On the Convergence of the Inverses of Toeplitz Matrices and Its Applications

Feng-Wen Sun, *Member, IEEE*, Yimin Jiang, *Member, IEEE*, and John S. Baras, *Fellow, IEEE*

***Abstract*—Many issues in signal processing involve the inverses of Toeplitz matrices. One widely used technique is to replace Toeplitz matrices with their associated circulant matrices, based on the well-known fact that Toeplitz matrices asymptotically converge to their associated circulant matrices in the weak sense. This often leads to considerable simplification. However, it is well known that such a weak convergence cannot be strengthened into strong convergence. It is this fact that severely limits the usefulness of the close relation between Toeplitz matrices and circulant matrices. Observing that communication receiver design often needs to seek optimality in regard to a data sequence transmitted within finite duration, we define the *finite-term strong convergence* regarding two families of matrices. We present a condition under which the inverses of a Toeplitz matrix converges in the strong sense to a circulant matrix for finite-term quadratic forms. This builds a critical link in the application of the convergence theorems for the inverses of Toeplitz matrices since the weak convergence generally finds its usefulness in issues associated with minimum mean squared error and the finite-term strong convergence is useful in issues associated with the maximum-likelihood or maximum *a posteriori* principles.**

***Index Terms*—Circulant matrix, maximum a posteriori, maximum likelihood, strong convergence, Toeplitz matrix.**

I. INTRODUCTION

A FAMILY of Toeplitz matrices $\boldsymbol{T}_n$ is defined by a sequence of complex numbers $\{t_i; i = \ldots, -1, 0, 1, \ldots\}$ such that the entry of $\boldsymbol{T}_n$ at the ith row and jth column is equal to t_{i-j}, i.e., $\boldsymbol{T}_n = \{t_{i-j}\}$. We restrict our discussion to the case that $t_{-i} = t_i^*$, where t_i^* is the complex conjugate of t_i. With this restriction, $\boldsymbol{T}_n$ becomes Hermitian. Toeplitz Hermitian matrices play a pivotal role in signal processing. In fact, what is really relevant is the inverse of such a matrix rather than the matrix itself for many applications. For instance, if t_i represents the autocorrelation of a stationary random process, the inverse of $\boldsymbol{T}_n$ is associated with the joint probability density function of n consecutive samples of the random process. In *filtering* problems, such an inverse appears in the Wiener–Hopf equation [1], [2].

One of the difficulties in analyzing the inverse matrices arises from the fact that the inverse of a Toeplitz matrix is no longer Toeplitz, though it was shown in [3] and [4] that such an inverse can be decomposed into multiplication and summation of Toeplitz matrices.

One technique to tackle the problem is to exploit the relation between Toeplitz matrices and their associated circulant matrices. An $n \times n$ matrix is called a *circulant* matrix if its (i, j)th entry is only a function of $(i - j) \bmod n$. In particular, for the family of Toeplitz matrices defined by the sequence $\{t_i\}$, a family of their associated circulant matrices can be defined through the discrete-time Fourier transform (DTFT) of the sequence $\{t_i\}$.Let $\mathcal{F}(\lambda)$ denote the DTFT of $\{t_i\}$, i.e.,

$$\mathcal{F}(\lambda) = \sum_{k=-\infty}^{\infty} t_k e^{-j\lambda k}.$$

Note that $\mathcal{F}(\lambda)$ is real due to the Hermitian constraint. Let $\boldsymbol{U}_n$ denote the unitary matrix defined as

$$\boldsymbol{U}_n = \frac{1}{\sqrt{n}} \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & e^{-j(2\pi/n)} & \cdots & e^{-j(2\pi(n-1)/n)} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & e^{-j(2\pi(n-1)/n)} & \cdots & e^{-j(2\pi(n-1)(n-1)/n)} \end{bmatrix} \tag{1}$$

and $\boldsymbol{D}_n$ denote the diagonal matrix with the ith diagonal entry equal to $\mu_{i,n} = \mathcal{F}(2\pi i/n)$, i.e.,

$$\boldsymbol{D}_n = \mathrm{diag}\{\mu_{0,n}, \mu_{1,n}, \ldots, \mu_{n-1,n}\}. \tag{2}$$

The matrix

$$\boldsymbol{C}_n = \boldsymbol{U}_n^H \boldsymbol{D}_n \boldsymbol{U}_n \tag{3}$$

is a circulant matrix [5], [6].

It has been observed that in many applications substituting $\boldsymbol{T}_n$ with $\boldsymbol{C}_n$ often leads to very useful and dramatic simplification to the problems at hand. This is due to the fact that the inverse of a circulant matrix is still circulant, which can be diagonalized by the discrete Fourier transform (DFT). The DFT-based eigendecomposition of $\boldsymbol{C}_n$ usually provides additional insight into the frequency domain. Apparently, in order to make such a substitution meaningful, the inverses of Toeplitz matrices need to converge to their associated circulant matrices. Depending on applications, there are many different ways to define matrix convergence. We first examine the known convergence of Toeplitz matrices to circulant matrices and typical applications associated with them.

Manuscript received October 20, 2000; revised August 1, 2002.

F.-W. Sun and Y. Jiang are with Hughes Network Systems, Inc., Germantown, MD 20876 USA (e-mail: fsun@hns.com; yjiang@hns.com).

J. S. Baras is with the Department of Electrical and Computer Engineering and the Institute for Systems Research, University of Maryland, College Park, MD 20742 USA (e-mail: baras@isr.umd.edu).

Communicated by J. A. O'Sullivan, Associate Editor for Detection and Estimation.

Digital Object Identifier 10.1109/TIT.2002.806157

The best known convergence is the weak convergence, which is based on the weak norm defined for an $n \times n$ matrix $\boldsymbol{A} = \{a_{ij}\}$ as

$$\sqrt{n^{-1} \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} |a_{ij}|^2}. \tag{4}$$

It can be shown that the Toeplitz matrix $\boldsymbol{T}_n$ converges to $\boldsymbol{C}_n$ in the weak sense as long as $|\mathcal{F}(\lambda)|$ is bounded [5], [6].[1] Note that $\boldsymbol{T}_n$ converging to $\boldsymbol{C}_n$ may not necessarily mean that $\boldsymbol{T}_n^{-1}$ converges to $\boldsymbol{C}_n^{-1}$ even if $\boldsymbol{T}_n^{-1}$ does exist. A sufficient condition for the weak convergence of the inverse is that the strong norm of $\boldsymbol{T}_n^{-1}$ and $\boldsymbol{C}_n^{-1}$ is uniformly bounded [6]. The strong norm for a Hermitian matrix $\boldsymbol{A}$ can be defined as

$$\|\boldsymbol{A}\| = \max_{\boldsymbol{x}} \left| \frac{\boldsymbol{x}^H \boldsymbol{A} \boldsymbol{x}}{\boldsymbol{x}^H \boldsymbol{x}} \right|$$

where the maximum is over all the vectors of the same dimension as $\boldsymbol{A}$. The strong norm is also called the *spectrum norm* for a Hermitian matrix. In [7], the weak convergence was extended to a different family of circulant matrices $\tilde{\boldsymbol{C}}_n$ for the case that $t_k = t_{-k}$, where $\tilde{\boldsymbol{C}}_n$ depends only on t_k, $k = 0, \ldots, n-1$ rather than the entire sequence. Specifically, the diagonal entries of $\tilde{\boldsymbol{D}}_n$, a diagonal matrix, are taken as those of $\boldsymbol{U}_n \boldsymbol{T}_n \boldsymbol{U}_n^H$ and again $\tilde{\boldsymbol{C}}_n = \boldsymbol{U}_n^H \tilde{\boldsymbol{D}}_n \boldsymbol{U}_n$. In this case, the (i, j)th entry of $\tilde{\boldsymbol{C}}_n$ equals [7]

$$t_{|i-j|} + \frac{|i-j|}{n} \left[t_{n-|i-j|} - t_{|i-j|} \right].$$

When the condition $t_k = t_{-k}$ does not hold, a different family of $\hat{\boldsymbol{C}}_n$ with the (i, j)th entry equal to $t_{-|i-j|} + t_{n-|i-j|}$ can be defined [6]. It can be shown that $\boldsymbol{T}_n$ converges to $\hat{\boldsymbol{C}}_n$ and $\tilde{\boldsymbol{C}}_n$ in the weak sense. The convergence can be readily extended to the inverse of these matrices [6], [7].

Examining the definition of the weak norm in (4), we can see that the weak convergence is in the mean sense due to the division factor $1/n$. Indeed, several successful applications of the weak approximation theory relate to the evaluation of the mean of some quantities, such as source coding and filtering problems based on the minimum mean squared error (MMSE) criterion, or computing the mean of a quadratic form associated with a random process [6]–[9].

However, the usefulness of all these convergence theorems is severely limited due to the fact that many applications actually involve quadratic forms of $\boldsymbol{T}_n^{-1}$. Even if $\boldsymbol{T}_n^{-1}$ converges to $\boldsymbol{C}_n^{-1}$ in the weak sense, substituting $\boldsymbol{T}_n^{-1}$ with $\boldsymbol{C}_n^{-1}$ may not yield correct results since the convergence of a quadratic form can only be guaranteed if the convergence is in the strong sense. Note that in the literature, replacing the inverse of a Toeplitz matrix with a circulant matrix in evaluating quadratic forms associated with likelihood functions or Wiener filtering problems has been widely used [10]–[15]. However, such an approximation was used in the cited references without theoretical basis. It may potentially lead to erroneous results.

[1]In [5], $\boldsymbol{C}_n$ is defined through the inverse DFT, i.e., the ith diagonal entry of $\boldsymbol{D}_n$ is equal to $\mathcal{F}(-2\pi i/n)$ and $\boldsymbol{U}_n$ is replaced by $\boldsymbol{U}_n^T$. The current notation is more consistent with engineering conventions.

When the strong convergence is of interest, it is worth mentioning that Baxter [16], [17] showed that all the entries of any *fixed* row of $\boldsymbol{T}_n^{-1}$ uniformly converge under the Paley–Wiener condition. The standard spectral factorization theorem can be used to give a closed-form formula for the entries of the converged inverse matrix. Actually, Baxter's results can be used to show that $\boldsymbol{T}_n^{-1}$ can converge to a circulant matrix in the strong sense only when each of the $\boldsymbol{T}_n$'s is an identity matrix.

In [12], it is shown that the inverse of a *bidirectional* infinite-dimension Toeplitz matrix $\boldsymbol{T}_n$ with its (i, j)th entry equal to t_{i-j} has a circulant inverse. Bidirectional infinity means that i and j range from $-\infty$ to ∞. Note that the result in [12] does not indicate in any way whether the finite-dimension matrices $\boldsymbol{T}_n^{-1}$ converge to the infinite-dimension matrices. This result is often mistakenly quoted as the theoretical basis for replacing the inverse of a Toeplitz matrix with a circulant matrix for quadratic forms.

The association between a Toeplitz matrix and a circulant matrix was also exploited for iterative computation of the inverse of the original Toeplitz matrix, where a circulant matrix is used as a preconditioner to reduce the computation load and improve the stability of the numerical algorithms [18]–[20]. For preconditioning, the major issues include the selection of the preconditioning circulant matrix and the distribution of the eigenvalues of the preconditioned matrix $\boldsymbol{C}^{-1}\boldsymbol{T}$ ($\boldsymbol{T}$ is a Toeplitz matrix, $\boldsymbol{C}$ is its preconditioner). In order to speed up the convergence of an iterative algorithm, it is desirable to make the eigenvalues of the preconditioned matrix cluster around a single value. For details, see [18] and the references therein. As pointed out by Chan and Ng [18], the circulant approximation in preconditioning is *not* to replace the Toeplitz matrix with a circulant matrix in the subsequent computation, and the preconditioning does not alter the solution to the Toeplitz systems.

The objective of this paper is to extend the convergence theorems to a form of strong convergence such that a large class of communication receiver design problems involving the maximum-likelihood (ML) and maximum *a posteriori* (MAP) principles can benefit from the relation between Toeplitz matrices and circulant matrices.

Designers of communication systems often seek optimality for a data sequence transmitted within finite duration, such as a finite-length training sequence for synchronization, or a finite-length spreading sequence for spread spectrum. Based on this observation, in Section II, we introduce the definition of the *finite-term strong convergence* regarding two families of matrices. We demonstrate that the design schemes based on the ML and MAP principles can benefit from such a convergence. The finite-term strong convergence separates the length of the transmission window from that of the observation window. By increasing the observation window, the receiver design approaches the optimal solution when the noise incurred in the system is correlated. Therefore, we can often obtain a closed-form formula for the optimal receiver design through substituting Toeplitz matrices with their associated circulant matrices when the condition of the finite-term strong convergence is met. Such a convergence builds a critical link in the application of the convergence theorems for the inverses of Toeplitz matrices since the well-known weak convergence generally finds its usefulness in issues associated with the MMSE criterion and the finite-term strong convergence is useful in issues associated with the ML and MAP rules. In Section III, we further present a condition under which the inverse of Toeplitz matrices converges to a circulant matrix in the finite-term strong sense. In Section IV, we demonstrate a typical application of the finite-term strong convergence by deriving a novel timing/phase estimator which does not require an integer number of samples per symbol. This can significantly reduce the sampling rate requirement for high-speed modem design. Section V concludes the paper.

II. Finite-Term Strong Convergence

Consider the quadratic form $\boldsymbol{x}^H\boldsymbol{T}_n^{-1}\boldsymbol{x}$ with $\boldsymbol{x}$ having only a finite number of nonzero terms, without loss of generality, assuming in the middle of the vector, i.e.,

$$\boldsymbol{x} = (0, \ldots, 0, x_{-L}, \ldots, x_0, \ldots, x_L, 0, \ldots, 0)$$

where L does not increase with n, the dimension of the vector. We denote such a quadratic form as a *finite-term quadratic form*. The finite-term strong convergence for two families of Hermitian matrices is defined as follows.

Definition 1: For two families of Hermitian matrices $\boldsymbol{A}_n$ and $\boldsymbol{B}_n$, consider the quadratic form

$$\max_{\boldsymbol{x}} \frac{||(\boldsymbol{A_n} - \boldsymbol{B_n})\boldsymbol{x}||}{||\boldsymbol{x}||}, \tag{5}$$

where $||\boldsymbol{x}|| \triangleq \sqrt{\boldsymbol{x}^H\boldsymbol{x}}$ is the vector norm for a vector $\boldsymbol{x}$, the maximum is over all the n-dimension vectors of the form

$$\boldsymbol{x} = (0, \ldots, 0, x_{-L}, \ldots, x_0, \ldots, x_L, 0, \ldots, 0). \tag{6}$$

If (5) converges to zero for any given L as $n \to \infty$, we say that $\boldsymbol{A}_n$ converges to $\boldsymbol{B}_n$ in the finite-term strong sense. The following well-known equation for a Hermitian matrix $\boldsymbol{A}$ establishes the link between the quadratic form and the spectral norm [6]

$$\max_{\boldsymbol{x}} \left| \frac{\boldsymbol{x}^H\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^H\boldsymbol{x}} \right| = \max_{\boldsymbol{x}} \frac{||\boldsymbol{A}\boldsymbol{x}||}{||\boldsymbol{x}||} = |\lambda_M|$$

where $|\lambda_M|$ is the largest absolute eigenvalue of $\boldsymbol{A}$.

As pointed out earlier, the weak convergence theorems can be used in solving linear filtering and coding problems based on the MMSE criterion [7], [9]. However, they are not very useful for designing the ML and MAP algorithms that are widely adopted in digital receivers. We shall show that the finite-term strong convergence can play a pivotal role in design schemes based on the ML or MAP principles.

The received signals in many digital communication systems can be modeled by a desired signal carrying user data and transmission parameters embedded in a Gaussian noise [10], i.e.,

$$\boldsymbol{y} = \boldsymbol{s}(\boldsymbol{I}_u, \boldsymbol{P}) + \boldsymbol{N} \tag{7}$$

where $\boldsymbol{y}$ is the received signal vector of length n, $\boldsymbol{s}(\boldsymbol{I}_u, \boldsymbol{P})$ is the transmitted signal vector, $\boldsymbol{I}_u$ is the user data vector, $\boldsymbol{P}$ is the synchronization and channel parameters, $\boldsymbol{N}$ is the channel noise vector, which is a zero-mean and correlated wide-sense stationary Gaussian process. The objective of receiver design is to detect the user data $\boldsymbol{I}_u$ and/or estimate the parameters $\boldsymbol{P}$ based on the observation $\boldsymbol{y}$. We further assume that the signal $\boldsymbol{s}(\boldsymbol{I}_u, \boldsymbol{P})$ has a finite number of nonzero terms. The finite-length assumption is valid for many receiver design problems in digital communications. Actually, communication signals are always both time and bandwidth limited in an engineering sense. For instance, the number of nonzero terms can be equal to the length of a training sequence for data-aided synchronization, or the length of a spread sequence for spread spectrum systems, etc. Due to the correlation of the noise, optimal receiving schemes require that the observation window be larger than the signal transmission window. Therefore, the observation vector $\boldsymbol{y}$ typically is longer than the transmitted data for optimal performance, i.e., $\boldsymbol{s}(\boldsymbol{I}_u, \boldsymbol{P})$ can be modeled as the form in (6). Note that it is unlikely to transmit an isolated data segment in engineering practice. For instance, a training sequence is typically followed by user data, and a spreading sequence is followed by another spreading sequence. However, it is a common practice to analyze and design a communication system based on one-shot observation of the designated data since it can often avoid unnecessary complication of the following or preceding data which are not used in the receiver design. Separating the observation window and the transmission window can largely enhance the usefulness of the convergence theorem to be presented. In [10], [14], [15], besides ignoring the condition for the convergence, the authors assume that only for long transmission is it possible to replace the Toeplitz matrices with their associated circulant matrices.

Following the Gaussian distribution assumption, the likelihood function of $\boldsymbol{s}(\boldsymbol{I}_u, \boldsymbol{P})$ can be written as

$$\begin{aligned} &f(\boldsymbol{y}|\boldsymbol{s}(\boldsymbol{I}_u, \boldsymbol{P})) \\ &\quad = C\exp\left(-\frac{1}{2}\,(\boldsymbol{y} - \boldsymbol{s}(\boldsymbol{I}_u, \boldsymbol{P}))^H\boldsymbol{T}_n^{-1}(\boldsymbol{y} - \boldsymbol{s}(\boldsymbol{I}_u, \boldsymbol{P}))\right) \end{aligned} \tag{8}$$

where $\boldsymbol{T}_n$ is the autocovariance matrix of the noise process $\boldsymbol{N}$, which is Toeplitz and Hermitian, and C is a constant independent of the transmitted signal. The ML algorithm is to find the $\boldsymbol{I}_u$ and $\boldsymbol{P}$ that maximize the likelihood function. It is difficult to evaluate (8) because the inverse matrix $\boldsymbol{T}_n^{-1}$ is hard to obtain analytically. Replacing $\boldsymbol{T}_n^{-1}$ with its associated circulant matrix $\boldsymbol{C}_n^{-1}$ naturally leads into the frequency-domain approach since a circulant matrix can be eigendecomposed by the DFT. The weak norm approximation is not a sufficient tool to warrant such a replacement because the transmitted signal $\boldsymbol{s}(\boldsymbol{I}_u, \boldsymbol{P})$ can be arbitrary.

Assume that $\boldsymbol{T}_n^{-1}$ converges to $\boldsymbol{C}_n^{-1}$ in the finite-term strong sense, i.e.,

$$\lim_{n\to\infty} ||\boldsymbol{T}_n^{-1}\boldsymbol{s}(\boldsymbol{I}_u, \boldsymbol{P})|| = \lim_{n\to\infty} ||\boldsymbol{C}_n^{-1}\boldsymbol{s}(\boldsymbol{I}_u, \boldsymbol{P})||. \tag{9}$$

Let us examine the condition for replacing $\boldsymbol{T}_n^{-1}$ with $\boldsymbol{C}_n^{-1}$ in (8).The exponent of (8) consists of the following four terms:

$$\begin{gathered} \boldsymbol{y}^H\boldsymbol{T}_n^{-1}\boldsymbol{y} \\ \boldsymbol{y}^H\boldsymbol{T}_n^{-1}\boldsymbol{s}(\boldsymbol{I}_u, \boldsymbol{P}) \\ \boldsymbol{s}(\boldsymbol{I}_u, \boldsymbol{P})^H\boldsymbol{T}_n^{-1}\boldsymbol{y} \\ \boldsymbol{s}(\boldsymbol{I}_u, \boldsymbol{P})^H\boldsymbol{T}_n^{-1}\boldsymbol{s}(\boldsymbol{I}_u, \boldsymbol{P}). \end{gathered}$$

The convergence of

$$\lim_{n\to\infty} \boldsymbol{s}(\boldsymbol{I}_u, \boldsymbol{P})^H \boldsymbol{T}_n^{-1} \boldsymbol{s}(\boldsymbol{I}_u, \boldsymbol{P}) = \lim_{n\to\infty} \boldsymbol{s}(\boldsymbol{I}_u, \boldsymbol{P})^H \boldsymbol{C}_n^{-1} \boldsymbol{s}(\boldsymbol{I}_u, \boldsymbol{P})$$

follows directly from the assumed convergence (9). The convergence of $\boldsymbol{y}^H \boldsymbol{T}_n^{-1} \boldsymbol{s}(\boldsymbol{I}_u, \boldsymbol{P})$ and $\boldsymbol{s}(\boldsymbol{I}_u, \boldsymbol{P})^H \boldsymbol{T}_n^{-1} \boldsymbol{y}$ for normalized received vector $\boldsymbol{y}$ follows from the fact that

$$\|\boldsymbol{y}^H(\boldsymbol{T}_n^{-1} - \boldsymbol{C}_n^{-1})\boldsymbol{s}(\boldsymbol{I}_u, \boldsymbol{P})\| \le \|\boldsymbol{y}\| \|(\boldsymbol{T}_n^{-1} - \boldsymbol{C}_n^{-1})\boldsymbol{s}(\boldsymbol{I}_u, \boldsymbol{P})\|.$$

The term $\boldsymbol{y}^H \boldsymbol{T}_n^{-1} \boldsymbol{y}$ may not converge to $\boldsymbol{y}^H \boldsymbol{C}_n^{-1} \boldsymbol{y}$ since $\boldsymbol{y}$ cannot be modeled as a finite-term vector due to the noise. However, this does not affect the derivation of the ML (or MAP) solution since $\boldsymbol{y}$ does not include the useful $\boldsymbol{I}_u$ and $\boldsymbol{P}$ terms. This means that we can replace this term with $\boldsymbol{y}^H \boldsymbol{C}_n^{-1} \boldsymbol{y}$ regardless of the convergence. Therefore, the finite-term strong convergence is sufficient for replacing $\boldsymbol{T}_n^{-1}$ with $\boldsymbol{C}_n^{-1}$ in (8) for a large observation window.

Note that we did not impose any condition on the transmitted signal $\boldsymbol{s}(\boldsymbol{I}_u, \boldsymbol{P})$ as long as it has a finite number of nonzero terms. Since the optimal solution calls for a long observation window, this implies that the solution obtained by increasing the observation window and using circulant matrices will yield the true optimal ML or MAP solutions.

In the next section, we will present a condition for the inverses of Toeplitz matrices to converge to their associated circulant matrices in the finite-term strong sense.

III. A Condition for the Convergence

For presenting the finite-term strong convergence theorem, the concept of a *partial DTFT* is helpful in shortening the notation. For any integer w, the partial DTFT of the sequence $\{t_i\}$ is defined as

$$\mathcal{PF}(w, \lambda) = \sum_{k=w}^{\infty} t_k e^{-jk\lambda} \tag{10}$$

which is based on the observation that $\mathcal{PF}(w, \lambda)$ is actually the DTFT of $\{\ldots, 0, t_w, t_{w+1}, \ldots\}$. Note that $\mathcal{PF}(-\infty, \lambda)$ is the DTFT of $\{t_i\}$, i.e., $\mathcal{PF}(-\infty, \lambda) = \mathcal{F}(\lambda)$. We further denote the ratio of the partial DTFT to the DTFT as $\mathcal{R}_f(w, \lambda)$, i.e.,

$$\mathcal{R}_f(w, \lambda) \triangleq \frac{\mathcal{PF}(w, \lambda)}{\mathcal{PF}(-\infty, \lambda)} = \frac{\mathcal{PF}(w, \lambda)}{\mathcal{F}(\lambda)}. \tag{11}$$

Since the difference between $\boldsymbol{T}_n^{-1}$ and $\boldsymbol{C}_n^{-1}$ is equal to $\boldsymbol{T}_n^{-1}(\boldsymbol{I}_n - \boldsymbol{T}_n \boldsymbol{C}_n^{-1})$, where $\boldsymbol{I}_n$ is the n-dimension identity matrix, we start with the evaluation of the matrix $\boldsymbol{I}_n - \boldsymbol{T}_n \boldsymbol{C}_n^{-1}$. The following lemma expresses the entries of $(\boldsymbol{I}_n - \boldsymbol{T}_n \boldsymbol{C}_n^{-1})$ by means of the partial DTFT.

Lemma 1: The (w, v)th entry of $\boldsymbol{I}_n - \boldsymbol{T}_n \boldsymbol{C}_n^{-1}$ is equal to

$$(\boldsymbol{I}_n - \boldsymbol{T}_n \boldsymbol{C}_n^{-1})_{w,v} = \frac{1}{n} \sum_{s=0}^{n-1} e^{j2\pi(w-v)s/n} \left(\mathcal{R}_f(w+1, 2\pi s/n) + \mathcal{R}_f(n-w, 2\pi s/n)^*\right). \tag{12}$$

Furthermore, $(\boldsymbol{I}_n - \boldsymbol{T}_n \boldsymbol{C}_n^{-1})_{w,v}$ is upper-bounded by

$$\begin{aligned} |(\boldsymbol{I}_n - \boldsymbol{T}_n \boldsymbol{C}_n^{-1})_{w,v}| &\le \max_{0\le s\le n-1} \frac{1}{2} |\mathcal{R}_f(n-w, 2\pi(s+1)/n) \\ &\qquad - \mathcal{R}_f(n-w, 2\pi s/n)| \\ &\quad + \max_{0\le s\le n-1} \frac{1}{2} |\mathcal{R}_f(w+1, 2\pi(s+1)/n) \\ &\qquad - \mathcal{R}_f(w+1, 2\pi s/n)| \\ &\quad + \Big(\max_{0\le s\le n-1} |\mathcal{R}_f(w+1, 2\pi s/n)| \\ &\qquad + \max_{0\le s\le n-1} |\mathcal{R}_f(n-w, 2\pi s/n)| \Big) \\ &\quad \cdot \frac{|2\pi(w-v+n/2)/n \bmod 2\pi| n + 2}{2n}. \end{aligned} \tag{13}$$

Proof: See the Appendix. □

Lemma 2: Let $\boldsymbol{x}$ be of the form defined in (6). Further assume that every entry of the middle $2L+1$ columns of $\boldsymbol{I}_n - \boldsymbol{T}_n \boldsymbol{C}_n^{-1}$ is bounded by B, i.e.,

$$B = \max_{|v-n/2|\le L} |(\boldsymbol{I}_n - \boldsymbol{T}_n \boldsymbol{C}_n^{-1})_{w,v}|$$

and there are at most M rows of $\boldsymbol{I}_n - \boldsymbol{T}_n \boldsymbol{C}_n^{-1}$ containing nonzero entries, then

$$\|(\boldsymbol{T}_n^{-1} - \boldsymbol{C}_n^{-1})\boldsymbol{x}\| \le \|\boldsymbol{T}_n^{-1}\| \sqrt{(2L+1)M} B \|\boldsymbol{x}\| \tag{14}$$

where $\|\boldsymbol{T}_n^{-1}\|$ is the spectral norm of the matrix.

Proof: The following inequality:

$$\|(\boldsymbol{T}_n^{-1} - \boldsymbol{C}_n^{-1})\boldsymbol{x}\| \le \|\boldsymbol{T}_n^{-1}\| \cdot \|(\boldsymbol{I}_n - \boldsymbol{T}_n \boldsymbol{C}_n^{-1})\boldsymbol{x}\|$$

reduces the proof to show that

$$\|(\boldsymbol{I}_n - \boldsymbol{T}_n \boldsymbol{C}_n^{-1})\boldsymbol{x}\| \le \sqrt{(2L+1)M} B \|\boldsymbol{x}\|.$$

Let the matrix $\boldsymbol{Q}_n$ be obtained by setting all the columns of $\boldsymbol{I}_n - \boldsymbol{T}_n \boldsymbol{C}_n^{-1}$ to zero except the middle $2L+1$ columns, then

$$\boldsymbol{Q}_n \boldsymbol{x} = (\boldsymbol{I}_n - \boldsymbol{T}_n \boldsymbol{C}_n^{-1})\boldsymbol{x}$$

for any $\boldsymbol{x}$ of the form

$$(0, \ldots, 0, x_{-L}, \ldots, x_0, \ldots, x_L, 0, \ldots, 0).$$

Therefore, if we can show that for any $\boldsymbol{x}$

$$\|\boldsymbol{Q}_n \boldsymbol{x}\| \le \sqrt{(2L+1)M} B \|\boldsymbol{x}\| \tag{15}$$

the lemma will follow. Inequality (15) implies that

$$\|\boldsymbol{Q}_n\| \le \sqrt{(2L+1)M} B$$

where $\|\boldsymbol{Q}_n\| = \max\{\sqrt{\lambda}: \lambda \text{ is an eigenvalue of } \boldsymbol{Q}_n^H \boldsymbol{Q}_n\}$ is the spectrum norm. Since the nonzero entries of $\boldsymbol{Q}_n$ are bounded by B, the nonzero entries of $\boldsymbol{Q}_n^H \boldsymbol{Q}_n$ are bound by MB^2. The matrix $\boldsymbol{Q}_n^H \boldsymbol{Q}_n$ has at most $2L+1$ nonzero entries for any row or column. This shows that the largest eigenvalue of $\boldsymbol{Q}_n^H \boldsymbol{Q}_n$ is bounded by $(2L+1)MB^2$. Thus,

$$\|\boldsymbol{Q}_n\| \le \sqrt{(2L+1)M}\,B.$$

This completes the proof of the lemma. □

Lemma 3: For a continuous $\mathcal{R}_f(w, \lambda)$, if both the $\mathcal{R}_f(w, \lambda)$ and $d\mathcal{R}_f(w, \lambda)/d\lambda$ are bounded, then

$$\begin{aligned} &|(\boldsymbol{I}_n - \boldsymbol{T}_n \boldsymbol{C}_n^{-1})_{w,v}| \\ &\quad \le \frac{\pi(B_2(w+1)+B_2(n-w))}{n} + (B_1(w+1)+B_1(n-w)) \\ &\qquad \cdot \frac{|2\pi(w-v+n/2)/n \bmod 2\pi|n+2}{2n} \end{aligned} \tag{16}$$

where

$$B_1(w) \triangleq \max_\lambda |\mathcal{R}_f(w, \lambda)|$$

and

$$B_2(w) \triangleq \max_\lambda |d\mathcal{R}_f(w, \lambda)/d\lambda|.$$

Proof: For any $w \in [0, n-1]$, there exists a

$$\lambda_0 \in [2\pi s/n, 2\pi(s+1)/n]$$

such that

$$\begin{aligned} &|\mathcal{R}_f(w, 2\pi(s+1)/n) - \mathcal{R}_f(w, 2\pi s/n)| \\ &\quad = \left| \frac{2\pi}{n} \frac{d\mathcal{R}_f(w,\lambda)}{d\lambda} \right|_{\lambda=\lambda_0} \le \max_\lambda \left| \frac{d\mathcal{R}_f(w,\lambda)}{d\lambda} \right| \cdot \frac{2\pi}{n}. \end{aligned} \tag{17}$$

The inequality of (16) follows (13) and (17). □

Lemma 4: If the sequence $\{t_i\}$ satisfies the following conditions:

$$\sum_{k=-\infty}^{\infty} |kt_k| < \infty$$

and its DTFT $\mathcal{F}(\lambda) \neq 0, \ \forall \lambda \in [0, 2\pi]$, then

$$B_1(w) \le C \sum_{k=w}^{\infty} |t_k| \tag{18}$$

and

$$B_2(w) \le D \sum_{k=w}^{\infty} |kt_k| \tag{19}$$

where C, D are constants.

Proof: Under the condition that

$$\sum_{k=-\infty}^{\infty} |kt_k| < \infty$$

the DTFT of $\{t_i\}$ is continuous [6]. Therefore, there is a minimum for $\mathcal{F}(\lambda)$ over $[0, 2\pi]$. Based on the assumption of the lemma, this minimum is nonzero. Clearly

$$|B_1(w)| = \max_\lambda \left| \frac{\mathcal{P}\mathcal{F}(w,\lambda)}{\mathcal{F}(\lambda)} \right| \le \frac{1}{\min_{\lambda\in[0,2\pi]} |\mathcal{F}(\lambda)|} \cdot \sum_{k=w}^{\infty} |t_k|.$$

Similarly, based on the definition of $B_2(w)$, it follows that

$$\begin{aligned} |B_2(w)| &\le \frac{\sum_{k=w}^{\infty} |kt_k| \sum_{l=-\infty}^{\infty} |t_l| + \sum_{k=w}^{\infty} |t_k| \sum_{l=-\infty}^{\infty} |lt_l|}{\min_{\lambda\in[0,2\pi]} |\mathcal{F}(\lambda)|^2} \\ &\le \frac{2\sum_{l=-\infty}^{\infty} |t_l|}{\min_{\lambda\in[0,2\pi]} |\mathcal{F}(\lambda)|^2} \cdot \sum_{k=w}^{\infty} |kt_k|, \end{aligned}$$

where the second inequality follows the fact that $|t_k| = |t_{-k}|$ and

$$\begin{aligned} &\sum_{k=w}^{\infty} |kt_k| \sum_{l=-\infty}^{\infty} |t_l| - \sum_{k=w}^{\infty} |t_k| \sum_{l=-\infty}^{\infty} |lt_l| \\ &\quad = \sum_{k=w}^{\infty} \sum_{l=-(w-1)}^{w-1} (|kt_k||t_l| - |t_k||lt_l|) \\ &\quad = \sum_{k=w}^{\infty} \sum_{l=-(w-1)}^{w-1} |t_k t_l|(|k| - |l|) \ge 0. \end{aligned}$$ □

Theorem 1: Let $\boldsymbol{T}_n$ be a family of Toeplitz Hermitian matrices associated with the sequence $\{t_i\}$, and $\mathcal{F}(\lambda)$ be the DTFT of $\{t_i\}$. If $|\mathcal{F}(\lambda)| \neq 0$ for $\lambda \in [0, 2\pi]$ and $\sum_{k=-\infty}^{\infty} |kt_k| < \infty$, $\boldsymbol{T}_n^{-1}$ converges to $\boldsymbol{C}_n^{-1}$ in the finite-term strong sense.

Furthermore, for a vector $\boldsymbol{x}$ with the form defined in (6), the quadratic form is bounded by

$$\frac{\|(\boldsymbol{T}_n^{-1} - \boldsymbol{C}_n^{-1})\boldsymbol{x}\|}{\|\boldsymbol{x}\|} \le O\left(1/\sqrt{n}\right). \tag{20}$$

Proof: Since all the rows of $\boldsymbol{I}_n - \boldsymbol{T}_n \boldsymbol{C}_n^{-1}$ are in general nonzero, the M in Lemma 2 becomes n and

$$\begin{aligned} \frac{\|(\boldsymbol{T}_n^{-1} - \boldsymbol{C}_n^{-1})\boldsymbol{x}\|}{\|\boldsymbol{x}\|} &\le \|\boldsymbol{T}_n^{-1}\| \sqrt{(2L+1)n} \\ &\quad \cdot \max_{|v-n/2|\le L} |(\boldsymbol{I}_n - \boldsymbol{T}_n \boldsymbol{C}_n^{-1})_{w,v}|. \end{aligned} \tag{21}$$

It is known that all the eigenvalues (defined as $\tau_{s,n}$, $s \in [0, n-1]$) of $\boldsymbol{T}_n$ are between m_f and M_f [5, p. 64], where m_f is the greatest lower bound of $\mathcal{F}(\lambda)$ and M_f is the least upper bound of $\mathcal{F}(\lambda)$, respectively, i.e.,

$$m_f \le \tau_{s,n} \le M_f.$$

Since $\mathcal{F}(\lambda)$ is continuous and nonzero, m_f and M_f have the same sign, and all the eigenvalues of $\boldsymbol{T}_n$ are bounded by

$$|\tau_{s,n}| \ge \min\{|m_f|, |M_f|\} > 0$$

which means that $\|\boldsymbol{T}_n^{-1}\|$, the strong norm of $\boldsymbol{T}_n^{-1}$, is bounded by $1/\min\{|m_f|, |M_f|\} < \infty$. Now we turn to

$$\max_{|v-n/2|\le L} |(\boldsymbol{I}_n - \boldsymbol{T}_n \boldsymbol{C}_n^{-1})_{w,v}|.$$

Combining Lemma 3, Lemma 4, and $|v-n/2|\le L$, we can easily see that

$$\begin{aligned}|(\boldsymbol{I}_n-\boldsymbol{T}_n\boldsymbol{C}_n^{-1})_{w,v}| \le{}& \frac{\pi D}{n}\cdot\left(\sum_{k=w+1}^{\infty}|kt_k|+\sum_{k=n-w}^{\infty}|kt_k|\right)\\ &+C\left(\sum_{k=w+1}^{\infty}|t_k|+\sum_{k=n-w}^{\infty}|t_k|\right)\\ &\cdot\frac{|2\pi w/n \bmod 2\pi|n+2\pi L+2}{2n}\end{aligned}\tag{22}$$

which follows that

$$\begin{aligned}|2\pi(w-v+n/2)/n \bmod 2\pi|n &= |2\pi(w-l)/n \bmod 2\pi|n\\ &\le |2\pi w/n \bmod 2\pi|n+2\pi L\end{aligned}$$

where $l=v-n/2$ and $-L\le l\le L$. Note that $(x \bmod 2\pi)$ takes value within $[-\pi,\pi)$. Furthermore

$$|2\pi w/n \bmod 2\pi|n\le 2\pi\cdot\min\{w,n-w\}\tag{23}$$

and

$$\begin{aligned}\min\{w,n-w\}&\left(\sum_{k=w+1}^{\infty}|t_k|+\sum_{k=n-w}^{\infty}|t_k|\right)\\ &\le w\sum_{k=w+1}^{\infty}|t_k|+(n-w)\sum_{k=n-w}^{\infty}|t_k|.\end{aligned}\tag{24}$$

With the assumption that $\sum_{k=-\infty}^{\infty}|kt_k|$ converges

$$w\sum_{k=w}^{\infty}|t_k|\le\sum_{k=w}^{\infty}|kt_k|<\infty.$$

Combining (22)–(24), we can see that

$$\max_{|v-n/2|\le L}|(\boldsymbol{I}_n-\boldsymbol{T}_n\boldsymbol{C}_n^{-1})_{w,v}|\le O(1/n).\tag{25}$$

Substituting (25) into (21) completes the proof. □

If the sequence $\{t_i\}$ is of finite order, the convergence rate can be strengthened.

Theorem 2: Let $\boldsymbol{T}_n$ be a family of Toeplitz Hermitian matrices associated with the sequence $\{t_i\}$ of finite order, i.e., $t_i=0$ for $|i|>W$ [6, p. 23], and let $\mathcal{F}(\lambda)$ be the DTFT of $\{t_i\}$. If $|\mathcal{F}(\lambda)|\neq 0$ for $\lambda\in[0,2\pi]$, $\boldsymbol{T}_n^{-1}$ converges to $\boldsymbol{C}_n^{-1}$ in the finite-term strong sense. Furthermore, for a vector $\boldsymbol{x}$ with the form defined in (6), the quadratic form is bounded by

$$\frac{\|(\boldsymbol{T}_n^{-1}-\boldsymbol{C}_n^{-1})\boldsymbol{x}\|}{\|\boldsymbol{x}\|}\le O(1/n).\tag{26}$$

Proof: Note that $\mathcal{PF}(w,\lambda)$ is equal to zero for $w>W$. When $W<w<n-W$, (12) shows that $(\boldsymbol{I}_n-\boldsymbol{T}_n\boldsymbol{C}_n^{-1})_{w,v}$ is equal to zero. Therefore, in (21), $\sqrt{(2L+1)n}$ can be replaced by $\sqrt{(2L+1)2W}$. The nonzero entries of $\boldsymbol{I}_n-\boldsymbol{T}_n\boldsymbol{C}_n^{-1}$ can be bounded exactly the same way as in Theorem 1. □

Example: In order to illustrate the theorems derived here, let us consider a family of finite-order Toeplitz matrices $\boldsymbol{T}_n$ by limiting the nonzero terms to t_0, t_1, t_{-1} and $t_1=t_{-1}^*$. It is easy to derive the closed-form formula of the inverse matrices. Therefore, these theorems can be examined.

Let d_n be the determinant of $\boldsymbol{T}_n$. It can be shown that the entry at the uth row and vth column of $\boldsymbol{T}_n^{-1}$ is equal to

$$(-1)^{u-v}\frac{t_1^{u-v}d_v d_{n-1-u}}{d_n},\qquad u\ge v\tag{27}$$

$$(-1)^{u-v}\frac{t_{-1}^{v-u}d_u d_{n-1-v}}{d_n},\qquad u<v.\tag{28}$$

It can be readily seen that the determinant d_n of $\boldsymbol{T}_n$ satisfies the recursive relation

$$d_n=t_0d_{n-1}-t_1t_{-1}d_{n-2}\tag{29}$$

with the initial condition $d_0=1$, $d_1=t_0$. Solving the difference equation (29), we have

$$d_n=\frac{\lambda_2^{n+1}-\lambda_1^{n+1}}{\lambda_2-\lambda_1}$$

where λ_1 and λ_2 are the solutions of the equation

$$x^2-t_0x+t_1t_{-1}=0.$$

Note that as long as $\lambda_1\neq\lambda_2$, $\boldsymbol{T}_n$ is nonsingular for any n.

We consider the case $t_0<2|t_1|$. In this case, λ_2 and λ_1 are complex conjugates with the same magnitude equal to $|t_1|$. Let $\lambda_1=|t_1|e^{j\theta}$ and $\lambda_2=|t_1|e^{-j\theta}$. We further limit $\theta/(2\pi)$ to be an irrational number. Under this condition, $e^{(n+1)\theta}\neq e^{-(n+1)\theta}$ for any $n>0$. This guarantees that $\boldsymbol{T}_n$ is nonsingular. With these notations, (27) can be rewritten as

$$\begin{aligned}(-1)^{u-v}&e^{j((u-v)\theta_t)}\frac{\sin[(v+1)\theta]\sin[(n-u)\theta]}{|t_1|\sin\theta\sin[(n+1)\theta]}\\ &=(-1)^{u-v}\frac{\cos[(n-u-v-1)\theta]-\cos[(n-u+v+1)\theta]}{2|t_1|\sin\theta\sin[(n+1)\theta]}\\ &\quad\cdot e^{j((u-v)\theta_t)}\end{aligned}\tag{30}$$

where θ_t is the phase of t_1. For fixed n and $u-v$, the denominator and the last term of the nominator are constants. In other words, (30) varies only with $\cos[(n-u-v-1)\theta]$. It can be shown that $\{\cos(n\theta); n=1,2,\ldots\}$ are dense over the interval $(-1,1)$ for irrational $\theta/(2\pi)$. This means that $\cos[(n-u-v-1)\theta]$ oscillates with u, v. Taking the diagonal entries as an example, with $u=v$, $\cos[(n-2v-1)\theta]$ oscillates with v for any fixed n. It is generally true for any fixed $u-v$. Fig. 1 illustrates this oscillation of the diagonal entries of $\boldsymbol{T}_n^{-1}$ with $t_0=1$, $t_{-1}=t_1=\sqrt{2}$, and $n=200$. The oscillation indicates that $\boldsymbol{T}_n^{-1}$ cannot converge to a Toeplitz matrix in any sense. Note that the DTFT of the sequence $\{t_i\}$ has zero within $[0,2\pi]$ if and only if $t_0<2|t_1|$, i.e., which violates the condition of the weak norm convergence theorem presented in [6].

Now we consider the case $t_0>2|t_1|$. Under this condition, λ_1 and λ_2 are real, assume $\lambda_2>\lambda_1$. Without loss of generality, let n be even. The (v,v)th ($v\in[0,n-1]$) diagonal entry of $\boldsymbol{T}_n^{-1}$ is equal to

$$\frac{d_v d_{n-1-v}}{d_n}$$

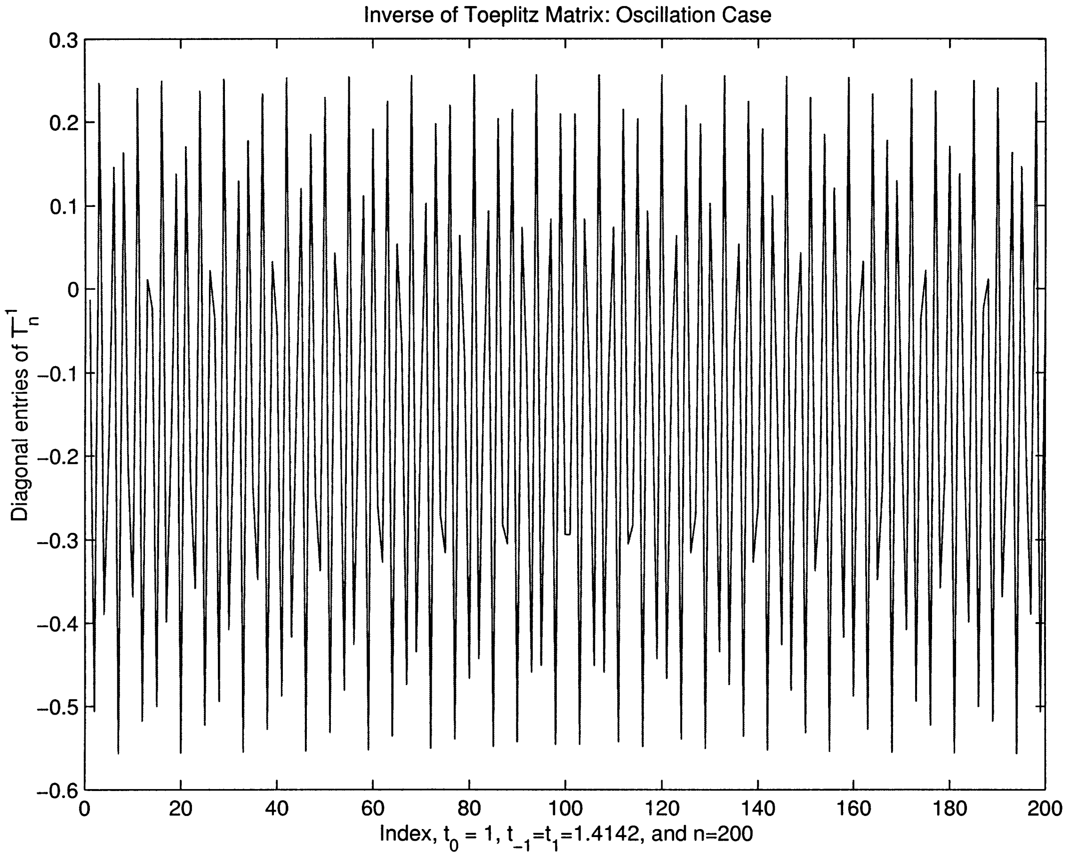

Fig. 1. Diagonal entries of $\boldsymbol{T}_n^{-1}$ in the oscillation case $(t_0 < 2|t_1|)$ with $t_0 = 1, t_{-1} = t_1 = \sqrt{2}$, and $n = 200$.

whose ratio to the $(n/2, n/2)$th entry is

$$\gamma_{v,n} \triangleq \frac{d_v d_{n-1-v}}{d_{n/2} d_{n/2-1}} = \frac{1 - \left(\frac{\lambda_1}{\lambda_2}\right)^{n-v} - \left(\frac{\lambda_1}{\lambda_2}\right)^{v+1} + \left(\frac{\lambda_1}{\lambda_2}\right)^{n+1}}{1 - \left(\frac{\lambda_1}{\lambda_2}\right)^{n/2} - \left(\frac{\lambda_1}{\lambda_2}\right)^{n/2+1} + \left(\frac{\lambda_1}{\lambda_2}\right)^{n+1}} \tag{31}$$

with the $(n/2, n/2)$th entry equal to

$$\frac{1}{\lambda_2 - \lambda_1}$$

as $n \to \infty$. Because $\lambda_2 > \lambda_1$, it is easy to see that for large n and intermediate v (i.e., v is close to $n/2$), (31) converges to 1 as $n \to \infty$. However, for the v on the boundary (i.e., v is close to 0 or $n-1$)

$$\lim_{n\to\infty} \gamma_{v,n} = \begin{cases} 1 - \left(\frac{\lambda_1}{\lambda_2}\right)^{v+1}, & v \text{ is close to } 0 \\ 1 - \left(\frac{\lambda_1}{\lambda_2}\right)^{n-v}, & v \text{ is close to } n-1. \end{cases} \tag{32}$$

This means that the middle segment of the diagonal entries of $\boldsymbol{T}_n^{-1}$ converges to $1/(\lambda_2 - \lambda_1)$, however, the entries on the boundary are always smaller than $1/(\lambda_2 - \lambda_1)$. Fig. 2 shows the diagonal entries of $\boldsymbol{T}_n^{-1}$ with $t_0 = 1, t_{-1} = t_1 = 0.35$, and $n = 200$. In other words, the entries on the boundary of the inverse matrix do not converge to the entries of a circulant matrix.

According to the property of the strong norm for a Hermitian matrix $\boldsymbol{A}$

$$\|\boldsymbol{A}\| \geq |a_{ij}|, \qquad 0 \leq i, j \leq n-1 \tag{33}$$

where a_{ij} is an arbitrary entry of $\boldsymbol{A}$, which implies that $\boldsymbol{T}_n^{-1}$ does not converge to a Toeplitz or circulant matrix in the strong sense. However, the central entries of $\boldsymbol{T}_n^{-1}$ do converge to those of a circulant matrix, which makes the finite-term quadratic form converge.

IV. Application

Substituting the inverse of a Toeplitz matrix by a circulant matrix has been widely used in the literature and yielded many useful results, e.g., [10]–[15]. The theorems derived in this paper fill the gap in these applications and avoid potential erroneous results that ignore the conditions of the convergence. The convergence rate of the finite-term quadratic form ($O(1/\sqrt{n})$ or $O(1/n)$) also provides an upper bound on the residue error, giving system designers guidance for the accuracy of the approximation.

In this section, we further illustrate the idea presented in Section II, i.e., the finite-term strong convergence allows us to derive the ML or MAP algorithms and analyze the performance for finite-length signals. As we pointed out earlier, the DFT associated with the eigendecomposition of a circulant matrix leads naturally to the frequency-domain analysis, which is also a good approximation of the Karhunen–Loeve expansion [7] that decorrelates a random process. The most basic communication

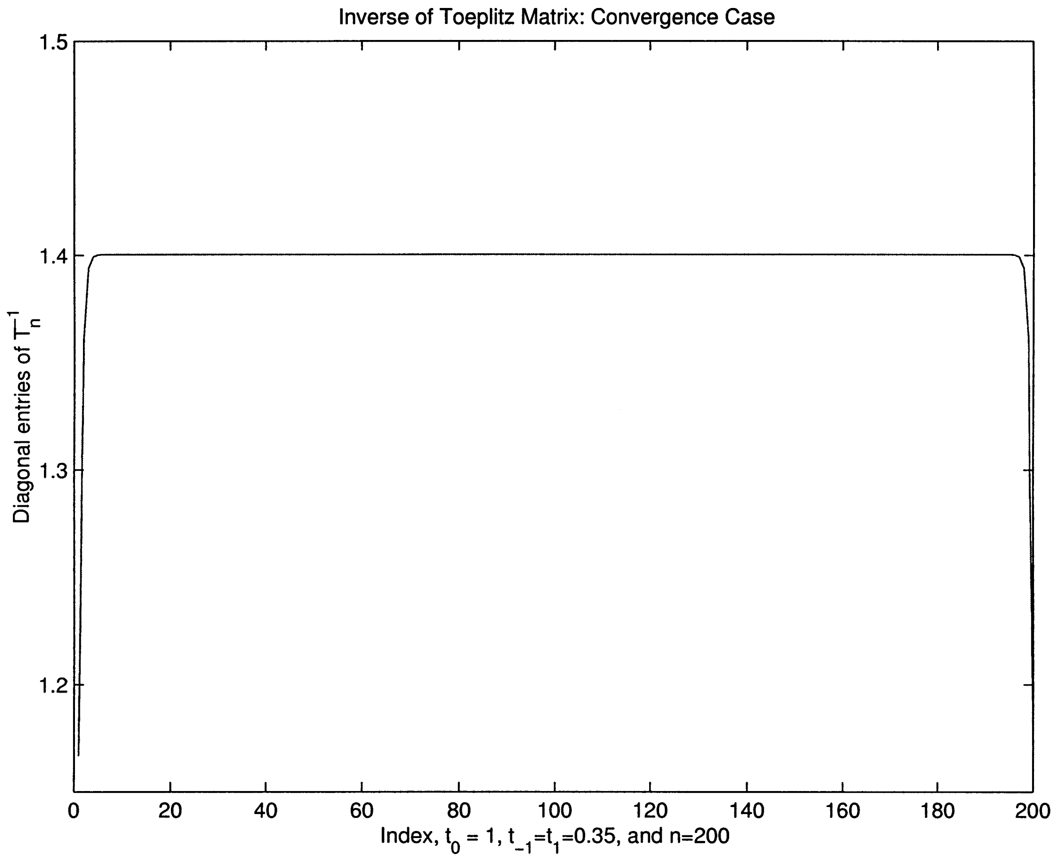

Fig. 2. Diagonal entries of $\boldsymbol{T}_n^{-1}$ in the convergence case $(t_0 > 2|t_1|)$ with $t_0 = 1, t_{-1} = t_1 = 0.35$, and $n = 200$.

channel model is the additive white Gaussian noise (AWGN) model. Even for an AWGN channel, due to the prefiltering in the receiver front end, the noise process $\boldsymbol{N}$ is no longer white when multiple samples per symbol period are used for receiver front-end design. In order to simplify the analysis, it was often simply assumed that the noise is white, or is prewhitened. However, the white noise assumption is often oversimplified and the prewhitening operation may lead to large intersymbol interference and needs the statistics of the noise process. The DFT eigendecomposition decorrelates the noise process in the frequency domain without the knowledge of the noise process, which is based on the circulant matrix approximation. This is especially useful in designing robust estimation algorithms [13].

The sequel illustrates this methodology by applying the frequency-domain approach to design the data-aided ML joint carrier phase and symbol timing offsets estimator. Using the theorems to compute the performance limits (e.g., the Cramer–Rao lower bound) for parameter estimations in colored Gaussian noise can been found in [22].

Following the signal model defined in (7), we assume that the entries of $\boldsymbol{s}(\boldsymbol{I}_u, \boldsymbol{P})$ are the digital samples of the desired (i.e., no noise) receiver matched filter output, whose kth $(k \in [-n/2, n/2-1])$ entry is equal to

$$s_k\sqrt{E_s}\sum_{m=-K/2}^{K/2-1} a_m r(kT_s - mT - \tau T)e^{j\phi} \tag{34}$$

where $r(t)$ is the pulse-shaping function, e.g., the raised-cosine function, $\{a_m\}$ is the training sequence of length K, T_s is the sampling period, T is the symbol period with $T = R_s T_s$, $\tau \in [0, T)$ and $\phi \in [-\pi, \pi)$ are the symbol timing offset and carrier phase offset, respectively. We further assume that the sampling rate is not lower than the Nyquist sampling rate. Note that R_s does not need to be an integer. The (k, l)th entry of the autocovariance matrix $\boldsymbol{T}_n$ is

$$t_{k,l} = \frac{N_0}{2}\, r((k-l)T_s)$$

i.e., the noise process $\boldsymbol{N}$ is colored due to the matched filter. Because of the correlation of the noise, the length of the observation window n should be longer than the length of the transmitted signal for optimal reception. Following the ML rule, the ML estimates of τ and ϕ are the arguments that maximize the likelihood function, i.e.,

$$\left(\hat{\tau}, \hat{\phi}\right) = \arg\max_{\tau,\phi}\left\{-\frac{1}{2}\left[-\boldsymbol{y}^H\boldsymbol{T}_n^{-1}\boldsymbol{s}(\boldsymbol{I}_u, \boldsymbol{P}) - \boldsymbol{s}(\boldsymbol{I}_u, \boldsymbol{P})^H\boldsymbol{T}_n^{-1}\boldsymbol{y} + \boldsymbol{s}(\boldsymbol{I}_u, \boldsymbol{P})^H\boldsymbol{T}_n^{-1}\boldsymbol{s}(\boldsymbol{I}_u, \boldsymbol{P})\right]\right\}. \tag{35}$$

The pulse shape $r(kT_s)$ usually decreases faster than $O(1/|k|^2)$, e.g., the raised cosine pulse has $r(kT_s) \leq O(1/|k|^3)$. Therefore, the condition of the finite-term strong convergence theory is met for the positive-definite $\boldsymbol{T}_n$. Replacing $\boldsymbol{T}_n^{-1}$ with $\boldsymbol{C}_n^{-1}$ defined in (3), we obtain the ith diagonal entry of $\boldsymbol{D}_n$ equal to

$$\mathcal{F}_r(2\pi i/K) = \frac{N_0}{2T_s}\sum_{k=-\infty}^{\infty}\mathcal{R}\left(\frac{2\pi i}{KT_s} - \frac{2\pi k}{T_s}\right) \tag{36}$$

where $\mathcal{F}_r(\lambda)$ is the DTFT of $\{r(kT_s)\}$, $\mathcal{R}(\lambda)$ is the Fourier transform of $r(t)$.

After some arithmetic, we obtain

$$\left(\hat{\tau},\, \hat{\phi}\right) = \arg\max_{\tau,\phi} \left\{ \frac{\sqrt{E_s}}{N_0 n} \Re \left(\sum_{m=-n/2}^{n/2-1} \mathcal{F}_y\left(\frac{2\pi m}{n}\right) \cdot \mathcal{A}\left(\frac{2\pi m R_s}{n}\right)^* e^{j(2\pi\tau m R_s/n - \phi)} \right) \right\} \quad (37)$$

where $\Re(\cdot)$ is the real part,

$$\mathcal{F}_y(\lambda) \triangleq \sum_{k=-n/2}^{n/2-1} y_k e^{-j\lambda k}$$

and

$$\mathcal{A}(\lambda) \triangleq \sum_{k=-K/2}^{K/2-1} a_k e^{-j\lambda k}.$$

Define $\mu(\tau)$ as

$$\mu(\tau) \triangleq \frac{1}{n} \sum_{m=-n/2}^{n/2-1} \mathcal{F}_y\left(\frac{2\pi m}{n}\right) \mathcal{A}\left(\frac{2\pi m R_s}{n}\right)^* e^{j2\pi\tau m R_s/n}. \quad (38)$$

The two-dimension maximization can be downsized to a one-dimension search

$$(\tau,\, \phi) = \arg\max_{\tau,\phi} \left\{ |\mu(\tau)| \Re\left(e^{-j(\phi - \arg(\mu(\tau)))}\right) \right\}. \quad (39)$$

Therefore, the ML estimate of τ is

$$\hat{\tau} = \arg\max_{\tau} |\mu(\tau)| \quad (40)$$

and the ML estimate of ϕ is

$$\hat{\phi} = \arg\left(\mu\left(\hat{\tau}\right)\right). \quad (41)$$

The inverse DFT of $\mathcal{F}_y(2\pi m/n) e^{j2\pi\tau m R_s/n}$ with $m \in [-n/2,\, n/2-1]$ is equal to

$$y(kT_s + \tau T), \qquad k \in [-n/2,\, n/2-1]$$

where $y(t)$ is the continuous time signal output of the matched filter, and the inverse DFT of $\mathcal{A}(2\pi m R_s/n)$, $m \in [-n/2,\, n/2-1]$ is equal to

$$\sum_{m=-K/2}^{K/2-1} a_m \frac{e^{-j\pi(k-mR_s)} - e^{j\pi(k-mR_s)}}{n(1 - e^{j2\pi(k-mR_s)/n})}, \qquad k \in [-n/2,\, n/2-1]. \quad (42)$$

In the case that R_s is an integer, (42) becomes

$$\sum_{m=-K/2}^{K/2-1} a_m \delta[k - mR_s], \qquad k \in [-n/2,\, n/2-1]$$

where $\delta[k] = 1$ if $k = 0$ and $\delta[k] = 0$ if else. Based on Parseval's relation, when R_s is an integer, $\mu(\tau)$ is equal to

$$\mu(\tau) = \sum_{m=-K/2}^{K/2-1} y(mT + \tau T) a_m^* \quad (43)$$

which provides a time-domain implementation. In fact, the time-domain estimator (43) is derived in [10, p. 297] using a different method that is only applicable to the case that the sampling rate is a multiple of the symbol rate [10]. In order to satisfy the Nyquist sampling condition, the sampling rate of the time-domain estimator has to be at least two samples per symbol period even for a signaling with 20% excessive bandwidth. The frequency-domain estimator proposed here can reduce the sampling rate to be the exact Nyquist sampling rate (i.e., 1.2 symbol rate), which reduces the sampling speed by more than 40%. For high-speed broad-band modem design, increasing the sampling rate can be extremely difficult and costly.

V. Conclusion

This paper closes a critical link in the application of the convergence theorems for the inverses of Toeplitz matrices: strengthening the well-known weak convergence theorem into the strong sense convergence for finite-term quadratic forms. We showed that this convergence is conditional. Prior literature essentially ignores the possibility of erroneous results by simply applying the weak convergence to compute quadratic forms. We further showed that the strong sense convergence theorem can naturally lead to frequency-domain solutions due to the fact that the eigendecomposition of Toeplitz matrices can be approximated by the DFT. Demonstrating the application of the convergence theorem, we derived a novel timing and carrier phase offsets estimator. This estimator takes the frequency-domain approach. For applying the estimator, we do not require the sampling rate to be an integer multiple of the symbol rate. This can significantly reduce the sampling rate requirement for high-speed modems. Due to the pivotal role of the inverses of Toeplitz matrices for a stationary random process, the strong sense convergence theorem presented in this paper can be applied to a wide range of detection and estimation problems. In a separate paper, by applying the finite-term strong convergence theorem, we derive the true Cramer–Rao lower bound for data-aided synchronization [22].

Appendix

A. Proof of Lemma 1

Proof: Following the definition of $\boldsymbol{C}_n$ in (3), the $(k,\, l)$th entry of $\boldsymbol{C}_n^{-1}$ is equal to

$$(\boldsymbol{C}_n^{-1})_{k,\,l} = \frac{1}{n} \sum_{s=0}^{n-1} \mu_{s,\,n}^{-1} e^{j2\pi(k-l)s/n} \quad (44)$$

then the $(w,\, v)$th entry of $\boldsymbol{T}_n \boldsymbol{C}_n^{-1}$ is equal to

$$\begin{aligned}(\boldsymbol{T}_n \boldsymbol{C}_n^{-1})_{w,\,v} &= \frac{1}{n} \sum_{m=0}^{n-1} t_{w-m} \sum_{s=0}^{n-1} \mu_{s,\,n}^{-1} e^{j2\pi(m-v)s/n} \\ &= \frac{1}{n} \sum_{s=0}^{n-1} \mu_{s,\,n}^{-1} e^{j2\pi(w-v)s/n} \sum_{m=0}^{n-1} t_{w-m} e^{-j2\pi(w-m)s/n}. \quad (45)\end{aligned}$$

Based on the definition of $\mu_{s,\,n}$

$$\mu_{s,\,n} = \sum_{k=-\infty}^{\infty} t_k e^{-j2\pi ks/n}$$

we have

$$\begin{aligned}\sum_{m=0}^{n-1} & t_{w-m} e^{-j2\pi(w-m)s/n} \\ &= \mu_{s,\,n} - \sum_{k=-\infty}^{w-n} t_k e^{-j2\pi ks/n} - \sum_{k=w+1}^{\infty} t_k e^{-j2\pi ks/n} \\ &= \mu_{s,\,n} - \mathcal{PF}(n-w,\, 2\pi s/n)^* - \mathcal{PF}(w+1,\, 2\pi s/n) \qquad (46)\end{aligned}$$

where the second equality follows the Hermitian assumption $t_k^* = t_{-k}$. Substituting (46) into (45), we obtain that

$$\begin{aligned}\frac{1}{n} & \sum_{s=0}^{n-1} \mu_{s,\,n}^{-1} e^{j2\pi(w-v)s/n} (\mu_{s,\,n} - \mathcal{PF}(w+1,\, 2\pi s/n) \\ & - \mathcal{PF}(n-w,\, 2\pi s/n)^*) \\ &= \delta[w-v] - \frac{1}{n}\sum_{s=0}^{n-1} \mu_{s,\,n}^{-1} e^{j2\pi(w-v)s/n} (\mathcal{PF}(w+1,\, 2\pi s/n) \\ & + \mathcal{PF}(n-w,\, 2\pi s/n)^*). \qquad (47)\end{aligned}$$

The second equality follows

$$\frac{1}{n}\sum_{s=0}^{n-1} e^{-j2\pi(w-v)s/n} = \delta[w-v] \triangleq \begin{cases} 1, & w = v \\ 0, & \text{otherwise.}\end{cases}$$

Therefore, the first term of (47) corresponds to an identify matrix for $w,\, v = 0,\, \ldots,\, n-1$. This shows that the $(w,\, v)$th entry of $\boldsymbol{I}_n - \boldsymbol{T}_n \boldsymbol{C}_n^{-1}$ can be expressed as

$$\begin{aligned}(\boldsymbol{I}_n - \boldsymbol{T}_n \boldsymbol{C}_n^{-1})_{w,\,v} &= \frac{1}{n}\sum_{s=0}^{n-1} \mu_{s,\,n}^{-1} e^{j2\pi(w-v)s/n} (\mathcal{PF}(w+1,\, 2\pi s/n) \\ & + \mathcal{PF}(n-w,\, 2\pi s/n)^*). \qquad (48)\end{aligned}$$

Actually, $\mathcal{PF}(w,\, 2\pi s/n)\mu_{s,\,n}^{-1}$ represents the ratio of the partial DTFT to the DTFT, i.e.,

$$\mathcal{PF}(w,\, 2\pi s/n)\mu_{s,\,n}^{-1} = \left.\frac{\mathcal{PF}(w,\lambda)}{\mathcal{F}(\lambda)}\right|_{\lambda=2\pi s/n}. \qquad (49)$$

For convenience, let $\mathcal{X}(w,\, \lambda)$ denote the following:

$$\mathcal{X}(w,\, \lambda) \triangleq \frac{\mathcal{PF}(w+1, \lambda)}{\mathcal{F}(\lambda)} + \frac{\mathcal{PF}(n-w, \lambda)^*}{\mathcal{F}(\lambda)} \qquad (50)$$

thus, (50) is equal to

$$\mathcal{X}(w,\, \lambda) = \mathcal{R}_f(w+1,\, \lambda) + \mathcal{R}_f(n-w,\, \lambda)^*.$$

With this notation, (48) can be written as

$$(\boldsymbol{I}_n - \boldsymbol{T}_n \boldsymbol{C}_n^{-1})_{w,\,v} = \frac{1}{n}\sum_{s=0}^{n-1} \mathcal{X}(w,\, 2\pi s/n) e^{j2\pi(w-v)s/n} \qquad (51)$$

which proves (12).

Now consider the following summation obtained by replacing s in (51) with $2\lfloor s/2 \rfloor$:

$$\frac{1}{n}\sum_{s=0}^{n-1} \mathcal{X}(w,\, (2\pi)\cdot 2\lfloor s/2\rfloor/n) e^{j2\pi(w-v)s/n}. \qquad (52)$$

The terms of (52) with even index s are equal to those of (51). It can be readily verified that the difference between (51) and (52) is equal to

$$\frac{1}{n}\sum_{s=0}^{\lfloor (n-2)/2 \rfloor} (\mathcal{X}(w,\, (2\pi)\cdot(2s+1)/n) - \mathcal{X}(w,\, (2\pi)\cdot 2s/n)) \cdot e^{j2\pi(w-v)(2s+1)/n}. \qquad (53)$$

Thus, the difference can be upper-bounded by

$$\max_{0\le s\le n-1} \tfrac{1}{2}|\mathcal{X}(w,\, (2\pi)\cdot(s+1)/n) - \mathcal{X}(w,\, (2\pi)\cdot s/n)|. \qquad (54)$$

For even and odd n, (52) is, respectively, equal to

$$\frac{1}{n}\sum_{s=0}^{n/2-1} \mathcal{X}(w,\, 4\pi s/n) e^{j2\pi(w-v)2s/n}(1 + e^{j2\pi(w-v)/n}) \qquad (55)$$

and

$$\begin{aligned}\frac{1}{n} & \sum_{s=0}^{(n-1)/2-1} \mathcal{X}(w,\, 4\pi s/n) e^{j2\pi(w-v)2s/n}(1 + e^{j2\pi(w-v)/n}) \\ & + \frac{1}{n}\mathcal{X}(w,\, (2\pi)(n-1)/n) e^{j2\pi(w-v)(n-1)/n}. \qquad (56)\end{aligned}$$

By using the following inequality:

$$|1 - e^{jx}| = |e^{-jx/2} - e^{jx/2}| = |2\sin(x/2)| \le |x \bmod 2\pi|$$

where $x \bmod 2\pi \triangleq x - 2\pi\lfloor x/2\pi + 1/2\rfloor$, which is within $[-\pi,\, \pi)$, we obtain that

$$\begin{aligned}|1 + e^{j2\pi(w-v)/n}| &= |1 - e^{j2\pi(w-v+n/2)/n}| \\ &\le |2\pi(w-v+n/2)/n \bmod 2\pi|. \qquad (57)\end{aligned}$$

Substituting (57) into (55), (55) can be upper-bounded by

$$\frac{1}{2}\max_{0\le s\le n-1} |\mathcal{X}(w,\, 2\pi s/n)| \cdot |2\pi(w-v+n/2)/n \bmod 2\pi|. \qquad (58)$$

Similarly, (56) can be upper-bounded by

$$\begin{aligned}\frac{n-1}{2n} & \max_{0\le s\le n-1} |\mathcal{X}(w,\, 2\pi s/n)| \cdot |2\pi(w-v+n/2)/n \bmod 2\pi| \\ & + \frac{1}{n}\max_{0\le s\le n-1} |\mathcal{X}(w,\, 2\pi s/n)| \\ &= \max_{0\le s\le n-1} |\mathcal{X}(w,\, 2\pi s/n)| \\ & \cdot \frac{|2\pi(w-v+n/2)/n \bmod 2\pi|(n-1)+2}{2n}. \qquad (59)\end{aligned}$$

Both (58) and (59) are smaller than

$$\max_{0\le s\le n-1} |\mathcal{X}(w,\, 2\pi s/n)| \cdot \frac{|2\pi(w-v+n/2)/n \bmod 2\pi|n+2}{2n}. \qquad (60)$$

Therefore, $(\boldsymbol{I}_n - \boldsymbol{T}_n \boldsymbol{C}_n^{-1})_{w,\,v}$ is upper-bounded by the summation of (54) and (60). □

REFERENCES

[1] S. Haykin, *Adaptive Filter Theory*, 3rd ed. Englewood Cliffs, NJ: Prentice-Hall, 1996.

[2] H. V. Poor, *An Introduction to Signal Detection and Estimation*, 2nd ed. New York: Springer-Verlag, 1994.

[3] T. Kailath, A. Vieira, and M. Morf, "Inverses of Toeplitz operators, innovations, and orthogonal polynominals," *SIAM Rev.*, vol. 20, pp. 106–119, Jan. 1978.

[4] T. Kailath, B. Levy, L. Ljung, and M. Morf, "Fast time-invariant implementations of Gaussian signal detectors," *IEEE Trans. Inform. Theory*, vol. IT-24, pp. 469–476, July 1978.

[5] U. Grenander and G. Szego, *Toeplitz Forms and Their Applications*. Berkeley, CA: Univ. Calif. Press, 1958.

[6] R. M. Gray, "Toeplitz and circulant matrices: A review," Information Systems Lab., Center for Systems Research, Stanford Univ., Stanford, CA, Tech. Rep., 2000.

[7] J. Pearl, "On coding and filtering stationary signals by discrete Fourier transforms," *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 229–232, Mar. 1973.

[8] ——, "Basis-restricted transformations and performance measures for spectral representation," *IEEE Trans. Inform. Theory*, vol. IT-17, pp. 751–752, Nov. 1971.

[9] P. J. Sherman, "Circulant approximations of the inverses of Toeplitz matrices and related quantites with applications to stationary random processes," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-33, pp. 1630–1632, Dec. 1985.

[10] H. Meyr, M. Moeneclaey, and S. Fechtel, *Digital Communication Receivers, Synchronization, Channel Estimation, and Signal Processing*. New York: Wiley, 1998.

[11] S. Kay, *Modern Spectral Estimation Theory and Application*. Englewood Cliffs, NJ: Prentice-Hall, 1988.

[12] H. Kobayashi, "Simultaneous adaptive estimation and decision algorithms for carrier modulated data transmission systems," *IEEE Trans. Commun.*, vol. COM-19, pp. 268–280, June 1971.

[13] Y. Li, L. J. Cimini, and N. Sollenberger, "Robust channel estimation for OFDM systems with rapid dispersive fading channels," *IEEE Trans. Commun.*, vol. 46, pp. 902–915, July 1998.

[14] C. Georghiades and M. Moeneclaey, "Sequence estimation and synchronization from nonsynchronized samples," *IEEE Trans. Inform. Theory*, vol. 37, pp. 1649–1657, Nov. 1991.

[15] F. Gini, M. Luise, and R. Reggiannini, "Cramer–Rao bounds in the parametric estimation of fading radiotransmission channels," *IEEE Trans. Commun.*, vol. 46, pp. 1390–1398, Oct. 1998.

[16] G. Baxter, "An asymptotic results for the finite predictor," *Math. Scand.*, vol. 10, pp. 137–144, 1962.

[17] ——, "A norm inequality for a finite section Wiener–Hopf equation," *Illinois J. Math.*, pp. 97–103, 1962.

[18] R. Chan and M. K. Ng, "Conjugate gradient methods for Toeplitz system," *SIAM Rev.*, vol. 38, pp. 427–482, 1996.

[19] M. Hanke and J. Nagy, "Inverse Toeplitz preconditioners for ill-posed problems," *Linear Alg. Its Applic.*, vol. 284, pp. 137–156, 1998.

[20] T. Ku and C. Kuo, "Preconditioned iteratiove methods for solving Toeplitz-plus-Hankel systems," *SIAM J. Numer. Anal.*, vol. 30, pp. 824–845, 1993.

[21] Y. Jiang, "Synchronization and channel parameter estimation in wireless communications," Ph.D. dissertation, Univ. Maryland, College Park, July 2000.

[22] Y. Jiang, F.-W. Sun, and J. S. Baras, "On the performance limits of the data-aided synchronization," *IEEE Trans. Inform. Theory*, vol. 49, pp. 191–203, Jan. 2003.