Toeplitz and Circulant Matrices: A review

Robert M. Gray

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

now

the essence of knowledge

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

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

$$= \begin{pmatrix} x_0 t_0 \\ t_1 x_0 + t_0 x_1 \\ \sum_{i=0}^{2} t_{2-i} x_i \\ \vdots \\ \sum_{i=0}^{n-1} t_{n-1-i} x_i \end{pmatrix}$$

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 $K_n = [K_X(k,j);\, k,j = 0,1,\dots,n-1]$ 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

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

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.

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

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

and eigenvector

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

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,

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

where we have used the orthogonality of the complex exponentials:

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

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

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

is the matrix composed of the eigenvectors as columns, and $\Psi = \mathrm{diag}(\psi_k)$ 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 $y^{(m)} = \frac{1}{\sqrt{n}}\left(1, e^{-2\pi im/n}, \cdots, e^{-2\pi i(n-1)/n}\right)'$, $m = 0, 1, \ldots, n-1$, 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 $\mathrm{diag}(\psi_k)$. 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.