Toeplitz and Circulant Matrices: A review

Toeplitz and Circulant Matrices: A review

Robert M. Gray

Deptartment of Electrical Engineering
Stanford University
Stanford 94305, USA
rmgray@stanford.edu

the essence of knowledge

Contents

Chapter 1 Introduction **1**

1.1 Toeplitz and Circulant Matrices 1
1.2 Examples 5
1.3 Goals and Prerequisites 9

Chapter 2 The Asymptotic Behavior of Matrices **11**

2.1 Eigenvalues 11
2.2 Matrix Norms 14
2.3 Asymptotically Equivalent Sequences of Matrices 17
2.4 Asymptotically Absolutely Equal Distributions 24

Chapter 3 Circulant Matrices **31**

3.1 Eigenvalues and Eigenvectors 32
3.2 Matrix Operations on Circulant Matrices 34

Chapter 4 Toeplitz Matrices **37**

4.1 Sequences of Toeplitz Matrices 37
4.2 Bounds on Eigenvalues of Toeplitz Matrices 41
4.3 Banded Toeplitz Matrices 43
4.4 Wiener Class Toeplitz Matrices 48

Chapter 5 Matrix Operations on Toeplitz Matrices 61

5.1 Inverses of Toeplitz Matrices 62
5.2 Products of Toeplitz Matrices 67
5.3 Toeplitz Determinants 70

Chapter 6 Applications to Stochastic Time Series 73

6.1 Moving Average Processes 74
6.2 Autoregressive Processes 77
6.3 Factorization 80

Acknowledgements 83

References 85

Abstract

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

The fundamental theorems on the asymptotic behavior of eigenvalues, inverses, and products of banded Toeplitz matrices and Toeplitz matrices with absolutely summable elements are derived in a tutorial manner. Mathematical elegance and generality are sacrificed for conceptual simplicity and insight in the hope 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.

1

Introduction

1.1 Toeplitz and Circulant Matrices

A Toeplitz matrix is an $n \times n$ matrix $T_n = [t_{k,j};\ k,j = 0,1,\ldots,n-1]$ where $t_{k,j} = t_{k-j}$, i.e., a matrix of the form

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

Such matrices arise in many applications. For example, suppose that

$$x = (x_0, x_1, \ldots, x_{n-1})' = \begin{pmatrix} x_0 \\ x_1 \\ \vdots \\ x_{n-1} \end{pmatrix}$$

is a column vector (the prime denotes transpose) denoting an "input" and that t_k is zero for $k < 0$. Then the vector

$$\begin{aligned} y &= T_n x = \begin{bmatrix} t_0 & 0 & 0 & \cdots & 0 \\ t_1 & t_0 & 0 & & \\ t_2 & t_1 & t_0 & & \vdots \\ \vdots & & & \ddots & \\ t_{n-1} & & & \cdots & t_0 \end{bmatrix} \begin{pmatrix} x_0 \\ x_1 \\ x_2 \\ \vdots \\ x_{n-1} \end{pmatrix} \\ &= \begin{pmatrix} x_0 t_0 \\ t_1 x_0 + t_0 x_1 \\ \sum_{i=0}^{2} t_{2-i} x_i \\ \vdots \\ \sum_{i=0}^{n-1} t_{n-1-i} x_i \end{pmatrix} \end{aligned}$$

with entries

$$y_k = \sum_{i=0}^{k} t_{k-i} x_i$$

represents the the output of the discrete time causal time-invariant filter h with "impulse response" t_k. Equivalently, this is a matrix and vector formulation of a discrete-time convolution of a discrete time input with a discrete time filter.

As another example, suppose that $\{X_n\}$ is a discrete time random process with mean function given by the expectations $m_k = E(X_k)$ and covariance function given by the expectations $K_X(k,j) = E[(X_k - m_k)(X_j - m_j)]$. Signal processing theory such as prediction, estimation, detection, classification, regression, and communcations and information theory are most thoroughly developed under the assumption that the mean is constant and that the covariance is Toeplitz, i.e., $K_X(k,j) = K_X(k-j)$, in which case the process is said to be weakly stationary. (The terms "covariance stationary" and "second order stationary" also are used when the covariance is assumed to be Toeplitz.) In this case the $n \times n$ covariance matrices $K_n = [K_X(k,j);\ k,j = 0,1,\ldots,n-1]$ are Toeplitz matrices. Much of the theory of weakly stationary processes involves applications of

Toeplitz matrices. Toeplitz matrices also arise in solutions to differential and integral equations, spline functions, and problems and methods in physics, mathematics, statistics, and signal processing.

A common special case of Toeplitz matrices — which will result in significant simplification and play a fundamental role in developing more general results — results when every row of the matrix is a right cyclic shift of the row above it so that $t_k = t_{-(n-k)} = t_{k-n}$ for $k = 1, 2, \ldots, n-1$. In this case the picture becomes

$$C_n = \begin{bmatrix} t_0 & t_{-1} & t_{-2} & \cdots & t_{-(n-1)} \\ t_{-(n-1)} & t_0 & t_{-1} & & \\ t_{-(n-2)} & t_{-(n-1)} & t_0 & & \vdots \\ \vdots & & & \ddots & \\ t_{-1} & t_{-2} & & \cdots & t_0 \end{bmatrix}. \qquad (1.2)$$

A matrix of this form is called a *circulant* matrix. Circulant matrices arise, for example, in applications involving the discrete Fourier transform (DFT) and the study of cyclic codes for error correction.

A great deal is known about the behavior of Toeplitz matrices — the most common and complete references being Grenander and Szegö [16] and Widom [33]. A more recent text devoted to the subject is Böttcher and Silbermann [5]. Unfortunately, however, the necessary level of mathematical sophistication for understanding reference [16] 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 [16] are here made explicit.

The most famous and arguably the most important result describing Toeplitz matrices is Szegö's theorem for sequences of Toeplitz matrices $\{T_n\}$ which deals with the behavior of the eigenvalues as n goes to infinity. A complex scalar α is an eigenvalue of a matrix A if there is a nonzero vector x such that

$$Ax = \alpha x, \tag{1.3}$$

in which case we say that x is a (right) eigenvector of A. If A is Hermitian, that is, if $A^* = A$, where the asterisk denotes conjugate transpose, then the eigenvalues of the matrix are real and hence $\alpha^* = \alpha$, where the asterisk denotes the conjugate in the case of a complex scalar. When this is the case we assume that the eigenvalues $\{\alpha_i\}$ are ordered in a nondecreasing manner so that $\alpha_0 \geq \alpha_1 \geq \alpha_2 \cdots$. This eases the approximation of sums by integrals and entails no loss of generality. Szegö's theorem deals with the asymptotic behavior of the eigenvalues $\{\tau_{n,i};\ i = 0, 1, \ldots, n-1\}$ of a sequence of Hermitian Toeplitz matrices $T_n = [t_{k-j}; k, j = 0, 1, 2, \ldots, n-1]$. The theorem requires that several technical conditions be satisfied, including the existence of the Fourier series with coefficients t_k related to each other by

$$f(\lambda) = \sum_{k=-\infty}^{\infty} t_k e^{ik\lambda};\ \lambda \in [0, 2\pi] \tag{1.4}$$

$$t_k = \frac{1}{2\pi} \int_0^{2\pi} f(\lambda) e^{-ik\lambda}\, d\lambda. \tag{1.5}$$

Thus the sequence $\{t_k\}$ determines the function f and vice versa, hence the sequence of matrices is often denoted as $T_n(f)$. If $T_n(f)$ is Hermitian, that is, if $T_n(f)^* = T_n(f)$, then $t_{-k} = t_k^*$ and f is real-valued.

Under suitable assumptions the Szegö theorem states that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} F(\tau_{n,k}) = \frac{1}{2\pi} \int_0^{2\pi} F(f(\lambda))\, d\lambda \tag{1.6}$$

for any function F that is continuous on the range of f. Thus, for example, choosing $F(x) = x$ results in

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \tau_{n,k} = \frac{1}{2\pi} \int_0^{2\pi} f(\lambda)\, d\lambda, \tag{1.7}$$

so that the arithmetic mean of the eigenvalues of $T_n(f)$ converges to the integral of f. The trace $\operatorname{Tr}(A)$ of a matrix A is the sum of its diagonal elements, which in turn from linear algebra is the sum of the eigenvalues of A if the matrix A is Hermitian. Thus (1.7) implies that

$$\lim_{n\to\infty}\frac{1}{n}\mathrm{Tr}(T_n(f)) = \frac{1}{2\pi}\int_0^{2\pi} f(\lambda)\,d\lambda. \tag{1.8}$$

Similarly, for any power s

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}\tau_{n,k}^s = \frac{1}{2\pi}\int_0^{2\pi} f(\lambda)^s\,d\lambda. \tag{1.9}$$

If f is real and such that the eigenvalues $\tau_{n,k} \geq m > 0$ for all n, k, then $F(x) = \ln x$ is a continuous function on $[m, \infty)$ and the Szegö theorem can be applied to show that

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}\ln\tau_{n,i} = \frac{1}{2\pi}\int_0^{2\pi}\ln f(\lambda)\,d\lambda. \tag{1.10}$$

From linear algebra, however, the determinant of a matrix $T_n(f)$ is given by the product of its eigenvalues,

$$\det(T_n(f)) = \prod_{i=0}^{n-1}\tau_{n,i},$$

so that (1.10) becomes

$$\begin{aligned}\lim_{n\to\infty}\ln\det(T_n(f))^{1/n} &= \lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}\ln\tau_{n,i}\\ &= \frac{1}{2\pi}\int_0^{2\pi}\ln f(\lambda)\,d\lambda. \end{aligned}\tag{1.11}$$

As we shall later see, if f has a lower bound $m > 0$, than indeed all the eigenvalues will share the lower bound and the above derivation applies. Determinants of Toeplitz matrices are called *Toeplitz determinants* and (1.11) describes their limiting behavior.

1.2 Examples

A few examples from statistical signal processing and information theory illustrate the the application of the theorem. These are described with a minimum of background in order to highlight how the asymptotic eigenvalue distribution theorem allows one to evaluate results for processes using results from finite-dimensional vectors.

The differential entropy rate of a Gaussian process

Suppose that $\{X_n;\ n = 0, 1, \ldots\}$ is a random process described by probability density functions $f_{X^n}(x^n)$ for the random vectors $X^n = (X_0, X_1, \ldots, X_{n-1})$ defined for all $n = 0, 1, 2, \ldots$. The Shannon differential entropy $h(X^n)$ is defined by the integral

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

and the differential entropy rate of the random process 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[7].)

A stationary zero mean Gaussian random process is completely described by its mean correlation function $r_{k,j} = r_{k-j} = E[X_k X_j]$ or, equivalently, by its power spectral density function f, the Fourier transform of the covariance function:

$$\begin{aligned} f(\lambda) &= \sum_{n=-\infty}^{\infty} r_n e^{in\lambda}, \\ r_k &= \frac{1}{2\pi} \int_0^{2\pi} f(\lambda) e^{-i\lambda k}\, d\lambda \end{aligned}$$

For a fixed positive integer n, the probability density function is

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

where R_n is the $n \times n$ covariance matrix with entries r_{k-j}. A straightforward multidimensional integration using the properties of Gaussian random vectors yields the differential entropy

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

The problem at hand is to evaluate the entropy rate

$$h(X) = \lim_{n\to\infty} \frac{1}{n} h(X^n) = \frac{1}{2}\ln(2\pi e) + \lim_{n\to\infty}\frac{1}{n}\ln\det(R_n).$$

The matrix R_n is the Toeplitz matrix T_n generated by the power spectral density f and $\det(R_n)$ is a Toeplitz determinant and we have immediately from (1.11) that

$$h(X) = \frac{1}{2}\log\left(2\pi e\frac{1}{2\pi}\int_0^{2\pi}\ln f(\lambda)\,d\lambda\right). \qquad (1.12)$$

This is a typical use of (1.6) to evaluate the limit of a sequence of finite-dimensional qualities, in this case specified by the determinants of of a sequence of Toeplitz matrices.

The Shannon rate-distortion function of a Gaussian process

As a another example of the application of (1.6), consider the evaluation of the rate-distortion function of Shannon information theory for a stationary discrete time Gaussian random process with 0 mean, covariance $K_X(k,j) = t_{k-j}$, and power spectral density $f(\lambda)$ given by (1.4). The rate-distortion function characterizes the optimal tradeoff of distortion and bit rate in data compression or source coding systems. The derivation details can be found, e.g., in Berger [3], Section 4.5, but the point here is simply to provide an example of an application of (1.6). The result is found by solving an n-dimensional optimization in terms of the eigenvalues $\tau_{n,k}$ of $T_n(f)$ and then taking limits to obtain parametric expressions for distortion and rate:

$$\begin{aligned} D_\theta &= \lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}\min(\theta,\tau_{n,k}) \\ R_\theta &= \lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}\max(0,\frac{1}{2}\ln\frac{\tau_{n,k}}{\theta}). \end{aligned}$$

The theorem can be applied to turn this limiting sum involving eigenvalues into an integral involving the power spectral density:

$$\begin{aligned} D_\theta &= \int_0^{2\pi} \min(\theta, f(\lambda))\, d\lambda \\ R_\theta &= \int_0^{2\pi} \max\left(0, \frac{1}{2}\ln\frac{f(\lambda)}{\theta}\right) d\lambda. \end{aligned}$$

Again an infinite dimensional problem is solved by first solving a finite dimensional problem involving the eigenvalues of matrices, and then using the asymptotic eigenvalue theorem to find an integral expression for the limiting result.

One-step prediction error

Another application with a similar development is the one-step prediction error problem. Suppose that X_n is a weakly stationary random process with covariance t_{k-j}. A classic problem in estimation theory is to find the best linear predictor based on the previous n values of X_i, $i = 0, 1, 2, \ldots, n-1$,

$$\hat{X}_n = \sum_{i=1}^{n} a_i X_{n-i},$$

in the sense of minimizing the mean squared error $E[(X_n - \hat{X}_n)^2]$ over all choices of coefficients a_i. It is well known (see, e.g., [14]) that the minimum is given by the ratio of Toeplitz determinants $\det T_{n+1}/\det T_n$. The question is to what this ratio converges in the limit as n goes to ∞. This is not quite in a form suitable for application of the theorem, but we have already evaluated the limit of $\det T_n^{1/n}$ in (1.11) and for large n we have that

$$(\det T_n)^{1/n} \approx \exp\left(\frac{1}{2\pi}\int_0^{2\pi} \ln f(\lambda)\, d\lambda\right) \approx (\det T_{n+1})^{1/(n+1)}$$

and hence in particular that

$$(\det T_{n+1})^{1/(n+1)} \approx (\det T_n)^{1/n}$$

so that

$$\frac{\det T_{n+1}}{\det T_n} \approx (\det T_n)^{1/n} \to \exp\left(\frac{1}{2\pi}\int_0^{2\pi} \ln f(\lambda)\, d\lambda\right),$$

providing the desired limit. These arguments can be made exact, but it is hoped they make the point that the asymptotic eigenvalue distribution theorem for Hermitian Toeplitz matrices can be quite useful for evaluating limits of solutions to finite-dimensional problems.

Further examples

The Toeplitz distribution theorems have also found application in more complicated information theoretic evaluations, including the channel capacity of Gaussian channels [30, 29] and the rate-distortion functions of autoregressive sources [11]. The examples described here were chosen because they were in the author's area of competence, but similar applications crop up in a variety of areas. A Google$^{\mathrm{TM}}$ search using the title of this document shows diverse applications of the eigenvalue distribution theorem and related results, including such areas of coding, spectral estimation, watermarking, harmonic analysis, speech enhancement, interference cancellation, image restoration, sensor networks for detection, adaptive filtering, graphical models, noise reduction, and blind equalization.

1.3 Goals and Prerequisites

The primary goal of this work is to prove a special case of Szegö's asymptotic eigenvalue distribution theorem in Theorem 4.2. The assumptions used here are less general than Szegö's, but this permits more straightforward proofs which require far less mathematical background. In addition to the fundamental theorems, several related results that naturally follow but do not appear to be collected together anywhere are presented. We do not attempt to survey the fields of applications of these results, as such a survey would be far beyond the author's stamina and competence. A few applications are noted by way of examples.

The essential prerequisites are a knowledge of matrix theory, an engineer's knowledge of Fourier series and random processes, and calculus (Riemann integration). A first course in analysis would be helpful, but it is not assumed. 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, measure theory, functional analysis, and harmonic analysis are not used.

Our approach is to relate the properties of Toeplitz matrices to those of their simpler, more structured special case — 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 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 [16]. The basic results for the special case of a banded Toeplitz matrix appeared in [12], 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 [16].

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.

We sacrifice 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.

2

The Asymptotic Behavior of Matrices

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 major use of the theorems of this chapter is to relate the asymptotic behavior of a sequence of complicated matrices to that of a simpler asymptotically equivalent sequence of matrices.

2.1 Eigenvalues

Any complex matrix A can be written as

$$A = URU^*, \tag{2.1}$$

where the asterisk $*$ denotes conjugate transpose, U is unitary, i.e., $U^{-1} = U^*$, and $R = \{r_{k,j}\}$ is an upper triangular matrix ([18], p. 79). The eigenvalues of A are the principal diagonal elements of R. If A is normal, i.e., if $A^*A = AA^*$, then R is a diagonal matrix, which we denote as $R = \mathrm{diag}(\alpha_k; k = 0, 1, \ldots, n-1)$ or, more simply, $R = \mathrm{diag}(\alpha_k)$. If A is Hermitian, then it is also normal and its eigenvalues are real.

A matrix A is nonnegative definite if $x^*Ax \geq 0$ for all nonzero vectors x. The matrix is positive definite if the inequality is strict for all nonzero vectors x. (Some books refer to these properties as positive definite and strictly positive definite, respectively.) If a Hermitian matrix is nonnegative definite, then its eigenvalues are all nonnegative. If the matrix is positive definite, then the eigenvalues are all (strictly) positive.

The extreme values of the eigenvalues of a Hermitian matrix H can be characterized in terms of the Rayleigh quotient $R_H(x)$ of the matrix and a complex-valued vector x defined by

$$R_H(x) = (x^*Hx)/(x^*x). \tag{2.2}$$

As the result is both important and simple to prove, we state and prove it formally. The result will be useful in specifying the interval containing the eigenvalues of a Hermitian matrix.

Usually in books on matrix theory it is proved as a corollary to the variational description of eigenvalues given by the Courant-Fischer theorem (see, e.g., [18], p. 116, for the case of real symmetric matrices), but the following result is easily demonstrated directly.

Lemma 2.1. Given a Hermitian matrix H, let η_M and η_m be the maximum and minimum eigenvalues of H, respectively. Then

$$\eta_m = \min_x R_H(x) = \min_{z:z^*z=1} z^*Hz \tag{2.3}$$

$$\eta_M = \max_x R_H(x) = \max_{z:z^*z=1} z^*Hz. \tag{2.4}$$

Proof. Suppose that e_m and e_M are eigenvectors corresponding to the minimum and maximum eigenvalues η_m and η_M, respectively. Then $R_H(e_m) = \eta_m$ and $R_H(e_M) = \eta_M$ and therefore

$$\eta_m \geq \min_x R_H(x) \tag{2.5}$$

$$\eta_M \leq \max_x R_H(x). \tag{2.6}$$

Since H is Hermitian we can write $H = UAU^*$, where U is unitary and

A is the diagonal matrix of the eigenvalues η_k, and therefore

$$\begin{aligned}\frac{x^*Hx}{x^*x} &= \frac{x^*UAU^*x}{x^*x}\\ &= \frac{y^*Ay}{y^*y} = \frac{\sum_{k=1}^n |y_k|^2\eta_k}{\sum_{k=1}^n |y_k|^2},\end{aligned}$$

where $y = U^*x$ and we have taken advantage of the fact that U is unitary so that $x^*x = y^*y$. But for all vectors y, this ratio is bound below by η_m and above by η_M and hence for all vectors x

$$\eta_m \leq R_H(x) \leq \eta_M \tag{2.7}$$

which with (2.5–2.6) completes the proof of the left-hand equalities of the lemma. The right-hand equalities are easily seen to hold since if x minimizes (maximizes) the Rayleigh quotient, then the normalized vector x/x^*x satisfies the constraint of the minimization (maximization) to the right, hence the minimum (maximum) of the Rayleigh quotion must be bigger (smaller) than the constrained minimum (maximum) to the right. Conversely, if x achieves the rightmost optimization, then the same x yields a Rayleigh quotient of the the same optimum value. □

The following lemma is useful when studying non-Hermitian matrices and products of Hermitian matrices. First note that if A is an arbitrary complex matrix, then the matrix A^*A is both Hermitian and nonnegative definite. It is Hermitian because $(A^*A)^* = A^*A$ and it is nonnegative definite since if for any complex vector x we define the complex vector $y = Ax$, then

$$x^*(A^*A)x = y^*y = \sum_{k=1}^n |y_k|^2 \geq 0.$$

Lemma 2.2. Let A be a matrix with eigenvalues α_k. Define the eigenvalues of the Hermitian nonnegative definite matrix A^*A to be $\lambda_k \geq 0$. Then

$$\sum_{k=0}^{n-1} \lambda_k \geq \sum_{k=0}^{n-1} |\alpha_k|^2, \tag{2.8}$$

with equality iff (if and only if) A is 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.,

$$\mathrm{Tr}\{A^*A\} = \sum_{k=0}^{n-1} (A^*A)_{k,k} = \sum_{k=0}^{n-1} \lambda_k. \tag{2.9}$$

From (2.1), $A = URU^*$ and hence

$$\begin{aligned}\mathrm{Tr}\{A^*A\} &= \mathrm{Tr}\{R^*R\} = \sum_{k=0}^{n-1}\sum_{j=0}^{n-1} |r_{j,k}|^2 \\ &= \sum_{k=0}^{n-1} |\alpha_k|^2 + \sum_{k \neq j} |r_{j,k}|^2 \\ &\geq \sum_{k=0}^{n-1} |\alpha_k|^2 \end{aligned} \tag{2.10}$$

Equation (2.10) will hold with equality iff R is diagonal and hence iff A is normal. □

Lemma 2.2 is a direct consequence of Shur's theorem ([18], pp. 229-231) and is also proved in [16], p. 106.

2.2 Matrix Norms

To study the asymptotic equivalence of matrices we require a metric on the space of linear space of matrices. A convenient metric for our purposes is a norm of the difference of two matrices. A *norm* $N(A)$ on the space of $n \times n$ matrices satisfies the following properties:

(1) $N(A) \geq 0$ with equality if and only if $A = 0$, is the all zero matrix.
(2) For any two matrices A and B,

$$N(A + B) \leq N(A) + N(B). \tag{2.11}$$

(3) For any scalar c and matrix A, $N(cA) = |c|N(A)$.

The triangle inequality in (2.11) will be used often as is the following direct consequence:

$$N(A-B) \geq |N(A)-N(B)|. \tag{2.12}$$

Two norms — the operator or strong norm and the Hilbert-Schmidt or weak norm (also called the Frobenius norm or Euclidean norm when the scaling term is removed) — will be used here ([16], pp. 102–103).

Let A be a matrix with eigenvalues α_k and let $\lambda_k \geq 0$ be the eigenvalues of the Hermitian nonnegative definite matrix A^*A. The strong norm $\parallel A \parallel$ is defined by

$$\parallel A \parallel = \max_x R_{A^*A}(x)^{1/2} = \max_{z:z^*z=1} [z^*A^*Az]^{1/2}. \tag{2.13}$$

From Lemma 2.1

$$\parallel A \parallel^2 = \max_k \lambda_k \overset{\Delta}{=} \lambda_M. \tag{2.14}$$

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

$$\parallel A \parallel^2 = \max_{z:z^*z=1} z^*A^*Az \geq (e_M^*A^*)(Ae_M) = |\alpha_M|^2. \tag{2.15}$$

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

$$\parallel A \parallel = \max_k |\alpha_k| = |\alpha_M|. \tag{2.16}$$

The weak norm (or Hilbert-Schmidt norm) of an $n \times n$ matrix $A = [a_{k,j}]$ is defined by

$$\begin{aligned} |A| &= \left(\frac{1}{n} \sum_{k=0}^{n-1} \sum_{j=0}^{n-1} |a_{k,j}|^2 \right)^{1/2} \\ &= (\frac{1}{n} \mathrm{Tr}[A^*A])^{1/2} = \left(\frac{1}{n} \sum_{k=0}^{n-1} \lambda_k \right)^{1/2}. \end{aligned} \tag{2.17}$$

The quantity $\sqrt{n}|A|$ is sometimes called the Frobenius norm or Euclidean norm. From Lemma 2.2 we have

$$|A|^2 \geq \frac{1}{n}\sum_{k=0}^{n-1} |\alpha_k|^2, \text{with equality iff } A \text{ is normal.} \tag{2.18}$$

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

$$\| A \|^2 = \max_k \lambda_k \geq \frac{1}{n}\sum_{k=0}^{n-1} \lambda_k = |A|^2. \tag{2.19}$$

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

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

Lemma 2.3. Given two $n \times n$ matrices $G = \{g_{k,j}\}$ and $H = \{h_{k,j}\}$, then

$$|GH| \leq \| G \| \, |H|. \tag{2.20}$$

Proof. Expanding terms yields

$$\begin{aligned} |GH|^2 &= \frac{1}{n}\sum_i \sum_j |\sum_k g_{i,k} h_{k,j}|^2 \\ &= \frac{1}{n}\sum_i \sum_j \sum_k \sum_m g_{i,k} g_{i,m}^* h_{k,j} h_{m,j}^* \\ &= \frac{1}{n}\sum_j h_j^* G^* G h_j, \end{aligned} \tag{2.21}$$

where h_j is the j^{th} column of H. From (2.13),

$$\frac{h_j^* G^* G h_j}{h_j^* h_j} \leq \| G \|^2$$

and therefore

$$|GH|^2 \leq \frac{1}{n} \| G \|^2 \sum_j h_j^* h_j = \| G \|^2 \, |H|^2.$$

□

Lemma 2.3 is the matrix equivalent of (7.3a) of ([16], p. 103). Note that the lemma does not require that G or H be Hermitian.

2.3 Asymptotically Equivalent Sequences of Matrices

We will be considering sequences of $n \times n$ matrices that approximate each other as n becomes 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:

$$\parallel A_n \parallel, \parallel B_n \parallel \leq M < \infty, n = 1, 2, \ldots \tag{2.22}$$

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 the sequences $\{A_n\}$ and $\{B_n\}$ will be abbreviated $A_n \sim B_n$.

We can immediately prove several properties of asymptotic equivalence which are collected in the following theorem.

Theorem 2.1. Let $\{A_n\}$ and $\{B_n\}$ be sequences of matrices with eigenvalues $\{\alpha_n, i\}$ and $\{\beta_n, i\}$, respectively.

(1) If $A_n \sim B_n$, then

$$\lim_{n\to\infty} |A_n| = \lim_{n\to\infty} |B_n|. \tag{2.23}$$

(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 $A_nC_n \sim B_nD_n$.

(4) If $A_n \sim B_n$ and $\parallel A_n^{-1} \parallel, \parallel B_n^{-1} \parallel \leq K < \infty$, all n, then $A_n^{-1} \sim B_n^{-1}$.

(5) If $A_nB_n \sim C_n$ and $\parallel A_n^{-1} \parallel \leq K < \infty$, then $B_n \sim A_n^{-1}C_n$.

(6) If $A_n \sim B_n$, then there are finite constants m and M such that

$$m \leq \alpha_{n,k}, \beta_{n,k} \leq M \ , \quad n = 1, 2, \ldots \quad k = 0, 1, \ldots, n-1. \tag{2.24}$$

Proof.

(1) Eq. (2.23) follows directly from (2.12).

(2) $|A_n - C_n| = |A_n - B_n + B_n - C_n| \le |A_n - B_n| + |B_n - C_n| \overrightarrow{{}_{n\to\infty}} 0$

(3) Applying Lemma 2.3 yields

$$\begin{aligned}|A_nC_n - B_nD_n| &= |A_nC_n - A_nD_n + A_nD_n - B_nD_n| \\ &\le \parallel A_n \parallel |C_n - D_n| + \parallel D_n \parallel |A_n - B_n| \\ &\overrightarrow{{}_{n\to\infty}} 0.\end{aligned}$$

(4)

$$\begin{aligned}|A_n^{-1} - B_n^{-1}| &= |B_n^{-1}B_nA_n^{-1} - B_n^{-1}A_nA_n^{-1}| \\ &\le \parallel B_n^{-1} \parallel \cdot \parallel A_n^{-1} \parallel \cdot |B_n - A_n| \\ &\overrightarrow{{}_{n\to\infty}} 0.\end{aligned}$$

(5)

$$\begin{aligned}B_n - A_n^{-1}C_n &= A_n^{-1}A_nB_n - A_n^{-1}C_n \\ &\le \parallel A_n^{-1} \parallel |A_nB_n - C_n| \\ &\overrightarrow{{}_{n\to\infty}} 0.\end{aligned}$$

(6) If $A_n \sim B_n$ then they are uniformly bounded in strong norm by some finite number M and hence from (2.15), $|\alpha_{n,k}| \le M$ and $|\beta_{n,k}| \le M$ and hence $-M \le \alpha_{n,k}, \beta_{n,k} \le M$. So the result holds for $m = -M$ and it may hold for larger m, e.g., $m = 0$ if the matrices are all nonnegative definite.

□

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 results to follow. It shows that if the weak norm of the difference of the two matrices is small, then the sums of the eigenvalues of each must be close.

Lemma 2.4. Given two matrices A and B with eigenvalues $\{\alpha_k\}$ and $\{\beta_k\}$, respectively, then

$$|\frac{1}{n}\sum_{k=0}^{n-1}\alpha_k - \frac{1}{n}\sum_{k=0}^{n-1}\beta_k| \le |A-B|.$$

Proof: Define the difference matrix $D = A - B = \{d_{k,j}\}$ so that

$$\begin{aligned}\sum_{k=0}^{n-1}\alpha_k - \sum_{k=0}^{n-1}\beta_k &= \mathrm{Tr}(A) - \mathrm{Tr}(B)\\ &= \mathrm{Tr}(D).\end{aligned}$$

Applying the Cauchy-Schwarz inequality (see, e.g., [22], p. 17) to $\mathrm{Tr}(D)$ yields

$$\begin{aligned}|\mathrm{Tr}(D)|^2 &= \left|\sum_{k=0}^{n-1} d_{k,k}\right|^2 \le n\sum_{k=0}^{n-1}|d_{k,k}|^2\\ &\le n\sum_{k=0}^{n-1}\sum_{j=0}^{n-1}|d_{k,j}|^2 = n^2|D|^2. \qquad (2.25)\end{aligned}$$

Taking the square root and dividing by n proves the lemma. □

An immediate consequence of the lemma is the following corollary.

Corollary 2.1. Given two sequences of asymptotically equivalent matrices $\{A_n\}$ and $\{B_n\}$ with eigenvalues $\{\alpha_{n,k}\}$ and $\{\beta_{n,k}\}$, respectively, then

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}(\alpha_{n,k} - \beta_{n,k}) = 0, \qquad (2.26)$$

and hence if either limit exists individually,

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}\alpha_{n,k} = \lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}\beta_{n,k}. \qquad (2.27)$$

Proof. Let $D_n = \{d_{k,j}\} = A_n - B_n$. Eq. (2.27) is equivalent to

$$\lim_{n\to\infty}\frac{1}{n}\mathrm{Tr}(D_n) = 0. \qquad (2.28)$$

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

$$0 \leq |\frac{1}{n}\mathrm{Tr}(D_n)|^2 \leq |D_n|^2 \underset{n\to\infty}{\longrightarrow} 0 \tag{2.29}$$

from the lemma, which implies (2.28) and hence (2.27). □

The previous corollary can be interpreted as saying the sample or arithmetic means of the eigenvalues of two matrices are asymptotically equal if the matrices are asymptotically equivalent. It is easy to see that if the matrices are Hermitian, a similar result holds for the means of the squared eigenvalues. From (2.12) and (2.18),

$$\begin{aligned} |D_n| &\geq \big|\,|A_n| - |B_n|\,\big| \\ &= \left|\sqrt{\frac{1}{n}\sum_{k=0}^{n-1}\alpha_{n,k}^2} - \sqrt{\frac{1}{n}\sum_{k=0}^{n-1}\beta_{n,k}^2}\right| \\ &\underset{n\to\infty}{\longrightarrow} 0 \end{aligned}$$

if $|D_n| \underset{n\to\infty}{\longrightarrow} 0$, yielding the following corollary.

Corollary 2.2. Given two sequences of asymptotically equivalent Hermitian matrices $\{A_n\}$ and $\{B_n\}$ with eigenvalues $\{\alpha_{n,k}\}$ and $\{\beta_{n,k}\}$, respectively, then

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}(\alpha_{n,k}^2 - \beta_{n,k}^2) = 0, \tag{2.30}$$

and hence if either limit exists individually,

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}\alpha_{n,k}^2 = \lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}\beta_{n,k}^2. \tag{2.31}$$

Both corollaries relate limiting sample (arithmetic) averages of eigenvalues or moments of an eigenvalue distribution rather than individual eigenvalues. Equations (2.27) and (2.31) 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. Then for any positive integer s the sequences of matrices $\{A_n^s\}$ and $\{B_n^s\}$ are also asymptotically equivalent,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} (\alpha_{n,k}^s - \beta_{n,k}^s) = 0, \tag{2.32}$$

and hence if either separate limit exists,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \alpha_{n,k}^s = \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \beta_{n,k}^s. \tag{2.33}$$

Proof. Let $A_n = B_n + D_n$ as in the proof of Corollary 2.1 and consider $A_n^s - B_n^s \overset{\Delta}{=} \Delta_n$. Since the eigenvalues of A_n^s are $\alpha_{n,k}^s$, (2.32) can be written in terms of Δ_n as

$$\lim_{n\to\infty} \frac{1}{n} \mathrm{Tr}(\Delta_n) = 0. \tag{2.34}$$

The matrix Δ_n is a sum of several terms each being a product of D_n's and B_n's, but containing at least one D_n (to see this use the binomial theorem applied to matrices to expand A_n^s). Repeated application of Lemma 2.3 thus gives

$$|\Delta_n| \le K|D_n| \underset{n\to\infty}{\longrightarrow} 0, \tag{2.35}$$

where K does not depend on n. Equation (2.35) allows us to apply Corollary 2.1 to the matrices A_n^s and D_n^s to obtain (2.34) and hence (2.32). □

Theorem 2.2 is the fundamental theorem concerning asymptotic eigenvalue behavior of asymptotically equivalent sequences of matrices. Most of the succeeding results on eigenvalues will be applications or specializations of (2.33).

Since (2.33) holds for any positive integer s we can add sums corresponding to different values of s to each side of (2.33). This observation leads to the following corollary.

Corollary 2.3. Suppose that $\{A_n\}$ and $\{B_n\}$ are asymptotically equivalent sequences of matrices with eigenvalues $\{\alpha_{n,k}\}$ and $\{\beta_{n,k}\}$, respectively, and let $f(x)$ be any polynomial. Then

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} (f(\alpha_{n,k}) - f(\beta_{n,k})) = 0 \tag{2.36}$$

and hence if either limit exists separately,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(\alpha_{n,k}) = \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(\beta_{n,k}) . \tag{2.37}$$

Proof. Suppose that $f(x) = \sum_{s=0}^{m} a_s x^s$. Then summing (2.32) over s yields (2.36). If either of the two limits exists, then (2.36) implies that both exist and that they are equal. □

Corollary 2.3 can be used to show that (2.37) can hold for any analytic function $f(x)$ since such functions can be expanded into complex Taylor series, which can be viewed as polynomials with a possibly infinite number of terms. Some effort is needed, however, to justify the interchange of limits, which can be accomplished if the Taylor series converges uniformly. 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 Weierstrass approximation theorem ([6], p. 66) to immediately generalize Corollary 2.3. This theorem, our one real excursion into analysis, is stated below for reference.

Theorem 2.3. (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 and uniformly by a polynomial. Applying Theorem 2.3 to Corollary 2.3 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. From Theorem 2.1 there exist finite numbers m and M such that

$$m \le \alpha_{n,k}, \beta_{n,k} \le M \ , \quad n = 1, 2, \ldots \quad k = 0, 1, \ldots, n-1. \tag{2.38}$$

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

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} (F(\alpha_{n,k}) - F(\beta_{n,k})) = 0, \tag{2.39}$$

and hence if either of the limits exists separately,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} F(\alpha_{n,k}) = \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} F(\beta_{n,k}) \tag{2.40}$$

Theorem 2.4 is the matrix equivalent of Theorem 7.4a of [16]. When two real sequences $\{\alpha_{n,k}; k = 0, 1, \ldots, n-1\}$ and $\{\beta_{n,k}; k = 0, 1, \ldots, n-1\}$ satisfy (2.38) and (2.39), they are said to be *asymptotically equally distributed* ([16], p. 62, where the definition is attributed to Weyl).

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

Corollary 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, such that $\alpha_{n,k}, \beta_{n,k} \ge m > 0$. Then if either limit exists,

$$\lim_{n\to\infty} (\det A_n)^{1/n} = \lim_{n\to\infty} (\det B_n)^{1/n}. \tag{2.41}$$

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

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

and hence

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

or equivalently

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

from which (2.41) follows. □

With suitable mathematical care the above corollary can be extended to cases where $\alpha_{n,k}, \beta_{n,k} > 0$ provided additional constraints are imposed on the matrices. For example, if the matrices are assumed to be Toeplitz matrices, then the result holds even if the eigenvalues can get arbitrarily small but remain strictly positive. (See the discussion on p. 66 and in Section 3.1 of [16] for the required technical conditions.) The difficulty with allowing the eigenvalues to approach 0 is that their logarithms are not bounded. Furthermore, the function $\ln x$ is not continuous at $x = 0$, so Theorem 2.4 does not apply. Nonetheless, it is possible to say something about the asymptotic eigenvalue distribution in such cases and this issue is revisited in Theorem 5.2(d).

In this section the concept of asymptotic equivalence of matrices was defined and its implications studied. The main consequences are 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 $|A_n^{-1} - B_n^{-1}| \underset{n\to\infty}{\longrightarrow} 0$ 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 [1, 2]. Baxter's results are beyond the scope of this work.

2.4 Asymptotically Absolutely Equal Distributions

It is possible to strengthen Theorem 2.4 and some of the interim results used in its derivation using reasonably elementary methods. The key additional idea required is the Wielandt-Hoffman theorem [34], a result from matrix theory that is of independent interest. The theorem is stated and a proof following Wilkinson [35] is presented for completeness. This section can be skipped by readers not interested in the stronger notion of equal eigenvalue distributions as it is not needed in the sequel. The bounds of Lemmas 2.5 and 2.5 are of interest in their own right and are included as they strengthen the the traditional bounds.

Theorem 2.5. (Wielandt-Hoffman theorem) Given two Hermitian matrices A and B with eigenvalues α_k and β_k, respectively, then

$$\frac{1}{n}\sum_{k=0}^{n-1}|\alpha_k-\beta_k|^2 \leq |A-B|^2.$$

Proof: Since A and B are Hermitian, we can write them as $A = U\text{diag}(\alpha_k)U^*$, $B = W\text{diag}(\beta_k)W^*$, where U and W are unitary. Since the weak norm is not effected by multiplication by a unitary matrix,

$$\begin{aligned}
|A-B| &= |U\text{diag}(\alpha_k)U^* - W\text{diag}(\beta_k)W^*| \\
&= |\text{diag}(\alpha_k)U^* - U^*W\text{diag}(\beta_k)W^*| \\
&= |\text{diag}(\alpha_k)U^*W - U^*W\text{diag}(\beta_k)| \\
&= |\text{diag}(\alpha_k)Q - Q\text{diag}(\beta_k)|,
\end{aligned}$$

where $Q = U^*W = \{q_{i,j}\}$ is also unitary. The (i,j) entry in the matrix $\text{diag}(\alpha_k)Q - Q\text{diag}(\beta_k)$ is $(\alpha_i - \beta_j)q_{i,j}$ and hence

$$|A-B|^2 = \frac{1}{n}\sum_{i=0}^{n-1}\sum_{j=0}^{n-1}|\alpha_i-\beta_j|^2|q_{i,j}|^2 \stackrel{\Delta}{=} \sum_{i=0}^{n-1}\sum_{j=0}^{n-1}|\alpha_i-\beta_j|^2 p_{i,j} \tag{2.42}$$

where we have defined $p_{i,j} = (1/n)|q_{i,j}|^2$. Since Q is unitary, we also have that

$$\sum_{i=0}^{n-1}|q_{i,j}|^2 = \sum_{j=0}^{n-1}|q_{i,j}|^2 = 1 \tag{2.43}$$

or

$$\sum_{i=0}^{n-1}p_{i,j} = \sum_{j=0}^{n-1}p_{i,j} = \frac{1}{n}. \tag{2.44}$$

This can be interpreted in probability terms: $p_{i,j} = (1/n)|q_{i,j}|^2$ is a probability mass function or pmf on $\{0,1,\ldots,n-1\}^2$ with uniform marginal probability mass functions. Recall that it is assumed that the eigenvalues are ordered so that $\alpha_0 \geq \alpha_1 \geq \alpha_2 \geq \cdots$ and $\beta_0 \geq \beta_1 \geq \beta_2 \geq \cdots$.

We claim that for all such matrices P satisfying (2.44), the right-hand side of (2.42) is minimized by $P = (1/n)I$, where I is the identity matrix, so that

$$\sum_{i=0}^{n-1}\sum_{j=0}^{n-1} |\alpha_i - \beta_j|^2 p_{i,j} \geq \sum_{i=0}^{n-1} |\alpha_i - \beta_i|^2,$$

which will prove the result. To see this suppose the contrary. Let ℓ be the smallest integer in $\{0, 1, \ldots, n-1\}$ such that P has a nonzero element off the diagonal in either row ℓ or in column ℓ. If there is a nonzero element in row ℓ off the diagonal, say $p_{\ell,a}$ then there must also be a nonzero element in column ℓ off the diagonal, say $p_{b,\ell}$ in order for the constraints (2.44) to be satisfied. Since ℓ is the smallest such value, $\ell < a$ and $\ell < b$. Let x be the smaller of $p_{l,a}$ and $p_{b,l}$. Form a new matrix P' by adding x to $p_{\ell,\ell}$ and $p_{b,a}$ and subtracting x from $p_{b,\ell}$ and $p_{\ell,a}$. The new matrix still satisfies the constraints and it has a zero in either position (b, ℓ) or (ℓ, a). Furthermore the norm of P' has changed from that of P by an amount

$$x\left((\alpha_\ell - \beta_\ell)^2 + (\alpha_b - \beta_a)^2 - (\alpha_\ell - \beta_a)^2 - (\alpha_b - \beta_\ell)^2\right)$$
$$= -x(\alpha_\ell - \alpha_b)(\beta_\ell - \beta_a) \leq 0$$

since $\ell > b$, $\ell > a$, the eigenvalues are nonincreasing, and x is positive. Continuing in this fashion all nonzero offdiagonal elements can be zeroed out without increasing the norm, proving the result. □

From the Cauchy-Schwarz inequality

$$\sum_{k=0}^{n-1} |\alpha_k - \beta_k| \leq \sqrt{\sum_{k=0}^{n-1}(\alpha_k - \beta_k)^2}\sqrt{\sum_{k=0}^{n-1} 1^2} = \sqrt{n\sum_{k=0}^{n-1}(\alpha_k - \beta_k)^2},$$

which with the Wielandt-Hoffman theorem yields the following strengthening of Lemma 2.4,

$$\frac{1}{n}\sum_{k=0}^{n-1} |\alpha_k - \beta_k| \leq \sqrt{\frac{1}{n}\sum_{k=0}^{n-1}(\alpha_k - \beta_k)^2} \leq |A_n - B_n|,$$

which we formalize as the following lemma.

Lemma 2.5. Given two Hermitian matrices A and B with eigenvalues α_n and β_n in nonincreasing order, respectively, then

$$\frac{1}{n}\sum_{k=0}^{n-1}|\alpha_k - \beta_k| \le |A - B|.$$

Note in particular that the absolute values are outside the sum in Lemma 2.4 and inside the sum in Lemma 2.5. As was done in the weaker case, the result can be used to prove a stronger version of Theorem 2.4. This line of reasoning, using the Wielandt-Hoffman theorem, was pointed out by William F. Trench who used special cases in his paper [23]. Similar arguments have become standard for treating eigenvalue distributions for Toeplitz and Hankel matrices. See, for example, [32, 9, 4]. The following theorem provides the derivation. The specific statement result and its proof follow from a private communication from William F. Trench. See also [31, 24, 25, 26, 27, 28].

Theorem 2.6. Let A_n and B_n be asymptotically equivalent sequences of Hermitian matrices with eigenvalues $\alpha_{n,k}$ and $\beta_{n,k}$ in nonincreasing order, respectively. From Theorem 2.1 there exist finite numbers m and M such that

$$m \le \alpha_{n,k}, \beta_{n,k} \le M \ , \quad n = 1, 2, \ldots \quad k = 0, 1, \ldots, n-1. \tag{2.45}$$

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

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}|F(\alpha_{n,k}) - F(\beta_{n,k})| = 0. \tag{2.46}$$

The theorem strengthens the result of Theorem 2.4 because of the magnitude inside the sum. Following Trench [24] in this case the eigenvalues are said to be *asymptotically absolutely equally distributed.*
Proof: From Lemma 2.5

$$\frac{1}{n}\sum_{k=0}|\alpha_{n,k} - \beta_{n,k}| \le |A_n - B_n|, \tag{2.47}$$

which implies (2.46) for the case $F(r) = r$. For any nonnegative integer j

$$|\alpha_{n,k}^j - \beta_{n,k}^j| \le j \max(|m|, |M|)^{j-1} |\alpha_{n,k} - \beta_{n,k}|. \tag{2.48}$$

By way of explanation consider $a, b \in [m, M]$. Simple long division shows that

$$\frac{a^j - b^j}{a - b} = \sum_{l=1}^{j} a^{j-l} b^{l-1}$$

so that

$$\begin{aligned} |\frac{a^j - b^j}{a - b}| &= \frac{|a^j - b^j|}{|a - b|} \\ &= |\sum_{l=1}^{j} a^{j-l} b^{l-1}| \\ &\le \sum_{l=1}^{j} |a^{j-l} b^{l-1}| \\ &= \sum_{l=1}^{j} |a|^{j-l} |b|^{l-1} \\ &\le j \max(|m|, |M|)^{j-1}, \end{aligned}$$

which proves (2.48). This immediately implies that (2.46) holds for functions of the form $F(r) = r^j$ for positive integers j, which in turn means the result holds for any polynomial. If F is an arbitrary continuous function on $[m, M]$, then from Theorem 2.3 given $\epsilon > 0$ there is a polynomial P such that

$$|P(u) - F(u)| \le \epsilon, u \in [m, M].$$

Using the triangle inequality,

$$
\begin{aligned}
&\frac{1}{n}\sum_{k=0}^{n-1}|F(\alpha_{n,k})-F(\beta_{n,k})| \\
&\qquad = \frac{1}{n}\sum_{k=0}^{n-1}|F(\alpha_{n,k})-P(\alpha_{n,k})+P(\alpha_{n,k})-P(\beta_{n,k})+P(\beta_{n,k})-F(\beta_{n,k})| \\
&\qquad \le \frac{1}{n}\sum_{k=0}^{n-1}|F(\alpha_{n,k})-P(\alpha_{n,k})|+\frac{1}{n}\sum_{k=0}^{n-1}|P(\alpha_{n,k})-P(\beta_{n,k})| \\
&\qquad\quad +\frac{1}{n}\sum_{k=0}^{n-1}|P(\beta_{n,k})-F(\beta_{n,k})| \\
&\qquad \le 2\epsilon+\frac{1}{n}\sum_{k=0}^{n-1}|P(\alpha_{n,k})-P(\beta_{n,k})|
\end{aligned}
$$

As $n \to \infty$ the remaining sum goes to 0, which proves the theorem since ϵ can be made arbitrarily small. □

3

Circulant Matrices

A circulant matrix C is a Toeplitz matrix 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 structure can also be characterized by noting that the (k, j) entry of C, $C_{k,j}$, is given by

$$C_{k,j} = c_{(j-k) \bmod n}.$$

The properties of circulant matrices are well known and easily derived ([18], p. 267,[8]). 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.

3.1 Eigenvalues and Eigenvectors

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 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 prime denotes transpose and the normalization is chosen to give the eigenvector unit energy. Choosing ρ_m as the complex n^{th} root of unity, $\rho_m = e^{-2\pi i m/n}$, we have eigenvalue

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

and eigenvector

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

Thus from the definition of eigenvalues and eigenvectors,

$$Cy^{(m)} = \psi_m y^{(m)}, m = 0, 1, \ldots, n-1. \tag{3.8}$$

Equation (3.7) should be familiar to those with standard engineering backgrounds as simply the discrete Fourier transform (DFT) of the sequence $\{c_k\}$. Thus we can recover the sequence $\{c_k\}$ from the ψ_k by the Fourier inversion formula. In particular,

$$\begin{aligned}
\frac{1}{n}\sum_{m=0}^{n-1}\psi_m e^{2\pi i\ell m} &= \frac{1}{n}\sum_{m=0}^{n-1}\sum_{k=0}^{n-1}\left(c_k e^{-2\pi imk/n}\right)e^{2\pi i\ell m/n} \\
&= \sum_{k=0}^{n-1} c_k \frac{1}{n}\sum_{m=0}^{n-1} e^{2\pi i(\ell-k)m/n} = c_\ell, \qquad (3.9)
\end{aligned}$$

where we have used the orthogonality of the complex exponentials:

$$\sum_{m=0}^{n-1} e^{2\pi imk/n} = n\delta_{k \bmod n} = \begin{cases} n & k \bmod n = 0 \\ 0 & \text{otherwise} \end{cases}, \tag{3.10}$$

where δ is the Kronecker delta,

$$\delta_m = \begin{cases} 1 & m = 0 \\ 0 & \text{otherwise} \end{cases}.$$

Thus the eigenvalues of a circulant matrix comprise the DFT of the first row of the circulant matrix, and conversely first row of a circulant matrix is the inverse DFT of the eigenvalues.

Eq. (3.8) can be written as a single matrix equation

$$CU = U\Psi, \tag{3.11}$$

where

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

is the matrix composed of the eigenvectors as columns, and $\Psi = \operatorname{diag}(\psi_k)$ is the diagonal matrix with diagonal elements $\psi_0, \psi_1, \ldots, \psi_{n-1}$. Furthermore, (3.10) implies that U is unitary. By way of details, denote that the $(k,j)^{\text{th}}$ element of UU^* by $a_{k,j}$ and observe that $a_{k,j}$ will be the product of the kth row of U, which is $\{e^{-2\pi imk/n}/\sqrt{n};\ m = 0,1,\ldots,n-1\}$, times the jth column of U^*, which is $\{e^{2\pi imj/n}/\sqrt{n};\ m = 0,1,\ldots,n-1\}$ so that

$$a_{k,j} = \frac{1}{n}\sum_{m=0}^{n-1} e^{2\pi im(j-k)/n} = \delta_{(k-j) \bmod n}$$

and hence $UU^* = I$. Similarly, $U^*U = I$. Thus (3.11) implies that

$$C = U\Psi U^* \tag{3.12}$$

$$\Psi = U^*CU. \tag{3.13}$$

Since C is unitarily similar to a diagonal matrix it is normal.

3.2 Matrix Operations on Circulant Matrices

The following theorem summarizes the properties derived in the previous section regarding eigenvalues and eigenvectors of circulant matrices and provides some easy implications.

Theorem 3.1. Every circulant matrix C has eigenvectors $y^{(m)} = \frac{1}{\sqrt{n}}\left(1, e^{-2\pi im/n}, \cdots, e^{-2\pi im(n-1)/n}\right)'$, $m = 0,1,\ldots,n-1$, and corresponding eigenvalues

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

and can be expressed in the form $C = U\Psi U^*$, where U has the eigenvectors as columns in order and Ψ is $\operatorname{diag}(\psi_k)$. In particular all circulant matrices share the same eigenvectors, the same matrix U works for all circulant matrices, and any matrix of the form $C = U\Psi U^*$ is circulant.

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

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

respectively. Then

1. C and B commute and

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

where $\gamma = \mathrm{diag}(\psi_m\beta_m)$, 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

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

Proof. We have $C = U\Psi U^*$ and $B = U\Phi U^*$ where $\Psi = \mathrm{diag}(\psi_m)$ and $\Phi = \mathrm{diag}(\beta_m)$.

1. $CB = U\Psi U^* U\Phi U^* = U\Psi\Phi U^* = U\Phi\Psi U^* = BC$. Since $\Psi\Phi$ is diagonal, the first part of the theorem implies that CB is circulant.
2. $C + B = U(\Psi + \Phi)U^*$.
3. If Ψ is nonsingular, then

$$\begin{aligned} CU\Psi^{-1}U^* &= U\Psi U^* U\Psi^{-1}U^* = U\Psi\Psi^{-1}U^* \\ &= UU^* = I. \end{aligned}$$

□

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

We shall see that suitably chosen sequences of circulant matrices asymptotically approximate sequences of Toeplitz matrices and hence results similar to those in Theorem 3.1 will hold asymptotically for sequences of Toeplitz matrices.

4

Toeplitz Matrices

4.1 Sequences of Toeplitz Matrices

Given the simplicity of sums, products, eigenvalues,, inverses, and determinants of circulant matrices, an obvious approach to the study of asymptotic properties of sequences of Toeplitz matrices is to approximate them by sequences asymptotically equivalent of circulant matrices and then applying the results developed thus far. Such results are most easily derived when strong assumptions are placed on the sequence of Toeplitz matrices which keep the structure of the matrices simple and allow them to be well approximated by a natural and simple sequence of related circulant matrices. Increasingly general results require corresponding increasingly complicated constructions and proofs.

Consider the infinite sequence $\{t_k\}$ and define the corresponding sequence of $n \times n$ Toeplitz matrices $T_n = [t_{k-j};\ k, j = 0, 1, \ldots, n-1]$ as in (1.1). Toeplitz matrices can be classified by the restrictions placed on the sequence t_k. The simplest class results if there is a finite m for which $t_k = 0$, $|k| > m$, in which case T_n is said to be a *banded* Toeplitz matrix. A banded Toeplitz matrix has the appearance of the of (4.1), possessing a finite number of diagonals with nonzero entries and zeros everywhere else, so that the nonzero entries lie within a “band” including the main diagonal:

$$T_n = \left[\begin{array}{ccccccccccc}
t_0 & t_{-1} & \cdots & t_{-m} & & & & & & & \\
t_1 & t_0 & & & & & & & & & \\
\vdots & & & & & & & 0 & & & \\
 & & \ddots & & & & \ddots & & & & \\
t_m & & & & & & & & & & \\
 & \ddots & & & & & & & & & \\
 & & t_m & \cdots & t_1 & t_0 & t_{-1} & \cdots & t_{-m} & & \\
 & & & & & & & & & \ddots & \\
 & & & & \ddots & & & \ddots & & & t_{-m} \\
 & & & & & & & & & & \vdots \\
 & 0 & & & & & & & & t_0 & t_{-1} \\
 & & & & & & & t_m & \cdots & t_1 & t_0
\end{array}\right]. \tag{4.1}$$

In the more general case where the t_k are not assumed to be zero for large k, there are two common constraints placed on the infinite sequence $\{t_k;\ k = \ldots, -2, -1, 0, 1, 2, \ldots\}$ which defines all of the matrices T_n in the sequence. The most general is to assume that the t_k are square summable, i.e., that

$$\sum_{k=-\infty}^{\infty} |t_k|^2 < \infty. \tag{4.2}$$

Unfortunately this case requires mathematical machinery beyond that assumed here; 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., that

$$\sum_{k=-\infty}^{\infty} |t_k| < \infty. \tag{4.3}$$

Note that (4.3) is indeed a stronger constraint than (4.2) since

$$\sum_{k=-\infty}^{\infty} |t_k|^2 \le \left\{ \sum_{k=-\infty}^{\infty} |t_k| \right\}^2. \tag{4.4}$$

The assumption of absolute summability greatly simplifies the mathematics, but does not alter the fundamental concepts of Toeplitz and circulant matrices involved. As the main purpose here is tutorial and we wish chiefly to relay the flavor and an intuitive feel for the results, we will confine interest to the absolutely summable case. The main advantage of (4.3) over (4.2) is that it ensures the existence and of the Fourier series $f(\lambda)$ defined by

$$f(\lambda) = \sum_{k=-\infty}^{\infty} t_k e^{ik\lambda} = \lim_{n\to\infty} \sum_{k=-n}^{n} t_k e^{ik\lambda}. \tag{4.5}$$

Not only does the limit in (4.5) converge if (4.3) 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 right-hand side does not depend on λ and it goes to zero as $n \to \infty$ from (4.3). Thus given ϵ there is a single N, not depending on λ, such that

$$\left| f(\lambda) - \sum_{k=-n}^{n} t_k e^{ik\lambda} \right| \le \epsilon \ , \ \text{all } \lambda \in [0, 2\pi] \ , \ \text{if } n \ge N. \tag{4.6}$$

Furthermore, if (4.3) holds, then $f(\lambda)$ is Riemann integrable and the t_k can be recovered from f from the ordinary Fourier inversion formula:

$$t_k = \frac{1}{2\pi} \int_0^{2\pi} f(\lambda) e^{-ik\lambda}\, d\lambda. \tag{4.7}$$

As a final useful property of this case, $f(\lambda)$ is a continuous function of $\lambda \in [0, 2\pi]$ except possibly at a countable number of points.

A sequence of Toeplitz matrices $T_n = [t_{k-j}]$ for which the t_k are absolutely summable is said to be in the *Wiener class*,. Similarly, a function $f(\lambda)$ defined on $[0, 2\pi]$ is said to be in the Wiener class if it has a Fourier series with absolutely summable Fourier coefficients. It will often be of interest to begin with a function f in the Wiener class and then define the sequence of of $n \times n$ Toeplitz matrices

$$T_n(f) = \left[\frac{1}{2\pi}\int_0^{2\pi} f(\lambda)e^{-i(k-j)\lambda}d\lambda\ ; \quad k, j = 0, 1, \cdots, n-1\right], \tag{4.8}$$

which will then also be in the Wiener class. The Toeplitz matrix $T_n(f)$ will be Hermitian if and only if f is real. More specifically, $T_n(f) = T_n^*(f)$ if and only if $t_{k-j} = t_{j-k}^*$ for all k, j or, equivalently, $t_k^* = t_{-k}$ all k. If $t_k^* = t_{-k}$, however,

$$\begin{aligned} f^*(\lambda) &= \sum_{k=-\infty}^{\infty} t_k^* e^{-ik\lambda} = \sum_{k=-\infty}^{\infty} t_{-k}e^{-ik\lambda} \\ &= \sum_{k=-\infty}^{\infty} t_k e^{ik\lambda} = f(\lambda), \end{aligned}$$

so that f is real. Conversely, if f is real, then

$$\begin{aligned} t_k^* &= \frac{1}{2\pi}\int_0^{2\pi} f^*(\lambda)e^{ik\lambda}\,d\lambda \\ &= \frac{1}{2\pi}\int_0^{2\pi} f(\lambda)e^{ik\lambda}\,d\lambda = t_{-k}. \end{aligned}$$

It will be of interest to characterize the maximum and minimum magnitude of the eigenvalues of Toeplitz matrices and how these relate to the maximum and minimum values of the corresponding functions f. Problems arise, however, if the function f has a maximum or minimum at an isolated point. To avoid such difficulties we define the *essential supremum* $M_f = \text{ess sup} f$ of a real valued function f as the smallest number a for which $f(x) \le a$ except on a set of total length or measure 0. In particular, if $f(x) > a$ only at isolated points x and not on any interval of nonzero length, then $M_f \le a$. Similarly, the *essential infimum* $m_f = \text{ess inf} f$ is defined as the largest value of a for which

$f(x) \geq a$ except on a set of total length or measure 0. The key idea here is to view M_f and m_f as the maximum and minimum values of f, where the extra verbiage is to avoid technical difficulties arising from the values of f on sets that do not effect the integrals. Functions f in the Wiener class are bounded since

$$|f(\lambda)| \leq \sum_{k=-\infty}^{\infty} |t_k e^{ik\lambda}| \leq \sum_{k=-\infty}^{\infty} |t_k| \tag{4.9}$$

so that

$$m_{|f|}, M_{|f|} \leq \sum_{k=-\infty}^{\infty} |t_k|. \tag{4.10}$$

4.2 Bounds on Eigenvalues of Toeplitz Matrices

In this section Lemma 2.1 is used to obtain bounds on the eigenvalues of Hermitian Toeplitz matrices and an upper bound bound to the strong norm for general Toeplitz matrices.

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

$$m_f \leq \tau_{n,k} \leq M_f. \tag{4.11}$$

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

$$\parallel T_n(f) \parallel \leq 2M_{|f|}, \tag{4.12}$$

so that the sequence of Toeplitz matrices $\{T_n(f)\}$ is uniformly bounded over n if the essential supremum of $|f|$ is finite.

Proof. From Lemma 2.1,

$$\max_k \ \tau_{n,k} = \max_x (x^* T_n(f) x)/(x^* x) \tag{4.13}$$

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

so that

$$\begin{aligned} x^*T_n(f)x &= \sum_{k=0}^{n-1}\sum_{j=0}^{n-1} t_{k-j}x_kx_j^* \\ &= \sum_{k=0}^{n-1}\sum_{j=0}^{n-1}\left[\tfrac{1}{2\pi}\int_0^{2\pi} f(\lambda)e^{i(k-j)\lambda}\,d\lambda\right]x_kx_j^* \qquad (4.14) \\ &= \tfrac{1}{2\pi}\int_0^{2\pi}\left|\sum_{k=0}^{n-1}x_ke^{ik\lambda}\right|^2 f(\lambda)\,d\lambda \end{aligned}$$

and likewise

$$x^*x = \sum_{k=0}^{n-1}|x_k|^2 = \frac{1}{2\pi}\int_0^{2\pi}|\sum_{k=0}^{n-1}x_ke^{ik\lambda}|^2\,d\lambda. \qquad (4.15)$$

Combining (4.14)–(4.15) results in

$$m_f \le \frac{\displaystyle\int_0^{2\pi} f(\lambda)\left|\sum_{k=0}^{n-1}x_ke^{ik\lambda}\right|^2 d\lambda}{\displaystyle\int_0^{2\pi}\left|\sum_{k=0}^{n-1}x_ke^{ik\lambda}\right|^2 d\lambda} = \frac{x^*T_n(f)x}{x^*x} \le M_f, \qquad (4.16)$$

which with (4.13) yields (4.11).

We have already seen in (2.16) that if $T_n(f)$ is Hermitian, then $\| T_n(f) \| = \max_k |\tau_{n,k}| \overset{\Delta}{=} |\tau_{n,M}|$. Since $|\tau_{n,M}| \le \max(|M_f|, |m_f|) \le M_{|f|}$, (4.12) holds 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 $f_i = (f - f^*)/2i$. From the triangle inequality for norms,

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

Since $|(f \pm f^*)/2 \le (|f|+|f^*|)/2 \le M_{|f|}$, $M_{|f_r|}+M_{|f_i|} \le 2M_{|f|}$, proving (4.12). □

Note for later use that the weak norm of a Toeplitz matrix takes a particularly simple form. Let $T_n(f) = \{t_{k-j}\}$, then by collecting equal terms we have

$$\begin{aligned} |T_n(f)|^2 &= \frac{1}{n}\sum_{k=0}^{n-1}\sum_{j=0}^{n-1}|t_{k-j}|^2 \\ &= \frac{1}{n}\sum_{k=-(n-1)}^{n-1}(n-|k|)|t_k|^2 \\ &= \sum_{k=-(n-1)}^{n-1}(1-|k|/n)|t_k|^2. \end{aligned} \tag{4.17}$$

We are now ready to put all the pieces together to study the asymptotic behavior of $T_n(f)$. If we can find an asymptotically equivalent sequence of circulant matrices, then all of the results regarding circulant matrices and asymptotically equivalent sequences of matrices apply. The main difference between the derivations for simple sequence of banded Toeplitz matrices and the more general case is the sequence of circulant matrices chosen. Hence to gain some feel for the matrix chosen, we first consider the simpler banded case where the answer is obvious. The results are then generalized in a natural way.

4.3 Banded Toeplitz Matrices

Let T_n be a sequence of banded Toeplitz matrices of order $m+1$, that is, $t_i = 0$ unless $|i| \le m$. Since we are interested in the behavior or T_n for large n we choose $n >> m$. As is easily seen from (4.1), T_n looks like a circulant matrix except for the upper left and lower right-hand corners, i.e., each row is the row above shifted to the right one place. We can make a banded Toeplitz 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_n in just this way, i.e.,

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

$$= \begin{bmatrix}
c_0^{(n)} & \cdots & & & c_{n-1}^{(n)} \\
c_{n-1}^{(n)} & c_0^{(n)} & \cdots & & \\
\vdots & & & \ddots & \vdots \\
c_1^{(n)} & \cdots & & c_{n-1}^{(n)} & c_0^{(n)}
\end{bmatrix}. \qquad (4.18)$$

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

$$c_k^{(n)} = \begin{cases} t_{-k} & k = 0, 1, \cdots, m \\ t_{n-k} & k = n-m, \cdots, n-1 \\ 0 & \text{otherwise} \end{cases} \qquad (4.19)$$

If a Toeplitz matrix is specified by a function f and hence denoted by $T_n(f)$, then the circulant matrix defined by (4.18–4.19) is similarly denoted $C_n(f)$. The function f will be explicitly shown when it is useful to do so, for example when the results being developed specifically involve f.

The matrix C_n is intuitively a candidate for a simple matrix asymptotically equivalent to T_n — we need only demonstrate that it is indeed both asymptotically equivalent and simple.

Lemma 4.2. The matrices T_n and C_n defined in (4.1) and (4.18) are asymptotically equivalent, i.e., both are bounded in the strong norm and

$$\lim_{n\to\infty} |T_n - C_n| = 0. \tag{4.20}$$

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

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

□

The above lemma is almost trivial since the matrix $T_n - C_n$ has fewer than m^2 non-zero entries and hence the $1/n$ 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 (4.1) and (4.18) and let their eigenvalues be $\tau_{n,k}$ and $\psi_{n,k}$, respectively, then for any positive integer s

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1}\left(\tau_{n,k}^s - \psi_{n,k}^s\right) = 0. \tag{4.21}$$

In fact, for finite n,

$$\left|\frac{1}{n}\sum_{k=0}^{n-1}\left(\tau_{n,k}^s - \psi_{n,k}^s\right)\right| \le Kn^{-1/2}, \tag{4.22}$$

where K is not a function of n.

Proof. Equation (4.21) is direct from Lemma 4.2 and Theorem 2.2. Equation (4.22) follows from Corollary 2.1 and Lemma 4.2. □

The lemma implies that if either of the separate limits converges, then both will and

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \tau_{n,k}^s = \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \psi_{n,k}^s. \tag{4.23}$$

The next lemma shows that the second limit indeed converges, and in fact provides an evaluation for the limit.

Lemma 4.4. Let $C_n(f)$ be constructed from $T_n(f)$ as in (4.18) and let $\psi_{n,k}$ be the eigenvalues of $C_n(f)$, then for any positive integer s we have

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \psi_{n,k}^s = \frac{1}{2\pi} \int_0^{2\pi} f^s(\lambda)\, d\lambda. \tag{4.24}$$

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

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} F(\psi_{n,k}) = \frac{1}{2\pi} \int_0^{2\pi} F(f(\lambda))\, d\lambda. \tag{4.25}$$

Proof. From Theorem 3.1 we have exactly

$$\begin{aligned}
\psi_{n,j} &= \sum_{k=0}^{n-1} c_k^{(n)} e^{-2\pi ijk/n} \\
&= \sum_{k=0}^{m} t_{-k} e^{-2\pi ijk/n} + \sum_{k=n-m}^{n-1} t_{n-k} e^{-2\pi ijk/n} \\
&= \sum_{k=-m}^{m} t_k e^{-2\pi ijk/n} = f(\frac{2\pi j}{n})
\end{aligned} \tag{4.26}$$

Note that the eigenvalues of $C_n(f)$ 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

$$\begin{aligned}
\lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}\psi_{n,k}^s &= \lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}f(2\pi k/n)^s \\
&= \lim_{n\to\infty}\sum_{k=0}^{n-1}f(\lambda_k)^s\Delta\lambda/(2\pi) \\
&= \frac{1}{2\pi}\int_0^{2\pi} f(\lambda)^s d\lambda, \qquad (4.27)
\end{aligned}$$

where the continuity of $f(\lambda)$ guarantees the existence of the limit of (4.27) as a Riemann integral. If $T_n(f)$ and $C_n(f)$ are Hermitian, than the $\psi_{n,k}$ and $f(\lambda)$ are real and application of the Weierstrass theorem to (4.27) yields (4.25). Lemma 4.2 and (4.26) ensure that $\psi_{n,k}$ and $\tau_{n,k}$ are in the 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 banded Toeplitz matrix with eigenvalues $\tau_{n,k}$, then for any positive integer s

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}\tau_{n,k}^s = \frac{1}{2\pi}\int_0^{2\pi} f(\lambda)^s\, d\lambda. \qquad (4.28)$$

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

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}F(\tau_{n,k}) = \frac{1}{2\pi}\int_0^{2\pi} F(f(\lambda))\, d\lambda; \qquad (4.29)$$

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 $T_n(f)x = \tau x$ and $C_n(f)x = \psi x$, $n > 2m+1$, 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 results for circulant matrices and asymptotically equivalent sequences of matrices. Since these theorems are identical in statement and proof with the more general case of functions f in the Wiener class, we defer these theorems momentarily and generalize Theorem 4.1 to more general Toeplitz matrices with no assumption of bandedeness.

4.4 Wiener Class Toeplitz Matrices

Next consider the case of f in the Wiener class, i.e., the case where the sequence $\{t_k\}$ is absolutely summable. As in the case of sequences of banded Toeplitz matrices, the basic approach is to find a sequence of circulant matrices $C_n(f)$ that is asymptotically equivalent to the sequence of Toeplitz matrices $T_n(f)$. In the more general case under consideration, the construction of $C_n(f)$ is necessarily more complicated. Obviously the choice of an appropriate sequence of circulant matrices to approximate a sequence of Toeplitz matrices 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. Since f is assumed to be in the Wiener class, we have the Fourier series representation

$$f(\lambda) = \sum_{k=-\infty}^{\infty} t_k e^{ik\lambda} \tag{4.30}$$

$$t_k = \frac{1}{2\pi}\int_0^{2\pi} f(\lambda) e^{-ik\lambda}\, d\lambda. \tag{4.31}$$

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

$$c_k^{(n)} = \frac{1}{n}\sum_{j=0}^{n-1} f(2\pi j/n) e^{2\pi ijk/n}. \tag{4.32}$$

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

$$\begin{aligned}\lim_{n\to\infty} c_k^{(n)} &= \lim_{n\to\infty} \tfrac{1}{n}\sum_{j=0}^{n-1} f(2\pi j/n)e^{2\pi ijk/n} \\ &= \tfrac{1}{2\pi}\int_0^{2\pi} f(\lambda)e^{ik\lambda}d\lambda = t_{-k}\end{aligned} \tag{4.33}$$

and hence the $c_k^{(n)}$ are simply the sum approximations to the Riemann integrals giving t_{-k}. Equations (4.32), (3.7), and (3.9) show that the eigenvalues $\psi_{n,m}$ of $C_n(f)$ are simply $f(2\pi m/n)$; that is, from (3.7) and (3.9)

$$\begin{aligned}\psi_{n,m} &= \sum_{k=0}^{n-1} c_k^{(n)} e^{-2\pi imk/n} \\ &= \sum_{k=0}^{n-1}\left(\frac{1}{n}\sum_{j=0}^{n-1} f(2\pi j/n)e^{2\pi ijk/n}\right)e^{-2\pi imk/n} \\ &= \sum_{j=0}^{n-1} f(2\pi j/n)\left\{\frac{1}{n}\sum_{k=0}^{n-1} e^{2\pi ik(j-m)/n}\right\} \\ &= f(2\pi m/n).\end{aligned} \tag{4.34}$$

Thus, $C_n(f)$ has the useful property (4.26) of the circulant approximation (4.19) used in the banded case. As a result, the conclusions of Lemma 4.4 hold for the more general case with $C_n(f)$ constructed as in (4.32). Equation (4.34) in turn defines $C_n(f)$ since, if we are told that $C_n(f)$ is a circulant matrix with eigenvalues $f(2\pi m/n), m = 0, 1, \cdots, n-1$, then from (3.9)

$$\begin{aligned}c_k^{(n)} &= \frac{1}{n}\sum_{m=0}^{n-1}\psi_{n,m}e^{2\pi imk/n} \\ &= \frac{1}{n}\sum_{m=0}^{n-1} f(2\pi m/n)e^{2\pi imk/n},\end{aligned} \tag{4.35}$$

as in (4.32). Thus, either (4.32) or (4.34) 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. Given a function f satisfying (4.30–4.31) and define the circulant matrix $C_n(f)$ by (4.32).

(1) Then

$$c_k^{(n)} = \sum_{m=-\infty}^{\infty} t_{-k+mn} \ , \qquad k = 0, 1, \cdots, n-1. \tag{4.36}$$

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

(2) If $f(\lambda)$ is real and $m_f = \text{ess inf } f > 0$, 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) Applying (4.31) to $\lambda = 2\pi j/n$ gives

$$f(2\pi \frac{j}{n}) = \sum_{\ell=-\infty}^{\infty} t_\ell e^{i\ell 2\pi j/n}$$

which when inserted in (4.32) yields

$$\begin{aligned} c_k^{(n)} &= \frac{1}{n}\sum_{j=0}^{n-1} f(2\pi\frac{j}{n}) e^{2\pi ijk/n} \\ &= \frac{1}{n}\sum_{j=0}^{n-1}\left(\sum_{\ell=-\infty}^{\infty} t_\ell e^{i\ell 2\pi j/n}\right) e^{2\pi ijk/n} \qquad (4.37) \\ &= \sum_{\ell=-\infty}^{\infty} t_\ell \frac{1}{n}\sum_{j=0}^{n-1} e^{i2\pi(k+\ell)j/n} = \sum_{\ell=-\infty}^{\infty} t_\ell \delta_{(k+\ell) \bmod n}, \end{aligned}$$

where the final step uses (3.10). The term $\delta_{(k+\ell) \bmod n}$ will be 1 whenever $\ell = -k$ plus a multiple mn of n, which yields (4.36).

(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 (4.35) 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 (4.36) 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 [19] 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

$$\hat{f}_n(\lambda) = \sum_{m=-(n-1)}^{n-1} t_m e^{im\lambda}, \tag{4.38}$$

which depends only on $\{t_m; m = 0, \pm 1, \cdots, \pm n-1\}$, and then define the circulant matrix $C_n(\hat{f}_n)$; that is, the circulant matrix having as top row $(\hat{c}_0^{(n)}, \cdots, \hat{c}_{n-1}^{(n)})$ where analogous to the derivation of (4.37)

$$\begin{aligned}
\hat{c}_k^{(n)} &= \frac{1}{n}\sum_{j=0}^{n-1} \hat{f}_n(\frac{2\pi j}{n}) e^{2\pi ijk/n} \\
&= \frac{1}{n}\sum_{j=0}^{n-1} \left(\sum_{\ell=-(n-1)}^{n-1} t_\ell e^{i\ell 2\pi j/n} \right) e^{2\pi ijk/n} \\
&= \sum_{\ell=-(n-1)}^{n-1} t_\ell \frac{1}{n} \sum_{j=0}^{n-1} e^{i2\pi(k+\ell)j/n} \\
&= \sum_{\ell=-(n-1)}^{n-1} t_\ell \delta_{(k+\ell) \bmod n}.
\end{aligned}$$

Now, however, we are only interested in values of ℓ which have the form $-k$ plus a multiple mn of n for which $-(n-1) \le -k + mn \le n-1$. This will always include the $m = 0$ term for which $\ell = -k$. If $k = 0$, then only the $m = 0$ term lies within the range. If $k = 1, 2, \ldots, n-1$, then $m = -1$ results in $-k+n$ which is between 1 and $n-1$. No other multiples lie within the range, so we end up with

$$\hat{c}_k^{(n)} = \begin{cases} t_0 & k = 0 \\ t_{-k} + t_{n-k} & k = 1, 2, \ldots, n-1 \end{cases}. \tag{4.39}$$

Since $C_n(\hat{f}_n)$ is also a Toeplitz matrix, define $C_n(\hat{f}_n) = T'_n = \{t'_{k-j}\}$ with

$$t'_k = \begin{cases} \hat{c}_{-k}^{(n)} = t_k + t_{n+k} & k = -(n-1), \ldots, -1 \\ \hat{c}_0^{(n)} = t_0 & k = 0 \\ \hat{c}_{n-k}^{(n)} = t_{-(n-k)} + t_k & k = 1, 2, \ldots, n-1 \end{cases}, \tag{4.40}$$

which can be pictured as

$$T'_n =$$

$$\begin{bmatrix} t_0 & t_{-1} + t_{n-1} & t_{-2} + t_{n-2} & \cdots & t_{-(n-1)} + t_1 \\ t_1 + t_{-(n-1)} & t_0 & t_{-1} + t_{n-1} & & \\ t_2 + t_{-(n-2)} & t_1 + t_{-(n-1)} & t_0 & & \vdots \\ \vdots & & & \ddots & \\ t_{n-1} + t_1 & & & \cdots & t_0 \end{bmatrix} \tag{4.41}$$

Like the original approximation $C_n(f)$, the approximation $C_n(\hat{f}_n)$ reduces to the $C_n(f)$ of (4.19) for a banded Toeplitz matrix of order m if $n > 2m+1$. The following lemma shows that these circulant matrices are asymptotically equivalent to each other and to T_m.

Lemma 4.6. Let $T_n(f) = \{t_{k-j}\}$ where

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

and

$$f(\lambda) = \sum_{k=-\infty}^{\infty} t_k e^{ik\lambda}, \quad \hat{f}_n(\lambda) = \sum_{k=-(n-1)}^{n-1} t_k e^{ik\lambda}.$$

Define the circulant matrices $C_n(f)$ and $C_n(\hat{f}_n)$ as in (4.32) and (4.38)–(4.39). Then,

$$C_n(f) \sim C_n(\hat{f}_n) \sim T_n. \tag{4.42}$$

Proof. Since both $C_n(f)$ and $C_n(\hat{f}_n)$ are circulant matrices with the same eigenvectors (Theorem 3.1), we have from part 2 of Theorem 3.1 and (2.17) that

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

Recall from (4.6) 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 $n \geq N$

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

Since ϵ is arbitrary,

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

proving that

$$C_n(f) \sim C_n(\hat{f}_n). \tag{4.43}$$

Application of (4.40) and (4.17) results in

$$\begin{aligned}
|T_n(f) - C_n(\hat{f}_n)|^2 &= \sum_{k=-(n-1)}^{n-1} (1 - |k|/n)|t_k - t'_k|^2 \\
&= \sum_{k=-(n-1)}^{-1} \frac{n+k}{n}|t_{n+k}|^2 + \sum_{k=1}^{n-1} \frac{n-k}{n}|t_{-(n-k)}|^2 \\
&= \sum_{k=-(n-1)}^{-1} \frac{k}{n}|t_k|^2 + \sum_{k=1}^{n-1} \frac{k}{n}|t_{-k}|^2 \\
&= \sum_{k=1}^{n-1} \frac{k}{n}\left(|t_k|^2 + |t_{-k}|^2\right) \qquad (4.44)
\end{aligned}$$

Since the $\{t_k\}$ are absolutely summable, they are also square summable from (4.4) and hence given $\epsilon > 0$ we can choose an N large enough so that

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

Therefore

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

Since ϵ is arbitrary,

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

and hence

$$T_n(f) \sim C_n(\hat{f}_n), \tag{4.45}$$

which with (4.43) and Theorem 2.1 proves (4.42). □

Pearl [19] develops a circulant matrix similar to $C_n(\hat{f}_n)$ (depending only on the entries of $T_n(f)$) such that (4.45) holds in the more general case where (4.2) instead of (4.3) holds.

We now have a sequence of circulant matrices $\{C_n(f)\}$ asymptotically equivalent to the sequence $\{T_n(f)\}$ and the eigenvalues, inverses and products of the circulant matrices are known exactly. Therefore Lemmas 4.2–4.4 and Theorems 2.2–2.2 can be applied to generalize Theorem 4.1.

Theorem 4.2. Let $T_n(f)$ be a sequence of Toeplitz matrices such that $f(\lambda)$ is in the Wiener class or, equivalently, that $\{t_k\}$ is absolutely summable. Let $\tau_{n,k}$ be the eigenvalues of $T_n(f)$ and s be any positive integer. Then

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \tau_{n,k}^s = \frac{1}{2\pi} \int_0^{2\pi} f(\lambda)^s \, d\lambda. \tag{4.46}$$

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

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} F(\tau_{n,k}) = \frac{1}{2\pi} \int_0^{2\pi} F(f(\lambda)) \, d\lambda. \tag{4.47}$$

Theorem 4.2 is the fundamental eigenvalue distribution theorem of Szegö (see [16]). The approach used here is essentially a specialization of Grenander and Szegö ([16], ch. 7).

Theorem 4.2 yields the following two corollaries.

Corollary 4.1. Given the assumptions of the theorem, define the eigenvalue distribution function $D_n(x) =$ (number of $\tau_{n,k} \le x)/$n. 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

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

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. It can be interpreted as not allowing $f(\lambda)$ to have a flat region around the point x. The limiting distribution function evaluated at x describes the fraction of the eigenvalues that smaller than x in the limit as $n \to \infty$, which in turn implies that the fraction of eigenvalues between two values a and $b > a$ is $D(b) - D(a)$. This is similar to the role of a cumulative distribution function (cdf) in probability theory.

Proof. Define the indicator function

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

We have

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

Unfortunately, $1_x(\alpha)$ is not a continuous function and hence Theorem 4.2 cannot be immediately applied. 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 4.2 can be used to conclude that

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

and

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

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

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

and

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

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

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

and

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

Thus if

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

then

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

□

Corollary 4.2. Assume that the conditions of Theorem 4.2 hold and let m_f and M_f denote the essential infimum and the essential supremum of f, respectively. Then

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

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

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

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

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

5

Matrix Operations on Toeplitz Matrices

Applications of Toeplitz matrices like those of matrices in general involve matrix operations such as addition, inversion, products and the computation of eigenvalues, eigenvectors, and determinants. The properties of Toeplitz matrices particular to these operations are based primarily on three fundamental results that have been described earlier:

(1) matrix operations are simple when dealing with circulant matrices,
(2) given a sequence of Toeplitz matrices, we can instruct asymptotically equivalent sequences of circulant matrices, and
(3) asymptotically equivalent sequences of matrices have equal asymptotic eigenvalue distributions and other related properties.

In the next few sections some of these operations are explored in more depth for sequences of Toeplitz matrices. Generalizations and related results can be found in Tyrtyshnikov [31].

5.1 Inverses of Toeplitz Matrices

In some applications we wish to study the asymptotic distribution of a function $F(\tau_{n,k})$ of the eigenvalues that is not continuous at the minimum or maximum value of f. For example, in order for the results derived thus far to apply to the function $F(f(\lambda)) = 1/f(\lambda)$ which arises when treating inverses of Toeplitz matrices, it has so far been necessary to require that the essential infimum $m_f > 0$ because the function $F(1/x)$ is not continuous at $x = 0$. If $m_f = 0$, the basic asymptotic eigenvalue distribution Theorem 4.2 breaks down and the limits and the integrals involved might not exist. The limits might exist and equal something else, or they might simply fail to exist. In order to treat the inverses of Toeplitz matrices when f has zeros, we state without proof an intuitive extension of the fundamental Toeplitz result that shows how to find asymptotic distributions of suitably truncated functions. To state the result, define the mid function

$$\mathrm{mid}(x,y,z) \triangleq \begin{cases} z & y \geq z \\ y & x \leq y \leq z \\ x & y \leq z \end{cases} \tag{5.1}$$

$x < z$. This function can be thought of as having input y and thresholds z and X and it puts out y if y is between z and x, z if y is smaller than z, and x if y is greater than x. The following result was proved in [13] and extended in [25]. See also [26, 27, 28].

Theorem 5.1. Suppose that f is in the Wiener class. Then for any function $F(x)$ continuous on $[\psi, \theta] \subset [m_f, M_f]$

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} F(\mathrm{mid}(\psi, \tau_{n,k}, \theta) = \frac{1}{2\pi} \int_0^{2\pi} F(\mathrm{mid}(\psi, f(\lambda), \theta)\, d\lambda. \tag{5.2}$$

Unlike Theorem 4.2 we pick arbitrary points ψ and θ such that F is continuous on the closed interval $[\psi, \theta]$. These need not be the minimum and maximum of f.

Theorem 5.2. Assume that f is in the Wiener class and is real and that $f(\lambda) \geq 0$ with equality holding at most at a countable number of points. Then **(a)** $T_n(f)$ is nonsingular

(b) If $f(\lambda) \geq m_f > 0$, then

$$T_n(f)^{-1} \sim C_n(f)^{-1}, \tag{5.3}$$

where $C_n(f)$ is defined in (4.35). Furthermore, if we define $T_n(f) - C_n(f) = D_n$ then $T_n(f)^{-1}$ has the expansion

$$\begin{aligned} T_n(f)^{-1} & \\ &= [C_n(f) + D_n]^{-1} \\ &= C_n(f)^{-1}\left[I + D_nC_n(f)^{-1}\right]^{-1} \\ &= C_n(f)^{-1}\left[I + D_nC_n(f)^{-1} + \left(D_nC_n(f)^{-1}\right)^2 + \cdots\right], \end{aligned} \tag{5.4}$$

and the expansion converges (in weak norm) for sufficiently large n.
(c) If $f(\lambda) \geq m_f > 0$, then

$$T_n(f)^{-1} \sim T_n(1/f) = \left[\frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{e^{i(k-j)\lambda}}{f(\lambda)}\, d\lambda\right]; \tag{5.5}$$

that is, if the spectrum is strictly positive, then the inverse of a sequence of Toeplitz matrices 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

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1} F(\rho_{n,k}) = \frac{1}{2\pi}\int_{-\pi}^{\pi} F((1/f(\lambda))\, d\lambda. \tag{5.6}$$

(d) Suppose that $m_f = 0$ and that the derivative of $f(\lambda)$ exists and is bounded for all λ. 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 (5.2) may not exist. For any finite θ, however, the following similar fact is true: If $F(x)$ is a continuous function on $[1/M_f, \theta]$, then

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1} F(\min(\rho_{n,k}, \theta)) = \frac{1}{2\pi}\int_{0}^{2\pi} F(\min(1/f(\lambda), \theta))\, d\lambda. \tag{5.7}$$

Proof. (a) Since $f(\lambda) > 0$ except at possibly countably many points, we have from (4.14)

$$x^* T_n(f) x = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left| \sum_{k=0}^{n-1} x_k e^{ik\lambda} \right|^2 f(\lambda) d\lambda > 0.$$

Thus for all n

$$\min_k \tau_{n,k} > 0$$

and hence

$$\det T_n(f) = \prod_{k=0}^{n-1} \tau_{n,k} \neq 0$$

so that $T_n(f)$ is nonsingular.

(b) From Lemma 4.6, $T_n \sim C_n$ and hence (5.1) follows from Theorem 2.1 since $f(\lambda) \geq m_f > 0$ ensures that

$$\parallel T_n(f)^{-1} \parallel, \parallel C_n(f)^{-1} \parallel \leq 1/m_f < \infty.$$

The series of (5.4) will converge in weak norm if

$$|D_n C_n(f)^{-1}| < 1. \tag{5.8}$$

Since

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

Eq. (5.8) must hold for large enough n.

(c) We have from the triangle inequality that

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

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

$$|T_n(f)^{-1} - C_n(f)^{-1}| \leq \frac{\epsilon}{2}. \tag{5.9}$$

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

$$|C_n(f)^{-1} - T_n(1/f)| \leq \epsilon/2 \tag{5.10}$$

so that for any $\epsilon > 0$ from (5.7)–(5.8) can choose n such that

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

which implies (5.5). Equation (5.6) follows from (5.5) and Theorem 2.4. Alternatively, if $G(x)$ is any continuous function on $[1/M_f, 1/m_f]$ and (5.4) follows directly from Lemma 4.6 and Theorem 2.3 applied to $G(1/x)$.

(d) 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

$$\| T_n^{-1} \| = \max_k \rho_{n,k} = 1/\min_k \tau_{n,k} \tag{5.11}$$

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, consider the disjoint sets

$$\begin{aligned} E_k &= \{\lambda : 1/k \geq f(\lambda)/M_f > 1/(k+1)\} \\ &= \{\lambda : k \leq M_f/f(\lambda) < k+1\} \end{aligned} \tag{5.12}$$

and let $|E_k|$ denote the length of the set E_k, that is,

$$|E_k| = \int_{\lambda: M_f/k \geq f(\lambda) > M_f/(k+1)} d\lambda.$$

From (5.12)

$$\begin{aligned} \int_{-\pi}^{\pi} \frac{1}{f(\lambda)}\, d\lambda &= \sum_{k=1}^{\infty} \int_{E_k} \frac{1}{f(\lambda)}\, d\lambda \\ &\geq \sum_{k=1}^{\infty} \frac{|E_k| k}{M_f}. \end{aligned} \tag{5.13}$$

For a given k, E_k will comprise a union of disjoint intervals of the form (a, b) where for all $\lambda \in (a, b)$ we have that $1/k \geq f(\lambda)/M_f > 1/(k+1)$. There must be at least one such nonempty interval, so $|E_k|$ will be bound below by the length of this interval, $b - a$. Then for any $x, y \in (a, b)$

$$|f(y) - f(x)| = |\int_x^y \frac{df}{d\lambda}\, d\lambda| \leq \eta |y - x|.$$

By assumption there is some finite value η such that

$$\left|\frac{df}{d\lambda}\right| \leq \eta, \tag{5.14}$$

so that

$$|f(y) - f(x)| = \leq \eta |y - x|.$$

Pick x and y so that $f(x) = M_f/(k+1)$ and $f(y) = M_f/k$ (since f is continuous at almost all points, this argument works almost everywhere – it needs more work if these end points are not points of continuity of f), then

$$b - a \geq |y - x| \geq M_f(\frac{1}{k} - \frac{1}{k+1}) = \frac{M_f}{k+1}.$$

Combining this with (5.13) yields

$$\begin{aligned} \int_{-\pi}^{\pi} d\lambda/f(\lambda) &\geq \sum_{k=1}^{\infty} (k/M_f)(\frac{M_f}{k(k+1})))/\eta & (5.15) \\ &= \sum_{k=1}^{\infty} \frac{1}{k+1}, & (5.16) \end{aligned}$$

which diverges so that $1/f(\lambda)$ is not integrable. To prove (5.5) 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 (5.5). Note that (5.5) implies that the eigenvalues of $T_n(f)^{-1}$ 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 (d) is when $T_n(f)$ is banded 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 (b) is due to Rino [20]. The proof of (d) is motivated by one of Widom [33]. Further results along the lines of (d) regarding unbounded Toeplitz matrices may be found in [13]. Related results considering asymptotically equal distributions of unbounded sequences can be found in Tyrtyshnikov [32] and Trench [25]. These works extend Weyl's definition of asymptotically equal distributions to unbounded sequences using the mid function used here to treat inverses. This leads to conditions for equal distributions and their implications.

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 [33] and the cited references. The results of Baxter [1] can also be applied to consider the asymptotic behavior of inverses in quite general cases.

5.2 Products of Toeplitz Matrices

We next combine Theorem 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. A key point is that while the product of Toeplitz matrices is not Toeplitz, a sequence of products of Toeplitz matrices $\{T_n(f)T_n(g)\}$ is asymptotically equivalent to a sequence of Toeplitz matrices $\{T_n(fg)\}$.

Theorem 5.3. Let $T_n(f)$ and $T_n(g)$ be defined as in (4.8) where $f(\lambda)$ and $g(\lambda)$ are two functions in the Wiener class. Define $C_n(f)$ and $C_n(g)$ as in (4.35) and let $\rho_{n,k}$ be the eigenvalues of $T_n(f)T_n(g)$

(a)

$$T_n(f)T_n(g) \sim C_n(f)C_n(g) = C_n(fg). \tag{5.17}$$

$$T_n(f)T_n(g) \sim T_n(g)T_n(f). \tag{5.18}$$

$$\lim_{n\to\infty} n^{-1}\sum_{k=0}^{n-1}\rho_{n,k}^s = \frac{1}{2\pi}\int_0^{2\pi}[f(\lambda)g(\lambda)]^s\, d\lambda \quad s = 1, 2, \ldots. \tag{5.19}$$

(b) If $T_n(f)$ and $T_n(g)$ are Hermitian, then for any $F(x)$ continuous on

$[m_f m_g, M_f M_g]$

$$\lim_{n\to\infty} n^{-1} \sum_{k=0}^{n-1} F(\rho_{n,k}) = \frac{1}{2\pi} \int_0^{2\pi} F(f(\lambda)g(\lambda))\, d\lambda. \tag{5.20}$$

(c)

$$T_n(f)T_n(g) \sim T_n(fg). \tag{5.21}$$

(d) Let $f_1(\lambda), ., f_m(\lambda)$ be in the Wiener class. Then if the $C_n(f_i)$ are defined as in (4.35)

$$\prod_{i=1}^{m} T_n(f_i) \sim C_n\left(\prod_{i=1}^{m} f_i\right) \sim T_n\left(\prod_{i=1}^{m} f_i\right). \tag{5.22}$$

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

$$\lim_{n\to\infty} n^{-1} \sum_{k=0}^{n-1} \rho_{n,k}^s = \frac{1}{2\pi} \int_0^{2\pi} \left(\prod_{i=1}^{m} f_i(\lambda)\right)^s d\lambda \tag{5.23}$$

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

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} (\mathrm{Re}[\rho_{n,k}])^s = \frac{1}{2\pi} \int_0^{2\pi} \left(\prod_{i=1}^{m} f_i(\lambda)\right)^s d\lambda. \tag{5.24}$$

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} (\Im[\rho_{n,k}])^2 = 0. \tag{5.25}$$

Proof. (a) Equation (5.14) follows from Lemmas 4.5 and 4.6 and Theorems 2.1 and 2.3. Equation (5.16) follows from (5.14). Note that while Toeplitz matrices do not in general commute, asymptotically they do. Equation (5.17) follows from (5.14), Theorem 2.2, and Lemma 4.4.
(b) Proof follows from (5.14) and Theorem 2.4. Note that the eigenvalues of the product of two Hermitian matrices are real ([18], p. 105).

(c) Applying Lemmas 4.5 and 4.6 and Theorem 2.1

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

(d) Follows from repeated application of (5.14) and part (c).
(e) Equation (5.22) 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 in the sense that the imaginary part vanishes in the limit. 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

$$\begin{aligned} \lim_{n\to\infty} n^{-1}\sum_{k=0}^{n-1}(\alpha_{n,k} + i\beta_{n,k})^s &= \lim_{n\to\infty} n^{-1}\sum_{k=0}^{n-1}\psi_{n,k}^s \\ &= \frac{1}{2\pi}\int_0^{2\pi}\left[\prod_{i=1}^{m} f_i(\lambda)\right]^s d\lambda, \quad (5.26) \end{aligned}$$

where $\psi_{n,k}$ are the eigenvalues of $C_n\left(\prod_{i=1}^{m} f_i\right)$. From (2.17)

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

From (4.57), Theorem 2.1 and Lemma 4.4

$$\begin{aligned} \lim_{n\to\infty}\left|\prod_{i=1}^{m} T_n(f_i)\right|^2 &= \lim_{n\to\infty}\left|C_n\left(\prod_{i=1}^{m} f_i\right)\right|^2 \\ &= (2\pi)^{-1}\int_0^{2\pi}\left(\prod_{i=1}^{m} f_i(\lambda)\right)^2 d\lambda. \quad (5.27) \end{aligned}$$

Subtracting (5.26) for $s = 2$ from (5.27) yields

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

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

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

□

Parts (d) and (e) are here proved as in Grenander and Szegö ([16], 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 could be directly related to the Toeplitz matrices using the results for asymptotically equivalent sequences of matrices. We began with the banded 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 functions $f(\lambda)$ or Toeplitz matrices in the Wiener class and hence to absolutely summable coefficients for simplicity. The more general case of square summable t_k is treated in Chapter 7 of [16] and requires significantly more mathematical care, but can be interpreted as an extension of the approach taken here.

We did not treat sums of Toeplitz matrices as no additional consideration is needed: a sum of Toeplitz matrices of equal size is also a Toeplitz matrix, so the results immediately apply. We also did not consider the asymptotic behavior of eigenvectors for the simple reason that there do not exist results along the lines that intuition suggests, that is, that show that in some sense the eigenvectors for circulant matrices also work for Toeplitz matrices.

5.3 Toeplitz Determinants

We close the consideration of matrix operations on Toeplitz matrices by returning to a problem mentioned in the introduction and formalize the behavior of limits of Toeplitz determinants. Suppose now that $T_n(f)$ is a sequence of Hermitian Toeplitz matrices such that that $f(\lambda) \geq m_f > 0$. Let $C_n(f)$ denote the sequence of circulant matrices constructed from f as in (4.32). Then from (4.34) the eigenvalues of $C_n(f)$ are $f(2\pi m/n)$ for $m = 0, 1, \ldots, n-1$ and hence $\det(C_n(f)) = \prod_{m=0}^{n-1} f(2\pi m/n)$. This in turn implies that

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

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

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

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

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

This integral, the asymptotic equivalence of $C_n(f)$ and $T_n(f)$ (Lemma 4.6), and Corollary 2.4 together yield the following result ([16], p. 65).

Theorem 5.4. Let $T_n(f)$ be a sequence of Hermitian Toeplitz matrices in the Wiener class such that $\ln f(\lambda)$ is Riemann integrable and $f(\lambda) \geq m_f > 0$. Then

$$\lim_{n\to\infty} \left(\det(T_n(f))\right)^{\frac{1}{n}} = \exp\left(\frac{1}{2\pi} \int_0^{2\pi} \ln f(\lambda)\, d\lambda\right). \tag{5.28}$$

6

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 of system identification algorithms and Wiener-Hopf factorization.

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) $X^n = (X_0, X_1, \ldots, X_{n-1})'$, 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

$$R_n = E[(X^n - m^n)(X^n - m^n)^*]. \tag{6.1}$$

Covariance matrices are Hermitian since

$$R_n^* = E[(X^n - m^n)(X^n - m^n)^*]^* = E[(X^n - m^n)(X^n - m^n)^*]. \tag{6.2}$$

Setting $m = 0$ yields the This is the autocorrelation matrix. Subscripts will be dropped when they are clear from context. If the matrix R_n is Toeplitz for all n, 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 $f(\lambda) = \sum_{k=-\infty}^{\infty} r_k e^{ik\lambda}$ is 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 $f(\lambda)$ is 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.

We now proceed to investigate the behavior of two common linear models for random processes, both of which model a complicated process as the result of passing a simple process through a linear filter. For simplicity we will assume the process means are zero.

6.1 Moving Average Processes

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

$$U_n = \begin{cases} \sum_{k=0}^{n} b_k W_{n-k} = \sum_{k=0}^{n} b_{n-k} W_k & n = 0, 1, \dots \\ 0 & n < 0 \end{cases}. \tag{6.3}$$

We assume that $b_0 = 1$ with no loss of generality since otherwise we can incorporate b_0 into σ^2. Note that (6.3) 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 (6.3) by

$$U_n = \sum_{k=-\infty}^{\infty} b_k W_{n-k} = \sum_{k=-\infty}^{\infty} b_{n-k} W_k. \tag{6.4}$$

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 (6.3) 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 (6.3) converges in the mean-squared sense. The weakest possible assumption that will guarantee convergence of (6.3) is that

$$\sum_{k=0}^{\infty} |b_k|^2 < \infty. \tag{6.5}$$

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

$$\sum_{k=0}^{\infty} |b_k| < \infty. \tag{6.6}$$

As previously this will result in simpler mathematics.

Equation (6.3) can be rewritten as a matrix equation by defining the lower triangular Toeplitz matrix

$$B_n = \begin{bmatrix} 1 & & & & 0 & \\ b_1 & 1 & & & & \\ b_2 & b_1 & & & & \\ \vdots & b_2 & \ddots & \ddots & & \\ & & & & & \\ b_{n-1} & \dots & & b_2 & b_1 & 1 \end{bmatrix} \tag{6.7}$$

so that

$$U^n = B_n W^n. \tag{6.8}$$

If the filter b_n were not causal, then B_n would not be triangular. If in addition (6.4) held, i.e., we looked at the entire process at each time instant, then (6.8) would require infinite vectors and matrices as in Grenander and Rosenblatt [15]. 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:

$$\begin{aligned} R_U^{(n)} &= EU^n(U^n)^* = EB_nW^n(W^n)^*B_n^* \\ &= \sigma^2 B_n B_n^* \\ &= \left[\sigma^2 \sum_{\ell=0}^{\min(k,j)} b_{\ell-k} b_{\ell-j}^*\right] \end{aligned}$$

The matrix $R_U^{(n)} = [r_{k,j}]$ is not Toeplitz. For example, the upper left entry is 1 and the second diagonal entry is $1 + b_1^2$. However, as we next show, the sequence $R_U^{(n)}$ becomes asymptotically Toeplitz as $n \to \infty$. If we define

$$b(\lambda) = \sum_{k=0}^{\infty} b_k e^{ik\lambda} \tag{6.9}$$

then

$$B_n = T_n(b) \tag{6.10}$$

so that

$$R_U^{(n)} = \sigma^2 T_n(b) T_n(b)^*. \tag{6.11}$$

Observe that $R_U^{(n)}$ is Hermitian, as all covariance matrices must be. We can now apply the results of the previous sections to obtain the following theorem.

Theorem 6.1. Let U_n be a moving average process with covariance matrix $R_{U_n}(n)$ given by (6.9)–(6.11). Let $\rho_{n,k}$ be the eigenvalues of $R_U^{(n)}$. Then

$$R_U^{(n)} \sim \sigma^2 T_n(|b|^2) = T_n(\sigma^2|b|^2) \tag{6.12}$$

so that U_n is asymptotically stationary. If $m = \text{ess inf } \sigma^2|b(\lambda)|^2$ and $M = \text{ess sup } \sigma^2|b(\lambda)|^2$ and $F(x)$ is any continuous function on $[m, M]$, then

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} F(\rho_{n,k}) = \frac{1}{2\pi}\int_0^{2\pi} F(\sigma^2|b(\lambda)|^2)\, d\lambda. \tag{6.13}$$

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

$$R_U^{(n)^{-1}} \sim \sigma^{-2} T_n(1/|b|^2). \tag{6.14}$$

Proof. Since $R_U^{(n)}$ is Hermitian, the results follow from Theorems 4.2 and 5.2 and (2.3). □

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 5.2, 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 $R_U^{(n)}$ and $\sigma^2 T_n(|b|^2)$.

6.2 Autoregressive Processes

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

$$X_n = \begin{cases} -\sum_{k=1}^{n} a_k X_{n-k} + W_n & n = 0, 1, \ldots \\ 0 & n < 0. \end{cases} \tag{6.15}$$

Autoregressive process include nonstationary processes such as the Wiener process. Equation (6.15) can be rewritten as a vector equation by defining the lower triangular matrix.

$$A_n = \begin{bmatrix} 1 & & & & & \\ a_1 & 1 & & & & 0 \\ & a_1 & 1 & & & \\ & & \ddots & \ddots & & \\ & & & & & \\ a_{n-1} & & & & a_1 & 1 \end{bmatrix} \tag{6.16}$$

so that

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

Since

$$R_W^{(n)} = A_n R_X^{(n)} A_n^* \tag{6.17}$$

and det $A_n = 1 \neq 0$, A_n is nonsingular. Hence

$$R_X^{(n)} = \sigma^2 A_n^{-1} A_n^{-1*} \tag{6.18}$$

or

$$(R_X^{(n)})^{-1} = \sigma^{-2} A_n^* A_n. \tag{6.19}$$

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

$$t_{k,j} = \sum_{m=0}^{\min(k,j)} a_{m-k}^* a_{m-j}.$$

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

$$a(\lambda) = \sum_{k=0}^{\infty} a_k e^{ik\lambda} \tag{6.20}$$

we have that

$$(R_X^{(n)})^{-1} = \sigma^{-2} T_n(a)^* T_n(a). \tag{6.21}$$

Observe that $(R_X^{(n)})^{-1}$ is Hermitian.

Theorem 6.2. Let X_n be an autoregressive process with absolutely summable $\{a_k\}$ and covariance matrix $R_X^{(n)}$ with eigenvalues $\rho_{n,k}$. Then

$$(R_X^{(n)})^{-1} \sim \sigma^{-2} T_n(|a|^2). \tag{6.22}$$

If $m = \text{ess inf } \sigma^{-2}|a(\lambda)|^2$ and $M = \text{ess sup } \sigma^{-2}|a(\lambda)|^2$, then for any function $F(x)$ on $[m, M]$ we have

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} F(1/\rho_{n,k}) = \frac{1}{2\pi} \int_0^{2\pi} F(\sigma^2 |a(\lambda)|^2)\, d\lambda, \tag{6.23}$$

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

$$R_X^{(n)} \sim \sigma^2 T_n(1/|a|^2) \tag{6.24}$$

so that the process is asymptotically stationary.

Proof. Theorem 5.3. □

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 5.2 (d).

Corollary 6.1. Given the assumptions of the theorem, then for any finite θ and any function $F(x)$ continuous on $[1/m, \theta]$

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} F(\min(\rho_{n,k}, \theta)) = \frac{1}{2\pi} \int_0^{2\pi} F(\min(1/|a(\gamma)|^2, \theta))\, d\lambda. \tag{6.25}$$

Proof. Theorems 6.2 and 5.1. □

If we consider two models of a random process to be asymptotically equivalent if their covariances are asymptotically equivalent, then from Theorems 6.1 and 6.2 we have the following corollary.

Corollary 6.2. Given the assumptions of Theorems 6.1 and 6.2, 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

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

Proof. Follows from Theorems 5.2 and 5.3 and

$$\begin{aligned} R_X^{(n)} &= \sigma^2 T_n(a)^{-1} T_n^{-1}(a)^* \\ &\sim \sigma^2 T_n(1/a) T_n(1/a)^* \\ &\sim \sigma^2 T_n(1/a)^* T_n(1/a). \end{aligned} \tag{6.26}$$

Comparison of (6.26) with (6.11) completes the proof. □

The methods above can also easily be applied to study the mixed autoregressive-moving average linear models [33].

6.3 Factorization

Consider the problem of the asymptotic behavior of triangular factors of a sequence of Hermitian covariance matrices $T_n(f)$ in the Wiener class. It is well known that any such matrix can be factored into the product of a lower triangular matrix and its conjugate transpose ([15], p. 37), in particular

$$T_n(f) = \{t_{k,j}\} = B_n B_n^*, \tag{6.27}$$

where B_n is a lower triangular matrix with entries

$$b_{k,j}^{(n)} = \{(\det T_k)\det(T_{k-1})\}^{-1/2}\gamma(j,k), \tag{6.28}$$

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

$$b_{k,k}^{(n)} = \{(\det T_k)/(\det T_{k-1})\}^{1/2}. \tag{6.29}$$

Equation (6.28) is the result of a Gaussian elimination or 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?

Suppose that $f(\lambda)$ has the form

$$\begin{aligned} f(\lambda) &= \sigma^2 |b(\lambda)|^2 \qquad (6.30) \\ b^*(\lambda) &= b(-\lambda) \\ b(\lambda) &= \sum_{k=0}^{\infty} b_k e^{ik\lambda} \\ b_0 &= 1. \end{aligned}$$

The decomposition of a nonnegative function into a product with this form is known as a *Wiener-Hopf factorization* . For a current survey see the discussion and references in Kailath et al. [17] We have already constructed functions of this form when considering moving average and autoregressive models. It is a classic result that a necessary and sufficient condition for f to have such a factorization is that $\ln f$ have a finite integral.

From (6.27) and Theorem 5.2 we have

$$B_n B_n^* = T_n(f) \sim T_n(\sigma b) T_n(\sigma b)^*. \qquad (6.31)$$

We wish to show that (6.31) implies that

$$B_n \sim T_n(\sigma b). \qquad (6.32)$$

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 5.2 (a) that $\det T_n(f) \neq 0$ so that B_n is invertible. Thus from Theorem 2.1 (e) and (6.31) we have

$$T_n^{-1} B_n = [B_n^{-1} T_n]^{-1} \sim T_n^* B_n^{*-1} = [B_n^{-1} T_n]^*. \qquad (6.33)$$

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 (6.33) 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 (6.33) we have $G_n \sim G_n^{*-1}$

$$\left\{|g_{k,k}^{(n)}|^2\delta_{k,j}\right\} \sim I_n. \tag{6.34}$$

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 (6.34) yields

$$\lim_{n\to\infty} \sigma^{-2}\frac{1}{n}\sum_{k=0}^{n-1}\left(b_{k,k}^{(n)}\right)^2 = 1. \tag{6.35}$$

From (6.29) and Corollary 2.4, and the fact that $T_n(\sigma b)$ is triangular we have that

$$\begin{aligned}
\lim_{n\to\infty} \sigma^{-1}\frac{1}{n}\sum_{k=0}^{n-1} b_{k,k}^{(n)} &= \sigma^{-1}\lim_{n\to\infty}\{(\det T_n(f))/(\det T_{n-1}(f))\}^{1/2} \\
&= \sigma^{-1}\lim_{n\to\infty}\{\det T_n(f)\}^{1/2n}\sigma^{-1}\lim_{n\to\infty}\{\det T_n(\sigma b)\}^{1/n} \\
&= \sigma^{-1}\sigma = 1. \qquad (6.36)
\end{aligned}$$

Combining (6.35) and (6.36) yields

$$\lim_{n\to\infty} |B_n^{-1}T_n - I_n| = 0. \tag{6.37}$$

Applying Theorem 2.1 yields (6.32). □

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 [1] and is similar to a result of Rissanen and Barbosa [21].

Acknowledgements

The author would like to thank his brother, Augustine Heard Gray, Jr., for his assistance long ago in finding the eigenvalues of the inverse covariance matrices of discrete time Wiener processes, his first encounter with Toeplitz and asymptotically Toeplitz matrices. He would like to thank Adriano Garsia and Tom Pitcher for helping him struggle through Grenander and Szegö's book during summer lunches in 1967 when the author was a summer employee at JPL during his graduate student days at USC. This manuscript first appeared as a technical report in 1971 as an expanded version of the tutorial paper [12] and was revised in 1975. After laying dormant for many years, it was revised and converted to LaTeXand posted on the World Wide Web. That resulted in significant feedback, corrections, and suggestions and in many revisions through the years. Particular thanks go to Ronald M. Aarts of the Philips Research Labs for 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. For corrections, comments, and improvements to the 2001 revision thanks are due to William Trench, John Dattorro, and Young-Han Kim. In particular, Professor Trench brought the Wielandt-Hoffman theorem and its use to prove strengthened results to my attention. Section 2.4 largely follows his suggestions, although I take the blame for any introduced errors. For the 2002 revision, particular thanks to Cynthia Pozun of ENST for several corrections. For the 2005–2006 revisions, special thanks to Jean-François Chamberland-Tremblay, Lee Patton, Sergio Verdu and two very preceptive and helpful anonymous reviewers. Finally, the author would like to thank the National Science Foundation for the support of the author's research involving Toeplitz matrices which led to the original paper and report.

References

[1] G. Baxter, “A Norm Inequality for a ‘Finite-Section’ Wiener-Hopf Equation,” *Illinois J. Math.,* 1962, pp. 97–103.

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

[3] T. Berger, *Rate Distortion Theory: A Mathematical Basis for Data Compression*, Prentice Hall, Englewood Cliffs, New Jersey, 1971.

[4] A. Böttcher and S.M. Grudsky, *Toeplitz Matrices, Asymptotic Linear Algebra, and Functional Analysis*, Birkhäuser, 2000.

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

[6] W. Cheney, *Introduction to Approximation theory*, McGraw-Hill, 1966.

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

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

[9] D. Fasino and P. Tilli, “Spectral clustering properties of block multilevel Hankel matrices, *Linear Algebra and its Applications*, Vol. 306, pp. 155–163, 2000.

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

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

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

[13] 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.

[14] R.M. Gray and L.D. Davisson, *An Introduction to Statistical Signal Processing*, Cambridge University Press, London, 2005.

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

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

[17] T. Kailath, A. Sayed, and B. Hassibi, *Linear Estimation*, Prentice Hall, New Jersey, 2000.

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

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

[20] 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.

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

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

[23] W. F. Trench, "Asymptotic distribution of the even and odd spectra of real symmetric Toeplitz matrices," *Linear Algebra Appl.*, Vol. 302-303, pp. 155–162, 1999.

[24] W. F. Trench, "Absolute equal distribution of the spectra of Hermitian matrices," *Lin. Alg. Appl.*, 366 (2003), 417–431.

[25] W. F. Trench, "Absolute equal distribution of families of finite sets," *Lin. Alg. Appl.* 367 (2003), 131–146.

[26] W. F. Trench, "A note on asymptotic zero distribution of orthogonal polynomials," *Lin. Alg. Appl.* 375 (2003) 275–281

[27] W. F. Trench, "Simplification and strengthening of Weyl's definition of asymptotic equal distribution of two families of finite sets," *Cubo A Mathematical Journal* Vol. 06 N 3 (2004), 47–54.

[28] W. F. Trench, "Absolute equal distribution of the eigenvalues of discrete Sturm–Liouville problems," *J. Math. Anal. Appl.*, Volume 321, Issue 1 , 1 September 2006, Pages 299–307.

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

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

[31] E.E. Tyrtyshnikov, "Influence of matrix operations on the distribution of eigenvalues and singular values of Toeplitz matrices," *Linear Algebra and its Applications*, Vol. 207, pp. 225–249, 1994.

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

[33] 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.

[34] A.J. Hoffman and H. W. Wielandt, "The variation of the spectrum of a normal matrix," *Duke Math. J.*, Vol. 20, pp. 37–39, 1953.

[35] James H. Wilkinson, “Elementary proof of the Wielandt-Hoffman theorem and of its generalization,” Stanford University, Department of Computer Science Report Number CS-TR-70-150, January 1970 .

Index

absolutely summable, 32, 38, 48
absolutely summable Toeplitz matrices, 41
analytic function, 16
asymptotic equivalence, 38
asymptotically absolutely equally distributed, 21
asymptotically equally distributed, 17, 56
asymptotically equivalent matrices, 11
asymptotically weakly stationary, 64
autocorrelation matrix, 64
autoregressive process, 68

bounded matrix, 10
bounded Toeplitz matrices, 31

Cauchy-Schwartz inequality, 13, 20
characteristic equation, 5
circulant matrix, 2, 25
conjugate transpose, 6, 70
continuous, 17, 21, 22, 33, 39, 41, 48, 49, 52, 54–57, 67, 69
continuous complex function, 16
convergence
  uniform, 32
Courant-Fischer theorem, 6
covariance matrix, 1, 63, 64
cyclic matrix, 2
cyclic shift, 25

determinant, 17, 31, 60, 71
DFT, 28
diagonal, 8
differential entropy, 73

differential entropy rate, 73
discrete time, 63

eigenvalue, 5, 26, 31
eigenvalue distribution theorem, 40, 48
eigenvector, 5, 26
Euclidean norm, 9

factorization, 70
filter, 1
    linear time invariant, 63
finite order, 31
finite order Toeplitz matrix, 36
Fourier series, 32
    truncated, 44
Fourier transform
    discrete, 28
Frobenius norm, 9
function, analytic, 16

Gaussian process, 73

Hermitian, 6
Hilbert-Schmidt norm, 8, 9

identity matrix, 19
impulse respone, 65
information theory, 73
inverse, 29, 31, 53

Kronecker delta, 27
Kronecker delta response, 65

linear difference equation, 26

matrix
    bounded, 10
    circulant, 2, 25
    covariance, 1
    cyclic, 2
    Hermitian, 6
    normal, 6
    Toeplitz, 26, 31
matrix, Toeplitz, 1
mean, 64
metric, 8
moments, 15
moving average, 65

noncausal, 65
nonnegative definite, 6
nonsingular, 53
norm, 8
    axioms, 10
    Euclidean, 9
    Frobenius, 9
    Hilbert-Schmidt, 8, 9
    operator, 8
    strong, 8, 9
    weak, 8, 9
normal, 6, 8

one-sided, 64
operator norm, 8

polynomials, 16
positive definite, 6
power specral density, 64
power spectral density, 63, 64, 73
probability mass function, 19
product, 29, 31

random process, 63

  discrete time, 64
Rayleigh quotient, 6
Riemann integrable, 48, 53, 61

Shannon information theory, 73
Shur's theorem, 8
spectrum, 53
square summable, 31
Stone-Weierstrass approximation theorem, 16
Stone-Weierstrass theorem, 40
strictly positive definite, 6
sum, 29

Taylor series, 16
time series, 63
Toeplitz determinant, 60
Toeplitz matrix, 1, 26, 31
Toeplitz matrix, finite order, 31
trace, 8
transpose, 26
triangle inequality, 10
triangular, 63, 66, 68, 70, 72
two-sided, 64

uniform convergence, 32
uniformaly bounded, 38
unitary, 6
upper triangular, 6

weak norm, 9
weakly stationary, 64
  asymptotically, 64
white noise, 65
Wielandt-Hoffman theorem, 18, 20
Wiener-Hopf factorization, 63