Toeplitz and Circulant Matrices: A review

$$\begin{bmatrix} t_0 & t_{-1} & t_{-2} & \cdots & t_{-(n-1)} \\ t_1 & t_0 & t_{-1} & & \\ t_2 & t_1 & t_0 & & \vdots \\ \vdots & & & \ddots & \\ t_{n-1} & & & \cdots & t_0 \end{bmatrix}$$

Robert M. Gray
Information Systems Laboratory
Department of Electrical Engineering
Stanford University
Stanford, California 94305

Revised March 2000
This document available as an
Adobe portable document format (pdf) file at
`http://www-isl.stanford.edu/~gray/toeplitz.pdf`

The preparation of the original report was financed in part by the National Science Foundation and by the Joint Services Program at Stanford. Since then it has been done as a hobby.

Abstract

In this tutorial report the fundamental theorems on the asymptotic behavior of eigenvalues, inverses, and products of "finite section" Toeplitz matrices and Toeplitz matrices with absolutely summable elements are derived. Mathematical elegance and generality are sacrificed for conceptual simplicity and insight in the hopes of making these results available to engineers lacking either the background or endurance to attack the mathematical literature on the subject. By limiting the generality of the matrices considered the essential ideas and results can be conveyed in a more intuitive manner without the mathematical machinery required for the most general cases. As an application the results are applied to the study of the covariance matrices and their factors of linear models of discrete time random processes.

Acknowledgements

The author gratefully acknowledges the assistance of Ronald M. Aarts of the Philips Research Labs in correcting many typos and errors in the 1993 revision, Liu Mingyu in pointing out errors corrected in the 1998 revision, Paolo Tilli of the Scuola Normale Superiore of Pisa for pointing out an incorrect corollary and providing the correction, and to David Neuhoff of the University of Michigan for pointing out several typographical errors and some confusing notation.

Contents

Chapter 1

Introduction

A toeplitz matrix is an $n \times n$ matrix $T_n = t_{k,j}$ where $t_{k,j} = t_{k-j}$, i.e., a matrix of the form

$$T_n = \begin{bmatrix} t_0 & t_{-1} & t_{-2} & \cdots & t_{-(n-1)} \\ t_1 & t_0 & t_{-1} & & \\ t_2 & t_1 & t_0 & & \vdots \\ \vdots & & & \ddots & \\ t_{n-1} & & & \cdots & t_0 \end{bmatrix}. \tag{1.1}$$

Examples of such matrices are covariance matrices of weakly stationary stochastic time series and matrix representations of linear time-invariant discrete time filters. There are numerous other applications in mathematics, physics, information theory, estimation theory, etc. A great deal is known about the behavior of such matrices — the most common and complete references being Grenander and Szegö [1] and Widom [2]. A more recent text devoted to the subject is Böttcher and Silbermann [15]. Unfortunately, however, the necessary level of mathematical sophistication for understanding reference [1] is frequently beyond that of one species of applied mathematician for whom the theory can be quite useful but is relatively little understood. This caste consists of engineers doing relatively mathematical (for an engineering background) work in any of the areas mentioned. This apparent dilemma provides the motivation for attempting a tutorial introduction on Toeplitz matrices that proves the essential theorems using the simplest possible and most intuitive mathematics. Some simple and fundamental methods that are deeply buried (at least to the untrained mathematician) in [1] are here made explicit.

In addition to the fundamental theorems, several related results that naturally follow but do not appear to be collected together anywhere are presented.

The essential prerequisites for this report are a knowledge of matrix theory, an engineer's knowledge of Fourier series and random processes, calculus (Riemann integration), and hopefully a first course in analysis. Several of the occasional results required of analysis are usually contained in one or more courses in the usual engineering curriculum, e.g., the Cauchy-Schwarz and triangle inequalities. Hopefully the only unfamiliar results are a corollary to the Courant-Fischer Theorem and the Weierstrass Approximation Theorem. The latter is an intuitive result which is easily believed even if not formally proved. More advanced results from Lebesgue integration, functional analysis, and Fourier series are not used.

The main approach of this report is to relate the properties of Toeplitz matrices to those of their simpler, more structured cousin — the circulant or cyclic matrix. These two matrices are shown to be asymptotically equivalent in a certain sense and this is shown to imply that eigenvalues, inverses, products, and determinants behave similarly. This approach provides a simplified and direct path (to the author's point of view) to the basic eigenvalue distribution and related theorems. This method is implicit but not immediately apparent in the more complicated and more general results of Grenander in Chapter 7 of [1]. The basic results for the special case of a finite order Toeplitz matrix appeared in [16], a tutorial treatment of the simplest case which was in turn based on the first draft of this work. The results were subsequently generalized using essentially the same simple methods, but they remain less general than those of [1].

As an application several of the results are applied to study certain models of discrete time random processes. Two common linear models are studied and some intuitively satisfying results on covariance matrices and their factors are given. As an example from Shannon information theory, the Toeplitz results regarding the limiting behavior of determinants is applied to find the differential entropy rate of a stationary Gaussian random process.

We sacrifices mathematical elegance and generality for conceptual simplicity in the hope that this will bring an understanding of the interesting and useful properties of Toeplitz matrices to a wider audience, specifically to those who have lacked either the background or the patience to tackle the mathematical literature on the subject.

Chapter 2

The Asymptotic Behavior of Matrices

In this chapter we begin with relevant definitions and a prerequisite theorem and proceed to a discussion of the asymptotic eigenvalue, product, and inverse behavior of sequences of matrices. The remaining chapters of this report will largely be applications of the tools and results of this chapter to the special cases of Toeplitz and circulant matrices.

The eigenvalues λ_k and the eigenvectors (n-tuples) x_k of an $n \times n$ matrix M are the solutions to the equation

$$Mx = \lambda x \tag{2.1}$$

and hence the eigenvalues are the roots of the characteristic equation of M:

$$\det(M - \lambda I) = 0 \quad . \tag{2.2}$$

If M is Hermitian, i.e., if $M = M^*$, where the asterisk denotes conjugate transpose, then a more useful description of the eigenvalues is the variational description given by the Courant-Fischer Theorem [3, p. 116]. While we will not have direct need of this theorem, we will use the following important corollary which is stated below without proof.

Corollary 2.1 *Define the Rayleigh quotient of an Hermitian matrix H and a vector (complex $n-$tuple) x by*

$$R_H(x) = (x^* H x)/(x^* x). \tag{2.3}$$

Let η_M and η_m be the maximum and minimum eigenvalues of H, respectively. Then

$$\eta_m = \min_x \; R_H(x) = \min_{x^*x=1} \; x^*Hx \tag{2.4}$$

$$\eta_M = \max_x \; R_H(x) = \max_{x^*x=1} \; x^*Hx \tag{2.5}$$

This corollary will be useful in specifying the interval containing the eigenvalues of an Hermitian matrix.

The following lemma is useful when studying non-Hermitian matrices and products of Hermitian matrices. Its proof is given since it introduces and manipulates some important concepts.

Lemma 2.1 *Let A be a matrix with eigenvalues α_k. Define the eigenvalues of the Hermitian matrix A^*A to be λ_k. Then*

$$\sum_{k=0}^{n-1} \lambda_k \geq \sum_{k=0}^{n-1} |\alpha_k|^2, \tag{2.6}$$

with equality iff (if and only if) A is normal, that is, iff $A^*A = AA^*$. (If A is Hermitian, it is also normal.)

Proof.

The trace of a matrix is the sum of the diagonal elements of a matrix. The trace is invariant to unitary operations so that it also is equal to the sum of the eigenvalues of a matrix, i.e.,

$$\mathrm{Tr}\{A^*A\} = \sum_{k=0}^{n-1} (A^*A)_{k,k} = \sum_{k=0}^{n-1} \lambda_k. \tag{2.7}$$

Any complex matrix A can be written as

$$A = WRW^*. \tag{2.8}$$

where W is unitary and $R = \{r_{k,j}\}$ is an upper triangular matrix [3, p. 79]. The eigenvalues of A are the principal diagonal elements of R. We have

$$\begin{aligned} \mathrm{Tr}\{A^*A\} &= \mathrm{Tr}\{R^*R\} = \sum_{k=0}^{n-1}\sum_{j=0}^{n-1} |r_{j,k}|^2 \\ &= \sum_{k=0}^{n-1} |\alpha_k|^2 + \sum_{k \neq j} |r_{j,k}|^2 \geq \sum_{k=0}^{n-1} |\alpha_k|^2 \end{aligned} . \tag{2.9}$$

Equation (2.9) will hold with equality iff R is diagonal and hence iff A is normal.

Lemma 2.1 is a direct consequence of Shur's Theorem [3, pp. 229-231] and is also proved in [1, p. 106].

To study the asymptotic equivalence of matrices we require a metric or equivalently a norm of the appropriate kind. Two norms — the operator or strong norm and the Hilbert-Schmidt or weak norm — will be used here [1, pp. 102-103].

Let A be a matrix with eigenvalues α_k and let λ_k be the eigenvalues of the Hermitian matrix A^*A. The strong norm $\parallel A \parallel$ is defined by

$$\parallel A \parallel = \max_x R_{A^*A}(x)^{1/2} = \max_{x^*x=1} [x^*A^*Ax]^{1/2}. \tag{2.10}$$

From Corollary 2.1

$$\parallel A \parallel^2 = \max_k \lambda_k \triangleq \lambda_M. \tag{2.11}$$

The strong norm of A can be bounded below by letting e_M be the eigenvector of A corresponding to α_M, the eigenvalue of A having largest absolute value:

$$\parallel A \parallel^2 = \max_{x^*x=1} x^*A^*Ax \geq (e_M^*A^*)(Ae_M) = |\alpha_M|^2. \tag{2.12}$$

If A is itself Hermitian, then its eigenvalues α_k are real and the eigenvalues λ_k of A^*A are simply $\lambda_k = \alpha_k^2$. This follows since if $e^{(k)}$ is an eigenvector of A with eigenvalue α_k, then $A^*Ae^{(k)} = \alpha_k A^*e^{(k)} = \alpha_k^2 e^{(k)}$. Thus, in particular, if A is Hermitian then

$$\parallel A \parallel = \max_k |\alpha_k| = |\alpha_M|. \tag{2.13}$$

The weak norm of an $n \times n$ matrix $A = \{a_{k,j}\}$ is defined by

$$|A| = \left(n^{-1} \sum_{k=0}^{n-1} \sum_{j=0}^{n-1} |a_{k,j}|^2 \right)^{1/2} = (n^{-1}\mathrm{Tr}[A^*A])^{1/2} = \left(n^{-1} \sum_{k=0}^{n-1} \lambda_k \right)^{1/2}. \tag{2.14}$$

From Lemma 2.1 we have

$$|A|^2 \geq n^{-1} \sum_{k=0}^{n-1} |\alpha_k|^2, \tag{2.15}$$

with equality iff A is normal.

The Hilbert-Schmidt norm is the "weaker" of the two norms since

$$\| A \|^2 = \max_k \lambda_k \geq n^{-1} \sum_{k=0}^{n-1} \lambda_k = |A|^2. \tag{2.16}$$

A matrix is said to be bounded if it is bounded in both norms.

Note that both the strong and the weak norms are in fact norms in the linear space of matrices, i.e., both satisfy the following three axioms:

$$\begin{aligned} &1. \quad \| A \| \geq 0 \;, \quad \text{with equality iff} \quad A = 0 \;, \quad \text{the all zero matrix.} \\ &2. \quad \| A + B \| \leq \| A \| + \| B \| \\ &3. \quad \| cA \| = |c| \cdot \| A \| \end{aligned} \tag{2.17}$$

.

The triangle inequality in (2.17) will be used often as is the following direct consequence:

$$\| A - B \| \geq | \, \| A \| - \| B \| \, | . \tag{2.18}$$

The weak norm is usually the most useful and easiest to handle of the two but the strong norm is handy in providing a bound for the product of two matrices as shown in the next lemma.

Lemma 2.2 *Given two $n \times n$ matrices $G = \{g_{k,j}\}$ and $H = \{h_{k,j}\}$, then*

$$|GH| \leq \| G \| \cdot |H|. \tag{2.19}$$

Proof.

$$\begin{aligned} |GH|^2 &= n^{-1} \sum_i \sum_j \left| \sum_k g_{i,k} h_{k,j} \right|^2 \\ &= n^{-1} \sum_i \sum_j \sum_k \sum_m g_{i,k} \bar{g}_{i,m} h_{k,j} \bar{h}_{m,j} \\ &= n^{-1} \sum_j h_j^* G^* G h_j, \end{aligned} \tag{2.20}$$

where $*$ denotes conjugate transpose and h_j is the j^{th} column of H. From (2.10)

$$(h_j^* G^* G h_j)/(h_j^* h_j) \leq \parallel G \parallel^2$$

and therefore

$$|GH|^2 \leq n^{-1} \parallel G \parallel^2 \sum_j h_j^* h_j = \parallel G \parallel^2 \cdot |H|^2.$$

Lemma 2.2 is the matrix equivalent of 7.3a of [1, p. 103]. Note that the lemma does not require that G or H be Hermitian.

We will be considering $n \times n$ matrices that approximate each other when n is large. As might be expected, we will use the weak norm of the difference of two matrices as a measure of the "distance" between them. Two sequences of $n \times n$ matrices A_n and B_n are said to be asymptotically equivalent if

1. A_n and B_n are uniformly bounded in strong (and hence in weak) norm:

$$\parallel A_n \parallel \ , \quad \parallel B_n \parallel \leq M < \infty \tag{2.21}$$

   and

2. $A_n - B_n = D_n$ goes to zero in weak norm as $n \to \infty$:

$$\lim_{n\to\infty} |A_n - B_n| = \lim_{n\to\infty} |D_n| = 0.$$

Asymptotic equivalence of A_n and B_n will be abbreviated $A_n \sim B_n$. If one of the two matrices is Toeplitz, then the other is said to be asymptotically Toeplitz. We can immediately prove several properties of asymptotic equivalence which are collected in the following theorem.

Theorem 2.1

1. *If $A_n \sim B_n$, then*

$$\lim_{n\to\infty} |A_n| = \lim_{n\to\infty} |B_n|. \tag{2.22}$$

2. *If $A_n \sim B_n$ and $B_n \sim C_n$, then $A_n \sim C_n$.*

3. *If $A_n \sim B_n$ and $C_n \sim D_n$, then $A_n C_n \sim B_n D_n$.*

4. *If $A_n \sim B_n$ and $\| A_n^{-1} \|, \| B_n^{-1} \| \leq K < \infty$, i.e., A_n^{-1} and B_n^{-1} exist and are uniformly bounded by some constant independent of n, then $A_n^{-1} \sim B_n^{-1}$.*

5. *If $A_n B_n \sim C_n$ and $\| A_n^{-1} \| \leq K < \infty$, then $B_n \sim A_n^{-1} C_n$.*

Proof.

1. Eqs. (2.22) follows directly from (2.17).

2. $|A_n - C_n| = |A_n - B_n + B_n - C_n| \leq |A_n - B_n| + |B_n - C_n| \underset{n\to\infty}{\longrightarrow} 0$

3. Applying Lemma 2.2 yields

$$|A_n C_n - B_n D_n| = |A_n C_n - A_n D_n + A_n D_n - B_n D_n|$$

$$\leq \| A_n \| \cdot |C_n - D_n| + \| D_n \| \cdot |A_n - B_n| \underset{n\to\infty}{\longrightarrow} 0.$$

4. $|A_n^{-1} - B_n^{-1}| = |B_n^{-1} B_n A_n - B_n^{-1} A_n A_n^{-1}$

$$\leq \| B_n^{-1} \| \cdot \| A_n^{-1} \| \cdot |B_n - A_n| \underset{n\to\infty}{\longrightarrow} 0.$$

5. $B_n - A_n^{-1} C_n = A_n^{-1} A_n B_n - A_n^{-1} C_n$

$$\leq \| A_n^{-1} \| \cdot |A_n B_n - C_n| \underset{n\to\infty}{\longrightarrow} 0.$$

The above results will be useful in several of the later proofs.

Asymptotic equality of matrices will be shown to imply that eigenvalues, products, and inverses behave similarly. The following lemma provides a prelude of the type of result obtainable for eigenvalues and will itself serve as the essential part of the more general theorem to follow.

Lemma 2.3 *Given two sequences of asymptotically equivalent matrices A_n and B_n with eigenvalues $\alpha_{n,k}$ and $\beta_{n,k}$, respectively, then*

$$\lim_{n\to\infty} n^{-1} \sum_{k=0}^{n-1} \alpha_{n,k} = \lim_{n\to\infty} n^{-1} \sum_{k=0}^{n-1} \beta_{n,k}. \tag{2.23}$$

Proof.

Let $D_n = \{d_{k,j}\} = A_n - B_n$. Eq. (2.23) is equivalent to

$$\lim_{n\to\infty} n^{-1}\mathrm{Tr}(D_n) = 0. \tag{2.24}$$

Applying the Cauchy-Schwartz inequality [4, p. 17] to $\mathrm{Tr}(D_n)$ yields

$$\begin{aligned} |\mathrm{Tr}(D_n)|^2 &= \left|\sum_{k=0}^{n-1} d_{k,k}\right|^2 \le n\sum_{k=0}^{n-1}|d_{k,k}|^2 \\ &\le n\sum_{k=0}^{n-1}\sum_{j=0}^{n-1}|d_{k,j}|^2 = n^2|D_n|^2. \end{aligned}$$

Dividing by n^2, and taking the limit, results in

$$0 \le |n^{-1}\mathrm{Tr}(D_n)|^2 \le |D_n|^2 \underset{n\to\infty}{\longrightarrow} 0. \tag{2.25}$$

which implies (2.24) and hence (2.23).

Similarly to (2.23), if A_n and B_n are Hermitian then (2.22) and (2.15) imply that

$$\lim_{n\to\infty} n^{-1}\sum_{k=0}^{n-1}\alpha_{n,k}^2 = \lim_{n\to\infty} n^{-1}\sum_{k=0}^{n-1}\beta_{n,k}^2. \tag{2.26}$$

Note that (2.23) and (2.26) relate limiting sample (arithmetic) averages of eigenvalues or moments of an eigenvalue distribution rather than individual eigenvalues. Equations (2.23) and (2.26) are special cases of the following fundamental theorem of asymptotic eigenvalue distribution.

Theorem 2.2 *Let A_n and B_n be asymptotically equivalent sequences of matrices with eigenvalues $\alpha_{n,k}$ and $\beta_{n,k}$, respectively. Assume that the eigenvalue moments of either matrix converge, e.g., $\lim_{n\to\infty} n^{-1}\sum_{k=0}^{n-1}\alpha_{n,k}^s$ exists and is finite for any positive integer s. Then*

$$\lim_{n\to\infty} n^{-1}\sum_{k=0}^{n-1}\alpha_{n,k}^s = \lim_{n\to\infty} n^{-1}\sum_{k=0}^{n-1}\beta_{n,k}^s. \tag{2.27}$$

Proof.

Let $A_n = B_n + D_n$ as in Lemma 2.3 and consider $A_n^s - B_n^s \triangleq \Delta_n$. Since the eigenvalues of A_n^s are $\alpha_{n,k}^s$, (2.27) can be written in terms of Δ_n as

$$\lim_{n\to\infty} n^{-1}\mathrm{Tr}\Delta_n = 0. \tag{2.28}$$

The matrix Δ_n is a sum of several terms each being a product of $\Delta_n' s$ and $B_n' s$ but containing at least one D_n. Repeated application of Lemma 2.2 thus gives

$$|\Delta_n| \le K'|D_n| \underset{n\to\infty}{\longrightarrow} 0. \tag{2.29}$$

where K' does not depend on n. Equation (2.29) allows us to apply Lemma 2.3 to the matrices A_n^s and D_n^s to obtain (2.28) and hence (2.27).

Theorem 2.2 is the fundamental theorem concerning asymptotic eigenvalue behavior. Most of the succeeding results on eigenvalues will be applications or specializations of (2.27).

Since (2.26) holds for any positive integer s we can add sums corresponding to different values of s to each side of (2.26). This observation immediately yields the following corollary.

Corollary 2.2 *Let A_n and B_n be asymptotically equivalent sequences of matrices with eigenvalues $\alpha_{n,k}$ and $\beta_{n,k}$, respectively, and let $f(x)$ be any polynomial. Then*

$$\lim_{n\to\infty} n^{-1}\sum_{k=0}^{n-1} f\left(\alpha_{n,k}\right) = \lim_{n\to\infty} n^{-1}\sum_{k=0}^{n-1} f\left(\beta_{n,k}\right). \tag{2.30}$$

Whether or not A_n and B_n are Hermitian, Corollary 2.2 implies that (2.30) can hold for any analytic function $f(x)$ since such functions can be expanded into complex Taylor series, i.e., into polynomials. If A_n and B_n are Hermitian, however, then a much stronger result is possible. In this case the eigenvalues of both matrices are real and we can invoke the Stone-Weierstrass approximation Theorem [4, p. 146] to immediately generalize Corollary 2.3. This theorem, our one real excursion into analysis, is stated below for reference.

Theorem 2.3 *(Stone-Weierstrass) If $F(x)$ is a continuous complex function on $[a,b]$, there exists a sequence of polynomials $p_n(x)$ such that*

$$\lim_{n\to\infty} p_n(x) = F(x)$$

uniformly on $[a,b]$.

Stated simply, any continuous function defined on a real interval can be approximated arbitrarily closely by a polynomial. Applying Theorem 2.3 to Corollary 2.2 immediately yields the following theorem:

Theorem 2.4 *Let* A_n *and* B_n *be asymptotically equivalent sequences of Hermitian matrices with eigenvalues* $\alpha_{n,k}$ *and* $\beta_{n,k}$*, respectively. Since* A_n *and* B_n *are bounded there exist finite numbers* m *and* M *such that*

$$m \le \alpha_{n,k}, \beta_{n,k} \le M \ , \quad n = 1, 2, \ldots \quad k = 0, 1, \ldots, n-1. \tag{2.31}$$

Let $F(x)$ *be an arbitrary function continuous on* $[m, M]$*. Then*

$$\lim_{n\to\infty} n^{-1} \sum_{k=0}^{n-1} F[\alpha_{n,k}] = \lim_{n\to\infty} n^{-1} \sum_{k=0}^{n-1} F[\beta_{n,k}] \tag{2.32}$$

if either of the limits exists.

Theorem 2.4 is the matrix equivalent of Theorem (7.4a) of [1]. When two real sequences $\{\alpha_{n,k}; k = 0, 1, \ldots, n-1\}$ and $\{\beta_{n,k}; k = 0, 1, \ldots, n-1\}$ satisfy (2.31)-(2.32), they are said to be *asymptotically equally distributed* [1, p. 62].

As an example of the use of Theorem 2.4 we prove the following corollary on the determinants of asymptotically equivalent matrices.

Corollary 2.3 *Let* A_n *and* B_n *be asymptotically equivalent Hermitian matrices with eigenvalues* $\alpha_{n,k}$ *and* $\beta_{n,k}$*, respectively, such that* $\alpha_{n,k}, \beta_{n,k} \ge m > 0$*. Then*

$$\lim_{n\to\infty} (\det A_n)^{1/n} = \lim_{n\to\infty} (\det B_n)^{1/n}. \tag{2.33}$$

Proof.

From Theorem 2.4 we have for $F(x) = \ln x$

$$\lim_{n\to\infty} n^{-1} \sum_{k=0}^{n-1} \ln \alpha_{n,k} = \lim_{n\to\infty} n^{-1} \sum_{k=0}^{n-1} \ln \beta_{n,k}$$

and hence

$$\lim_{n\to\infty} \exp\left[n^{-1} \ln \prod_{k=0}^{n-1} \alpha_{n,k}\right] = \lim_{n\to\infty} \exp\left[n^{-1} \ln \prod_{k=0}^{n-1} \beta_{n,k}\right]$$

or equivalently

$$\lim_{n\to\infty} \exp[n^{-1} \ln \det A_n] = \lim_{n\to\infty} \exp[n^{-1} \ln \det B_n],$$

from which (2.33) follows.

With suitable mathematical care the above corollary can be extended to the case where $\alpha_{n,k}, \beta_{n,k} > 0$, but there is no m satisfying the hypothesis of the corollary, i.e., where the eigenvalues can get arbitrarily small but are still strictly positive.

In the preceding chapter the concept of asymptotic equivalence of matrices was defined and its implications studied. The main consequences have been the behavior of inverses and products (Theorem 2.1) and eigenvalues (Theorems 2.2 and 2.4). These theorems do not concern individual entries in the matrices or individual eigenvalues, rather they describe an "average" behavior. Thus saying $A_n^{-1} \sim B_n^{-1}$ means that that $|A_n^{-1} - B_n^{-1}| \underset{n\to\infty}{\longrightarrow} 0$ and says nothing about convergence of individual entries in the matrix. In certain cases stronger results on a type of elementwise convergence are possible using the stronger norm of Baxter [7, 8]. Baxter's results are beyond the scope of this report.

The major use of the theorems of this chapter is that we can often study the asymptotic behavior of complicated matrices by studying a more structured and simpler asymptotically equivalent matrix.

Chapter 3

Circulant Matrices

The properties of circulant matrices are well known and easily derived [3, p. 267],[19]. Since these matrices are used both to approximate and explain the behavior of Toeplitz matrices, it is instructive to present one version of the relevant derivations here.

A circulant matrix C is one having the form

$$C = \begin{bmatrix} c_0 & c_1 & c_2 & & \cdots & c_{n-1} \\ c_{n-1} & c_0 & c_1 & c_2 & & \vdots \\ & c_{n-1} & c_0 & c_1 & \ddots & \\ \vdots & \ddots & \ddots & \ddots & & c_2 \\ & & & & & c_1 \\ c_1 & \cdots & & c_{n-1} & & c_0 \end{bmatrix}, \tag{3.1}$$

where each row is a cyclic shift of the row above it. The matrix C is itself a special type of Toeplitz matrix. The eigenvalues ψ_k and the eigenvectors $y^{(k)}$ of C are the solutions of

$$Cy = \psi\, y \tag{3.2}$$

or, equivalently, of the n difference equations

$$\sum_{k=0}^{m-1} c_{n-m+k} y_k + \sum_{k=m}^{n-1} c_{k-m} y_k = \psi\, y_m;\ m = 0, 1, \ldots, n-1. \tag{3.3}$$

Changing the summation dummy variable results in

$$\sum_{k=0}^{n-1-m} c_k y_{k+m} + \sum_{k=n-m}^{n-1} c_k y_{k-(n-m)} = \psi\, y_m;\ m = 0, 1, \ldots, n-1. \tag{3.4}$$

One can solve difference equations as one solves differential equations — by guessing an (hopefully) intuitive solution and then proving that it works. Since the equation is linear with constant coefficients a reasonable guess is $y_k = \rho^k$ (analogous to $y(t) = e^{s\tau}$ in linear time invariant differential equations). Substitution into (3.4) and cancellation of ρ^m yields

$$\sum_{k=0}^{n-1-m} c_k \rho^k + \rho^{-n} \sum_{k=n-m}^{n-1} c_k \rho^k = \psi.$$

Thus if we choose $\rho^{-n} = 1$, i.e., ρ is one of the n distinct complex n^{th} roots of unity, then we have an eigenvalue

$$\psi = \sum_{k=0}^{n-1} c_k \rho^k \tag{3.5}$$

with corresponding eigenvector

$$y = n^{-1/2} \left(1, \rho, \rho^2, \ldots, \rho^{n-1}\right), \tag{3.6}$$

where the normalization is chosen to give the eigenvector unit energy. Choosing ρ_j as the complex n^{th} root of unity, $\rho_j = e^{-2\pi ij/n}$, we have eigenvalue

$$\psi_m = \sum_{k=0}^{n-1} c_k e^{-2\pi imk/n} \tag{3.7}$$

and eigenvector

$$y^{(m)} = n^{-1/2} \left(1, e^{-2\pi im/n}, \cdots, e^{-2\pi i(n-1)/n}\right).$$

From (3.7) we can write

$$C = U^* \Psi U, \tag{3.8}$$

where

$$\begin{aligned} U &= \left\{ y^{(0)} | y^{(1)} | \cdots | y^{(n-1)} \right\} \\ &= n^{-1} \left\{ e^{-2\pi imk/n} \; ; \; m, k = 0, 1, \ldots, n-1 \right\} \end{aligned}$$

$$\Psi = \{\psi_k \delta_{k,j}\}$$

To verify (3.8) we note that the $(k,j)^{th}$ element of C, say $a_{k,j}$, is

$$\begin{aligned} a_{k,j} &= n^{-1}\sum_{m=0}^{n-1} e^{2\pi im(k-j)/n}\psi_m \\ &= n^{-1}\sum_{m=0}^{n-1} e^{2\pi im(k-j)/n}\sum_{r=0}^{n-1} c_r e^{2\pi imr/n} \\ &= n^{-1}\sum_{r=0}^{n-1} c_r \sum_{m=0}^{n-1} e^{2\pi im(k-j+r)/n}. \end{aligned} \tag{3.9}$$

But we have

$$\sum_{m=0}^{n-1} e^{2\pi im(k-j+r)/n} = \begin{cases} n & k-j = -r \bmod n \\ 0 & \text{otherwise} \end{cases}$$

so that $a_{k,j} = c_{-(k-j) \bmod n}$. Thus (3.8) and (3.1) are equivalent. Furthermore (3.9) shows that any matrix expressible in the form (3.8) is circulant.

Since C is unitarily similar to a diagonal matrix it is normal. Note that all circulant matrices have the same set of eigenvectors. This leads to the following properties.

Theorem 3.1 *Let $C = \{c_{k-j}\}$ and $B = \{b_{k-j}\}$ be circulant $n \times n$ matrices with eigenvalues*

$$\psi_m = \sum_{k=0}^{n-1} c_k e^{-2\pi imk/n}$$

$$\beta_m = \sum_{k=0}^{n-1} b_k e^{-2\pi imk/n} \quad ,$$

respectively.

1. *C and B commute and*

$$CB = BC = U^*\gamma U \quad ,$$

where $\gamma = \{\psi_m\beta_m\delta_{k,m}\}$, and CB is also a circulant matrix.

2. *$C + B$ is a circulant matrix and*

$$C + B = U^*\Omega U,$$

where $\Omega = \{(\psi_m + \beta_m)\delta_{k,m}\}$

3. *If* $\psi_m \neq 0;\ m = 0, 1, \ldots, n-1,$ *then C is nonsingular and*

$$C^{-1} = U^*\Psi^{-1}U$$

so that the inverse of C can be straightforwardly constructed.

Proof.

We have $C = U^*\Psi U$ and $B = U^*\Phi U$ where Ψ and Φ are diagonal matrices with elements $\psi_m\delta_{k,m}$ and $\beta_m\phi_{k,m}$, respectively.

1. $CB = U^*\Psi UU^*\Phi U$
   $= U^*\Psi\Phi U$
   $= U^*\Phi\Psi U = BC$
   Since $\Psi\Phi$ is diagonal, (3.9) implies that CB is circulant.

2. $C + B = U^*(\Psi + \Phi)U$

3. $C^{-1} = (U^*\Psi U)^{-1}$
   $= U^*\Psi^{-1}U$
   if Ψ is nonsingular.

Circulant matrices are an especially tractable class of matrices since inverses, products, and sums are also circulants and hence both straightforward to construct and normal. In addition the eigenvalues of such matrices can easily be found exactly.

In the next chapter we shall see that certain circulant matrices asymptotically approximate Toeplitz matrices and hence from Chapter 2 results similar to those in Theorem 3 will hold asymptotically for Toeplitz matrices.

Chapter 4

Toeplitz Matrices

In this chapter the asymptotic behavior of inverses, products, eigenvalues, and determinants of finite Toeplitz matrices is derived by constructing an asymptotically equivalent circulant matrix and applying the results of the previous chapters. Consider the infinite sequence $\{t_k; k = 0, \pm 1, \pm 2, \cdots\}$ and define the finite $(n \times n)$ Toeplitz matrix $T_n = \{t_{k-j}\}$ as in (1.1). Toeplitz matrices can be classified by the restrictions placed on the sequence t_k. If there exists a finite m such that $t_k = 0$, $|k| > m$, then T_n is said to be a finite order Toeplitz matrix. If t_k is an infinite sequence, then there are two common constraints. The most general is to assume that the t_k are square summable, i.e., that

$$\sum_{k=-\infty}^{\infty} |t_k|^2 < \infty \quad . \tag{4.1}$$

Unfortunately this case requires mathematical machinery beyond that assumed in this paper; i.e., Lebesgue integration and a relatively advanced knowledge of Fourier series. We will make the stronger assumption that the t_k are absolutely summable, i.e.,

$$\sum_{k=-\infty}^{\infty} |t_k| < \infty. \tag{4.2}$$

This assumption greatly simplifies the mathematics but does not alter the fundamental concepts involved. As the main purpose here is tutorial and we wish chiefly to relay the flavor and an intuitive feel for the results, this paper will be confined to the absolutely summable case. The main advantage of (4.2) over (4.1) is that it ensures the existence and continuity of the Fourier series $f(\lambda)$ defined by

$$f(\lambda) = \sum_{k=-\infty}^{\infty} t_k e^{ik\lambda} = \lim_{n\to\infty} \sum_{k=-n}^{n} t_k e^{ik\lambda} \quad . \tag{4.3}$$

Not only does the limit in (4.3) converge if (4.2) holds, it converges *uniformly* for all λ, that is, we have that

$$\begin{aligned}
\left| f(\lambda) - \sum_{k=-n}^{n} t_k e^{ik\lambda} \right| &= \left| \sum_{k=-\infty}^{-n-1} t_k e^{ik\lambda} + \sum_{k=n+1}^{\infty} t_k e^{ik\lambda} \right| \\
&\le \left| \sum_{k=-\infty}^{-n-1} t_k e^{ik\lambda} \right| + \left| \sum_{k=n+1}^{\infty} t_k e^{ik\lambda} \right| , \\
&\le \sum_{k=-\infty}^{-n-1} |t_k| + \sum_{k=n+1}^{\infty} |t_k|
\end{aligned}$$

where the righthand side does not depend on λ and it goes to zero as $n \to \infty$ from (4.2), thus given ϵ there is a single N, not depending on λ, such that

$$\left| f(\lambda) - \sum_{k=-n}^{n} t_k e^{ik\lambda} \right| \le \epsilon \ , \ \text{ all } \ \lambda \in [0, 2\pi] \ , \ \text{ if } \ n \ge N. \tag{4.4}$$

Note that (4.2) is indeed a stronger constraint than (4.1) since

$$\sum_{k=-\infty}^{\infty} |t_k|^2 \le \left\{ \sum_{k=-\infty}^{\infty} |t_k| \right\}^2 .$$

Note also that (4.2) implies that $f(\lambda)$ is bounded since

$$\begin{aligned}
|f(\lambda)| &\le \sum_{k=-\infty}^{\infty} |t_k e^{ik\lambda}| \\
&\le \sum_{k=-\infty}^{\infty} |t_k| \triangleq M_{|f|} < \infty \quad .
\end{aligned}$$

The matrix T_n will be Hermitian if and only if f is real, in which case we denote the least upper bound and greatest lower bound of $f(\lambda)$ by M_f and m_f, respectively. Observe that $\max(|m_f|, |M_f|) \le M_{|f|}$.

Since $f(\lambda)$ is the Fourier series of the sequence t_k, we could alternatively begin with a bounded and hence Riemann integrable function $f(\lambda)$ on $[0, 2\pi]$ ($|f(\lambda)| \leq M_{|f|} < \infty$ for all λ) and define the sequence of $n \times n$ Toeplitz matrices

$$T_n(f) = \left\{ (2\pi)^{-1} \int_0^{2\pi} f(\lambda) e^{-i(k-j)} d\lambda \ ; \quad k, j = 0, 1, \cdots, n-1 \right\} \ . \qquad (4.5)$$

As before, the Toeplitz matrices will be Hermitian iff f is real. The assumption that $f(\lambda)$ is Riemann integrable implies that $f(\lambda)$ is continuous except possibly at a countable number of points. Which assumption is made depends on whether one begins with a sequence t_k or a function $f(\lambda)$ — either assumption will be equivalent for our purposes since it is the Riemann integrability of $f(\lambda)$ that simplifies the bookkeeping in either case. Before finding a simple asymptotic equivalent matrix to T_n, we use Corollary 2.1 to find a bound on the eigenvalues of T_n when it is Hermitian and an upper bound to the strong norm in the general case.

Lemma 4.1 *Let $\tau_{n,k}$ be the eigenvalues of a Toeplitz matrix $T_n(f)$. If $T_n(f)$ is Hermitian, then*

$$m_f \leq \tau_{n,k} \leq M_f. \qquad (4.6)$$

Whether or not $T_n(f)$ is Hermitian,

$$\| T_n(f) \| \leq 2M_{|f|} \qquad (4.7)$$

so that the matrix is uniformly bounded over n if f is bounded.

Proof.

Property (4.6) follows from Corollary 2.1:

$$\max_k \ \tau_{n,k} = \max_x (x^* T_n x)/(x^* x) \qquad (4.8)$$

$$\min_k \ \tau_{n,k} = \min_x (x^* T_n x)/(x^* x)$$

so that

$$\begin{aligned} x^*T_nx &= \sum_{k=0}^{n-1}\sum_{j=0}^{n-1} t_{k-j}x_k\bar{x}_j \\ &= \sum_{k=0}^{n-1}\sum_{j=0}^{n-1}\left[(2\pi)^{-1}\int_0^{2\pi} f(\lambda)e^{i(k-j)\lambda}\,d\lambda\right]x_k\bar{x}_j \qquad (4.9) \\ &= (2\pi)^{-1}\int_0^{2\pi}\left|\sum_{k=0}^{n-1}x_ke^{ik\lambda}\right|^2 f(\lambda)\,d\lambda \end{aligned}$$

and likewise

$$x^*x = \sum_{k=0}^{n-1}|x_k|^2 = (2\pi)^{-1}\int_0^{2\pi} d\lambda|\sum_{k=0}^{n-1}x_ke^{ik\lambda}|^2. \tag{4.10}$$

Combining (4.9)-(4.10) results in

$$m_f \le \frac{\displaystyle\int_0^{2\pi} d\lambda f(\lambda)\left|\sum_{k=0}^{n-1}x_ke^{ik\lambda}\right|^2}{\displaystyle\int_0^{2\pi} d\lambda\left|\sum_{k=0}^{n-1}x_ke^{ik\lambda}\right|^2} = \frac{x^*T_nx}{x^*x} \le M_f, \tag{4.11}$$

which with (4.8) yields (4.6). Alternatively, observe in (4.11) that if $e^{(k)}$ is the eigenvector associated with $\tau_{n,k}$, then the quadratic form with $x = e^{(k)}$ yields $x^*T_nx = \tau_{n,k}\sum_{k=0}^{n-1}|x_k|^2$. Thus (4.11) implies (4.6) directly.

We have already seen in (2.13) that if $T_n(f)$ is Hermitian, then $\| T_n(f) \| = \max_k |\tau_{n,k}| \triangleq |\tau_{n,M}|$, which we have just shown satisfies $|\tau_{n,M}| \le \max(|M_f|, |m_f|)$ which in turn must be less than $M_{|f|}$, which proves (4.7) for Hermitian matrices.. Suppose that $T_n(f)$ is not Hermitian or, equivalently, that f is not real. Any function f can be written in terms of its real and imaginary parts, $f = f_r + if_i$, where both f_r and f_i are real. In particular, $f_r = (f + f^*)/2$ and $f_i = (f - f^*)/2i$. Since the strong norm is a norm,

$$\begin{aligned} \| T_n(f) \| &= \| T_n(f_r + if_i) \| \\ &= \| T_n(f_r) + iT_n(f_i) \| \\ &\le \| T_n(f_r) \| + \| T_n(f_i) \| \\ &\le M_{|f_r|} + M_{|f_i|}. \end{aligned}$$

Since $|(f \pm f^*)/2 \leq (|f| + |f^*|)/2 \leq M_{|f|}$, $M_{|f_r|} + M_{|f_i|} \leq 2M_{|f|}$, proving (4.7).

Note for later use that the weak norm between Toeplitz matrices has a simpler form than (2.14). Let $T_n = \{t_{k-j}\}$ and $T_n' = \{t'_{k-j}\}$ be Toeplitz, then by collecting equal terms we have

$$\begin{aligned}
|T_n - T_n'|^2 &= n^{-1}\sum_{k=0}^{n-1}\sum_{j=0}^{n-1}|t_{k-j} - t'_{k-j}|^2 \\
&= n^{-1}\sum_{k=-(n-1)}^{n-1}(n - |k|)|t_k - t'_k|^2 \\
&= \sum_{k=-(n-1)}^{n-1}(1 - |k|/n)|t_k - t'_k|^2
\end{aligned} \quad . \tag{4.12}$$

We are now ready to put all the pieces together to study the asymptotic behavior of T_n. If we can find an asymptotically equivalent circulant matrix then all of the results of Chapters 2 and 3 can be instantly applied. The main difference between the derivations for the finite and infinite order case is the circulant matrix chosen. Hence to gain some feel for the matrix chosen we first consider the simpler finite order case where the answer is obvious, and then generalize in a natural way to the infinite order case.

4.1 Finite Order Toeplitz Matrices

Let T_n be a sequence of finite order Toeplitz matrices of order $m+1$, that is, $t_i = 0$ unless $|i| \leq m$. Since we are interested in the behavior or T_n for large n we choose $n >> m$. A typical Toeplitz matrix will then have the appearance of the following matrix, possessing a band of nonzero entries down the central diagonal and zeros everywhere else. With the exception of the upper left and lower right hand corners that T_n looks like a circulant matrix, i.e. each row is the row above shifted to the right one place.

$$
T = \left[\begin{array}{ccccccccccc}
t_0 & t_{-1} & \cdots & t_{-m} & & & & & & & \\
t_1 & t_0 & & & & & & & & & \\
\vdots & & & & & & & 0 & & & \\
 & & \ddots & & & & \ddots & & & & \\
t_m & & & & & & & & & & \\
 & \ddots & & & & & & & & & \\
 & & t_m & \cdots & t_1 & t_0 & t_{-1} & \cdots & t_{-m} & & \\
 & & & & & & & \ddots & & & t_{-m} \\
 & & & & & & & & & \vdots & \\
 & 0 & & & & & & & & t_0 & t_{-1} \\
 & & & & & & & t_m & \cdots & t_1 & t_0
\end{array} \right]. \tag{4.13}
$$

We can make this matrix exactly into a circulant if we fill in the upper right and lower left corners with the appropriate entries. Define the circulant matrix C in just this way, i.e.

$$
T = \left[\begin{array}{ccccccccccccc}
t_0 & t_{-1} & \cdots & t_{-m} & & & & & & t_m & \cdots & & t_1 \\
t_1 & & & & & & & & & & \ddots & & \vdots \\
 & & & & & & & & & & & & t_m \\
\vdots & & & & & \ddots & & & & & & & \\
t_m & & & & & & & & & 0 & & & \\
 & \ddots & & & & & & & & & & & \\
 & & & t_m & t_1 & t_0 & t_{-1} & \cdots & & t_{-m} & & & \\
 & & & & & & & & & & \ddots & & \\
 & & & 0 & & & & & & & & & t_{-m} \\
t_{-m} & & & & & & & & & & & & \\
\vdots & \ddots & & & & & & & & & & & \vdots \\
 & & & & & & & & & & t_0 & & t_{-1} \\
t_{-1} & \cdots & & t_{-m} & & & & & t_m & & \cdots & t_1 & t_0
\end{array} \right]
$$

$$= \begin{bmatrix} c_0^{(n)} & \cdots & & c_{n-1}^{(n)} \\ c_{n-1}^{(n)} & c_0^{(n)} & \cdots & \\ & & \ddots & \\ c_1^{(n)} & & c_{n-1}^{(n)} & c_0^{(n)} \end{bmatrix}. \tag{4.14}$$

Equivalently, C, consists of cyclic shifts of $(c_0^{(n)}, \cdots, c_{n-1}^{(n)})$ where

$$c_k^{(n)} = \begin{cases} t_{-k} & k = 0, 1, \cdots, m \\ t_{(n-k)} & k = n-m, \cdots, n-1 \\ 0 & \text{otherwise} \end{cases} \tag{4.15}$$

Given $T_n(f)$, the circulant matrix defined as in (4.14)-(4.15) is denoted $C_n(f)$.

The matrix $C_n(f)$ is intuitively a candidate for a simple matrix asymptotically equivalent to $T_n(f)$ — we need only prove that it is indeed both asymptotically equivalent and simple.

Lemma 4.2 *The matrices T_n and C_n defined in (4.13) and (4.14) are asymptotically equivalent, i.e., both are bounded in the strong norm and.*

$$\lim_{n\to\infty} |T_n - C_n| = 0. \tag{4.16}$$

Proof. The t_k are obviously absolutely summable, so T_n are uniformly bounded by $2M_{|f|}$ from Lemma 4.1. The matrices C_n are also uniformly bounded since $C_n^* C_n$ is a circulant matrix with eigenvalues $|f(2\pi k/n)|^2 \le 4M_{|f|}^2$. The weak norm of the difference is

$$\begin{aligned} |T_n - C_n|^2 &= n^{-1} \sum_{k=0}^{m} k(|t_k|^2 + |t_{-k}|^2) \\ &\le mn^{-1} \sum_{k=0}^{m} (|t_k|^2 + |t_{-k}|^2) \underset{n\to\infty}{\longrightarrow} 0 \end{aligned}.$$

The above Lemma is almost trivial since the matrix $T_n - C_n$ has fewer than m^2 non-zero entries and hence the n^{-1} in the weak norm drives $|T_n - C_n|$ to zero.

From Lemma 4.2 and Theorem 2.2 we have the following lemma:

Lemma 4.3 *Let T_n and C_n be as in (4.13) and (4.14) and let their eigenvalues be $\tau_{n,k}$ and $\psi_{n,k}$, respectively, then for any positive integer s*

$$\lim_{n\to\infty} n^{-1}\sum_{k=0}^{n-1}\tau_{n,k}^s = \lim_{n\to\infty} n^{-1}\sum_{k=0}^{n-1}\psi_{n,k}^s. \tag{4.17}$$

In fact, for finite n,

$$\left| n^{-1}\sum_{k=0}^{n-1}\tau_{n,k}^s - n^{-1}\sum_{k=0}^{n-1}\psi_{n,k}^s \right| \le K n^{-1/2}, \tag{4.18}$$

where K is not a function of n.

Proof.

Equation (4.17) is direct from Lemma 4.2 and Theorem 2.2. Equation (4.18) follows from Lemma 2.3 and Lemma 4.2.

Lemma 4.3 is of interest in that for finite order Toeplitz matrices one can find the rate of convergence of the eigenvalue moments. It can be shown that $k \le sM_f^{s-1}$.

The above two lemmas show that we can immediately apply the results of Section II to T_n and C_n. Although Theorem 2.1 gives us immediate hope of fruitfully studying inverses and products of Toeplitz matrices, it is not yet clear that we have simplified the study of the eigenvalues. The next lemma clarifies that we have indeed found a useful approximation.

Lemma 4.4 *Let $C_n(f)$ be constructed from $T_n(f)$ as in (4.14) and let $\psi_{n,k}$ be the eigenvalues of $C_n(f)$, then for any positive integer s we have*

$$\lim_{n\to\infty} n^{-1}\sum_{k=0}^{n-1}\psi_{n,k}^s = (2\pi)^{-1}\int_0^{2\pi} f^s(\lambda)\,d\lambda. \tag{4.19}$$

If $T_n(f)$ and hence $C_n(f)$ are Hermitian, then for any function $F(x)$ continuous on $[m_f, M_f]$ we have

$$\lim_{n\to\infty} n^{-1}\sum_{k=0}^{n-1} F(\psi_{n,k}) = (2\pi)^{-1}\int_0^{2\pi} F[f(\lambda)]\,d\lambda. \tag{4.20}$$

Proof.

From Chapter 3 we have exactly

$$\begin{aligned}
\psi_{n,j} &= \sum_{k=0}^{n-1} c_k^{(n)} e^{-2\pi ijk/n} \\
&= \sum_{k=0}^{m} t_{-k} e^{-2\pi ijk/n} + \sum_{k=n-m}^{n-1} t_{n-k} e^{-2\pi ijk/n} \ . \\
&= \sum_{k=-m}^{m} t_k e^{-2\pi ijk/n} = f(2\pi j n^{-1})
\end{aligned} \tag{4.21}$$

Note that the eigenvalues of C_n are simply the values of $f(\lambda)$ with λ uniformly spaced between 0 and 2π. Defining $2\pi k/n = \lambda_k$ and $2\pi/n = \Delta\lambda$ we have

$$\begin{aligned}
\lim_{n\to\infty} n^{-1}\sum_{k=0}^{n-1}\psi_{n,k}^s &= \lim_{n\to\infty} n^{-1}\sum_{k=0}^{n-1} f(2\pi k/n)^s \\
&= \lim_{n\to\infty}\sum_{k=0}^{n-1} f(\lambda_k)^s \Delta\lambda/(2\pi) \\
&= (2\pi)^{-1}\int_0^{2\pi} f(\lambda)^s d\lambda,
\end{aligned} \tag{4.22}$$

where the continuity of $f(\lambda)$ guarantees the existence of the limit of (4.22) as a Riemann integral. If T_n and C_n are Hermitian than the $\psi_{n,k}$ and $f(\lambda)$ are real and application of the Stone-Weierstrass theorem to (4.22) yields (4.20). Lemma 4.2 and (4.21) ensure that $\psi_{n,k}$ and $\tau_{n,k}$ are in the real interval $[m_f, M_f]$.

Combining Lemmas 4.2-4.4 and Theorem 2.2 we have the following special case of the fundamental eigenvalue distribution theorem.

Theorem 4.1 *If $T_n(f)$ is a finite order Toeplitz matrix with eigenvalues $\tau_{n,k}$, then for any positive integer s*

$$\lim_{n\to\infty} n^{-1} \sum_{k=0}^{n-1} \tau_{n,k}^s = (2\pi)^{-1} \int_0^{2\pi} f(\lambda)^s \, d\lambda. \tag{4.23}$$

Furthermore, if $T_n(f)$ is Hermitian, then for any function $F(x)$ continuous on $[m_f, M_f]$

$$\lim_{n\to\infty} n^{-1} \sum_{k=0}^{n-1} F(\tau_{n,k}) = (2\pi)^{-1} \int_0^{2\pi} F[f(\lambda)] \, d\lambda; \tag{4.24}$$

i.e., the sequences $\tau_{n,k}$ and $f(2\pi k/n)$ are asymptotically equally distributed.

This behavior should seem reasonable since the equations $T_n x = \tau x$ and $C_n x = \psi x$, $n > 2m+1$, are essentially the same n^{th} order difference equation with different boundary conditions. It is in fact the "nice" boundary conditions that make ψ easy to find exactly while exact solutions for τ are usually intractable.

With the eigenvalue problem in hand we could next write down theorems on inverses and products of Toeplitz matrices using Lemma 4.2 and the results of Chapters 2-3. Since these theorems are identical in statement and proof with the infinite order absolutely summable Toeplitz case, we defer these theorems momentarily and generalize Theorem 4.1 to more general Toeplitz matrices with no assumption of fine order.

4.2 Toeplitz Matrices

Obviously the choice of an appropriate circulant matrix to approximate a Toeplitz matrix is not unique, so we are free to choose a construction with the most desirable properties. It will, in fact, prove useful to consider two slightly different circulant approximations to a given Toeplitz matrix. Say we have an absolutely summable sequence $\{t_k;\ k = 0, \pm 1, \pm 2, \cdots\}$ with

$$\begin{aligned} f(\lambda) &= \sum_{k=-\infty}^{\infty} t_k e^{ik\lambda} \\ t_k &= (2\pi)^{-1} \int_0^{2\pi} f(\lambda) e^{-ik\lambda} \end{aligned} . \tag{4.25}$$

Define $C_n(f)$ to be the circulant matrix with top row $(c_0^{(n)}, c_1^{(n)}, \cdots, c_{n-1}^{(n)})$ where

$$c_k^{(n)} = n^{-1} \sum_{j=0}^{n-1} f(2\pi j/n) e^{2\pi ijk/n} \quad . \tag{4.26}$$

Since $f(\lambda)$ is Riemann integrable, we have that for fixed k

$$\begin{aligned} \lim_{n\to\infty} c_k^{(n)} &= \lim_{n\to\infty} n^{-1} \sum_{j=0}^{n-1} f(2\pi j/n) e^{2\pi ijk/n} \\ &= (2\pi)^{-1} \int_0^{2\pi} f(\lambda) e^{ik\lambda} d\lambda = t_{-k} \end{aligned} \tag{4.27}$$

and hence the $c_k^{(n)}$ are simply the sum approximations to the Riemann integral giving t_{-k}. Equations (4.26), (3.7), and (3.9) show that the eigenvalues $\psi_{n,m}$ of C_n are simply $f(2\pi m/n)$; that is, from (3.7) and (3.9)

$$\begin{aligned} \psi_{n,m} &= \sum_{k=0}^{n-1} c_k^{(n)} e^{-2\pi imk/n} \\ &= \sum_{k=0}^{n-1} \left(n^{-1} \sum_{j=0}^{n-1} f(2\pi j/n) e^{2\pi ijk/n} \right) e^{-2\pi imk/n} \\ &= \sum_{j=0}^{n-1} f(2\pi j/n) \left\{ n^{-1} \sum_{k=0}^{n-1} 2^{2\pi ik(j-m)/n} \right\} \\ &= f(2\pi m/n) \end{aligned} \quad . \tag{4.28}$$

Thus, $C_n(f)$ has the useful property (4.21) of the circulant approximation (4.15) used in the finite case. As a result, the conclusions of Lemma 4.4 hold for the more general case with $C_n(f)$ constructed as in (4.26). Equation (4.28) in turn defines $C_n(f)$ since, if we are told that C_n is a circulant matrix with eigenvalues $f(2\pi m/n), m = 0, 1, \cdots, n-1$, then from (3.9)

$$\begin{aligned} c_k^{(n)} &= n^{-1} \sum_{m=0}^{n-1} \psi_{n,m} e^{2\pi imk/n} \\ &= n^{-1} \sum_{m=0}^{n-1} f(2\pi m/n) e^{2\pi imk/n} \end{aligned} \quad , \tag{4.29}$$

as in (4.26). Thus, either (4.26) or (4.28) can be used to define $C_n(f)$.

The fact that Lemma 4.4 holds for $C_n(f)$ yields several useful properties as summarized by the following lemma.

Lemma 4.5

1. *Given a function f of (4.25) and the circulant matrix $C_n(f)$ defined by (4.26), then*

$$c_k^{(n)} = \sum_{m=-\infty}^{\infty} t_{-k+mn} \ , \quad k = 0, 1, \cdots, n-1. \tag{4.30}$$

(Note, the sum exists since the t_k are absolutely summable.)

2. *Given $T_n(f)$ where $f(\lambda)$ is real and $f(\lambda) \geq m_f > 0$, then*

$$C_n(f)^{-1} = C_n(1/f).$$

3. *Given two functions $f(\lambda)$ and $g(\lambda)$, then*

$$C_n(f)C_n(g) = C_n(fg).$$

Proof.

1. Since $e^{-2\pi imk/n}$ is periodic with period n, we have that

$$\begin{aligned} f(2\pi j/n) &= \sum_{m=-\infty}^{\infty} t_m e^{i2\pi jm/n} \sum_{m=-\infty}^{\infty} t_{-m} e^{-i2\pi jm/n} \\ &= \sum_{l=0}^{n-1} \sum_{m=-\infty}^{\infty} t_{-1+mn} e^{-2\pi ijl/n} \end{aligned}$$

and hence from (4.26) and (3.9)

$$\begin{aligned}
c_k^{(n)} &= n^{-1}\sum_{j=0}^{n-1} f(2\pi j/n)e^{2\pi ijk/n} \\
&= n^{-1}\sum_{j=0}^{n-1}\sum_{l=0}^{n-1}\sum_{m=-\infty}^{\infty} t_{-1+mn}e^{2\pi ij(k-l)/n} \\
&= \sum_{l=0}^{n-1}\sum_{m=-\infty}^{\infty} t_{-1+mn}\left\{n^{-1}\sum_{j=0}^{n-1} e^{2\pi ij(k-l)/n}\right\} \\
&= \sum_{m=-\infty}^{\infty} t_{-k+mn}
\end{aligned}$$

2. Since $C_n(f)$ has eigenvalues $f(2\pi k/n) > 0$, by Theorem 3.1, $C_n(f)^{-1}$ has eigenvalues $1/f(2\pi k/n)$, and hence from (4.29) and the fact that $C_n(f)^{-1}$ is circulant we have $C_n(f)^{-1} = C_n(1/f)$.

3. Follows immediately from Theorem 3.1 and the fact that, if $f(\lambda)$ and $g(\lambda)$ are Riemann integrable, so is $f(\lambda)g(\lambda)$.

Equation (4.30) points out a shortcoming of $C_n(f)$ for applications as a circulant approximation to $T_n(f)$ — it depends on the entire sequence $\{t_k; k = 0, \pm 1, \pm 2, \cdots\}$ and not just on the finite collection of elements $\{t_k; k = 0, \pm 1, \cdots, \pm n-1\}$ of $T_n(f)$. This can cause problems in practical situations where we wish a circulant approximation to a Toeplitz matrix T_n when we *only* know T_n and not f. Pearl [13] discusses several coding and filtering applications where this restriction is necessary for practical reasons. A natural such approximation is to form the truncated Fourier series

$$\hat{f}_n(\lambda) = \sum_{k=-n}^{n} t_k e^{ik\lambda}, \tag{4.31}$$

which depends only on $\{t_k; k = 0, \pm 1, \cdots, \pm n-1\}$, and then define the circulant matrix

$$\hat{C}_n = C_n(\hat{f}_n); \tag{4.32}$$

that is, the circulant matrix having as top row $(\hat{c}_0^{(n)}, \cdots, \hat{c}_{n-1}^{(n)})$ where

$$\begin{aligned} \hat{c}_k^{(n)} &= n^{-1}\sum_{j=0}^{n-1}\hat{f}_n(2\pi j/n)e^{2\pi ijk/n} \\ &= n^{-1}\sum_{j=0}^{n-1}\left(\sum_{m=-n}^{n} t_m e^{2\pi ijk/n}\right)e^{2\pi ijk/n} \qquad (4.33) \\ &= \sum_{m=-n}^{n} t_m \left(n^{-1}\sum_{j=0}^{n-1} e^{2\pi ij(k+m)/n}\right). \end{aligned}$$

The last term in parentheses is from (3.9) 1 if $m = -k$ or $m = n - k$, and hence

$$\hat{c}_k^{(n)} = t_{-k} + t_{n-k} \quad , \quad k = 0, 1, \cdots, n-1.$$

Note that both $C_n(f)$ and $\hat{C}_n = C_n(\hat{f}_n)$ reduces to the $C_n(f)$ of (4.15) for an r^{th} order Toeplitz matrix if $n > 2r + 1$.

The matrix $\hat{C}_n$ does not have the property (4.28) of having eigenvalues $f(2\pi k/n)$ in the general case (its eigenvalues are $\hat{f}_n(2\pi k/n), k = 0, 1, \cdots, n-1$), but it does not have the desirable property to depending only on the entries of T_n. The following lemma shows that these circulant matrices are asymptotically equivalent to each other and T_m.

Lemma 4.6 *Let* $T_n = \{t_{k-j}\}$ *where*

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

and define as usual

$$f(\lambda) = \sum_{k=-\infty}^{\infty} t_k e^{ik\lambda}.$$

Define the circulant matrices $C_n(f)$ *and* $\hat{C}_n = C_n(\hat{f}_n)$ *as in (4.26) and (4.31)-(4.32). Then,*

$$C_n(f) \sim \hat{C}_n \sim T_n. \qquad (4.34)$$

Proof.

Since both $C_n(f)$ and $\hat{C}_n$ are circulant matrices with the same eigenvectors (Theorem 3.1), we have from part 2 of Theorem 3.1 and (2.14) and the comment following it that

$$|C_n(f) - \hat{C}_n|^2 = n^{-1} \sum_{k=0}^{n-1} |f(2\pi k/n) - \hat{f}_n(2\pi k/n)|^2.$$

Recall from (4.4) and the related discussion that $\hat{f}_n(\lambda)$ uniformly converges to $f(\lambda)$, and hence given $\epsilon > 0$ there is an N such that for $n \geq N$ we have for all k, n that

$$|f(2\pi k(n) - \hat{f}_n(2\pi k/n)|^2 \leq \epsilon$$

and hence for $n \geq N$

$$|C_n(f) - \hat{C}_n|^2 \leq n^{-1} \sum_{i=0}^{n-1} \epsilon = \epsilon.$$

Since ϵ is arbitrary,

$$\lim_{n\to\infty} |C_n(f) - \hat{C}_n| = 0$$

proving that

$$C_n(f) \sim \hat{C}_n. \tag{4.35}$$

Next, $\hat{C}_n = \{t'_{k-j}\}$ and use (4.12) to obtain

$$|T_n - \hat{C}_n|^2 = \sum_{k=-(n-1)}^{n-1} (1 - |k|/n)|t_k - t'_k|^2.$$

From (4.33) we have that

$$t'_k = \begin{cases} \hat{c}^{(n)}_{|k|} & = & t_{|k|} + t_{n-|k|} & k \leq 0 \\ & & & \\ \hat{c}^{(n)}_{n-k} & = & t_{-|n-k|} + t_k & k \geq 0 \end{cases} \tag{4.36}$$

and hence

$$\begin{aligned} |T_n - \hat{C}_n|^2 &= |t_{n-1}|^2 + \sum_{k=0}^{n-1} (1 - k/n)(|t_{n-k}|^2 + |t_{-(n-k)}|^2) \\ &= |t_{n-1}|^2 + \sum_{k=0}^{n-1} (k/n)(|t_k|^2 + |t_{-k}|^2) \end{aligned}$$

Since the $\{t_k\}$ are absolutely summable,

$$\lim_{n\to\infty} |t_{n-1}|^2 = 0$$

and given $\epsilon > 0$ we can choose an N large enough so that

$$\sum_{k=N}^{\infty} |t_k|^2 + |t_{-k}|^2 \le \epsilon$$

and hence

$$\begin{aligned}
\lim_{n\to\infty} |T_n - \hat{C}_n| &= \lim_{n\to\infty} \sum_{k=0}^{n-1} (k/n)(|t_k|^2 + |t_{-k}|^2) \\
&= \lim_{n\to\infty} \left\{ \sum_{k=0}^{N-1} (k/n)(|t_k|^2 + |t_{-k}|^2) + \sum_{k=N}^{n-1} (k/n)(|t_k|^2 + |t_{-k}|^2) \right\} . \\
&\le \lim_{n\to\infty} n^{-1} \left(\sum_{k=0}^{N-1} k(|t_k|^2 + |t_{-k}|^2) \right) + \sum_{k=N}^{\infty} (|t_k|^2 + |t_{-k}|^2) \le \epsilon
\end{aligned}$$

Since ϵ is arbitrary,

$$\lim_{n\to\infty} |T_n - \hat{C}_n| = 0$$

and hence

$$T_n \sim \hat{C}_n, \tag{4.37}$$

which with (4.35) and Theorem 2.1 proves (4.34).

We note that Pearl [13] develops a circulant matrix similar to $\hat{C}_n$ (depending only on the entries of T_n) such that (4.37) holds in the more general case where (4.1) instead of (4.2) holds.

We now have a circulant matrix $C_n(f)$ asymptotically equivalent to T_n and whose eigenvalues, inverses and products are known exactly. We can now use Lemmas 4.2-4.4 and Theorems 2.2-2.3 to immediately generalize Theorem 4.1

Theorem 4.2 *Let $T_n(f)$ be a sequence of Toeplitz matrices such that $f(\lambda)$ is Riemann integrable, e.g., $f(\lambda)$ is bounded or the sequence t_k is absolutely summable. Then if $\tau_{n,k}$ are the eigenvalues of T_n and s is any positive integer*

$$\lim_{n\to\infty} n^{-1} \sum_{k=0}^{n-1} \tau_{n,k}^s = (2\pi)^{-1} \int_0^{2\pi} f(\lambda)^s \, d\lambda. \tag{4.38}$$

Furthermore, if $T_n(f)$ is Hermitian ($f(\lambda)$ is real) then for any function $F(x)$ continuous on $[m_f, M_f]$

$$\lim_{n\to\infty} n^{-1}\sum_{k=0}^{n-1} F(\tau_{n,k}) = (2\pi)^{-1}\int_0^{2\pi} F[f(\lambda)]\,d\lambda. \tag{4.39}$$

Theorem 4.2 is the fundamental eigenvalue distribution theorem of Szegö [1]. The approach used here is essentially a specialization of Grenander's [1, ch. 7].

Theorem 4.2 yields the following two corollaries.

Corollary 4.1 *Let $T_n(f)$ be Hermitian and define the eigenvalue distribution function $D_n(x) = n^{-1}$ (number of $\tau_{n,k} \le x$). Assume that*

$$\int_{\lambda: f(\lambda)=x} d\lambda = 0.$$

Then the limiting distribution $D(x) = \lim_{n\to\infty} D_n(x)$ exists and is given by

$$D(x) = (2\pi)^{-1}\int_{f(\lambda)\le x} d\lambda.$$

The technical condition of a zero integral over the region of the set of λ for which $f(\lambda) = x$ is needed to ensure that x is a point of continuity of the limiting distribution.

Proof.

Define the characteristic function

$$1_x(\alpha) = \begin{cases} 1 & m_f \le \alpha \le x \\ \\ 0 & \text{otherwise} \end{cases}.$$

We have

$$D(x) = \lim_{n\to\infty} n^{-1}\sum_{k=0}^{n-1} 1_x(\tau_{n,k}) \quad .$$

Unfortunately, $1_x(\alpha)$ is not a continuous function and hence Theorem 4.2 cannot be immediately implied. To get around this problem we mimic Grenander and Szegö p. 115 and define two continuous functions that provide upper and lower bounds to 1_x and will converge to it in the limit. Define

$$1_x^+(\alpha) = \begin{cases} 1 & \alpha \le x \\ 1 - \frac{\alpha - x}{\epsilon} & x < \alpha \le x + \epsilon \\ 0 & x + \epsilon < \alpha \end{cases}$$

$$1_x^-(\alpha) = \begin{cases} 1 & \alpha \le x - \epsilon \\ 1 - \frac{\alpha - x + \epsilon}{\epsilon} & x - \epsilon < \alpha \le x \\ 0 & x < \alpha \end{cases}$$

The idea here is that the upper bound has an output of 1 everywhere 1_x does, but then it drops in a continuous linear fashion to zero at $x + \epsilon$ instead of immediately at x. The lower bound has a 0 everywhere 1_x does and it rises linearly from x to $x - \epsilon$ to the value of 1 instead of instantaneously as does 1_x. Clearly

$$1_x^-(\alpha) < 1_x(\alpha) < 1_x^+(\alpha)$$

for all α.

Since both 1_x^+ and 1_x^- are continuous, Theorem 4 can be used to conclude that

$$\begin{aligned} & \lim_{n\to\infty} n^{-1} \sum_{k=0}^{n-1} 1_x^+(\tau_{n,k}) \\ &= (2\pi)^{-1} \int 1_x^+(f(\lambda))\, d\lambda \\ &= (2\pi)^{-1} \int_{f(\lambda)\le x} d\lambda + (2\pi)^{-1} \int_{x<f(\lambda)\le x+\epsilon} (1 - \frac{f(\lambda) - x}{\epsilon})\, d\lambda \\ &\le (2\pi)^{-1} \int_{f(\lambda)\le x} d\lambda + (2\pi)^{-1} \int_{x<f(\lambda)\le x+\epsilon} d\lambda \end{aligned}$$

and

$$\begin{aligned} & \lim_{n\to\infty} n^{-1} \sum_{k=0}^{n-1} 1_x^-(\tau_{n,k}) \\ &= (2\pi)^{-1} \int 1_x^-(f(\lambda))\, d\lambda \\ &= (2\pi)^{-1} \int_{f(\lambda)\le x-\epsilon} d\lambda + (2\pi)^{-1} \int_{x-\epsilon<f(\lambda)\le x} (1 - \frac{f(\lambda) - (x-\epsilon)}{\epsilon})\, d\lambda \end{aligned}$$

$$\begin{aligned}
&= (2\pi)^{-1}\int_{f(\lambda)\le x-\epsilon} d\lambda + (2\pi)^{-1}\int_{x-\epsilon<f(\lambda)\le x}(x-f(\lambda))\,d\lambda \\
&\ge (2\pi)^{-1}\int_{f(\lambda)\le x-\epsilon} d\lambda \\
&= (2\pi)^{-1}\int_{f(\lambda)\le x} d\lambda - (2\pi)^{-1}\int_{x-\epsilon<f(\lambda)\le x} d\lambda
\end{aligned}$$

These inequalities imply that for any $\epsilon > 0$, as n grows the sample average $n^{-1}\sum_{k=0}^{n-1} 1_x(\tau_{n,k})$ will be sandwitched between

$$(2\pi)^{-1}\int_{f(\lambda)\le x} d\lambda + (2\pi)^{-1}\int_{x<f(\lambda)\le x+\epsilon} d\lambda$$

and

$$(2\pi)^{-1}\int_{f(\lambda)\le x} d\lambda - (2\pi)^{-1}\int_{x-\epsilon<f(\lambda)\le x} d\lambda.$$

Since ϵ can be made arbitrarily small, this means the sum will be sandwiched between

$$(2\pi)^{-1}\int_{f(\lambda)\le x} d\lambda$$

and

$$(2\pi)^{-1}\int_{f(\lambda)\le x} d\lambda - (2\pi)^{-1}\int_{f(\lambda)=x} d\lambda.$$

Thus if

$$\int_{f(\lambda)=x} d\lambda = 0,$$

then

$$\begin{aligned}
D(x) &= (2\pi)^{-1}\int_0^{2\pi} 1_x[f(\lambda)]d\lambda \\
&= (2\pi)^{-1} v\int_{f(\lambda)\le x} d\lambda
\end{aligned}.$$

Corollary 4.2 *For $T_n(f)$ Hermitian we have*

$$\lim_{n\to\infty}\max_k \tau_{n,k} = M_f$$

$$\lim_{n\to\infty}\min_k \tau_{n,k} = m_f.$$

Proof.

From Corollary 4.1 we have for any $\epsilon > 0$

$$D(m_f + \epsilon) = \int_{f(\lambda) \le m_f + \epsilon} d\lambda > 0.$$

The strict inequality follows from the continuity of $f(\lambda)$. Since

$$\lim_{n \to \infty} n^{-1} \{\text{number of } \tau_{n,k} \text{ in } [m_f, m_f + \epsilon]\} > 0$$

there must be eigenvalues in the interval $[m_f, m_f + \epsilon]$ for arbitrarily small ϵ. Since $\tau_{n,k} \ge m_f$ by Lemma 4.1, the minimum result is proved. The maximum result is proved similarly.

We next consider the inverse of an Hermitian Toeplitz matrix.

Theorem 4.3 *Let $T_n(f)$ be a sequence of Hermitian Toeplitz matrices such that $f(\lambda)$ is Riemann integrable and $f(\lambda) \ge 0$ with equality holding at most at a countable number of points.*

1. *$T_n(f)$ is nonsingular*
2. *If $f(\lambda) \ge m_f > 0$, then*

$$T_n(f)^{-1} \sim C_n(f)^{-1}, \tag{4.40}$$

where $C_n(f)$ is defined in (4.29). Furthermore, if we define $T_n(f) - C_n(f) = D_n$ then $T_n(f)^{-1}$ has the expansion

$$\begin{aligned} T_n(f)^{-1} &= [C_n(f) + D_n]^{-1} \\ &= C_n(f)^{-1}\left[I + D_n C_n(f)^{-1}\right]^{-1} \\ &= C_n(f)^{-1}\left[I + D_n C_n(f)^{-1} + \left(D_n C_n(f)^{-1}\right)^2 + \cdots\right] \end{aligned} \tag{4.41}$$

and the expansion converges (in weak norm) for sufficiently large n.

3. *If $f(\lambda) \ge m_f > 0$, then*

$$T_n(f)^{-1} \sim T_n(1/f) = \left\{(2\pi)^{-1} \int_{-\pi}^{\pi} d\lambda e^{i(k-j)\lambda} / f(\lambda)\right\}; \tag{4.42}$$

that is, if the spectrum is strictly positive then the inverse of a Toeplitz matrix is asymptotically Toeplitz. Furthermore if $\rho_{n,k}$ are the eigenvalues of $T_n(f)^{-1}$ and $F(x)$ is any continuous function on $[1/M_f, 1/m_f]$, then

$$\lim_{n\to\infty} n^{-1}\sum_{k=0}^{n-1} F(\rho_{n,k}) = (2\pi)^{-1}\int_{-\pi}^{\pi} F[(1/f(\lambda)]\, d\lambda. \tag{4.43}$$

4. *If $m_f = 0, f(\lambda)$ has at least one zero, and the derivative of $f(\lambda)$ exists and is bounded, then $T_n(f)^{-1}$ is not bounded, $1/f(\lambda)$ is not integrable and hence $T_n(1/f)$ is not defined and the integrals of (4.41) may not exist. For any finite θ, however, the following similar fact is true: If $F(x)$ is a continuous function of $[1/M_f, \theta]$, then*

$$\lim_{n\to\infty} n^{-1}\sum_{k=0}^{n-1} F[\min(\rho_{n,k},\theta)] = (2\pi)^{-1}\int_{0}^{2\pi} F[\min(1/f(\lambda),\theta)]\, d\lambda. \tag{4.44}$$

Proof.

1. Since $f(\lambda) > 0$ except at possible a finite number of points, we have from (4.9)
$$x^* T_n x = \frac{1}{2\pi}\int_{-\pi}^{\pi}\left|\sum_{k=0}^{n-1} x_k e^{ik\lambda}\right|^2 f(\lambda)d\lambda > 0$$
so that for all n
$$\min_k \tau_{n,k} > 0$$
and hence
$$\det T_n = \prod_{k=0}^{n-1}\tau_{n,k} \neq 0$$
so that $T_n(f)$ is nonsingular.

2. From Lemma 4.6, $T_n \sim C_n$ and hence (4.40) follows from Theorem 2.1 since $f(\lambda) \geq m_f > 0$ ensures that
$$\| T_n^{-1} \|, \| C_n^{-1} \| \leq 1/m_f < \infty.$$

The series of (4.41) will converge in weak norm if

$$|D_n C_n^{-1}| < 1 \tag{4.45}$$

since

$$|D_n C_n^{-1}| \leq \| C_n^{-1} \| \cdot |D_n| \leq (1/m_f)|D_n| \underset{n\to\infty}{\longrightarrow} 0$$

(4.45) must hold for large enough n. From (4.40), however, if n is large enough, then probably the first term of the series is sufficient.

3. We have

$$|T_n(f)^{-1} - T_n(1/f)| \leq |T_n(f)^{-1} - C_n(f)^{-1}| + |C_n(f)^{-1} - T_n(1/f)|.$$

From (b) for any $\epsilon > 0$ we can choose an n large enough so that

$$|T_n(f)^{-1} - C_n(f)^{-1}| \leq \frac{\epsilon}{2}. \tag{4.46}$$

From Theorem 3.1 and Lemma 4.5, $C_n(f)^{-1} = C_n(1/f)$ and from Lemma 4.6 $C_n(1/f) \sim T_n(1/f)$. Thus again we can choose n large enough to ensure that

$$|C_n(f)^{-1} - T_n(1/f)| \leq \epsilon/2 \tag{4.47}$$

so that for any $\epsilon > 0$ from (4.46)-(4.47) can choose n such that

$$|T_n(f)^{-1} - T_n(1/f)| \leq \epsilon$$

which is (4.42). Equation (4.43) follows from (4.42) and Theorem 2.4. Alternatively, if $G(x)$ is any continuous function on $[1/M_f, 1/m_f]$ and (4.43) follows directly from Lemma 4.6 and Theorem 2.4 applied to $G(1/x)$.

4. When $f(\lambda)$ has zeros ($m_f = 0$) then from Corollary 4.2 $\lim_{n\to\infty} \min_k \tau_{n,k} = 0$ and hence

$$\| T_n^{-1} \| = \max_k \rho_{n,k} = 1/ \min_k \tau_{n,k} \tag{4.48}$$

is unbounded as $n \to \infty$. To prove that $1/f(\lambda)$ is not integrable and hence that $T_n(1/f)$ does not exist we define the sets

$$\begin{aligned} E_k &= \{\lambda : 1/k \geq f(\lambda)/M_f > 1/(k+1)\} \\ &= \{\lambda : k \leq M_f/f(\lambda) < k+1\} \end{aligned} \tag{4.49}$$

since $f(\lambda)$ is continuous on $[0, M_f]$ and has at least one zero all of these sets are nonzero intervals of size, say, $|E_k|$. From (4.49)

$$\int_{-\pi}^{\pi} d\lambda/f(\lambda) \geq \sum_{k=1}^{\infty} |E_k| k/M_f \tag{4.50}$$

since $f(\lambda)$ is differentiable there is some finite value η such that

$$\left|\frac{df}{d\lambda}\right| \leq \eta. \tag{4.51}$$

From (4.50) and (4.51)

$$\begin{aligned} \int_{-\pi}^{\pi} d\lambda/f(\lambda) &\geq \sum_{k=1}^{\infty} (k/M_f)(1/k - 1/(k+1)/\eta \\ &= (M_f\eta)^{-1} \sum_{k=1}^{\infty} 1/(k+1) \end{aligned}, \tag{4.52}$$

which diverges so that $1/f(\lambda)$ is not integrable. To prove (4.44) let $F(x)$ be continuous on $[1/M_f, \theta]$, then $F[\min(1/x, \theta)]$ is continuous on $[0, M_f]$ and hence Theorem 2.4 yields (4.44). Note that (4.44) implies that the eigenvalues of T_n^{-1} are asymptotically equally distributed up to any finite θ as the eigenvalues of the sequence of matrices $T_n[\min(1/f, \theta)]$.

A special case of part 4 is when $T_n(f)$ is finite order and $f(\lambda)$ has at least one zero. Then the derivative exists and is bounded since

$$\begin{aligned} df/d\lambda &= \left|\sum_{k=-m}^{m} ikt_k e^{ik\lambda}\right| \\ &\leq \sum_{k=-m}^{m} |k||t_k| < \infty \end{aligned}.$$

The series expansion of part 2 is due to Rino [6]. The proof of part 4 is motivated by one of Widom [2]. Further results along the lines of part 4 regarding unbounded Toeplitz matrices may be found in [5].

Extending (a) to the case of non-Hermitian matrices can be somewhat difficult, i.e., finding conditions on $f(\lambda)$ to ensure that $T_n(f)$ is invertible.

Parts (a)-(d) can be straightforwardly extended if $f(\lambda)$ is continuous. For a more general discussion of inverses the interested reader is referred to Widom [2] and the references listed in that paper. It should be pointed out that when discussing inverses Widom is concerned with the asymptotic behavior of finite matrices. As one might expect, the results are similar. The results of Baxter [7] can also be applied to consider the asymptotic behavior of finite inverses in quite general cases.

We next combine Theorems 2.1 and Lemma 4.6 to obtain the asymptotic behavior of products of Toeplitz matrices. The case of only two matrices is considered first since it is simpler.

Theorem 4.4 *Let $T_n(f)$ and $T_n(g)$ be defined as in (4.5) where $f(\lambda)$ and $g(\lambda)$ are two bounded Riemann integrable functions. Define $C_n(f)$ and $C_n(g)$ as in (4.29) and let $\rho_{n,k}$ be the eigenvalues of $T_n(f)T_n(g)$*

1.

$$T_n(f)T_n(g) \sim C_n(f)C_n(g) = C_n(fg). \tag{4.53}$$

$$T_n(f)T_n(g) \sim T_n(g)T_n(f). \tag{4.54}$$

$$\lim_{n\to\infty} n^{-1}\sum_{k=0}^{n-1}\rho_{n,k}^s = (2\pi)^{-1}\int_0^{2\pi}[f(\lambda)g(\lambda)]^s\,d\lambda \quad s = 1, 2, \ldots. \tag{4.55}$$

2. If $T_n(t)$ and $T_n(g)$ are Hermitian, then for any $F(x)$ continuous on $[m_f m_g, M_f M_g]$

$$\lim_{n\to\infty} n^{-1}\sum_{k=0}^{n-1}F(\rho_{n,k}) = (2\pi)^{-1}\int_0^{2\pi}F[f(\lambda)g(\lambda)]\,d\lambda. \tag{4.56}$$

3.

$$T_n(f)T_n(g) \sim T_n(fg). \tag{4.57}$$

4. Let $f_1(\lambda), .., f_m(\lambda)$ be Riemann integrable. Then if the $C_n(f_i)$ are defined as in (4.29)

$$\prod_{i=1}^{m}T_n(f_i) \sim C_n\left(\prod_{i=1}^{m}f_i\right) \sim T_n\left(\prod_{i=1}^{m}f_i\right). \tag{4.58}$$

5. *If $\rho_{n,k}$ are the eigenvalues of $\prod_{i=1}^{m} T_n(f_i)$, then for any positive integer s*

$$\lim_{n\to\infty} n^{-1} \sum_{k=0}^{n-1} \rho_{n,k}^s = (2\pi)^{-1} \int_0^{2\pi} \left(\prod_{i=1}^{m} f_i(\lambda) \right)^s d\lambda \tag{4.59}$$

If the $T_n(f_i)$ are Hermitian, then the $\rho_{n,k}$ are asymptotically real, i.e., the imaginary part converges to a distribution at zero, so that

$$\lim_{n\to\infty} n^{-1} \sum_{k=0}^{n-1} (Re[\rho_{n,k}])^s = (2\pi)^{-1} \int_0^{2\pi} \left(\prod_{i=1}^{m} f_i(\lambda) \right)^s d\lambda. \tag{4.60}$$

$$\lim_{n\to\infty} n^{-1} \sum_{k=0}^{n-1} (\Im[\rho_{n,k}])^2 = 0. \tag{4.61}$$

Proof.

1. Equation (4.53) follows from Lemmas 4.5 and 4.6 and Theorems 2.1 and 3. Equation (4.54) follows from (4.53). Note that while Toeplitz matrices do not in general commute, asymptotically they do. Equation (4.55) follows from (4.53), Theorem 2.2, and Lemma 4.4.

2. Proof follows from (4.53) and Theorem 2.4. Note that the eigenvalues of the product of two Hermitian matrices are real [3, p. 105].

3. Applying Lemmas 4.5 and 4.6 and Theorem 2.1

$$\begin{aligned} |T_n(f)T_n(g) - T_n(fg)| &= |T_n(f)T_n(g) - C_n(f)C_n(g) \\ &\quad + C_n(f)C_n(g) - T_n(fg)| \\ &\le |T_n(f)T_n(g) - C_n(f)C_n(g)| \\ &\quad + |C_n(fg) - T_n(fg)| \underset{n\to\infty}{\longrightarrow} 0 \end{aligned}.$$

4. Follows from repeated application of (4.53) and part (c).

5. Equation (4.58) follows from (d) and Theorem 2.1. For the Hermitian case, however, we cannot simply apply Theorem 2.4 since the eigenvalues $\rho_{n,k}$ of $\prod_i T_n(f_i)$ may not be real. We can show, however, that they

are asymptotically real. Let $\rho_{n,k} = \alpha_{n,k} + i\beta_{n,k}$ where $\alpha_{n,k}$ and $\beta_{n,k}$ are real. Then from Theorem 2.2 we have for any positive integer s

$$\begin{aligned} \lim_{n\to\infty} n^{-1}\sum_{k=0}^{n-1}(\alpha_{n,k} + i\beta_{n,k})^s &= \lim_{n\to\infty} n^{-1}\sum_{k=0}^{n-1}\psi_{n,k}^s \\ &= (2\pi)^{-1}\int_0^{2\pi}\left[\prod_{i=1}^{m} f_i(\lambda)\right]^s d\lambda \end{aligned}, \tag{4.62}$$

where $\psi_{n,k}$ are the eigenvalues of $C_n\left(\prod_{i=1}^{m} f_i\right)$. From (2.14)

$$n^{-1}\sum_{k=0}^{n-1}|\rho_{n,k}|^2 = n^{-1}\sum_{k=0}^{n-1}\left(\alpha_{n,k}^2 + \beta_{n,k}^2\right) \le \left|\prod_{i=i}^{m} T_n(f_i)\right|^2.$$

From (4.57), Theorem 2.1 and Lemma 4.4

$$\begin{aligned} \lim_{n\to\infty}\left|\prod_{i=1}^{m} T_n(f_i)\right|^2 &= \lim_{n\to\infty}\left|C_n\left(\prod_{i=1}^{m} f_i\right)\right|^2 \\ &= (2\pi)^{-1}\int_0^{2\pi}\left(\prod_{i=1}^{m} f_i(\lambda)\right)^2 d\lambda \end{aligned}. \tag{4.63}$$

Subtracting (4.61) for $s = 2$ from (4.61) yields

$$\lim_{n\to\infty} n^{-1}\sum_{k=1}^{n-1}\beta_{n,k}^2 \le 0.$$

Thus the distribution of the imaginary parts tends to the origin and hence

$$\lim_{n\to\infty} n^{-1}\sum_{k=0}^{n-1}\alpha_{n,k}^s = (2\pi)^{-1}\int_0^{2\pi}\left[\prod_{i=1}^{m} f_i(\lambda)\right]^s d\lambda.$$

Parts (d) and (e) are here proved as in Grenander and Szegö [1, pp. 105-106].

We have developed theorems on the asymptotic behavior of eigenvalues, inverses, and products of Toeplitz matrices. The basic method has been to find an asymptotically equivalent circulant matrix whose special simple structure as developed in Chapter 3 could be directly related to the Toeplitz matrices using the results of Chapter 2. We began with the finite order case since the appropriate circulant matrix is there obvious and yields certain desirable properties that suggest the corresponding circulant matrix in the infinite case. We have limited our consideration of the infinite order case to absolutely summable coefficients or to bounded Riemann integrable functions $f(\lambda)$ for simplicity. The more general case of square summable t_k or bounded Lebesgue integrable $f(\lambda)$ treated in Chapter 7 of [1] requires significantly more mathematical care but can be interpreted as an extension of the approach taken here.

4.3 Toeplitz Determinants

The fundamental Toeplitz eigenvalue distribution theory has an interesting application for characterizing the limiting behavior of determinants. Suppose now that $T_n(f)$ is a sequence of Hermitian Toeplitz matrices such that that $f(\lambda) \geq m_f > 0$. Let $C_n = C_n(f)$ denote the sequence of circulant matrices constructed from f as in (4.26). Then from (4.28) the eigenvalues of C_n are $f(2\pi m/n)$ for $m = 0, 1, \ldots, n-1$ and hence $\det C_n = \prod_{m=0}^{n-1} f(2\pi m/n)$. This in turn implies that

$$\ln\left(\det(C_n)\right)^{\frac{1}{n}} = \frac{1}{n}\ln\det C_n = \frac{1}{n}\sum_{m=0}^{n-1}\ln f(2\pi\frac{m}{n}).$$

These sums are the Riemann approximations to the limiting integral, whence

$$\lim_{n\to\infty}\ln\left(\det(C_n)\right)^{\frac{1}{n}} = \int_0^1 \ln f(2\pi\lambda)\,d\lambda.$$

Exponentiating, using the continuity of the logarithm for strictly positive arguments, and changing the variables of integration yields

$$\lim_{n\to\infty}\left(\det(C_n)\right)^{\frac{1}{n}} = e^{\frac{1}{2\pi}\int_0^{2\pi}\ln f(\lambda)\,d\lambda}.$$

This integral, the asymptotic equivalence of C_n and $T_n(f)$ (Lemma 4.6), and Corollary 2.3 togther yield the following result ([1], p. 65)

Theorem 4.5 *Let $T_n(f)$ be a sequence of Hermitian Toeplitz matrices such that $\ln f(\lambda)$ is Riemann integrable and $f(\lambda) \geq m_f > 0$. Then*

$$\lim_{n\to\infty}\left(det(T_n(f))\right)^{\frac{1}{n}} = e^{\frac{1}{2\pi}\int_0^{2\pi}\ln f(\lambda)\,d\lambda}. \tag{4.64}$$

Chapter 5

Applications to Stochastic Time Series

Toeplitz matrices arise quite naturally in the study of discrete time random processes. Covariance matrices of weakly stationary processes are Toeplitz and triangular Toeplitz matrices provide a matrix representation of causal linear time invariant filters. As is well known and as we shall show, these two types of Toeplitz matrices are intimately related. We shall take two viewpoints in the first section of this chapter section to show how they are related. In the first part we shall consider two common linear models of random time series and study the asymptotic behavior of the covariance matrix, its inverse and its eigenvalues. The well known equivalence of moving average processes and weakly stationary processes will be pointed out. The lesser known fact that we can define something like a power spectral density for autoregressive processes even if they are nonstationary is discussed. In the second part of the first section we take the opposite tack — we start with a Toeplitz covariance matrix and consider the asymptotic behavior of its triangular factors. This simple result provides some insight into the asymptotic behavior or system identification algorithms and Wiener-Hopf factorization.

The second section provides another application of the Toeplitz distribution theorem to stationary random processes by deriving the Shannon information rate of a stationary Gaussian random process.

Let $\{X_k;\ k \in \mathcal{I}\}$ be a discrete time random process. Generally we take $\mathcal{I} = \mathcal{Z}$, the space of all integers, in which case we say that the process is *two-sided*, or $\mathcal{I} = \mathcal{Z}_+$, the space of all nonnegative integers, in which case we say that the process is *one-sided*. We will be interested in vector representations of the process so we define the column vector ($n-$tuple) $X^n = (X_0, X_1, \ldots, X_{n-1})^t$, that is, X^n is an n-dimensional column vector. The mean vector is defined by $m^n = E(X^n)$, which we usually assume is zero for convenience. The $n \times n$ covariance matrix $R_n = \{r_{j,k}\}$ is defined by

$$R_n = E[(X^n - m^n)(X^n - m^n)^*]. \tag{5.1}$$

This is the autocorrelation matrix when the mean vector is zero. Subscripts will be dropped when they are clear from context. If the matrix R_n is Toeplitz, say $R_n = T_n(f)$, then $r_{k,j} = r_{k-j}$ and the process is said to be *weakly stationary*. In this case we can define $f(\lambda) = \sum_{k=-\infty}^{\infty} r_k e^{ik\lambda}$ as the power spectral density of the process. If the matrix R_n is not Toeplitz but is asymptotically Toeplitz, i.e., $R_n \sim T_n(f)$, then we say that the process is asymptotically weakly stationary and once again define $f(\lambda)$ as the power spectral density. The latter situation arises, for example, if an otherwise stationary process is initialized with $X_k = 0$, $k \leq 0$. This will cause a transient and hence the process is strictly speaking nonstationary. The transient dies out, however, and the statistics of the process approach those of a weakly stationary process as n grows.

The results derived herein are essentially trivial if one begins and deals only with doubly infinite matrices. As might be hoped the results for asymptotic behavior of finite matrices are consistent with this case. The problem is of interest since one often has finite order equations and one wishes to know the asymptotic behavior of solutions or one has some function defined as a limit of solutions of finite equations. These results are useful both for finding theoretical limiting solutions and for finding reasonable approximations for finite order solutions. So much for philosophy. We now proceed to investigate the behavior of two common linear models. For simplicity we will assume the process means are zero.

5.1 Moving Average Sources

By a linear model of a random process we mean a model wherein we pass a zero mean, independent identically distributed (iid) sequence of random variables W_k with variance σ^2 through a linear time invariant discrete time filtered to obtain the desired process. The process W_k is discrete time "white" noise. The most common such model is called a moving average process and is defined by the difference equation

$$U_n = \sum_{k=0}^{n} b_k W_{n-k} = \sum_{k=0}^{n} b_{n-k} W_k \tag{5.2}$$

$$U_n = 0;\ n < 0.$$

We assume that $b_0 = 1$ with no loss of generality since otherwise we can incorporate b_0 into σ^2. Note that (5.2) is a discrete time convolution, i.e., U_n is the output of a filter with "impulse response" (actually Kronecker δ response) b_k and input W_k. We could be more general by allowing the filter b_k to be noncausal and hence act on future W_k's. We could also allow the W_k's and U_k's to extend into the infinite past rather than being initialized. This would lead to replacing of (5.2) by

$$U_n = \sum_{k=-\infty}^{\infty} b_k W_{n-k} = \sum_{k=-\infty}^{\infty} b_{n-k} W_k. \tag{5.3}$$

We will restrict ourselves to causal filters for simplicity and keep the initial conditions since we are interested in limiting behavior. In addition, since stationary distributions may not exist for some models it would be difficult to handle them unless we start at some fixed time. For these reasons we take (5.2) as the definition of a moving average.

Since we will be studying the statistical behavior of U_n as n gets arbitrarily large, some assumption must be placed on the sequence b_k to ensure that (5.2) converges in the mean-squared sense. The weakest possible assumption that will guarantee convergence of (5.2) is that

$$\sum_{k=0}^{\infty} |b_k|^2 < \infty. \tag{5.4}$$

In keeping with the previous sections, however, we will make the stronger assumption

$$\sum_{k=0}^{\infty} |b_k| < \infty. \tag{5.5}$$

As previously this will result in simpler mathematics.

Equation (5.2) can be rewritten as a matrix equation by defining the lower triangular Toeplitz matrix

$$B_n = \begin{bmatrix} 1 & & & & 0 & \\ b_1 & 1 & & & & \\ b_2 & b_1 & & & & \\ \vdots & b_2 & \ddots & \ddots & & \\ & & & & & \\ b_{n-1} & \dots & & b_2 & b_1 & 1 \end{bmatrix} \tag{5.6}$$

so that

$$U^n = B_n W^n. \tag{5.7}$$

If the filter b_n were not causal, then B_n would not be triangular. If in addition (5.3) held, i.e., we looked at the entire process at each time instant, then (5.7) would require infinite vectors and matrices as in Grenander and Rosenblatt [12]. Since the covariance matrix of W_k is simply $\sigma^2 I_n$, where I_n is the $n \times n$ identity matrix, we have for the covariance of U_n:

$$\begin{aligned} R_U^{(n)} &= EU^n(U^n)^* = EB_nW^n(W^n)^*B_n^* \\ &= \sigma B_n B_n^* \end{aligned} \tag{5.8}$$

or, equivalently

$$\begin{aligned} r_{k,j} &= \sigma^2 \sum_{\ell=0}^{n-1} b_{\ell-k}\bar{b}_{\ell-j} \\ &= \sigma^2 \sum_{\ell=0}^{\min(k,j)} b_{\ell+(k-j)}\bar{b}_\ell \end{aligned} . \tag{5.9}$$

From (5.9) it is clear that $r_{k,j}$ is not Toeplitz because of the $\min(k,j)$ in the sum. However, as we next show, as $n \to \infty$ the upper limit becomes large and $R_U^{(n)}$ becomes asymptotically Toeplitz. If we define

$$b(\lambda) = \sum_{k=0}^{\infty} b_k e^{ik\lambda} \tag{5.10}$$

then

$$B_n = T_n(b) \tag{5.11}$$

so that

$$R_U^{(n)} = \sigma^2 T_n(b) T_n(b)^*. \tag{5.12}$$

We can now apply the results of the previous sections to obtain the following theorem.

Theorem 5.1 *Let U_n be a moving average process with covariance matrix R_{U_n}. Let $\rho_{n,k}$ be the eigenvalues of $R_U^{(n)}$. Then*

$$R_U^{(n)} \sim \sigma^2 T_n(|b|^2) = T_n(\sigma^2|b|^2) \tag{5.13}$$

so that U_n is asymptotically stationary. If $m \leq |b(\gamma)|^2 \leq M$ and $F(x)$ is any continuous function on $[m, M]$, then

$$\lim_{n\to\infty} n^{-1} \sum_{k=0}^{n-1} F(\rho_{n,k}) = (2\pi)^{-1} \int_0^{2\pi} F(\sigma^2|b(\lambda)|^2)\, d\lambda. \tag{5.14}$$

If $|b(\lambda)|^2 \geq m > 0$, then

$$R_U^{(n)^{-1}} \sim \sigma^{-2} T_n(1/|b|^2). \tag{5.15}$$

Proof.

(Theorems 4.2-4.4 and 2.4.)

If the process U_n had been initiated with its stationary distribution then we would have had exactly

$$R_U^{(n)} = \sigma^2 T_n(|b|^2).$$

More knowledge of the inverse $R_U^{(n)^{-1}}$ can be gained from Theorem 4.3, e.g., circulant approximations. Note that the spectral density of the moving average process is $\sigma^2|b(\lambda)|^2$ and that sums of functions of eigenvalues tend to an integral of a function of the spectral density. In effect the spectral density determines the asymptotic density function for the eigenvalues of R_n and T_n.

5.2 Autoregressive Processes

Let W_k be as previously defined, then an autoregressive process X_n is defined by

$$X_n = -\sum_{k=1}^{n} a_k X_{n-k} + W_k \qquad n = 0, 1, \ldots$$

$$X_n = 0 \qquad\qquad n < 0. \tag{5.16}$$

Autoregressive process include nonstationary processes such as the Wiener process. Equation (5.16) can be rewritten as a vector equation by defining the lower triangular matrix.

$$A_n = \begin{bmatrix} 1 & & & & & \\ a_1 & 1 & & & & 0 \\ & a_1 & 1 & & & \\ & & \ddots & \ddots & & \\ & & & & & \\ a_{n-1} & & & & a_1 & 1 \end{bmatrix} \tag{5.17}$$

so that

$$A_n X^n = W^n.$$

We have

$$R_W^{(n)} = A_n R_X^{(n)} A_n^* \tag{5.18}$$

since det $A_n = 1 \neq 0$, A_n is nonsingular so that

$$R_X^{(n)} = \sigma^2 A_n^{-1} A_n^{-1*} \tag{5.19}$$

or

$$(R_X^{(n)})^{-1} = \sigma^2 A_n^* A_n \tag{5.20}$$

or equivalently, if $(R_X^{(n)})^{-1} = \{t_{k,j}\}$ then

$$t_{k,j} = \sum_{m=0}^{n} \bar{a}_{m-k} a_{m-j} = \sum_{m=0}^{n-\max(k,j)} a_m a_{m+(k-j)}.$$

Unlike the moving average process, we have that the inverse covariance matrix is the product of Toeplitz triangular matrices. Defining

$$a(\lambda) = \sum_{k=0}^{\infty} a_k e^{ik\lambda} \tag{5.21}$$

we have that

$$(R_X^{(n)})^{-1} = \sigma^{-2} T_n(a)^* T_n(a) \tag{5.22}$$

and hence the following theorem.

Theorem 5.2 *Let X_n be an autoregressive process with covariance matrix $R_X^{(n)}$ with eigenvalues $\rho_{n,k}$. Then*

$$(R_X^{(n)})^{-1} \sim \sigma^{-2} T_n(|a|^2). \tag{5.23}$$

If $m' \leq |a(\lambda)|^2 \leq m'$, then for any function $F(x)$ on $[m', M']$ we have

$$\lim_{n\to\infty} n^{-1} \sum_{k=0}^{n-1} F(1/\rho_{n,k}) = (2\pi)^{-1} \int_0^{2\pi} F(\sigma^2 |a(\lambda)|^2)\, d\lambda, \tag{5.24}$$

where $1/\rho_{n,k}$ are the eigenvalues of $(R_X^{(n)})^{-1}$. If $|a(\lambda)|^2 \geq m' > 0$, then

$$R_X^{(n)} \sim \sigma^2 T_n(1/|a|^2) \tag{5.25}$$

so that the process is asymptotically stationary.

Proof.

(Theorems 5.1.)

Note that if $|a(\lambda)|^2 > 0$, then $1/|a(\lambda)|^2$ is the spectral density of X_n. If $|a(\lambda)|^2$ has a zero, then $R_X^{(n)}$ may not be even asymptotically Toeplitz and hence X_n may not be asymptotically stationary (since $1/|a(\lambda)|^2$ may not be integrable) so that strictly speaking x_k will not have a spectral density. It is often convenient, however, to define $\sigma^2/|a(\lambda)|^2$ as the spectral density and it often is useful for studying the eigenvalue distribution of R_n. We can relate $\sigma^2/|a(\lambda)|^2$ to the eigenvalues of $R_X^{(n)}$ even in this case by using Theorem 4.3 part 4.

Corollary 5.1 *If X_k is an autoregressive process and $\rho_{n,k}$ are the eigenvalues of $R_X^{(n)}$, then for any finite θ and any function $F(x)$ continuous on $[1/m', \theta]$*

$$\lim_{n\to\infty} n^{-1} \sum_{k=0}^{n-1} F[\min(\rho_{n,k}, \theta)] = (2\pi)^{-1} \int_0^{2\pi} F[\min(1/|a(\gamma)|^2, \theta)]\, d\lambda. \tag{5.26}$$

Proof.

(Theorem 5.2 and Theorem 4.3.)

If we consider two models of a random process to be asymptotically equivalent if their covariances are asymptotically equivalent, then from Theorems 5.1d and 5.2 we have the following corollary.

Corollary 5.2 *Consider the moving average process defined by*

$$U^n = T_n(b)W^n$$

and the autoregressive process defined by

$$T_n(a)X^n = W^n.$$

Then the processes U_n and X_n are asymptotically equivalent if

$$a(\lambda) = 1/b(\lambda)$$

and $M \geq a(\lambda) \geq m > 0$ so that $1/b(\lambda)$ is integrable.

Proof.

(Theorem 4.3 and Theorem 4.5.)

$$\begin{aligned} R_X^{(n)} &= \sigma^2 T_n(a)^{-1} T_n^{-1}(a)^* \\ &\sim \sigma^2 T_n(1/a) T_n(1/a)^* \\ &\sim \sigma^2 T_n(1/a)^* T_n(1/a). \end{aligned} \tag{5.27}$$

Comparison of (5.27) with (5.12) completes the proof.

The methods above can also easily be applied to study the mixed autoregressive-moving average linear models [2].

5.3 Factorization

As a final example we consider the problem of the asymptotic behavior of triangular factors of a sequence of Hermitian covariance matrices $T_n(f)$. It is well known that any such matrix can be factored into the product of a lower triangular matrix and its conjugate transpose [12, p. 37], in particular

$$T_n(f) = \{t_{k,j}\} = B_n B_n^*, \tag{5.28}$$

where B_n is a lower triangular matrix with entries

$$b_{k,j}^{(n)} = \{(\det T_k)\det(T_{k-1})\}^{-1/2}\gamma(j,k), \tag{5.29}$$

where $\gamma(j,k)$ is the determinant of the matrix T_k with the right hand column replaced by $(t_{j,0}, t_{j,1}, \ldots, t_{j,k-1})^t$. Note in particular that the diagonal elements are given by

$$b_{k,k}^{(n)} = \{(\det T_k)/(\det T_{k-1})\}^{1/2}. \tag{5.30}$$

Equation (5.29) is the result of a Gaussian elimination of a Gram-Schmidt procedure. The factorization of T_n allows the construction of a linear model of a random process and is useful in system identification and other recursive procedures. Our question is how B_n behaves for large n; specifically is B_n asymptotically Toeplitz?

Assume that $f(\lambda) \geq m > 0$. Then $\ln f(\lambda)$ is integrable and we can perform a Wiener-Hopf factorization of $f(\lambda)$, i.e.,

$$\begin{aligned} f(\lambda) &= \sigma^2 |b(\lambda)|^2 \\ \bar{b}(\lambda) &= b(-\lambda) \\ b(\lambda) &= \sum_{k=0}^{\infty} b_k e^{ik\lambda} \\ b_0 &= 1 \end{aligned} \quad . \tag{5.31}$$

From (5.28) and Theorem 4.4 we have

$$B_n B_n^* = T_n(f) \sim T_n(\sigma b) T_n(\sigma b)^*. \tag{5.32}$$

We wish to show that (5.32) implies that

$$B_n \sim T_n(\sigma b). \tag{5.33}$$

Proof.

Since $\det T_n(\sigma b) = \sigma^n \neq 0$, $T_n(\sigma b)$ is invertible. Likewise, since $\det B_n = [\det T_n(f)]^{1/2}$ we have from Theorem 4.3 part 1 that $\det T_n(f) \neq 0$ so that B_n is invertible. Thus from Theorem 2.1 (e) and (5.32) we have

$$T_n^{-1} B_n = [B_n^{-1} T_n]^{-1} \sim T_n^* B_n^{*-1} = [B_n^{-1} T_n]^*. \tag{5.34}$$

Since B_n and T_n are both lower triangular matrices, so is B_n^{-1} and hence $B_n T_n$ and $[B_n^{-1} T_n]^{-1}$. Thus (5.34) states that a lower triangular matrix is asymptotically equivalent to an upper triangular matrix. This is only possible if both matrices are asymptotically equivalent to a diagonal matrix, say $G_n = \{g_{k,k}^{(n)} \delta_{k,j}\}$. Furthermore from (5.34) we have $G_n \sim G_n^{*-1}$

$$\left\{ |g_{k,k}^{(n)}|^2 \delta_{k,j} \right\} \sim I_n. \tag{5.35}$$

Since $T_n(\sigma b)$ is lower triangular with main diagonal element σ, $T_n(\sigma b)^{-1}$ is lower triangular with all its main diagonal elements equal to $1/\sigma$ even though the matrix $T_n(\sigma b)^{-1}$ is not Toeplitz. Thus $g_{k,k}^{(n)} = b_{k,k}^{(n)}/\sigma$. Since $T_n(f)$ is Hermitian, $b_{k,k}$ is real so that taking the trace in (5.35) yields

$$\lim_{n\to\infty} \sigma^{-2} n^{-1} \sum_{k=0}^{n-1} \left(b_{k,k}^{(n)} \right)^2 = 1. \tag{5.36}$$

From (5.30) and Corollary 2.3, and the fact that $T_n(\sigma b)$ is triangular we have that

$$\begin{aligned}
\lim_{n\to\infty} \sigma^{-1} n^{-1} \sum_{k=0}^{n-1} b_{k,k}^{(n)} &= \sigma^{-1} \lim_{n\to\infty} \{(\det T_n(f))/(\det T_{n-1}(f))\}^{1/2} \\
&= \sigma^{-1} \lim_{n\to\infty} \{\det T_n(f)\}^{1/2n} \sigma^{-1} \lim_{n\to\infty} \{\det T_n(\sigma b)\}^{1/n} \\
&= \sigma^{-1} \cdot \sigma = 1
\end{aligned} \tag{5.37}$$

Combining (5.36) and (5.37) yields

$$\lim_{n\to\infty} |B_n^{-1} T_n - I_n| = 0. \tag{5.38}$$

Applying Theorem 2.1 yields (5.33).

Since the only real requirements for the proof were the existence of the Wiener-Hopf factorization and the limiting behavior of the determinant, this result could easily be extended to the more general case that $\ln f(\lambda)$ is integrable. The theorem can also be derived as a special case of more general results of Baxter [8] and is similar to a result of Rissanen [11].

5.4 Differential Entropy Rate of Gaussian Processes

As a final application of the Toeplitz eigenvalue distribution theorem, we consider a property of a random process that arises in Shannon information theory. Given a random process $\{X_n\}$ for which a probability density function $f_{X^n}(x^n)$ is for the random vector $X^n = (X_0, X_1, \ldots, X_{n-1})$ defined for all positive integers n, the Shannon differential entropy $h(X^n)$ is defined by the integral

$$h(X^n) = -\int f_{X^n}(x^n) \log f_{X^n}(x^n)\, dx^n$$

and the differential entropy rate is defined by the limit

$$h(X) = \lim_{n\to\infty} \frac{1}{n} h(X^n)$$

if the limit exists. (See, for example, Cover and Thomas[14].) The logarithm is usually taken as base 2 and the units are *bits*. We will use the Toeplitz theorem to evaluate the differential entropy rate of a stationary Gaussian random process.

A stationary zero mean Gaussian random process is comletely described by its mean correlation function $R_X(k,m) = R_X(k-m) = E[(X_k - m)(X_k - m)]$ or, equivalently, by its power spectral density function

$$S(f) = \sum_{n=-\infty}^{\infty} R_X(n) e^{-2\pi i n f},$$

the Fourier transform of the covariance function. For a fixed positive integer n, the probability density function is

$$f_{X^n}(x^n) = \frac{1}{(2\pi)^{n/2} \det(R_n)^{1/2}} e^{-\frac{1}{2}(x^n - m^n)^t R_n^{-1}(x^n - m^n)},$$

where R_n is the $n \times n$ covariance matrix with entries $R_X(k,m)$, $k,m = 0,1,\ldots,n-1$. A straightforward multidimensional integration using the properties of Gaussian random vectors yields the differential entropy

$$h(X^n) = \frac{1}{2}\log(2\pi e)^n \det R_n.$$

If we now identify the the covariance matrix R_n as the Toeplitz matrix generated by the power spectral density, $T_n(S)$, then from Theorem 4.5 we have immediately that

$$h(X) = \frac{1}{2}\log(2\pi e)\sigma_\infty^2 \tag{5.39}$$

where

$$\sigma_\infty^2 = \frac{1}{2\pi}\int_0^{2\pi} \ln S(f)\, df. \tag{5.40}$$

The Toeplitz distribution theorems have also found application in more complicated information theoretic evaluations, including the channel capacity of Gaussian channels [17, 18] and the rate-distortion functions of autoregressive sources [9].

Bibliography

[1] U. Grenander and G. Szegö, *Toeplitz Forms and Their Applications,* University of Calif. Press, Berkeley and Los Angeles, 1958.

[2] H. Widom, "Toeplitz Matrices," in *Studies in Real and Complex Analysis,* edited by I.I. Hirschmann, Jr., *MAA Studies in Mathematics,* Prentice-Hall, Englewood Cliffs, NJ, 1965.

[3] P. Lancaster, *Theory of Matrices,* Academic Press, NY, 1969.

[4] W. Rudin, *Principles of Mathematical Analysis,* McGraw-Hill, NY, 1964.

[5] R.M. Gray, "On Unbounded Toeplitz Matrices and Nonstationary Time Series with an Application to Information Theory," *Information and Control,* **24**, pp. 181–196, 1974.

[6] C.L. Rino, "The Inversion of Covariance Matrices by Finite Fourier Transforms," *IEEE Trans. on Info. Theory,* **IT-16**, No. 2, March 1970, pp. 230–232.

[7] G. Baxter, "A Norm Inequality for a 'Finite-Section' Wiener-Hopf Equation," *Illinois J. Math.,* 1962, pp. 97–103.

[8] G. Baxter, "An Asymptotic Result for the Finite Predictor," *Math. Scand.,* **10**, pp. 137–144, 1962.

[9] R.M. Gray, "Information Rates of Autoregressive Processes," *IEEE Trans. on Info. Theory,* **IT-16**, No. 4, July 1970, pp. 412–421.

[10] F.R. Gantmacher, *The Theory of Matrices,* Chelsea Publishing Co., NY 1960.

[11] J. Rissanen and L. Barbosa, "Properties of Infinite Covariance Matrices and Stability of Optimum Predictors," *Information Sciences,* **1**, 1969, pp. 221–236.

[12] U. Grenander and M. Rosenblatt, *Statistical Analysis of Stationary Time Series,* Wiley and Sons, NY, 1966, Chapter 1.

[13] J. Pearl, "On Coding and Filtering Stationary Signals by Discrete Fourier Transform," *IEEE Trans. on Info. Theory,* **IT-19**, pp. 229–232, 1973.

[14] T. A. Cover and J. A. Thomas, *Elements of Information Theory*, Wiley, New York, 1991.

[15] A. Böttcher and B. Silbermann, *Introduction to Large Truncated Toeplitz Matrices*, Springer, New York, 1999.

[16] R. M. Gray, "On the asymptotic eigenvalue distribvution of Toeplitz matrices," *IEEE Transactions on Information Theory*, Vol. 18, November 1972, pp. 725–730.

[17] B.S. Tsybakov, "On the transmission capacity of a discrete-time Gaussian channel with filter," (in Russion),*Probl. Peredach. Inform.*, Vol 6, pp. 78–82, 1970.

[18] B.S. Tsybakov, "Transmission capacity of memoryless Gaussian vector channels," (in Russion),*Probl. Peredach. Inform.*, Vol 1, pp. 26–40, 1965.

[19] P. J. Davis, *Circulant Matrices*, Wiley-Interscience, NY, 1979.