolutely bounded imply

$$\lim_{N\to\infty} \frac{1}{N} \sum_{l=0}^{N-1} \vartheta(\lambda_l(\boldsymbol{H}_N)) = \int_0^1 \vartheta(\widetilde{h}(f))df,$$

and

$$\lim_{N\to\infty} \frac{1}{N} \sum_{l=0}^{N-1} \left(\vartheta(\lambda_l(\boldsymbol{H}_N)) - \vartheta(\lambda_l(\overline{\boldsymbol{C}}_N))\right) = 0$$

for all ϑ that are continuous on $\left[\text{ess}\inf \widetilde{h}, \text{ess}\sup \widetilde{h}\right]$. Finally, (4) follows from Theorem III.3 with $g = \widetilde{h}$, $u_{N,l} = \lambda_l(\boldsymbol{H}_N)$ and $v_{N,l} = \lambda_{\rho(l)}(\overline{\boldsymbol{C}}_N)$. This completes the proof of Theorem I.3. ∎

Remark. Theorem III.1 requires that g is continuous and that the extreme values of the sequences asymptotically converge to the extreme values of g (but meanwhile the extreme values of the sequences can be outside of the range of g). Theorem III.3 requires the sequences to be strictly inside the range of g.

D. Proof of Theorem I.4

Our proof of Theorem I.4 appears in Appendix E. ∎

Remark. Theorem I.4 works only for the circulant matrix $\overline{\boldsymbol{C}}_N$ and not $\widetilde{\boldsymbol{C}}_N$ or $\widehat{\boldsymbol{C}}_N$. This is closely related to the fact that the partial Cesàro sum has better convergence than the partial Fourier sum [23].

Remark. Theorem I.4 only requires $\widetilde{h}$ to be bounded and Riemann integrable, while Theorem I.3 requires the range of $\widetilde{h}$ to be connected.

IV. Simulations

In this section, we provide several examples to illustrate our theory. In the legends of Figures 1–3, we refer to the circulant approximations $\widetilde{\boldsymbol{C}}_N, \widehat{\boldsymbol{C}}_N$, and $\overline{\boldsymbol{C}}_N$ as Circulant1, Circulant2, and Circulant3, respectively.

A. $h[k] = W\left(\frac{\sin(\pi W k)}{\pi k}\right)^2, W = \frac{1}{4}$

In our first example, the sequence $h[k]$ is absolutely summable and the corresponding symbol

$$\widetilde{h}(f) = \text{tri}(\frac{f}{W}) = \begin{cases} 1 - \frac{f}{W}, & 0 \leq f \leq W \\ 1 - \frac{1-f}{W}, & 1-W \leq f \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

is a triangular signal, which is continuous on $[0,1]$. Figure 1(a) shows $\widetilde{h}$, Figure 1(b) shows $\lambda_l(\boldsymbol{H}_N)$, $\lambda_{\rho(l)}(\widetilde{\boldsymbol{C}}_N)$, $\lambda_{\rho(l)}(\widehat{\boldsymbol{C}}_N)$ and $\lambda_{\rho(l)}(\overline{\boldsymbol{C}}_N)$ for $N = 500$, and Figure 1(c) shows $\max_{l\in[N]} \left|\lambda_l(\boldsymbol{H}_N) - \lambda_{\rho(l)}(\boldsymbol{C}_N)\right|$ against the dimension N for all $\boldsymbol{C}_N \in \left\{\widetilde{\boldsymbol{C}}_N, \widehat{\boldsymbol{C}}_N, \overline{\boldsymbol{C}}_N\right\}$. As guaranteed by Theorem I.1, it can be observed in Figure 1(c) that the individual asymptotic convergence of eigenvalues holds for all $\widetilde{\boldsymbol{C}}_N, \widehat{\boldsymbol{C}}_N$, and $\overline{\boldsymbol{C}}_N$.

B. $h[k] = \frac{1+(-1)^k}{j2\pi k}$

In this case, the sequence $h[k]$ is not absolutely summable and the symbol

$$\widetilde{h}(f) = \begin{cases} 2f, & 0 < f \leq \frac{1}{2} \\ 2f - 1, & \frac{1}{2} < f \leq 1 \end{cases}$$

is not continuous, but its range is connected. Figure 2(a) shows $\widetilde{h}$, Figure 2(b) shows $\lambda_l(\boldsymbol{H}_N)$, $\lambda_{\rho(l)}(\widetilde{\boldsymbol{C}}_N)$, $\lambda_{\rho(l)}(\widehat{\boldsymbol{C}}_N)$ and $\lambda_{\rho(l)}(\overline{\boldsymbol{C}}_N)$ for $N = 500$, and Figure 2(c) shows $\max_{l\in[N]} \left|\lambda_l(\boldsymbol{H}_N) - \lambda_{\rho(l)}(\boldsymbol{C}_N)\right|$ against the dimension N for all $\boldsymbol{C}_N \in \left\{\widetilde{\boldsymbol{C}}_N, \widehat{\boldsymbol{C}}_N, \overline{\boldsymbol{C}}_N\right\}$. It is observed from Figure 2(c) that the individual asymptotic convergence of the eigenvalues holds for $\overline{\boldsymbol{C}}_N$—as guaranteed by Theorem I.3—but not for $\widetilde{\boldsymbol{C}}_N$ and $\widehat{\boldsymbol{C}}_N$. Figure 2(c) also shows that the errors $\max_{l\in[N]} \left|\lambda_l(\boldsymbol{H}_N) - \lambda_{\rho(l)}(\widetilde{\boldsymbol{C}}_N)\right|$ and $\max_{l\in[N]} \left|\lambda_l(\boldsymbol{H}_N) - \lambda_{\rho(l)}(\widehat{\boldsymbol{C}}_N)\right|$ converge to the size of the Gibbs jump (≈ 0.089).

C. $h[k] = \frac{\sin(2\pi W k)}{\pi k}, W = \frac{1}{4}$

In this example, the sequence $h[k]$ is not absolutely summable and the symbol

$$\widetilde{h}(f) = \begin{cases} 0, & W < f \leq 1 - W, \\ 1, & \text{otherwise}, \end{cases}$$

is a rectangular window function, which is not continuous and whose range is not connected. Figure 3(a) shows $\widetilde{h}$, Figure 3(b) shows $\lambda_l(\boldsymbol{H}_N)$, $\lambda_{\rho(l)}(\widetilde{\boldsymbol{C}}_N)$, $\lambda_{\rho(l)}(\widehat{\boldsymbol{C}}_N)$ and $\lambda_{\rho(l)}(\overline{\boldsymbol{C}}_N)$ for $N = 2048$, and Figure 3(c) shows $\max_{l\in[N]} \left|\lambda_l(\boldsymbol{H}_N) - \lambda_{\rho(l)}(\boldsymbol{C}_N)\right|$ against the dimension N for all $\boldsymbol{C}_N \in \left\{\widetilde{\boldsymbol{C}}_N, \widehat{\boldsymbol{C}}_N, \overline{\boldsymbol{C}}_N\right\}$. Figure 3(c) illustrates that the individual asymptotic convergence of eigenvalues does not hold for the circulant matrices $\widetilde{\boldsymbol{C}}_N$, $\widehat{\boldsymbol{C}}_N$, and $\overline{\boldsymbol{C}}_N$. Indeed, the sequence $h[k]$ does not meet the assumptions in either Theorem I.1 or Theorem I.3.

Due to the gap (between 0 to 1) in the range of the window function $\widetilde{h}$, the eigenvalues of $\boldsymbol{H}_N$ and $\overline{\boldsymbol{C}}_N$ have different behavior in the transition region. To better illustrate this, Figures 3(d) and 3(e), respectively, show $\lambda_l(\boldsymbol{H}_N)$ and $\lambda_{\rho(l)}(\overline{\boldsymbol{C}}_N)$ for $N = 2048$. We see that the eigenvalues of the Toeplitz matrix $\boldsymbol{H}_N$ cover the range $[0,1]$ somewhat uniformly, while the eigenvalues of the $\overline{\boldsymbol{C}}_N$ tend to cluster around 0, $1/2$, and 1 (there are none near $1/4$ or $3/4$). The following result formally explains the transition behavior of the eigenvalues of $\boldsymbol{H}_N$.

Lemma IV.1. *[6, 29, 30] Let $h[k] = \frac{\sin(2\pi W k)}{\pi k}$ with $W = \frac{1}{4}$. Fix $\epsilon \in (0, \frac{1}{2})$. Then there exist constants C_1, C_2 and N_1 such that the distance between any 2 consecutive eigenvalues of $\boldsymbol{H}_N$ inside $(\epsilon, 1-\epsilon)$ is bounded from below by $\frac{C_1}{\ln(N)}$ and from above by $\frac{C_2}{\ln(N)}$; that is*

$$\frac{C_1}{\ln(N)} \leq \lambda_l(\boldsymbol{H}_N) - \lambda_{l+1}(\boldsymbol{H}_N) \leq \frac{C_2}{\ln(N)}$$

for all $\epsilon \leq \lambda_{l+1}(\boldsymbol{H}_N) \leq \lambda_l(\boldsymbol{H}_N) \leq 1 - \epsilon$ and $N \geq N_1$. Also

$$\lambda_{\lfloor \frac{1}{2}N \rfloor - 1} \geq \frac{1}{2} \geq \lambda_{\lceil \frac{1}{2}N \rceil}$$

for all $N \in \mathbb{N}$.

On the other hand, we have the following result on the eigenvalues of $\overline{\boldsymbol{C}}_N$.

Lemma IV.2. *Let $h[k] = \frac{\sin(2\pi W k)}{\pi k}$ with $W = \frac{1}{4}$. Then*

$$\left|\lambda_l\left(\overline{\boldsymbol{C}}_N\right) - \frac{1}{2}\right| \begin{cases} = 0, & l = N/4, 3N/4, \\ \geq \alpha, & l \in [N] \text{ and } l \neq N/4, 3N/4, \end{cases}$$

with $\alpha = 0.4$ if N is a multiple of 4.

Proof. See Appendix F. ■

With more sophisticated analysis, we believe that the above result could be improved to $\alpha \approx 0.45$. This is suggested by Figure 3(e).

Combining Lemmas IV.1 and IV.2, we conclude that $\max_{l\in[N]} \left|\lambda_l(\boldsymbol{H}_N) - \lambda_{\rho(l)}(\overline{\boldsymbol{C}}_N)\right|$ approaches ≈ 0.2 as $N \to \infty$ and N is a multiple of 4.

Finally, Figure 3(f) plots $\left|\lambda_0(\boldsymbol{H}_N) - \lambda_{\rho(0)}(\overline{\boldsymbol{C}}_N)\right|$ and $\left|\lambda_{N-1}(\boldsymbol{H}_N) - \lambda_{\rho(N-1)}(\overline{\boldsymbol{C}}_N)\right|$ against the dimension N. As can be observed, the largest and smallest eigenvalues of $\overline{\boldsymbol{C}}_N$ converge to the largest and smallest eigenvalues of $\boldsymbol{H}_N$, respectively. This is as guaranteed by Theorem I.4.

V. Conclusions

It is well known that any sequence of uniformly bounded Hermitian Toeplitz matrices is asymptotically equivalent to certain sequences of circulant matrices derived from the Toeplitz matrices. We have provided conditions under which the asymptotic equivalence of the matrices implies the individual asymptotic convergence of the eigenvalues. Our results suggest that instead of directly computing the eigenvalues of a Toeplitz matrix, one can compute a fast spectrum approximation using the FFT. This is long known, but we provide new guarantees for the asymptotic convergence of the individual eigenvalues. Some numerical examples have demonstrated the dependence of the convergence behavior on the properties of the symbol of the Toeplitz matrix. An interesting question would be whether it is possible to extend our analysis to general (non-Hermitian) Toeplitz matrices, along the lines of the Avram-Parter theorem [8, 9]. In addition, it would also be of interest to extend our analysis to the asymptotic equivalence of block Toeplitz and block circulant matrices.

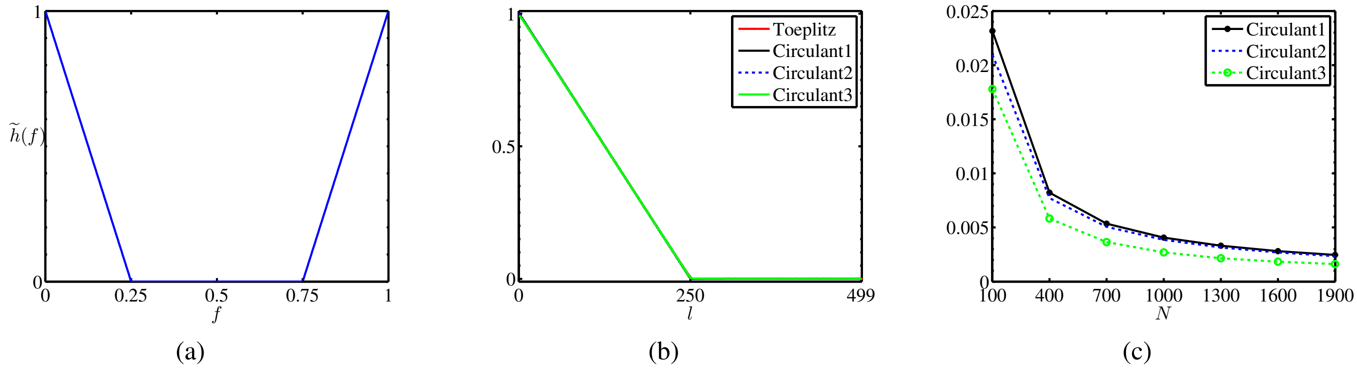

Fig. 1. (a) Illustration of a continuous symbol $\widetilde{h}(f)$. (b) The eigenvalues of the Toeplitz matrix $\boldsymbol{H}_N$ and the circulant approximations $\widetilde{\boldsymbol{C}}_N, \widehat{\boldsymbol{C}}_N$, and $\overline{\boldsymbol{C}}_N$, arranged in decreasing order. Here $N = 500$. (c) A plot of $\max_{l \in [N]} \left|\lambda_l(\boldsymbol{H}_N) - \lambda_{\rho(l)}(\boldsymbol{C}_N)\right|$ versus the dimension N for all $\boldsymbol{C}_N \in \left\{\widetilde{\boldsymbol{C}}_N, \widehat{\boldsymbol{C}}_N, \overline{\boldsymbol{C}}_N\right\}$.

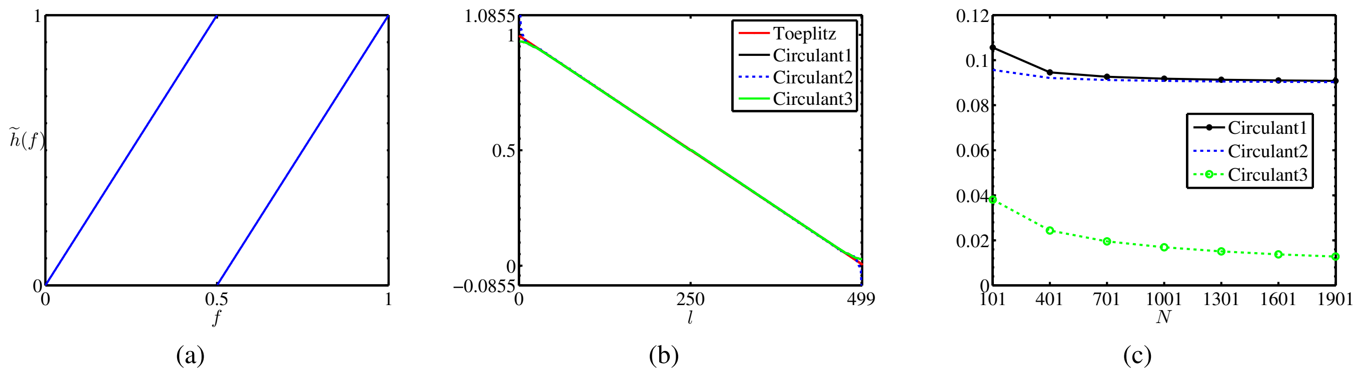

Fig. 2. (a) Illustration of a discontinuous symbol $\widetilde{h}(f)$. (b) The eigenvalues of the Toeplitz matrix $\boldsymbol{H}_N$ and the circulant approximations $\widetilde{\boldsymbol{C}}_N, \widehat{\boldsymbol{C}}_N$, and $\overline{\boldsymbol{C}}_N$, arranged in decreasing order. Here $N = 500$. (c) A plot of $\max_{l \in [N]} \left|\lambda_l(\boldsymbol{H}_N) - \lambda_{\rho(l)}(\boldsymbol{C}_N)\right|$ versus the dimension N for all $\boldsymbol{C}_N \in \left\{\widetilde{\boldsymbol{C}}_N, \widehat{\boldsymbol{C}}_N, \overline{\boldsymbol{C}}_N\right\}$.

Appendix A
Proof of Lemma II.4

It follows from the definition of $\widehat{\boldsymbol{C}}_N$ that

$$
\begin{aligned}
&\left\|\boldsymbol{H}_N - \widehat{\boldsymbol{C}}_N\right\|_F^2 \\
=&\sum_{k=1}^{\lfloor \frac{N-1}{2} \rfloor} k\left(\left|h[k] - h[-N+k]\right|^2 + \left|h[-k] - h[N-k]\right|^2\right) \\
&+ \sum_{k=\lfloor \frac{N-1}{2} \rfloor + 1}^{\lfloor N/2 \rfloor} k\left(\left|h[k]\right|^2 + \left|h[-k]\right|^2\right) \\
\leq &\sum_{k=1}^{\lfloor N/2 \rfloor} 2k\left(\left|h[k]\right|^2 + \left|h[-k]\right|^2 + \left|h[N-k]\right|^2 + \left|h[k-N]\right|^2\right) \\
\leq &\sum_{k=1}^{N-1} 2k\left(\left|h[k]\right|^2 + \left|h[-k]\right|^2\right).
\end{aligned}
$$

Fix $\epsilon > 0$. By assumption that the sequence $h[k]$ is square summable, there exists N_0 such that

$$
\sum_{k=N_0}^{\infty} \left|h[k]\right|^2 + \left|h[-k]\right|^2 \leq \epsilon.
$$

Thus we have

$$
\begin{aligned}
&\frac{1}{N}\left\|\boldsymbol{H}_N - \widehat{\boldsymbol{C}}_N\right\|_F^2 \\
\leq &\frac{1}{N}\sum_{k=1}^{N_0-1} 2k\left(\left|h[k]\right|^2 + \left|h[-k]\right|^2\right) \\
&+ \frac{1}{N}\sum_{k=N_0}^{N} 2k\left(\left|h[k]\right|^2 + \left|h[-k]\right|^2\right) \\
\leq &\frac{1}{N}\sum_{k=1}^{N_0-1} 2k\left(\left|h[k]\right|^2 + \left|h[-k]\right|^2\right) + 2\sum_{k=N_0}^{N}\left(\left|h[k]\right|^2 + \left|h[-k]\right|^2\right) \\
\leq &\epsilon + 2\epsilon = 3\epsilon
\end{aligned}
$$

when $N \geq \max\{N_0, N_1\}$ with $N_1 \geq \sum_{k=1}^{N_0-1} 2k\left(\left|h[k]\right|^2 + \left|h[-k]\right|^2\right)/\epsilon$. Since ϵ is arbitrary,

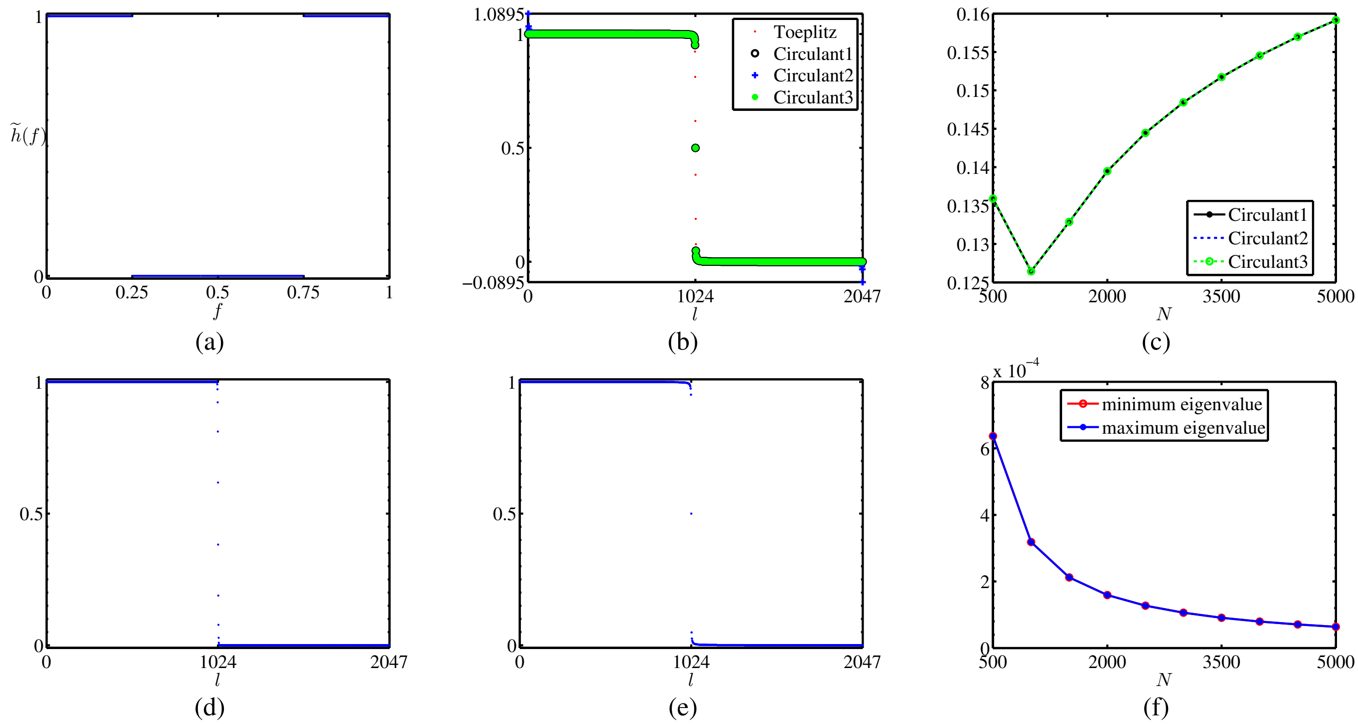

Fig. 3. (a) Illustration of a discontinuous symbol $\widetilde{h}(f)$ whose range is not connected. (b) The eigenvalues of the Toeplitz matrix $\boldsymbol{H}_N$ and the circulant approximations $\widetilde{\boldsymbol{C}}_N, \widehat{\boldsymbol{C}}_N$, and $\overline{\boldsymbol{C}}_N$, arranged in decreasing order. Here $N = 2048$. (c) A plot of $\max_{l\in[N]} \left|\lambda_l(\boldsymbol{H}_N) - \lambda_{\rho(l)}(\boldsymbol{C}_N)\right|$ versus the dimension N for all $\boldsymbol{C}_N \in \left\{\widetilde{\boldsymbol{C}}_N, \widehat{\boldsymbol{C}}_N, \overline{\boldsymbol{C}}_N\right\}$. (d) The eigenvalues of the Toeplitz matrix $\boldsymbol{H}_N$. (e) The eigenvalues of the circulant matrix $\overline{\boldsymbol{C}}_N$, arranged in decreasing order. (f) A plot of $\left|\lambda_0(\boldsymbol{H}_N) - \lambda_{\rho(0)}(\overline{\boldsymbol{C}}_N)\right|$ and $\left|\lambda_{N-1}(\boldsymbol{H}_N) - \lambda_{\rho(N-1)}(\overline{\boldsymbol{C}}_N)\right|$ versus the dimension N.

we obtain

$$\lim_{N\to\infty} \frac{1}{N}\left\|\boldsymbol{H}_N - \widehat{\boldsymbol{C}}_N\right\|_F^2 = 0.$$

Noting that $\boldsymbol{H}_N$ and $\widehat{\boldsymbol{C}}_N$ are absolutely bounded by assumption, we conclude $\boldsymbol{H}_N \sim \widehat{\boldsymbol{C}}_N$. The proofs of $\boldsymbol{H}_N \sim \widetilde{\boldsymbol{C}}_N$ and $\boldsymbol{H}_N \sim \overline{\boldsymbol{C}}_N$ follow from the same approach. ■

Appendix B
Proof of Theorem III.1

Set

$$F_g(\alpha) := \frac{1}{d-c}\mu\left\{x \in [c,d] : g(x) \le \alpha\right\}, \tag{9}$$

$$F_{u_N}(\alpha) := \frac{1}{N}\#\left\{l \in [N], u_{N,l} \le \alpha\right\},$$

$$F_{v_N}(\alpha) := \frac{1}{N}\#\left\{l \in [N], v_{N,l} \le \alpha\right\}.$$

Here, $\mu(E)$ is the Lebsegue measure of a subset $E \in \mathbb{R}$.

Definition II.1 states that the sequences $\{\{u_{N,l}\}_{l\in[N]}\}_{N=1}^{\infty}$ and $\{\{v_{N,l}\}_{l\in[N]}\}_{N=1}^{\infty}$ are asymptotically equally distributed if

$$\lim_{N\to\infty} \frac{1}{N}\sum_{l=0}^{N-1}\left(\vartheta\left(u_{N,l}\right) - \vartheta\left(v_{N,l}\right)\right) = 0$$

for all ϑ that are continuous on $[a,b]$. Here $[a,b]$ is the smallest interval that covers the sequences $\{\{u_{N,l}\}_{l\in[N]}\}_{N=1}^{\infty}$ and $\{\{v_{N,l}\}_{l\in[N]}\}_{N=1}^{\infty}$.

Trench [28] strengthens this definition by showing the following result.

Lemma B.1. *[28, Asymptotcially (absolutely) equal distribution] Assume that $b \ge u_{N,0} \ge u_{N,1} \ge \cdots \ge u_{N,N-1} \ge a$ and $b \ge v_{N,0} \ge v_{N,1} \ge \cdots \ge v_{N,N-1} \ge a$. The following are equivalent:*

1) $\lim_{N\to\infty} \frac{1}{N}\sum_{l=0}^{N-1}\left(\vartheta\left(u_{N,l}\right) - \vartheta\left(v_{N,l}\right)\right) = 0$ *for all ϑ that are continuous on $[a,b]$;*
2) $\lim_{N\to\infty} \frac{1}{N}\sum_{l=0}^{N-1}\left|\vartheta\left(u_{N,l}\right) - \vartheta\left(v_{N,l}\right)\right| = 0$ *for all ϑ that are continuous on $[a,b]$.*

Here the sequences $\{\{u_{N,l}\}_{l\in[N]}\}_{N=1}^{\infty}$ and $\{\{v_{N,l}\}_{l\in[N]}\}_{N=1}^{\infty}$ are said to be *absolutely asymptotically equally distributed* [28] if

$$\lim_{N\to\infty} \frac{1}{N}\sum_{l=0}^{N-1}\left|\vartheta\left(u_{N,l}\right) - \vartheta\left(v_{N,l}\right)\right| = 0$$

for all ϑ that are continuous on $[a,b]$.

Viewing $g : [c,d] \to \mathbb{R}$ as a random variable, in probabilistic language, F_g is the cumulative distribution function (CDF) associated to g. Also F_{u_N} and F_{v_N} can be viewed as the CDF of the discrete random variables $\boldsymbol{u}_N : \{0,1,\ldots,N-1\} \to \mathbb{R}$

defined by $\boldsymbol{u}_N(l) = u_{N,l}$ and $\boldsymbol{v}_N : \{0, 1, \ldots, N-1\} \to \mathbb{R}$ defined by $\boldsymbol{v}_N(l) = v_{N,l}$, respectively. It is well known that the CDF of a random variable is right continuous and non-decreasing. The following result, known as the Portmanteau Lemma, gives a number of equal descriptions of weak convergence in terms of the CDF and the means of the random variables.

Lemma B.2. *[31, Portmanteau Lemma] The following are equivalent:*

1) $\lim_{N\to\infty} \frac{1}{N} \sum_{l=0}^{N-1} \vartheta(u_{N,l}) = \frac{1}{d-c} \int_c^d \vartheta(g(x))dx$*, for all bounded, continuous functions* ϑ*;*
2) $\lim_{N\to\infty} F_{u_N}(\alpha) = F_g(\alpha)$ *for every point* α *at which* F_g *is continuous.*

Despite the fact that $F_g(\alpha)$ is right continuous and non-decreasing everywhere, some stronger results about $F_g(\alpha)$ can be obtained by utilizing the fact that g is continuous on $[c, d]$.

Lemma B.3. *Let* $F_g(\alpha)$ *be defined as in (9). Then* $F_g(\alpha)$ *is strictly increasing on* $\mathcal{R}(g)$*, i.e., for every* $\alpha \in int(\mathcal{R}(g))$*, there exists* $\epsilon > 0$ *such that, for each pair* (α_1, α_2) *satisfying*

$$\min_{x\in[c,d]} g(x) \le \alpha - \epsilon < \alpha_1 < \alpha < \alpha_2 < \alpha + \epsilon \le \max_{x\in[c,d]} g(x),$$

we have

$$F_g(\alpha_1) < F_g(\alpha) < F_g(\alpha_2).$$

Proof (of Lemma B.3). Since $g(x) : [c, d] \to \mathbb{R}$ is continuous, there exists ϵ such that $(\alpha - \epsilon, \alpha + \epsilon) \subset \mathcal{R}(g)$ for $\alpha \in \text{int}(\mathcal{R}(g))$. Let α_1 be an arbitrary value such that $\alpha - \epsilon < \alpha_1 < \alpha$ and let $\alpha_1' = \frac{\alpha+\alpha_1}{2} \in \mathcal{R}(g)$. Noting that g is continuous, we have

$$\mu\left\{x \in [c, d] : |g(x) - \alpha_1'| < \frac{\alpha - \alpha_1}{2}\right\} > 0.$$

Thus, we obtain

$$\begin{aligned} F_g(\alpha) - F_g(\alpha_1) &= \frac{1}{d-c}\mu\{x \in [c, d] : \alpha_1 < g(x) \le \alpha\} \\ &\ge \frac{1}{d-c}\mu\{x \in [c, d] : \alpha_1 < g(x) < \alpha\} > 0. \end{aligned}$$

Similarly, we have $F_g(\alpha) < F_g(\alpha_2)$ for $\alpha < \alpha_2 < \alpha + \epsilon$. ■

We are now ready to prove the main part. First we show that (6) implies (5). Fix ϑ being some continuous function on $[a, b]$ and $\epsilon > 0$. The Weierstrass approximation theorem states that there exists a polynomial p on $[a, b]$ such that

$$|\vartheta(t) - p(t)| \le \frac{\epsilon}{3}$$

for all $t \in [a, b]$. Since p is a polynomial, there exists a constant C such that

$$|p(t_2) - p(t_1)| \le C\,|t_2 - t_1|$$

for any $a \le t_1 \le t_2 \le b$. Also (6) implies that there exists an $N_0 \in \mathbb{N}$ such that

$$|u_{N,l} - v_{N,l}| \le \frac{\epsilon}{3C}, \ \ \forall\ l \in [N]$$

for all $N \ge N_0$. Therefore, we have

$$\begin{aligned} &|\vartheta(u_{N,l}) - \vartheta(v_{N,l})| \\ \le &|\vartheta(u_{N,l}) - p(u_{N,l})| + |p(u_{N,l}) - p(v_{N,l})| \\ &+ |\vartheta(v_{N,l}) - p(v_{N,l})| \\ \le &\frac{\epsilon}{3} + C\frac{\epsilon}{3C} + \frac{\epsilon}{3} = \epsilon \end{aligned}$$

for all $l \in [N]$ and $N \ge N_0$. Thus

$$\begin{aligned} \left|\frac{1}{N}\sum_{l=0}^{N-1}\left(\vartheta\left(u_{N,l}\right) - \vartheta\left(v_{N,l}\right)\right)\right| \le &\frac{1}{N}\sum_{l=0}^{N-1}|\vartheta(u_{N,l}) - \vartheta(v_{N,l})| \\ \le &\epsilon \end{aligned}$$

for all $N \ge N_0$. Since ϵ is arbitrary, this implies (5).

Now let us show that (5) implies (6). We prove the statement (5) $\Rightarrow$ (6) by contradiction. Suppose (6) is not true, i.e., there exists an increasing sequence $\{M_k\}_{k=1}^{\infty}$ and $\epsilon_1 > 0$ such that

$$\max_{l\in[M_k]} |u_{M_k,l} - v_{M_k,l}| \ge 2\epsilon_1$$

for all $k \ge 1$. Let $l_k = \arg\max_{l\in[M_k]} |u_{M_k,l} - v_{M_k,l}|$ denote any point at which $|u_{M_k,l} - v_{M_k,l}|$ achieves its maximum, which implies $|u_{M_k,l_k} - v_{M_k,l_k}| \ge 2\epsilon_1$. Without loss of generality, we suppose $u_{M_k,l_k} \le v_{M_k,l_k}$, i.e., $u_{M_k,l_k} \le v_{M_k,l_k} - 2\epsilon_1$.

1) Suppose $u_{M_k,l_k} \ge \max_{x\in[c,d]} g(x)$, which indicates $v_{M_k,l_k} \ge 2\epsilon_1 + \max_{x\in[c,d]} g(x)$. This contradicts the assumption that $\lim_{N\to\infty} v_{N,0} = \max_{x\in[c,d]} g(x)$.
2) Suppose $u_{M_k,l_k} < \max_{x\in[c,d]} g(x)$. By assumption that

$$\lim_{N\to\infty} u_{N,N-1} = \lim_{N\to\infty} v_{N,N-1} = \min_{x\in[c,d]} g(x),$$

there exist $k_0 \in \mathbb{N}$ and $\alpha_k \in \text{int}(\mathcal{R}(g))$ such that

$$0 \le \alpha_k - u_{M_k,l_k} < \frac{\epsilon_1}{2}$$

and F_g is continuous at α_k for all $k \ge k_0$. Noting that F_g is right continuous everywhere and strictly increasing at α_k (which is shown in Lemma B.3), there exist $\epsilon_2, \epsilon_3 > 0$ such that $\epsilon_2 \le \frac{\epsilon_1}{2}$, F_g is continuous at $\alpha_k + \epsilon_2$, and

$$F_g(\alpha_k + \epsilon_2) = F_g(\alpha_k) + 3\epsilon_3. \tag{10}$$

Lemma B.2 indicates that

$$\lim_{M_k\to\infty} F_{u_{M_k}}(\alpha) = F_g(\alpha)$$

for every point α at which F_g is continuous. Thus there exist $k_1 \in \mathbb{N}, k_1 \ge k_0$ such that

$$\begin{aligned} \left|F_{u_{M_k}}\left(\alpha_k\right) - F_g\left(\alpha_k\right)\right| &< \epsilon_3, \\ \left|F_{u_{M_k}}\left(\alpha_k + \epsilon_2\right) - F_g\left(\alpha_k + \epsilon_2\right)\right| &< \epsilon_3 \end{aligned} \tag{11}$$

for all $k \geq k_1$. Thus, we have

$$\begin{aligned}
&F_{u_{M_k}}(\alpha_k+\epsilon_2) - F_{u_{M_k}}(\alpha_k)\\
=&F_{u_{M_k}}(\alpha_k+\epsilon_2) - F_g(\alpha_k+\epsilon_2) + F_g(\alpha_k+\epsilon_2)\\
&- F_g(\alpha_k) + F_g(\alpha_k) - F_{u_{M_k}}(\alpha_k)\\
\geq&F_g(\alpha_k+\epsilon_2) - F_g(\alpha_k) - \left|F_g(\alpha_k) - F_{u_{M_k}}(\alpha_k)\right|\\
&- \left|F_{u_{M_k}}(\alpha_k+\epsilon_2) - F_g(\alpha_k+\epsilon_2)\right|\\
\geq&3\epsilon_3 - \epsilon_3 - \epsilon_3 = \epsilon_3
\end{aligned}$$

for all $k \geq k_1$, where the last line follows from (10) and (11). Noting that the above equation is equivalent to

$$\frac{1}{M_k}\#\left\{l \in [M_k], \alpha_k < u_{M_k,l} \leq \alpha_k + \epsilon_2\right\} \geq \epsilon_3,$$

we have

$$\begin{aligned}
&\frac{1}{M_k}\#\left\{l \in [M_k], u_{M_k,l_k} < u_{M_k,l} \leq \alpha_k + \epsilon_2\right\}\\
\geq&\frac{1}{M_k}\#\left\{l \in [M_k], \alpha_k < u_{M_k,l} \leq \alpha_k + \epsilon_2\right\} \geq \epsilon_3.
\end{aligned}$$

Thus, we obtain

$$0 \leq u_{M_k,l_k-\lceil\epsilon_3 M_k\rceil} - u_{M_k,l_k} \leq \alpha_k + \epsilon_2 - u_{M_k,l_k} \leq \epsilon_1,$$

which implies

$$\begin{aligned}
&v_{M_k,l_k} - u_{M_k,l_k-\lceil\epsilon_3 M_k\rceil}\\
\geq&v_{M_k,l_k} - u_{M_k,l_k} + u_{M_k,l_k} - u_{M_k,l_k-\lceil\epsilon_3 M_k\rceil}\\
\geq&2\epsilon_1 - \epsilon_1 \geq \epsilon_1.
\end{aligned}$$

Now taking $\vartheta(t) = t$, we obtain

$$\begin{aligned}
&\frac{1}{M_k}\sum_{l=0}^{M_k-1}\left|\vartheta(u_{M_k,l}) - \vartheta(v_{M_k,l})\right|\\
\geq&\frac{1}{M_k}\sum_{l=l_k-\lceil\epsilon_3 M_k\rceil}^{l_k}\left|\vartheta(u_{M_k,l}) - \vartheta(v_{M_k,l})\right|\\
\geq&\frac{1}{M_k}\sum_{l=l_k-\lceil\epsilon_3 M_k\rceil}^{l_k}\left|\vartheta(u_{M_k,l}) - \vartheta(v_{M_k,l_k})\right|\\
\geq&\epsilon_3\epsilon_1 > 0
\end{aligned}$$

for all $k \geq k_1$. This contradicts Lemma B.1. ■

APPENDIX C
PROOF OF THEOREM I.2

We first establish the following useful results.

Lemma C.1. *Let $u_0, u_1, \ldots, u_{N-1} \in \mathbb{R}$ be an unordered sequence of N elements. We decreasingly arrange this sequence so that $u_{\rho(0)} \geq u_{\rho(1)} \geq \cdots \geq u_{\rho(N-1)}$. Then for any $r \in \{1, 2\ldots, N-1\}$, we have*

$$\begin{aligned}
&\max_{1\leq r'\leq r}\max_{l\in[N-r'-1]} u_{\rho(l)} - u_{\rho(l+r')}\\
\leq&\max_{1\leq r'\leq r}\max_{l\in[N-r'-1]}\left|u_l - u_{l+r'}\right|.
\end{aligned}$$

Proof (of Lemma C.1). The proof is straightforward for the case $r = 1$. If the sequence is constant, then

$$\max_{l\in[N-2]} u_{\rho(l)} - u_{\rho(l+1)} = \max_{l\in[N-2]}\left|u_l - u_{l+1}\right| = 0.$$

Suppose the sequence is not constant, i.e., there exist at least $l_1, l_2 \in [N]$ so that $u_{l_1} \neq u_{l_2}$. Let

$$l' = \arg\max_{l\in[N-2]} u_{\rho(l)} - u_{\rho(l+1)}$$

denote any point at which $u_{\rho(l)} - u_{\rho(l+1)}$ achieves its maximum. Search the sequence $\{u_l\}_{l\in[N]}$ to find $u_{l''}$ that is smaller than $u_{\rho(l')}$ and its index l'' is closest to $\rho(l')$. Thus

$$\max\left\{\left|u_{l''} - u_{l''+1}\right|, \left|u_{l''} - u_{l''-1}\right|\right\} \geq \max_{l\in[N-2]} u_{\rho(l)} - u_{\rho(l+1)}.$$

Suppose $r \geq 2$. Similarly, the proof for a constant sequence is obvious. Suppose the sequence is not constant. Let

$$\{l', r'\} = \arg\max_{1\leq r''\leq r}\max_{l\in[N-r''-1]} u_{\rho(l)} - u_{\rho(l+r'')}.$$

If there are several pairs $\{l', r'\}$ have the same values, we choose the one that r' has the smallest value. If $r' = 1$, the proof is similar to the case $r = 1$. We suppose $r' \geq 2$. Thus there exist at least r' elements that are smaller than $u_{\rho(l')}$ and only $r'-1$ elements that are greater than $u_{\rho(l'+r')}$ and smaller than $u_{\rho(l')}$. Search the sequence $\{u_l\}_{l\in[N]}$ to find $u_{l''}$ that is smaller than $u_{\rho(l')}$ and its index l'' is the r'-th closest to $\rho(l')$. Without loss of generality, suppose $l'' < \rho(l')$.

If $u_{l''} \leq u_{\rho(l'+r')}$, we have

$$\max_{1\leq r''\leq r'}\left|u_{l''} - u_{l''+r''}\right| \geq u_{\rho(l')} - u_{\rho(l'+r')}$$

since there is at least one element in $\{u_l, l''+1 \leq l \leq l''+r'\}$ that is greater than or equal to $u_{\rho(l')}$.

If $u_{l''} > u_{\rho(l'+r')}$, there exists $l'' \leq l''' \leq 2\rho(l') - l''$ such that $u_{l'''}$ is smaller than or equal to $u_{\rho(l'+r')}$ (otherwise, there are r' elements that are greater than $u_{\rho(l'+r')}$ and smaller than $u_{\rho(l')}$). Also near $u_{l'''}$, there must exist at least one element that is not smaller than $u_{\rho(l')}$. Then

$$\begin{aligned}
&\max_{1\leq r''\leq r'}\max\left\{\left|u_{l'''} - u_{l'''+r''}\right|, \left|u_{l'''} - u_{l'''-r''}\right|\right\}\\
\geq&u_{\rho(l')} - u_{\rho(l'+r')}.
\end{aligned}$$

This completes the proof. ■

In words, the largest error between the contiguous elements of a sequence is not magnified when the sequence is rearranged in decreasing (or increasing) order.

The following result establishes that the largest error between two sequences is not magnified when both of the sequences are rearranged in decreasing (or increasing) order.

Lemma C.2. *Let $u_0, \ldots, u_{N-1} \in \mathbb{R}$ and $v_0, \ldots, v_{N-1} \in \mathbb{R}$ be two unordered sequences of N elements. We decreasingly* *arrange these sequences so that* $u_{\rho(0)} \geq u_{\rho(1)} \geq \cdots u_{\rho(N-1)}$ *and* $v_{\rho(0)} \geq v_{\rho(1)} \geq \cdots v_{\rho(N-1)}$. *Then*

$$\max_{l\in[N-1]} \left|u_{\rho(l)} - v_{\rho(l)}\right| \leq \max_{l\in[N-1]} |u_l - v_l| .$$

Proof (of Lemma C.2). Let

$$r' = \arg\max_{r\in[N-1]} \left|u_{\rho(r)} - v_{\rho(r)}\right|$$

denote any point at which $\left|u_{\rho(r)} - v_{\rho(r)}\right|$ achieves its maximum and let l' be the index of $u_{\rho(r')}$. Without loss of generality, we suppose $u_{\rho(r')} \geq v_{\rho(r')}$. If $v_{l'} \leq v_{\rho(r')}$, we have $u_{l'} - v_{l'} \geq u_{\rho(r')} - v_{\rho(r')}$. Otherwise suppose $v_{l'} > v_{\rho(r')}$, which implies $r' \geq 1$. Since there are only r' elements in $\{u_l\}_{l\in[N]}$ that are greater than $u_{\rho(r')}$ and r' elements in $\{v_l\}_{l\in[N]}$ that are greater than $v_{\rho(r')}$, there must exist l'' such that $u_{l''} \geq u_{\rho(r')}$ and $v_{l''} \leq v_{\rho(r')}$. Hence

$$u_{l''} - v_{l''} \geq u_{\rho(r')} - v_{\rho(r')}.$$

■

Lemma C.3. *[24, Sturmian separation theorem] Let* $\boldsymbol{A}_N$ *be an* $N\times N$ *Hermitian matrix and let* $[\boldsymbol{A}_N]_{N-1}$ *be the* $(N-1)\times(N-1)$ *matrix obtained by deleting the last column and the last row of* $\boldsymbol{A}_N$. *Also let* $\lambda_0(\boldsymbol{A}_N) \geq \cdots \geq \lambda_{N-1}(\boldsymbol{A}_N)$ *and* $\lambda_0([\boldsymbol{A}_N]_{N-1}) \geq \cdots \geq \lambda_{N-2}([\boldsymbol{A}_N]_{N-1})$ *respectively denote the descending eigenvalues of* $\boldsymbol{A}_N$ *and* $[\boldsymbol{A}_N]_{N-1}$. *Then*

$$\lambda_l(\boldsymbol{A}_N) \geq \lambda_l([\boldsymbol{A}_N]_{N-1}) \geq \lambda_{l+1}(\boldsymbol{A}_N)$$

for all $0 \leq l \leq N-2$.

The above Sturmian separation theorem forms the foundation of the following analysis. We note that Zizler et.al. [32] utilized the Sturmian separation theorem to prove a refinement of Szegő's asymptotic formula in terms of the number of eigenvalues inside a given interval.

Now we are well equipped to prove Theorem I.2. In what follows, we consider $N > 2r$. Note that in this case $\widetilde{\boldsymbol{C}}_N$ is equivalent to $\widehat{\boldsymbol{C}}_N$ and the eigenvalues of $\widetilde{\boldsymbol{C}}_N$ are the DFT samples of $S_{N-1}(f) = \widetilde{h}(f) = \sum_{k=-r}^{r} h[k]e^{j2\pi fk}$. Recall that $[\boldsymbol{H}_N]_{N-r}$ is the $(N-r)\times(N-r)$ matrix obtained by deleting the last r columns and the last r rows of $\boldsymbol{H}_N$. Similar notation holds for $[\widetilde{\boldsymbol{C}}_N]_{N-r}$.

Note that $[\boldsymbol{H}]_{N-r}$ is exactly the same as $[\widetilde{\boldsymbol{C}}_N]_{N-r}$ as they have the same elements when $N > 2r$. Thus $[\boldsymbol{H}]_{N-r}$ and $[\widetilde{\boldsymbol{C}}_N]_{N-r}$ have the same eigenvalues. Let $\lambda_l([\widetilde{\boldsymbol{C}}_N]_{N-r})$ be permuted such that

$$\lambda_{\rho(0)}([\widetilde{\boldsymbol{C}}_N]_{N-r}) \geq \cdots \geq \lambda_{\rho(N-r-1)}([\widetilde{\boldsymbol{C}}_N]_{N-r}).$$

We first consider the simple case when $r = 1$. It follows from the Sturmian separation theorem that

$$\lambda_l(\boldsymbol{H}_N) \geq \lambda_l([\boldsymbol{H}_N]_{N-1}) \geq \lambda_{l+1}(\boldsymbol{H}_N),$$
$$\lambda_{\rho(l)}(\widetilde{\boldsymbol{C}}_N) \geq \lambda_{\rho(l)}([\widetilde{\boldsymbol{C}}_N]_{N-1}) \geq \lambda_{\rho(l+1)}(\widetilde{\boldsymbol{C}}_N)$$

for all $0 \leq l \leq N-2$. This implies the following relationship between $\lambda_l(\boldsymbol{H}_N)$ and $\lambda_{\rho(l)}(\widetilde{\boldsymbol{C}}_N)$

$$\begin{aligned}\lambda_l(\boldsymbol{H}_N) \leq&\lambda_{l-1}([\boldsymbol{H}_N]_{N-1}) = \lambda_{\rho(l-1)}([\widetilde{\boldsymbol{C}}_N]_{N-1})\\ \leq&\lambda_{\rho(l-1)}(\widetilde{\boldsymbol{C}}_N),\ \forall\ l = 1,2,\ldots,N-1,\\ \lambda_l(\boldsymbol{H}_N) \geq&\lambda_l([\boldsymbol{H}_N]_{N-1}) = \lambda_{\rho(l)}([\widetilde{\boldsymbol{C}}_N]_{N-1})\\ \geq&\lambda_{\rho(l+1)}(\widetilde{\boldsymbol{C}}_N),\ \forall\ l = 0,1,\ldots,N-2\end{aligned} \tag{12}$$

which give

$$\begin{aligned}&\left|\lambda_l\left(\boldsymbol{H}_N\right) - \lambda_{\rho(l)}\left(\widetilde{\boldsymbol{C}}_N\right)\right|\\ &= \max\left\{\lambda_l\left(\boldsymbol{H}_N\right) - \lambda_{\rho(l)}\left(\widetilde{\boldsymbol{C}}_N\right),\right.\\ &\qquad\left.\lambda_{\rho(l)}\left(\widetilde{\boldsymbol{C}}_N\right) - \lambda_l\left(\boldsymbol{H}_N\right)\right\}\\ &\leq \max\left\{\lambda_{\rho(l-1)}\left(\widetilde{\boldsymbol{C}}_N\right) - \lambda_{\rho(l)}\left(\widetilde{\boldsymbol{C}}_N\right),\right.\\ &\qquad\left.\lambda_{\rho(l)}\left(\widetilde{\boldsymbol{C}}_N\right) - \lambda_{\rho(l+1)}\left(\widetilde{\boldsymbol{C}}_N\right)\right\}\end{aligned}$$

for all $1 \leq l \leq N-2$. Applying Lemma C.1 with $r = 1$, we obtain

$$\begin{aligned}&\max_{1\leq l\leq N-2}\left|\lambda_l\left(\boldsymbol{H}_N\right) - \lambda_{\rho(l)}\left(\widetilde{\boldsymbol{C}}_N\right)\right|\\ &\leq \max_{0\leq l\leq N-2}\lambda_{\rho(l)}\left(\widetilde{\boldsymbol{C}}_N\right) - \lambda_{\rho(l+1)}\left(\widetilde{\boldsymbol{C}}_N\right)\\ &\leq \max_{0\leq l\leq N-2}\left|\lambda_l\left(\widetilde{\boldsymbol{C}}_N\right) - \lambda_{l+1}\left(\widetilde{\boldsymbol{C}}_N\right)\right|.\end{aligned}$$

Note that $\widetilde{h}(f)$ is Lipschitz continuous since it is continuously differentiable. There exists a Lipschitz constant K such that, for all f_1 and f_2 in $[0,1]$,

$$\left|\widetilde{h}(f_1) - \widetilde{h}(f_2)\right| \leq K\left|f_1 - f_2\right|.$$

From the fact that the eigenvalues of $\widetilde{\boldsymbol{C}}_N$ are the DFT samples of $\widetilde{h}(f)$, i.e., $\lambda_l\left(\widetilde{\boldsymbol{C}}_N\right) = \widetilde{h}(\frac{l}{N})$, it follows that

$$\begin{aligned}&\max_{1\leq l\leq N-2}\left|\lambda_l\left(\boldsymbol{H}_N\right) - \lambda_{\rho(l)}\left(\widetilde{\boldsymbol{C}}_N\right)\right|\\ &\leq \max_{0\leq l\leq N-2}\left|\lambda_l\left(\widetilde{\boldsymbol{C}}_N\right) - \lambda_{l+1}\left(\widetilde{\boldsymbol{C}}_N\right)\right|\\ &\leq \max_{0\leq l\leq N-2}\left|\widetilde{h}(\frac{l}{N}) - \widetilde{h}(\frac{l+1}{N})\right| \leq K\frac{1}{N}.\end{aligned} \tag{13}$$

Utilizing the fact that $\lambda_0(\boldsymbol{H}_N) \leq \max_{f\in[0,1]}\widetilde{h}(f)$ and $\lambda_{N-1}(\boldsymbol{H}_N) \geq \min_{f\in[0,1]}\widetilde{h}(f)$ (see Lemma III.2) and applying (12) with $l = 0$ which gives

$$\lambda_0\left(\boldsymbol{H}_N\right) \geq \lambda_{\rho(1)}\left(\widetilde{\boldsymbol{C}}_N\right),$$

we have

$$\begin{aligned}&\left|\lambda_0\left(\boldsymbol{H}_N\right)-\lambda_{\rho(0)}\left(\widetilde{\boldsymbol{C}}_N\right)\right|\\=&\max\left\{\lambda_0\left(\boldsymbol{H}_N\right)-\lambda_{\rho(0)}\left(\widetilde{\boldsymbol{C}}_N\right),\right.\\&\left.\lambda_{\rho(0)}\left(\widetilde{\boldsymbol{C}}_N\right)-\lambda_0\left(\boldsymbol{H}_N\right)\right\}\\\le&\max\left\{\max_{f\in[0,1]}\widetilde{h}(f)-\lambda_{\rho(0)}\left(\widetilde{\boldsymbol{C}}_N\right),\right.\\&\left.\lambda_{\rho(0)}\left(\widetilde{\boldsymbol{C}}_N\right)-\lambda_{\rho(1)}\left(\widetilde{\boldsymbol{C}}_N\right)\right\}\\\le&K\frac{1}{N},\end{aligned}$$

where the second inequality follows because $\lambda_l\left(\widetilde{\boldsymbol{C}}_N\right)$ are uniform samples of $\widetilde{h}(f)$ with grid size $\frac{1}{N}$. Similarly, we have

$$\begin{aligned}&\left|\lambda_{N-1}\left(\boldsymbol{H}_N\right)-\lambda_{\rho(N-1)}\left(\widetilde{\boldsymbol{C}}_N\right)\right|\\=&\max\left\{\lambda_{\rho(N-1)}\left(\widetilde{\boldsymbol{C}}_N\right)-\lambda_{N-1}\left(\boldsymbol{H}_N\right),\right.\\&\left.\lambda_{N-1}\left(\boldsymbol{H}_N\right)-\lambda_{\rho(N-1)}\left(\widetilde{\boldsymbol{C}}_N\right)\right\}\\\le&\max\left\{\lambda_{\rho(N-1)}\left(\widetilde{\boldsymbol{C}}_N\right)-\min_{f\in[0,1]}\widetilde{h}(f),\right.\\&\left.\lambda_{\rho(N-2)}\left(\widetilde{\boldsymbol{C}}_N\right)-\lambda_{\rho(N-1)}\left(\widetilde{\boldsymbol{C}}_N\right)\right\}\\\le&K\frac{1}{N}.\end{aligned}$$

Along with (13), we conclude

$$\max_{0\le l\le N-1}\left|\lambda_l\left(\boldsymbol{H}_N\right)-\lambda_{\rho(l)}\left(\widetilde{\boldsymbol{C}}_N\right)\right|\le K\frac{1}{N}.$$

Now we consider the case $r>1$. Repeatedly applying the Sturmian separation theorem r times yields

$$\begin{aligned}&\lambda_l(\boldsymbol{H}_N)\ge\lambda_l([\boldsymbol{H}_N]_{N-r})\ge\lambda_{l+r}(\boldsymbol{H}_N),\\&\lambda_{\rho(l)}(\widetilde{\boldsymbol{C}}_N)\ge\lambda_{\rho(l)}([\widetilde{\boldsymbol{C}}_N]_{N-r})\ge\lambda_{\rho(l+r)}(\widetilde{\boldsymbol{C}}_N)\end{aligned}$$

for all $0\le l\le N-r-1$. Noting that $[\boldsymbol{H}_N]_{N-r}$ is the same as $[\widetilde{\boldsymbol{C}}_N]_{N-r}$, we have

$$\begin{aligned}\lambda_l(\boldsymbol{H}_N)\le&\lambda_{l-r}([\boldsymbol{H}_N]_{N-r})=\lambda_{\rho(l-r)}([\widetilde{\boldsymbol{C}}_N]_{N-r})\\\le&\lambda_{\rho(l-r)}(\widetilde{\boldsymbol{C}}_N),\ \forall\ l=r,r+1,\dots,N-1,\\\lambda_l(\boldsymbol{H}_N)\ge&\lambda_l([\boldsymbol{H}_N]_{N-r})=\lambda_{\rho(l)}([\widetilde{\boldsymbol{C}}_N]_{N-r})\\\ge&\lambda_{\rho(l+r)}(\widetilde{\boldsymbol{C}}_N),\ \forall\ l=0,1,\dots,N-r-1\end{aligned}$$

which give

$$\begin{aligned}&\max_{r\le l\le N-r-1}\left|\lambda_l\left(\boldsymbol{H}_N\right)-\lambda_{\rho(l)}\left(\widetilde{\boldsymbol{C}}_N\right)\right|\\=&\max_{r\le l\le N-r-1}\max\left\{\lambda_l\left(\boldsymbol{H}_N\right)-\lambda_{\rho(l)}\left(\widetilde{\boldsymbol{C}}_N\right),\right.\\&\left.\lambda_{\rho(l)}\left(\widetilde{\boldsymbol{C}}_N\right)-\lambda_l\left(\boldsymbol{H}_N\right)\right\}\\\le&\max_{r\le l\le N-r-1}\max\left\{\lambda_{\rho(l-r)}\left(\widetilde{\boldsymbol{C}}_N\right)-\lambda_{\rho(l)}\left(\widetilde{\boldsymbol{C}}_N\right),\right.\\&\left.\lambda_{\rho(l)}\left(\widetilde{\boldsymbol{C}}_N\right)-\lambda_{\rho(l+r)}\left(\widetilde{\boldsymbol{C}}_N\right)\right\}\\\le&\max_{0\le l\le N-r-1}\lambda_{\rho(l)}\left(\widetilde{\boldsymbol{C}}_N\right)-\lambda_{\rho(l+r)}\left(\widetilde{\boldsymbol{C}}_N\right)\\\le&\max_{1\le r'\le r}\ \max_{0\le l\le N-r'-1}\left|\lambda_l\left(\widetilde{\boldsymbol{C}}_N\right)-\lambda_{l+r'}\left(\widetilde{\boldsymbol{C}}_N\right)\right|\\\le&K\frac{r}{N},\end{aligned}$$

where the third inequality follows from Lemma C.1. Since

$$\lambda_{r-1}(\boldsymbol{H}_N)\le\cdots\le\lambda_0(\boldsymbol{H}_N)\le\max_{f\in[0,1]}\widetilde{h}(f),$$

we bound $\left|\lambda_{r'}\left(\boldsymbol{H}_N\right)-\lambda_{\rho(r')}\left(\widetilde{\boldsymbol{C}}_N\right)\right|, r'\le r-1$ by considering the following two cases: if $\lambda_{\rho(r')}\left(\widetilde{\boldsymbol{C}}_N\right)\le\lambda_{r'}\left(\boldsymbol{H}_N\right)$, we have

$$\begin{aligned}&\lambda_{r'}\left(\boldsymbol{H}_N\right)-\lambda_{\rho(r')}\left(\widetilde{\boldsymbol{C}}_N\right)\\\le&\max_{f\in[0,1]}\widetilde{h}(f)-\lambda_{\rho(r-1)}\left(\widetilde{\boldsymbol{C}}_N\right)\le K\frac{r}{N};\end{aligned}$$

if $\lambda_{\rho(r')}\left(\widetilde{\boldsymbol{C}}_N\right)>\lambda_{r'}\left(\boldsymbol{H}_N\right)$, we have

$$\begin{aligned}&\lambda_{\rho(r')}\left(\widetilde{\boldsymbol{C}}_N\right)-\lambda_{r'}\left(\boldsymbol{H}_N\right)\\\le&\lambda_{\rho(r')}\left(\widetilde{\boldsymbol{C}}_N\right)-\lambda_{\rho(r'+r)}\left(\widetilde{\boldsymbol{C}}_N\right)\\\le&\max_{1\le r''\le r}\max_{0\le l\le N-r''-1}\left|\lambda_l\left(\widetilde{\boldsymbol{C}}_N\right)-\lambda_{l+r''}\left(\widetilde{\boldsymbol{C}}_N\right)\right|\\\le&K\frac{r}{N}\end{aligned}$$

where the second line follows because $\lambda_{r'}\left(\boldsymbol{H}_N\right)\ge\lambda_{\rho(r'+r)}\left(\widetilde{\boldsymbol{C}}_N\right)$ and the third line follows from Lemma C.1. Thus we have

$$\left|\lambda_{r'}\left(\boldsymbol{H}_N\right)-\lambda_{\rho(r')}\left(\widetilde{\boldsymbol{C}}_N\right)\right|\le K\frac{r}{N}$$

for all $0\le r'\le r-1$. Similarly,

$$\begin{aligned}&\left|\lambda_{N-r'}\left(\boldsymbol{H}_N\right)-\lambda_{\rho(N-r')}\left(\widetilde{\boldsymbol{C}}_N\right)\right|\\\le&\max\left\{\lambda_{\rho(N-r')}\left(\widetilde{\boldsymbol{C}}_N\right)-\min_{f\in[0,1]}\widetilde{h}(f),\right.\\&\left.\lambda_{\rho(N-r'-r)}\left(\widetilde{\boldsymbol{C}}_N\right)-\lambda_{\rho(N-r')}\left(\widetilde{\boldsymbol{C}}_N\right)\right\}\\\le&\max\left\{K\frac{r}{N},K\frac{r}{N}\right\}=K\frac{r}{N},\end{aligned}$$

for all $1 \le r' \le r$. Therefore,

$$\max_{0\le l\le N-1}\left|\lambda_l\left(\boldsymbol{H}_N\right)-\lambda_{\rho(l)}\left(\widetilde{\boldsymbol{C}}_N\right)\right| \le Kr\frac{1}{N}.$$

for all $N > 2r$.

Note that $S_{r+1}(f) = S_{r+2}(f) = \cdots = S_{N-1}(f)$ which gives

$$\sigma_N(f) = \frac{\sum_{n=0}^{N-1} S_n(f)}{N} = \frac{\sum_{n=0}^{r} S_n(f)}{N} + \frac{N-r-1}{N} S_{r+1}(f).$$

Thus

$$\begin{aligned}|\sigma_N(f)-S_{N-1}(f)| &= \left|\frac{\sum_{n=0}^{r} S_n(f)}{N} - \frac{r+1}{N}S_{r+1}(f)\right| \\ &= \left|\sum_{n=0}^{r} S_n(f) - (r+1)S_{r+1}(f)\right|\frac{1}{N} \\ &= O(\frac{1}{N})\end{aligned}$$

uniformly on $[0,1]$ as $N \to \infty$. Therefore,

$$\begin{aligned}&\max_{0\le l\le N-1}\left|\lambda_l(\overline{\boldsymbol{C}}_N)-\lambda_l(\widetilde{\boldsymbol{C}}_N)\right| \\ =&\max_{0\le l\le N-1}\left|\sigma_N(\frac{l}{N})-S_{N-1}(\frac{l}{N})\right| = O(\frac{1}{N})\end{aligned}$$

as $N \to \infty$. Finally,

$$\begin{aligned}&\max_{0\le l\le N-1}\left|\lambda_{\rho(l)}(\overline{\boldsymbol{C}}_N)-\lambda_l(\boldsymbol{H}_N)\right| \\ =&\max_{0\le l\le N-1}\left|\lambda_{\rho(l)}(\overline{\boldsymbol{C}}_N)-\lambda_{\rho(l)}(\widetilde{\boldsymbol{C}}_N)+\lambda_{\rho(l)}(\widetilde{\boldsymbol{C}}_N)-\lambda_l(\boldsymbol{H}_N)\right| \\ \le&\max_{0\le l\le N-1}\left|\lambda_{\rho(l)}(\overline{\boldsymbol{C}}_N)-\lambda_{\rho(l)}(\widetilde{\boldsymbol{C}}_N)\right| \\ &+\max_{0\le l\le N-1}\left|\lambda_{\rho(l)}(\widetilde{\boldsymbol{C}}_N)-\lambda_l(\boldsymbol{H}_N)\right| \\ \le&\max_{0\le l\le N-1}\left|\lambda_l(\overline{\boldsymbol{C}}_N)-\lambda_l(\widetilde{\boldsymbol{C}}_N)\right| \\ &+\max_{0\le l\le N-1}\left|\lambda_{\rho(l)}(\widetilde{\boldsymbol{C}}_N)-\lambda_l(\boldsymbol{H}_N)\right| \\ =&O(\frac{1}{N})\end{aligned}$$

as $N \to \infty$, where the second inequality follows from Lemma C.2. ■

APPENDIX D
PROOF OF THEOREM III.3

Theorem D.1. *(Riemann-Lebesgue theorem [33, Theorem 7.48]) The function $g(x) \in L^\infty([a,b])$ is Riemann integrable over $[a,b]$ if and only if it is continuous almost everywhere in $[a,b]$.*

Despite the fact that $F_g(\alpha)$ is right continuous and non-decreasing everywhere, some stronger results about $F_g(\alpha)$ can be obtained at some point α since $g(x)$ is Riemann integrable.

Lemma D.2. *Suppose $g(x) : [c,d] \to [a,b]$ is Riemann integrable and let $F_g(\alpha)$ be defined as in (9). Then $F_g(\alpha)$ is strictly increasing at α if $\alpha \in \operatorname{int}(\operatorname{ess}\mathcal{R}(g))$, i.e., there exists $\epsilon > 0$ such that, for every pair (α_1, α_2) such that $\alpha - \epsilon < \alpha_1 < \alpha < \alpha_2 < \alpha + \epsilon$, $F_g(\alpha_1) < F_g(\alpha) < F_g(\alpha_2)$.*

Proof (of Lemma D.2). Since $g(x) : [c,d] \to [a,b]$ is Riemann integrable and $\alpha \in \operatorname{int}(\operatorname{ess}\mathcal{R}(g))$, there exists ϵ such that $(\alpha-\epsilon, \alpha+\epsilon) \subset \operatorname{ess}\mathcal{R}(g)$. Let α_1 be an arbitrary value such that $\alpha - \epsilon < \alpha_1 < \alpha$ and let $\alpha_1' = \frac{\alpha+\alpha_1}{2} \in \operatorname{ess}\mathcal{R}(g)$. It follows from the definition of essential range that

$$\mu\left\{x \in [c,d] : |g(x) - \alpha_1'| < \frac{\alpha-\alpha_1}{2}\right\} > 0.$$

Thus, we obtain

$$\begin{aligned}F_g(\alpha) - F_g(\alpha_1) =& \frac{1}{d-c}\mu\{x \in [c,d] : \alpha_1 < g(x) \le \alpha\} \\ \ge& \frac{1}{d-c}\mu\{x \in [c,d] : \alpha_1 < g(x) < \alpha\} > 0.\end{aligned}$$

Similarly, we have $F_g(\alpha) < F_g(\alpha_2)$ for $\alpha < \alpha_2 < \alpha + \epsilon$. ■

We are now ready to prove the main part using the same approach that was used to prove Theorem III.1.

1) First, we show that (8) implies (7). This part is the same as those in Appendix B.
2) Now let us show that (7) implies (8). We prove the statement (7) ⇒ (8) by contradiction. Suppose (8) is not true, i.e., there exists an increasing sequence $\{M_k\}_{k=1}^\infty$ and $\epsilon_1 > 0$ such that

$$\max_{l\in[M_k]} |u_{M_k,l} - v_{M_k,l}| \ge 2\epsilon_1$$

for all $k \ge 1$. Let $l_k = \arg\max_{l\in[M_k]} |u_{M_k,l} - v_{M_k,l}|$ denote any point at which $|u_{M_k,l} - v_{M_k,l}|$ achieves its maximum. This implies $|u_{M_k,l_k} - v_{M_k,l_k}| \ge 2\epsilon_1$. Without loss of generality, we suppose F_g is continuous at u_{M_k,l_k} and $u_{M_k,l_k} \le v_{M_k,l_k}$, i.e., $u_{M_k,l_k} \le v_{M_k,l_k} - 2\epsilon_1$. Otherwise, one can always pick a $\widehat{u}_{M_k,l_k}$ that is close enough to u_{M_k,l_k} and such that F_g is continuous at $\widehat{u}_{M_k,l_k}$ since F_g is continuous almost everywhere.
By assumption, $u_{M_k,l_k} \in \operatorname{int}(\operatorname{ess}\mathcal{R}(g))$. Noting that F_g is right continuous everywhere and strictly increasing at u_{M_k,l_k} (which is shown in Lemma D.2), there exist $\epsilon_2 > 0$ and $\epsilon_3 > 0$ such that $\epsilon_2 < \epsilon_1$, F_g is continuous at $u_{M_k,l_k} + \epsilon_2$, and

$$F_g(u_{M_k,l_k} + \epsilon_2) = F_g(u_{M_k,l_k}) + 3\epsilon_3. \tag{14}$$

Lemma B.2 indicates that

$$\lim_{M_k\to\infty} F_{u_{M_k}}(\alpha) = F_g(\alpha)$$

for every point α at which F_g is continuous. Thus there exists $k_0 \in \mathbb{N}$ such that

$$\begin{aligned}\left|F_{u_{M_k}}(u_{M_k,l_k}) - F_g(u_{M_k,l_k})\right| \le& \epsilon_3, \\ \left|F_{u_{M_k}}(u_{M_k,l_k}+\epsilon_2) - F_g(u_{M_k,l_k}+\epsilon_2)\right| \le& \epsilon_3\end{aligned} \tag{15}$$

for all $k \geq k_0$. Thus, we have

$$\begin{aligned}
&F_{u_{M_k}}(u_{M_k,l_k}+\epsilon_2) - F_{u_{M_k}}(u_{M_k,l_k}) \\
=&\left(F_{u_{M_k}}(u_{M_k,l_k}+\epsilon_2) - F_g(u_{M_k,l_k}+\epsilon_2)\right) \\
&+ \left(F_g(u_{M_k,l_k}+\epsilon_2) - F_g(u_{M_k,l_k})\right) \\
&+ \left(F_g(u_{M_k,l_k}) - F_{u_{M_k}}(u_{M_k,l_k})\right) \\
\geq& F_g(u_{M_k,l_k}+\epsilon_2) - F_g(u_{M_k,l_k}) \\
&- \left|F_{u_{M_k}}(u_{M_k,l_k}+\epsilon_2) - F_g(u_{M_k,l_k}+\epsilon_2)\right| \\
&- \left|F_g(u_{M_k,l_k}) - F_{u_{M_k}}(u_{M_k,l_k})\right| \\
\geq& 3\epsilon_3 - \epsilon_3 - \epsilon_3 = \epsilon_3
\end{aligned}$$

for all $k \geq k_0$, where the last line follows from (14) and (15). Note that the above equation is equivalent to

$$\frac{1}{M_k}\#\left\{l \in [M_k], u_{M_k,l_k} < u_{M_k,l} \leq u_{M_k,l_k}+\epsilon_2\right\} \geq \epsilon_3.$$

Then

$$0 \leq u_{M_k,l_k-\lceil \epsilon_3 M_k\rceil} - u_{M_k,l_k} \leq u_{M_k,l_k}+\epsilon_2 - u_{M_k,l_k} = \epsilon_2,$$

which implies

$$\begin{aligned}
&v_{M_k,l_k} - u_{M_k,l_k-\lceil \epsilon_3 M_k\rceil} \\
\geq& v_{M_k,l_k} - u_{M_k,l_k} + u_{M_k,l_k} - u_{M_k,l_k-\lceil \epsilon_3 M_k\rceil} \\
\geq& 2\epsilon_1 - \epsilon_2 \geq 2\epsilon_1 - \epsilon_1 \geq \epsilon_1.
\end{aligned}$$

Now taking $\vartheta(t) = t$, we obtain

$$\begin{aligned}
&\frac{1}{M_k}\sum_{l=0}^{M_k-1}\left|\vartheta(u_{M_k,l}) - \vartheta(v_{M_k,l})\right| \\
\geq& \frac{1}{M_k}\sum_{l=l_k-\lceil \epsilon_3 M_k\rceil}^{l_k}\left|\vartheta(u_{M_k,l}) - \vartheta(v_{M_k,l})\right| \\
\geq& \frac{1}{M_k}\sum_{l=l_k-\lceil \epsilon_3 M_k\rceil}^{l_k}\left|\vartheta(u_{M_k,l}) - \vartheta(v_{M_k,l_k})\right| \\
\geq& \epsilon_3\epsilon_1 > 0
\end{aligned}$$

for all $k \geq k_1$. This contradicts Lemma B.1. ■

Appendix E
Proof of Theorem I.4

Lemma E.1. *Let $D_N(f) := \frac{\sin(\pi N f)}{\sin(\pi f)}$ denote the Dirichlet kernel. Fix $0 < W < \frac{1}{2}$. We have*

$$\begin{aligned}
\int_0^1 |D_N(f)|^2\, df &= N, \ \ \forall\ \ N \in \mathbb{N}, \\
\int_W^{1-W} |D_N(f)|^2\, df &= O(1), \ \ \textit{when}\ \ N \to \infty.
\end{aligned}$$

Proof (of Lemma E.1). Noting that $D_N(f) = \frac{\sin(\pi N f)}{\sin(\pi f)} = \frac{e^{j2\pi f N}}{e^{j\pi f}}\sum_{n=0}^{N-1} e^{j2\pi f n}$, we have

$$\begin{aligned}
|D_N(f)|^2 &= \left|\sum_{n=0}^{N-1} e^{j2\pi f n}\right|^2 = \left(\sum_{n=0}^{N-1} e^{j2\pi f n}\right)\left(\sum_{m=0}^{N-1} e^{-j2\pi f m}\right) \\
&= \sum_{n=0}^{N-1}\sum_{m=0}^{N-1} e^{j2\pi f(n-m)}.
\end{aligned}$$

It follows that

$$\begin{aligned}
\int_0^1 |D_N(f)|^2\, df &= \int_0^1 \sum_{n=0}^{N-1}\sum_{m=0}^{N-1} e^{j2\pi f(n-m)} df \\
&= \sum_{n=0}^{N-1}\sum_{m=0}^{N-1}\int_0^1 e^{j2\pi f(n-m)} df = N.
\end{aligned}$$

Fix $0 < W < \frac{1}{2}$. For any $f \in [W, 1-W]$, $|D_N(f)|$ is bounded above by $\frac{1}{\sin(\pi W)}$. Therefore,

$$\int_W^{1-W} |D_N(f)|^2\, df \leq \int_W^{1-W} \frac{1}{\sin^2(\pi W)} df \leq \frac{1}{\sin^2(\pi W)}.$$

■

Since $\widetilde{h}$ is bounded and Riemann integrable over $[0,1]$, it follows from the Riemann-Lebesgue theorem that $\widetilde{h}$ is continuous almost everywhere in $[0,1]$. Thus we can select $f_0 \in [0,1]$ and a positive number W such that

$$\left|\widetilde{h}(f) - \operatorname{ess\,sup}\widetilde{h}\right| \leq \frac{\epsilon}{4}$$

holds almost everywhere for $|f - f_0| \leq W$. For any $\boldsymbol{v} \in \mathbb{C}^N$, we have

$$\begin{aligned}
&\langle \boldsymbol{H}_N\boldsymbol{v}, \boldsymbol{v}\rangle \\
=& \int_0^1 |\widetilde{\boldsymbol{v}}(f)|^2\, \widetilde{h}(f) df \\
=& \int_{f_0-W}^{f_0+W} |\widetilde{\boldsymbol{v}}(f)|^2\, \widetilde{h}(f) df + \int_{\substack{f\in[0,1] \\ f\notin[f_0-W, f_0+W]}} |\widetilde{\boldsymbol{v}}(f)|^2\, \widetilde{h}(f) df \\
\geq& \left(\operatorname{ess\,sup}\widetilde{h} - \frac{\epsilon}{4}\right)\int_{f_0-W}^{f_0+W} |\widetilde{\boldsymbol{v}}(f)|^2\, \widetilde{d}f \\
&- \max\left(\left|\operatorname{ess\,inf}\widetilde{h}\right|, \left|\operatorname{ess\,sup}\widetilde{h}\right|\right)\cdot\int_{\substack{f\in[0,1] \\ f\notin[f_0-W, f_0+W]}} |\widetilde{\boldsymbol{v}}(f)|^2\, \widetilde{d}f.
\end{aligned} \tag{16}$$

The DTFT of $\boldsymbol{e}_{l/N}$ is

$$\widetilde{\boldsymbol{e}}_{l/N}(f) = \frac{e^{-j\pi N\left(f-\frac{l}{\sqrt{N}}\right)}}{e^{-j\pi\left(f-\frac{l}{N}\right)}} D_N(f - \frac{l}{N}).$$

Fix $\widetilde{h}$, ϵ and W. If $N \geq \frac{1}{W}$, there always exists l' such that $\left|\frac{l'}{N} - f_0\right| \leq \frac{W}{2}$. It follows from Lemma E.1 that

$$\begin{aligned}\int_{\frac{l'}{N}-\frac{W}{2}}^{\frac{l'}{N}+\frac{W}{2}} \left|\frac{1}{\sqrt{N}}\widetilde{\boldsymbol{e}}_{l'/N}(f)\right|^2 df &= 1 - \frac{1}{N}\int_{\frac{W}{2}}^{1-\frac{W}{2}} |D_N(f)|^2\, df \\ &= 1 - o(1)\end{aligned}$$

as $N \to \infty$. Note that $[\frac{l'}{N} - \frac{W}{2}, \frac{l'}{N} + \frac{W}{2}] \subset [f_0 - W, f_0 + W]$. Thus there exists $N_1 \in \mathbb{N}$ such that for all $N \geq \max\left\{N_1, \frac{1}{W}\right\}$

$$\begin{aligned}&\int_{f_0-W}^{f_0+W} \left|\frac{1}{\sqrt{N}}\widetilde{\boldsymbol{e}}_{l'/N}(f)\right|^2 \widetilde{d}f \geq 1 - \frac{\epsilon}{4\left|\operatorname{ess\,sup}\widetilde{h}\right|}, \\ &\int_{\substack{f\in[0,1] \\ f\notin[f_0-W,f_0+W]}} \left|\frac{1}{\sqrt{N}}\widetilde{\boldsymbol{e}}_{l'/N}(f)\right|^2 \widetilde{d}f \\ &\quad \leq \frac{\epsilon}{2\cdot\max\left(\left|\operatorname{ess\,inf}\widetilde{h}\right|, \left|\operatorname{ess\,sup}\widetilde{h}\right|\right)}.\end{aligned} \tag{17}$$

Combining (16) and (17) yields

$$\begin{aligned}\lambda_{l'}\left(\overline{\boldsymbol{C}}_N\right) =& \langle \boldsymbol{H}_N \frac{1}{\sqrt{N}}\boldsymbol{e}_{l'/N}, \frac{1}{\sqrt{N}}\boldsymbol{e}_{l'/N}\rangle \\ \geq& \left(\operatorname{ess\,sup}\widetilde{h} - \frac{\epsilon}{4}\right)\left(1 - \frac{\epsilon}{4\left|\operatorname{ess\,sup}\widetilde{h}\right|}\right) - \frac{\epsilon}{2} \\ \geq& \operatorname{ess\,sup}\widetilde{h} - \frac{\epsilon}{4} - \frac{\epsilon}{4} + \frac{\epsilon^2}{16\left|\operatorname{ess\,sup}\widetilde{h}\right|} - \frac{\epsilon}{2} \\ \geq& \operatorname{ess\,sup}\widetilde{h} - \epsilon\end{aligned}$$

for all $N \geq \max\left\{N_1, \frac{1}{W}\right\}$. Noting that $\lambda_{l'}\left(\overline{\boldsymbol{C}}_N\right) \leq \lambda_{\rho(0)}\left(\overline{\boldsymbol{C}}_N\right) \leq \operatorname{ess\,sup}\widetilde{h}$, we have

$$\left|\lambda_{\rho(0)}\left(\boldsymbol{C}_N\right) - \operatorname{ess\,sup}\widetilde{h}\right| \leq \epsilon$$

for all $N \geq N_0$. Since ϵ is arbitrary, we conclude

$$\lim_{N\to\infty} \lambda_{\rho(0)}\left(\overline{\boldsymbol{C}}_N\right) = \operatorname{ess\,sup}\widetilde{h}.$$

With a similar argument, we have

$$\lim_{N\to\infty} \lambda_{\rho(N-1)}\left(\overline{\boldsymbol{C}}_N\right) = \operatorname{ess\,inf}\widetilde{h}.$$

Noting that $\lambda_{\rho(0)}\left(\overline{\boldsymbol{C}}_N\right) \leq \lambda_0\left(\boldsymbol{H}_N\right) \leq \operatorname{ess\,sup}\widetilde{h}$ and $\operatorname{ess\,inf}\widetilde{h} \leq \lambda_{N-1}(\boldsymbol{H}_N) < \lambda_{\rho(N-1)}(\overline{\boldsymbol{C}}_N)$ (see Lemma III.2), we obtain

$$\begin{aligned}\lim_{N\to\infty} \lambda_{\rho(0)}\left(\overline{\boldsymbol{C}}_N\right) &= \lim_{N\to\infty} \lambda_0\left(\boldsymbol{H}_N\right) = \operatorname{ess\,sup}\widetilde{h} \\ \lim_{N\to\infty} \lambda_{\rho(N-1)}\left(\overline{\boldsymbol{C}}_N\right) &= \lim_{N\to\infty} \lambda_{N-1}\left(\boldsymbol{H}_N\right) = \operatorname{ess\,inf}\widetilde{h}.\end{aligned}$$

■

APPENDIX F
PROOF OF LEMMA IV.2

The following result indicates that the main lobe of the Dirichlet kernel contains most of its energy.

Lemma F.1. *Let $D_N(f) = \frac{\sin(\pi N f)}{\sin(\pi f)}$ be the Dirichlet kernel. Then*

$$\int_0^{\frac{1}{N}} |D_N(f)|^2\, df \geq 0.45N.$$

Proof (of Lemma F.1). Noting that $|D_N(f)| = \frac{|\sin(\pi N f)|}{|\sin(\pi f)|} \geq \frac{|\sin(\pi N f)|}{|\pi f|}$, we have

$$\begin{aligned}\int_0^{\frac{1}{N}} |D_N(f)|^2\, df \geq \int_0^{\frac{1}{N}} \left|\frac{\sin(\pi N f)}{\pi f}\right|^2 df &= \frac{N}{\pi}\int_0^{\pi} \left|\frac{\sin(f)}{f}\right|^2 df \\ &= \frac{N}{\pi}\int_0^{\pi} \frac{\frac{1-\cos(2f)}{2}}{f^2} df \\ &= \frac{N}{2\pi}\int_0^{\pi} \sum_{k=1}^{\infty} \frac{(-1)^{k+1}(2f)^{2k}}{(2k)!f^2} df \\ &= \frac{N}{\pi}\sum_{k=1}^{\infty} \frac{(-1)^{k+1}(2\pi)^{2k-1}}{(2k)!(2k-1)} \\ &\geq \frac{N}{\pi}\sum_{k=1}^{8} \frac{(-1)^{(k+1)}(2\pi)^{2k-1}}{(2k)!\,(2k-1)} \\ &\geq 0.45N,\end{aligned}$$

where the third line follows from the common Taylor series $\cos(2f) = \sum_{k=0}^{\infty}(-1)^k\frac{(2f)^{2k}}{(2k)!}$. ■

Suppose N is a multiple of 4. Note that

$$\begin{aligned}\lambda_l(\overline{\boldsymbol{C}}_N) &= \int_0^1 \left|\frac{1}{\sqrt{N}}\widetilde{\boldsymbol{e}}_{l/N}\right|^2 \widetilde{h}(f)df \\ &= \int_0^{\frac{1}{4}} \left|\frac{1}{\sqrt{N}}\widetilde{\boldsymbol{e}}_{l/N}(f)\right|^2 df + \int_{\frac{3}{4}}^1 \left|\frac{1}{\sqrt{N}}\widetilde{\boldsymbol{e}}_{l/N}(f)\right|^2 df.\end{aligned}$$

If $l = N/4$, $\left|\frac{1}{\sqrt{N}}\widetilde{\boldsymbol{e}}_{l/N}(f)\right| = \frac{1}{\sqrt{N}}\left|D_N(f - \frac{1}{4})\right|$. Thus

$$\begin{aligned}&\lambda_{N/4}(\overline{\boldsymbol{C}}_N) \\ =& \frac{1}{N}\int_0^{\frac{1}{4}} \left|D_N(f - \frac{1}{4})\right|^2 df + \frac{1}{N}\int_{\frac{3}{4}}^1 \left|D_N(f - \frac{1}{4})\right|^2 df \\ =& \frac{1}{N}\int_0^{1/2} |D_N(f)|^2\, df = \frac{1}{N}\frac{1}{2}\int_0^1 |D_N(f)|^2\, df = \frac{1}{2}.\end{aligned}$$

Similarly, we have $\lambda_{3N/4}(\overline{\boldsymbol{C}}_N) = \frac{1}{2}$.

Now for any $l \in [N]$, $\frac{l}{N} \in [0, \frac{1}{4}) \cup (\frac{3}{4}, 1]$, the main lobe of

$D_N(f-\frac{l}{N})$ is inside the interval $[0,\frac{1}{4}]\cup[\frac{3}{4},1]$. Thus

$$
\begin{aligned}
&\lambda_l(\overline{\boldsymbol{C}}_N)\\
=&\frac{1}{N}\int_0^{\frac{1}{4}}\left|D_N(f-\frac{l}{N})\right|^2 df+\frac{1}{N}\int_{\frac{3}{4}}^{1}\left|D_N(f-\frac{l}{N})\right|^2 df\\
\geq&\frac{2}{N}\int_0^{\frac{1}{N}}\left|D_N(f)\right|^2 df\geq 0.9.
\end{aligned}
$$

Similarly, for any $l\in[N]$, $\frac{l}{N}\in(\frac{1}{4},\frac{3}{4})$, we have

$$
\begin{aligned}
&\lambda_l(\overline{\boldsymbol{C}}_N)\\
=&\frac{1}{N}\int_0^{\frac{1}{4}}\left|D_N(f-\frac{l}{N})\right|^2 df+\frac{1}{N}\int_{\frac{3}{4}}^{1}\left|D_N(f-\frac{l}{N})\right|^2 df\\
\leq&1-\frac{2}{N}\int_0^{\frac{1}{N}}\left|D_N(f)\right|^2 df\leq 0.1.
\end{aligned}
$$

The proof is completed by noting that $0\leq\lambda_l(\overline{\boldsymbol{C}}_N)\leq 1$ for all $l\in[N]$. ■

References

[1] U. Grenander and G. Szegő, *Toeplitz Forms and Their Applications*, vol. 321. Univ of California Press, 1958.

[2] R. Gray, "On the asymptotic eigenvalue distribution of Toeplitz matrices," *IEEE Trans. Inf. Theory*, vol. 18, no. 6, pp. 725–730, 1972.

[3] J. Pearl, "On coding and filtering stationary signals by discrete Fourier transforms," *IEEE Trans. Inf. Theory*, vol. 19, no. 2, pp. 229–232, 1973.

[4] J. Makhoul, "Linear prediction: A tutorial review," *Proc. IEEE*, vol. 63, no. 4, pp. 561–580, 1975.

[5] T. Kailath, A. Vieira, and M. Morf, "Inverses of Toeplitz operators, innovations, and orthogonal polynomials," *SIAM Rev.*, vol. 20, no. 1, pp. 106–119, 1978.

[6] P. Deift, A. Its, and I. Krasovsky, "Eigenvalues of Toeplitz matrices in the bulk of the spectrum," *Bull. Inst. Math., Acad. Sin. (N.S.)*, pp. 437–461, 2012.

[7] R. M. Gray, "Toeplitz and circulant matrices: A review," *Commun. Inf. Theory*, vol. 2, no. 3, pp. 155–239, 2005.

[8] F. Avram, "On bilinear forms in gaussian random variables and Toeplitz matrices," *Probab. Theory Related Fields*, vol. 79, no. 1, pp. 37–45, 1988.

[9] S. V. Parter, "On the distribution of the singular values of Toeplitz matrices," *Linear Algebra Appl.*, vol. 80, pp. 115–130, 1986.

[10] E. E. Tyrtyshnikov, "A unifying approach to some old and new theorems on distribution and clustering," *Linear Algebra Appl.*, vol. 232, pp. 1–43, 1996.

[11] N. L. Zamarashkin and E. E. Tyrtyshnikov, "Distribution of eigenvalues and singular values of Toeplitz matrices under weakened conditions on the generating function," *Sbornik: Mathematics*, vol. 188, no. 8, pp. 1191–1201, 1997.

[12] D. Sakrison, "An extension of the theorem of Kac, Murdock and Szegő to N dimensions," *IEEE Trans. Inf. Theory*, vol. 15, no. 5, pp. 608–610, 1969.

[13] H. Gazzah, P. A. Regalia, and J.-P. Delmas, "Asymptotic eigenvalue distribution of block Toeplitz matrices and application to blind SIMO channel identification," *IEEE Trans. Inf. Theory*, vol. 47, no. 3, pp. 1243–1251, 2001.

[14] J. Gutiérrez-Gutiérrez and P. M. Crespo, "Asymptotically equivalent sequences of matrices and hermitian block Toeplitz matrices with continuous symbols: Applications to MIMO systems," *IEEE Trans. Inf. Theory*, vol. 54, no. 12, pp. 5671–5680, 2008.

[15] J. Bogoya, A. Bttcher, S. Grudsky, and E. Maximenko, "Maximum norm versions of the Szegő and Avram Parter theorems for Toeplitz matrices," *J. Approx. Theory*, vol. 196, no. 0, pp. 79 – 100, 2015.

[16] A. Dembo, "Bounds on the extreme eigenvalues of positive-definite Toeplitz matrices," *IEEE Trans. Inf. Theory*, vol. 34, no. 2, pp. 352–355, 1988.

[17] T. Laudadio, N. Mastronardi, and M. Van Barel, "Computing a lower bound of the smallest eigenvalue of a symmetric positive-definite Toeplitz matrix," *IEEE Trans. Inf. Theory*, vol. 54, no. 10, pp. 4726–4731, 2008.

[18] J. Mitola III and G. Q. Maguire Jr, "Cognitive radio: Making software radios more personal," *IEEE Personal Commun.*, vol. 6, no. 4, pp. 13–18, 1999.

[19] S. Haykin, "Cognitive radio: Brain-empowered wireless communications," *IEEE J. Select. Areas Commun.*, vol. 23, no. 2, pp. 201–220, 2005.

[20] Y. Zeng and Y.-C. Liang, "Eigenvalue-based spectrum sensing algorithms for cognitive radio," *IEEE Trans Commun.*, vol. 57, no. 6, pp. 1784–1793, 2009.

[21] G. Strang, "A proposal for Toeplitz matrix calculations," *Stud. Appl. Math.*, vol. 74, no. 2, pp. 171–176, 1986.

[22] T. F. Chan, "An optimal circulant preconditioner for Toeplitz systems," *SIAM J. Sci. Statis. Comput.*, vol. 9, no. 4, pp. 766–771, 1988.

[23] T. W. Körner, *Fourier analysis*. Cambridge university press, 1989.

[24] R. A. Horn and C. R. Johnson, eds., *Matrix Analysis*. New York, NY, USA: Cambridge University Press, 1986.

[25] R. H. Chan, "Circulant preconditioners for hermitian Toeplitz systems," *SIAM J. Matrix Anal. Appl.*, vol. 10, no. 4, pp. 542–550, 1989.

[26] R. H. Chan and G. Strang, "Toeplitz equations by conjugate gradients with circulant preconditioner," *SIAM J. Sci. Statis. Comput.*, vol. 10, no. 1, pp. 104–119, 1989.

[27] R. H. Chan, X.-Q. Jin, and M.-C. Yeung, "The spectra of super-optimal circulant preconditioned Toeplitz systems," *SIAM J. Numer. Anal.*, vol. 28, no. 3, pp. 871–879, 1991.

[28] W. F. Trench, "An elementary view of Weyl's theory of equal distribution," *Amer. Math. Monthly*, vol. 119, no. 10, pp. 852–861, 2012.

[29] D. Slepian, "Prolate Spheroidal Wave Functions, Fourier analysis, and uncertainty. V- The discrete case," *Bell Syst. Tech. J.*, vol. 57, no. 5, pp. 1371–1430, 1978.

[30] Z. Zhu and M. B. Wakin, "Approximating sampled sinusoids and multiband signals using multiband modulated DPSS dictionaries," to appear in *J. Fourier Anal. Appl.*, 2016.

[31] A. W. Van der Vaart, *Asymptotic Statistics*, vol. 3. Cambridge university press, 2000.

[32] P. Zizler, R. A. Zuidwijk, K. F. Taylor, and S. Arimoto, "A finer aspect of eigenvalue distribution of selfadjoint band Toeplitz matrices," *SIAM J. Matrix Anal. Appl.*, vol. 24, no. 1, pp. 59–67, 2002.

[33] T. M. Apostol, *Mathematical Analysis (2nd ed.)*. Addison Wesley Publishing Company, 1974.