Optimum Array Processing: Part IV of Detection, Estimation, and Modulation Theory
Harry L. Van Trees

ISBNs: 0-471-09390-4 (Hardback); 0-471-22110-4 (Electronic)

Appendix A

Matrix Operations

A.1 Introduction

Most of the analysis that we do in the array processing area utilizes vectors and matrices extensively. In this appendix, we have summarized the definitions and properties that will be useful in the text. We also develop the ideas of matrix algebra and matrix calculus. We assume that the reader has some familiarity with vectors and matrices, and many results are stated without proof.

There are almost no new results in this appendix. We have compiled (and borrowed) results from three types of sources. The first class of sources are books and articles that deal with matrices and linear algebra without a particular application focus. Representative books and articles of this type include Bellman [Bel72], Marcus [Mar60], Hohn [Hoh73], Noble and Daniel [ND77], Lancaster and Tismenetsky [LT85], Golub and Van Loan [GVL89], Grenander and Szego [GS58], Graham [Gra81], Rao and Mitra [RM71].

The second class of sources are books on adaptive antennas, spectral estimation, adaptive filtering, automatic control, or system identification that have included a chapter or appendix on matrices because they need matrix results for their respective applications. Representative books of this type include Kay [Kay88], Hudson [Hud81], Marple [Mar87], Haykin [Hay96], and Scharf [Sch91].

The third class of sources include journal articles or reports where a matrix result was needed for a specific application and has wider usage. Representative sources in this area include Cantoni and Butler [CB76b], [CB76a], Makhoul [Mak81], Stoica and Nehorai [SN89], and Kuroc [Kur89].

When a result appears in multiple references and appears to be wellknown in the "community" we have not included a reference. When we have followed a particular reference closely, we indicate that source.

We have included examples to relate the matrix results to the array processing problem in order to provide motivation. The physical models are explained in more detail in the text.

In general, all vectors and matrices are assumed to be complex. Real matrices and vectors are treated as special cases. The labeling of the matrices; $\mathbf{A}, \mathbf{B}, \mathbf{C}, \cdots$, and vectors $\mathbf{a}, \mathbf{b}, \mathbf{c}, \cdots$, in the appendix does not have a physical significance unless specifically indicated. (This is in contrast to the text.)

A.2 Basic Definitions and Properties

A.2.1 Basic Definitions

We define an $N \times M$ matrix $\mathbf{A}$ by defining its elements $a_{ik}, i = 1, 2, \cdots, N, k = 1, 2, \cdots, M$. We write it in matrix form with N rows and M columns,

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1M} \\ a_{21} & a_{22} & \cdots & a_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ a_{N1} & a_{N2} & \cdots & a_{NM} \end{bmatrix}. \tag{A.1}$$

We also use the notation $[\mathbf{A}]_{ij}$ for the ijth element. The vector $\mathbf{a}$ is an $N \times 1$ matrix,

$$\mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_N \end{bmatrix}. \tag{A.2}$$

The vector in (A.2) is often referred to as a column vector. We will suppress the column designation. The **product** of an $N \times M$ matrix $\mathbf{A}$ and a $M \times L$ matrix $\mathbf{B}$ is an $N \times L$ matrix $\mathbf{C}$ whose elements are given by

$$c_{ij} = \sum_{k=1}^{M} a_{ik} b_{kj}. \tag{A.3}$$

The product of two matrices is defined if and only if the number of columns in the first matrix is equal to the number of rows in the second matrix. When this is true, the matrices are referred to as conformable matrices.

The **transpose** of $\mathbf{A}$ is denoted by the superscript T,

$$\left[\mathbf{A}^T\right]_{ij} = [\mathbf{A}]_{ji} = a_{ji}. \tag{A.4}$$

The transpose has the property

$$(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T. \tag{A.5}$$

The **Hermitian transpose** is denoted by the superscript H,

$$\left[\mathbf{A}^H\right]_{ij} = \left([\mathbf{A}]_{ji}\right)^* = a_{ji}^*. \tag{A.6}$$

The Hermitian transpose has the property

$$(\mathbf{AB})^H = \mathbf{B}^H\mathbf{A}^H. \tag{A.7}$$

The transpose or Hermitian transpose of a vector is a $1 \times N$ matrix, which is referred to as a **row vector**.

If $N = M$, then $\mathbf{A}$ is a **square matrix**. A **symmetric matrix** is a square matrix in which

$$\mathbf{A}^T = \mathbf{A}. \tag{A.8}$$

This implies

$$a_{ij} = a_{ji}, \tag{A.9}$$

so the matrix is symmetric around the principal diagonal. For example, if $N = 4$,

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{12} & a_{22} & a_{23} & a_{24} \\ a_{13} & a_{23} & a_{33} & a_{34} \\ a_{14} & a_{24} & a_{34} & a_{44} \end{bmatrix}. \tag{A.10}$$

A **Hermitian matrix** is a square matrix $\mathbf{A}$ with elements that have complex conjugate symmetry,

$$\mathbf{A}^H = \mathbf{A}. \tag{A.11}$$

This implies

$$a_{ij} = a_{ji}^*. \tag{A.12}$$

For example, if $N = 4$,

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{12}^* & a_{22} & a_{23} & a_{24} \\ a_{13}^* & a_{23}^* & a_{33} & a_{34} \\ a_{14}^* & a_{24}^* & a_{34}^* & a_{44} \end{bmatrix}. \tag{A.13}$$

The diagonal elements of a Hermitian matrix are real.

The **inner product** of two vectors is defined as

$$\alpha = \mathbf{x}^H \mathbf{y}\,; \tag{A.14}$$

when $\mathbf{x} = \mathbf{y}$, α corresponds to the square of the Euclidean norm of the vector (A.36):

$$\alpha = \mathbf{x}^H \mathbf{x}. \tag{A.15}$$

The outer product of complex $N \times 1$ vector $\mathbf{x}$ and a complex $M \times 1$ vector $\mathbf{y}$ is a $N \times M$ matrix $\mathbf{A}$,

$$\mathbf{A} = \mathbf{x}\mathbf{y}^H. \tag{A.16}$$

When $\mathbf{x} = \mathbf{y}$, $\mathbf{A}$ is an $N \times N$ matrix

$$\mathbf{A} = \mathbf{x}\mathbf{x}^H. \tag{A.17}$$

The matrix $\mathbf{A}$ in (A.17) is Hermitian for arbitrary $\mathbf{x}$.

Example A.2.1

The correlation matrix of a complex random vector $\mathbf{x}$ is the expectation of the outer product

$$\mathbf{R_x} \triangleq E\left[\mathbf{x}\mathbf{x}^H\right] = \begin{bmatrix} E[x_1x_1^*] & E[x_1x_2^*] & E[x_1x_3^*] & E[x_1x_4^*] \\ E[x_2x_1^*] & E[x_2x_2^*] & E[x_2x_3^*] & E[x_2x_4^*] \\ E[x_3x_1^*] & E[x_3x_2^*] & E[x_3x_3^*] & E[x_3x_4^*] \\ E[x_4x_1^*] & E[x_4x_2^*] & E[x_4x_3^*] & E[x_4x_4^*] \end{bmatrix}. \tag{A.18}$$

Thus, $\mathbf{R_x}$ is a Hermitian matrix.

If $\mathbf{X}$ is an $N \times K$ matrix ($K \geq N$), then

$$\mathbf{A} = \mathbf{X}\mathbf{X}^H, \tag{A.19}$$

is a Hermitian matrix.

If $\mathbf{A}^H = -\mathbf{A}$, then $\mathbf{A}$ is **skew Hermitian**. The diagonal elements of a skew Hermitian matrix are imaginary. If $\mathbf{A}$ is real and $\mathbf{A}^T = -\mathbf{A}$, then $\mathbf{A}$ is **skew symmetric**.

If $\mathbf{A} = \mathbf{B}_1 + j\mathbf{B}_2$ is a Hermitian matrix where $\mathbf{B}_1$ and $\mathbf{B}_2$ are real, then:

(i) $\mathbf{B}_1$ is real symmetric,

(ii) $\mathbf{B}_2$ is skew symmetric,

(iii) $j\mathbf{B}_2$ is Hermitian.

If $\mathbf{A} = \mathbf{B}_1 + j\mathbf{B}_2$ is skew Hermitian, then:

(i) $\mathbf{B}_1$ is skew symmetric,

(ii) $\mathbf{B}_2$ is symmetric,

(iii) $j\mathbf{B}_2$ is skew Hermitian.

The **determinant** of a square matrix $\mathbf{A}$ is denoted by $\det(\mathbf{A})$ or $|\mathbf{A}|$. To calculate the determinant of $\mathbf{A}$, we evaluate

$$\det(\mathbf{A}) = \sum_{k=1}^{N} a_{ik}C_{ik}, \tag{A.20}$$

where

$$C_{ik} = (-1)^{i+k}M_{ik}, \tag{A.21}$$

and M_{ik} is the determinant of the submatrix of $\mathbf{A}$, which is constructed by deleting the ith row and kth column of $\mathbf{A}$. It is called the **minor** of a_{ik} and C_{ik} is called the **cofactor** of $\mathbf{A}$. M_{ii} is referred to as the ith **principal minor**.

The determinant has the following properties:

$$\det(\mathbf{AB}) = \det(\mathbf{A})\det(\mathbf{B}), \tag{A.22}$$

$$\det(\mathbf{A}^T) = \det(\mathbf{A}), \tag{A.23}$$

$$\det(\mathbf{A}^H) = (\det(\mathbf{A}))^*, \tag{A.24}$$

$$\det(k\mathbf{A}) = k^N \det(\mathbf{A}), \quad k \text{ is a scalar.} \tag{A.25}$$

Other properties of the determinant are given in various sections as they arise.

The **trace** of a square matrix $\mathbf{A}$ is denoted by $\text{tr}(\mathbf{A})$. The trace is the sum of the diagonal elements,

$$\text{tr}(\mathbf{A}) = \sum_{i=1}^{N} a_{ii}. \tag{A.26}$$

The trace has the following properties:

$$\text{tr}(\mathbf{A}+\mathbf{B}) = \text{tr}(\mathbf{A}) + \text{tr}(\mathbf{B}), \tag{A.27}$$

$$\text{tr}(\mathbf{AB}) = \text{tr}(\mathbf{BA}), \tag{A.28}$$

$$\text{tr}(\mathbf{ABC}) = \text{tr}(\mathbf{CAB}) = \text{tr}(\mathbf{BCA}). \tag{A.29}$$

If $\mathbf{x}$ and $\mathbf{y}$ are $N \times 1$ vectors and $\mathbf{A}$ is a square matrix, then

$$\alpha = \mathbf{x}^H\mathbf{A}\mathbf{y}, \tag{A.30}$$

is a scalar and

$$\alpha = \text{tr}(\alpha) = \text{tr}(\mathbf{x}^H\mathbf{A}\mathbf{y}) = \text{tr}(\mathbf{y}\mathbf{x}^H\mathbf{A}) = \text{tr}(\mathbf{A}\mathbf{y}\mathbf{x}^H). \tag{A.31}$$

Two useful inequalities are

$$\left| \text{tr } (\mathbf{A}^H\mathbf{B})\right|^2 \leq \text{ tr } (\mathbf{A}^H\mathbf{A}) \text{ tr } (\mathbf{B}^H\mathbf{B}), \tag{A.32}$$

and

$$\text{tr } (\mathbf{B}^H\mathbf{A}^H\mathbf{A}\mathbf{B}) \leq \text{ tr } (\mathbf{A}^H\mathbf{A}) \text{ tr } (\mathbf{B}^H\mathbf{B}). \tag{A.33}$$

The **rank** of a matrix is the number of linearly independent columns or rows. The rank of a matrix has the following properties:

$$\text{rank } (\mathbf{A}+\mathbf{B}) \leq \text{ rank } (\mathbf{A}) + \text{ rank } (\mathbf{B}), \tag{A.34}$$

and

$$\text{rank } (\mathbf{A}\mathbf{B}) \leq \min(\text{ rank } \mathbf{A}, \text{ rank } \mathbf{B}). \tag{A.35}$$

There are several **norms** for vectors and matrices that are useful. The **2-norm** of an $N \times 1$ vector is defined as

$$\| \mathbf{x} \|_2 \triangleq \left(\sum_{i=1}^{N} |x_i|^2\right)^{\frac{1}{2}} = (\mathbf{x}^H\mathbf{x})^{\frac{1}{2}}. \tag{A.36}$$

The 2-norm is also referred to as the Euclidean norm. More generally, the **p-norm** of an $N \times 1$ vector is

$$\| \mathbf{x} \|_p \triangleq \left(\sum_{i=1}^{N} |x_i|^p\right)^{\frac{1}{p}}, \quad p \geq 1. \tag{A.37}$$

The p-norm is also referred to as the Hölder norm. We use the Euclidean norm throughout the text and for notational simplicity will drop the subscript 2. Thus,

$$\| \mathbf{x} \| \triangleq (\mathbf{x}^H\mathbf{x})^{\frac{1}{2}}. \tag{A.38}$$

The **Frobenius (or Euclidean) norm** of an $N \times N$ matrix is defined as,

$$\| \mathbf{A} \|_F = \left(\sum_{i=1}^{N}\sum_{j=1}^{N} |a_{ij}|^2\right)^{\frac{1}{2}} = \left(\text{tr}\left[\mathbf{A}^H\mathbf{A}\right]\right)^{\frac{1}{2}}. \tag{A.39}$$

A.2.2 Matrix Inverses

If $\mathbf{A}$ is a square $N \times N$ matrix whose rank is equal to N, then we can define the **inverse** of $\mathbf{A}$, which is denoted by $\mathbf{A}^{-1}$ and satisfies

$$\mathbf{A}^{-1}\mathbf{A} = \mathbf{A}\mathbf{A}^{-1} = \mathbf{I}, \tag{A.40}$$

where $\mathbf{I}$ is the **identity matrix**,

$$\mathbf{I} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ 0 & & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}. \tag{A.41}$$

If the inverse does not exist (i.e., rank $\mathbf{A}) < N$), then $\mathbf{A}$ is referred to as a **singular matrix.**

The matrix $\mathbf{A}$ will be singular if and only if

$$\det(\mathbf{A}) = 0. \tag{A.42}$$

The inverse has the following properties:

$$\left(\mathbf{A}^T\right)^{-1} = \left(\mathbf{A}^{-1}\right)^T, \tag{A.43}$$

$$\left(\mathbf{A}^H\right)^{-1} = \left(\mathbf{A}^{-1}\right)^H, \tag{A.44}$$

$$(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}, \tag{A.45}$$

$$\det\left(\mathbf{A}^{-1}\right) = (\det(\mathbf{A}))^{-1}. \tag{A.46}$$

To calculate the inverse of $\mathbf{A}$ we use the formula,

$$\mathbf{A}^{-1} = \frac{\mathbf{C}^T}{\det(\mathbf{A})}, \tag{A.47}$$

where $\mathbf{C}$ is the matrix of cofactors,

$$[\mathbf{C}]_{ik} = C_{ik} = (-1)^{i+k} M_{ik}, \tag{A.48}$$

where M_{ik} is the minor defined following (A.21).

A formula for the inverse that we use frequently in the text is referred to as the **matrix inversion lemma**,

$$(\mathbf{A}+\mathbf{BCD})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}\left(\mathbf{DA}^{-1}\mathbf{B}+\mathbf{C}^{-1}\right)^{-1}\mathbf{DA}^{-1}, \tag{A.49}$$

where $\mathbf{A}$ is $N \times N$, $\mathbf{B}$ is $N \times M$, $\mathbf{C}$ is $M \times M$, and $\mathbf{D}$ is $M \times N$ and the requisite inverses are assumed to exist.

A special case of (A.49) is referred to as **Woodbury's identity**. Here, $\mathbf{B}$ is an $N \times 1$ column vector $\mathbf{x}$, $\mathbf{C}$ is a scalar equal to unity, and $\mathbf{D} = \mathbf{x}^H$. Then

$$\left(\mathbf{A}+\mathbf{x}\mathbf{x}^H\right)^{-1} = \mathbf{A}^{-1} - \frac{\mathbf{A}^{-1}\mathbf{x}\mathbf{x}^H\mathbf{A}^{-1}}{1+\mathbf{x}^H\mathbf{A}^{-1}\mathbf{x}}. \tag{A.50}$$

Several other inverse relations that follow from (A.49) are:

$$\left(\sigma^2\mathbf{I}+\mathbf{VSV}^H\right)^{-1} = \frac{1}{\sigma^2}\left(\mathbf{I}-\mathbf{V}\left(\mathbf{V}^H\mathbf{V}+\sigma^2\mathbf{S}^{-1}\right)^{-1}\mathbf{V}^H\right), \tag{A.51}$$

$$\left[\mathbf{A}^{-1}+\mathbf{B}^H\mathbf{C}^{-1}\mathbf{B}\right]^{-1} = \mathbf{A} - \mathbf{AB}^H\left[\mathbf{BAB}^H+\mathbf{C}\right]^{-1}\mathbf{BA}, \tag{A.52}$$

$$\left[\mathbf{A}^{-1}+\mathbf{B}^H\mathbf{C}^{-1}\mathbf{B}\right]^{-1}\mathbf{B}^H\mathbf{C}^{-1} = \mathbf{AB}^H\left[\mathbf{BAB}^H+\mathbf{C}\right]^{-1}, \tag{A.53}$$

and

$$\mathbf{C}^{-1} - \left[\mathbf{BAB}^H+\mathbf{C}\right]^{-1} = \mathbf{C}^{-1}\mathbf{B}\left[\mathbf{A}^{-1}+\mathbf{B}^H\mathbf{C}^{-1}\mathbf{B}\right]^{-1}\mathbf{B}^H\mathbf{C}^{-1}. \tag{A.54}$$

A.2.3 Quadratic Forms

A **Hermitian quadratic form** Q is defined as

$$Q = \mathbf{x}^H\mathbf{A}\mathbf{x}, \tag{A.55}$$

where $\mathbf{A}$ is a Hermitian matrix and $\mathbf{x}$ is a complex $N \times 1$ vector. Because $\mathbf{A}$ is Hermitian, Q is a real scalar.

$\mathbf{A}$ is positive definite if

$$\mathbf{x}^H \mathbf{A} \mathbf{x} > 0, \tag{A.56}$$

for all $\mathbf{x} \neq 0$.

$\mathbf{A}$ is positive semidefinite (or nonnegative definite) if

$$\mathbf{x}^H \mathbf{A} \mathbf{x} \geq 0, \tag{A.57}$$

for all $\mathbf{x} \neq 0$.

A square $N \times N$ matrix is positive definite if and only if

$$\mathbf{A} = \mathbf{C}\mathbf{C}^H, \tag{A.58}$$

where $\mathbf{C}$ is $N \times N$ and has rank N or its principal minors are all positive.

If $\mathbf{A}$ is positive definite, then

$$\mathbf{A}^{-1} = \left(\mathbf{C}^{-1}\right)^H \mathbf{C}^{-1}. \tag{A.59}$$

The factoring in (A.58) plays an important role in many applications. In Section A.5, we develop techniques for finding $\mathbf{C}$.

If $\mathbf{A}$ is positive definite and $\mathbf{B}$ is an $M \times N$ matrix of rank $M (M \leq N)$, then

$$\mathbf{B}\mathbf{A}\mathbf{B}^H, \tag{A.60}$$

is positive definite.

If $\mathbf{A}$ is positive definite, then the diagonal elements are all positive and $\det(\mathbf{A})$ is positive.

If $\mathbf{A}$ satisfies (A.58) and $\mathbf{C}$ is an $M \times N$ matrix with $M < N$, then $\mathbf{A}$ is positive semi-definite.

A.2.4 Partitioned Matrices

A **partitioned matrix A** is an $N \times M$ matrix that can be written in terms of submatrices. For example,

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}, \tag{A.61}$$

where the dimensions of the submatrices are

$$\mathbf{A}_{11} \qquad N_1 \times M_1, \tag{A.62}$$

$$\mathbf{A}_{21} \qquad (N - N_1) \times M_1, \tag{A.63}$$

$$\mathbf{A}_{12} \qquad N_1 \times (M - M_1), \tag{A.64}$$

$$\mathbf{A}_{22} \qquad (N - N_1) \times (M - M_1). \tag{A.65}$$

The advantage of utilizing partitioned matrices is that they can be operated upon by treating the submatrices as elements.

Four common operations are:

(**i**) Multiplication

$$\begin{aligned}\mathbf{AB} &= \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{A}_{11}\mathbf{B}_{11} + \mathbf{A}_{12}\mathbf{B}_{21} & \mathbf{A}_{11}\mathbf{B}_{12} + \mathbf{A}_{12}\mathbf{B}_{22} \\ \mathbf{A}_{21}\mathbf{B}_{11} + \mathbf{A}_{22}\mathbf{B}_{21} & \mathbf{A}_{21}\mathbf{B}_{12} + \mathbf{A}_{22}\mathbf{B}_{22} \end{bmatrix},\end{aligned} \tag{A.66}$$

where matrices in the indicated products must be conformable.

(**ii**) Conjugate Transposition

$$\begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}^H = \begin{bmatrix} \mathbf{A}_{11}^H & \mathbf{A}_{21}^H \\ \mathbf{A}_{12}^H & \mathbf{A}_{22}^H \end{bmatrix}. \tag{A.67}$$

(**iii**) Inverse ($\mathbf{A}$ is $N \times N$)

$$\mathbf{A}^{-1} = \begin{bmatrix} \left(\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}\right)^{-1} & -\left(\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}\right)^{-1}\mathbf{A}_{12}\mathbf{A}_{22}^{-1} \\ -\left(\mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\right)^{-1}\mathbf{A}_{21}\mathbf{A}_{11}^{-1} & \left(\mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\right)^{-1} \end{bmatrix}, \tag{A.68}$$

where $\mathbf{A}$ is given by (A.61) and $\mathbf{A}_{11}^{-1}$ and $\mathbf{A}_{22}^{-1}$ exist.

A special case of (A.68) occurs when $\mathbf{A}_{11}$ is $(N-1)\times(N-1)$, $\mathbf{A}_{12}$ is a $(N-1)\times 1$ column vector $\mathbf{a}_2$, $\mathbf{A}_{21}$ is $1\times(N-1)$ row vector $\mathbf{a}_1^H$, and $\mathbf{A}_{22}$ is a scalar c. Then,

$$\mathbf{A}^{-1} = \begin{bmatrix} \mathbf{A}_{11}^{-1} + \beta\mathbf{A}_{11}^{-1}\mathbf{a}_2\mathbf{a}_1^H\mathbf{A}_{11}^{-1} & -\beta\mathbf{A}_{11}^{-1}\mathbf{a}_2 \\ -\beta\mathbf{a}_1^H\mathbf{A}_{11}^{-1} & \beta \end{bmatrix}, \tag{A.69}$$

where $\beta = (\alpha - \mathbf{a}_1^H\mathbf{A}_{11}^{-1}\mathbf{a}_2)$. The result in (A.69) is useful when a new matrix is created by bordering the original matrix $\mathbf{A}_{11}$ with an additional row and column. The inverse of the **border** matrix can be expressed in terms of $\mathbf{A}_{11}^{-1}$ and the new row and column.

(**iv**) Determinant

$$\begin{aligned} \det(\mathbf{A}) &= \det(\mathbf{A}_{22})\det(\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}) \\ &= \det(\mathbf{A}_{11})\det(\mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}). \end{aligned} \tag{A.70}$$

A.2.5 Matrix products

In this section we define several matrix products that we find useful in our analyses.

Hadamard Product

The **Hadamard product** of two $N \times M$ dimensional matrices $\mathbf{A}$ and $\mathbf{B}$ is defined as[1]

$$\mathbf{A} \odot \mathbf{B} \triangleq \begin{bmatrix} a_{11}b_{11} & a_{12}b_{12} & \cdots & a_{1M}b_{1M} \\ a_{21}b_{21} & a_{22}b_{22} & \ddots & a_{2M}b_{2M} \\ \vdots & \vdots & \cdots & \vdots \\ a_{N1}b_{N1} & a_{N2}b_{N2} & \cdots & a_{NM}b_{NM} \end{bmatrix}. \tag{A.71}$$

We see that the Hadamard product is obtained by element-by-element multiplication.

Several properties will be useful:

[1]The Hadamard product is sometimes referred to as the Hadamard-Schur product.

(i)

$$\mathbf{A} \odot \mathbf{B} = \mathbf{B} \odot \mathbf{A}, \tag{A.72}$$

(ii)

$$(\mathbf{A} \odot \mathbf{B}) \odot \mathbf{C} = \mathbf{A} \odot (\mathbf{B} \odot \mathbf{C}), \tag{A.73}$$

(iii)

$$(\mathbf{A} \odot \mathbf{B})^T = \mathbf{A}^T \odot \mathbf{B}^T, \tag{A.74}$$

(iv)

$$\mathbf{a}\mathbf{b}^T \odot \mathbf{c}\mathbf{d}^T = (\mathbf{a} \odot \mathbf{c})(\mathbf{b} \odot \mathbf{d})^T, \tag{A.75}$$

(v)

$$\mathbf{E}_{qr}\mathbf{A} \odot \mathbf{B}\mathbf{E}_{st} = [\mathbf{A}]_{rt}[\mathbf{B}]_{qs}\mathbf{E}_{qt}. \tag{A.76}$$

(The matrix $\mathbf{E}_{ij}$ is defined as a matrix whose ij component is 1 and all other components are 0.)

(**vi**) If $\mathbf{A}$ and $\mathbf{B}$ are positive semi-definite, then

$$\mathbf{A} \odot \mathbf{B} \text{ is positive semi-definite} \qquad \text{([SN89] [Bel72])}. \tag{A.77}$$

(**vii**) If $\mathbf{A}, \mathbf{B}$ and $\mathbf{C}$ are Hermitian positive semi-definite matrices, then

$$[Re(\mathbf{A} \odot \mathbf{B})]^{-1}[Re(\mathbf{A} \odot \mathbf{C})][Re(\mathbf{A} \odot \mathbf{B})]^{-1} \geq \left\{Re\left[\mathbf{A} \odot \mathbf{B}\mathbf{C}^{-1}\mathbf{B}\right]\right\}^{-1} \tag{A.78}$$

where the inverses are assumed to exist ([SN89]).

(**viii**) If $\mathbf{A}$ is a Hermitian positive definite matrix, then

$$\left(\mathbf{I} \odot \mathbf{A}^{-1}\right) \geq (\mathbf{I} \odot \mathbf{A})^{-1}. \tag{A.79}$$

Kronecker product

If $\mathbf{A}$ is a $N \times M$ matrix and $\mathbf{B}$ is a $K \times L$ matrix, the **Kronecker product** is defined to be the $NK \times ML$ matrix,

$$\mathbf{A} \otimes \mathbf{B} \triangleq \begin{bmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & \cdots & a_{1M}\mathbf{B} \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} & \cdots & a_{2M}\mathbf{B} \\ \vdots & \vdots & \ddots & \vdots \\ a_{N1}\mathbf{B} & a_{N2}\mathbf{B} & \cdots & a_{NM}\mathbf{B} \end{bmatrix}, \tag{A.80}$$

(e.g., Chapter 12 [LT85] or [Gra81]).

Several properties will be useful:

(**i**) If $\mathbf{B}$ is an $M \times M$ matrix, then,

$$\mathbf{I}_N \otimes \mathbf{B} = \mathrm{diag}\left[\mathbf{B}, \mathbf{B}, \cdots, \mathbf{B}\right]. \tag{A.81}$$

(**ii**) If $\mathbf{A}$ is an $M \times M$ matrix, then,

$$\mathbf{A} \otimes \mathbf{I}_N = \begin{bmatrix} a_{11}\mathbf{I}_N & a_{12}\mathbf{I}_N & \cdots & a_{1M}\mathbf{I}_N \\ \vdots & & & \\ a_{M1}\mathbf{I}_N & a_{M2}\mathbf{I}_N & \cdots & a_{MM}\mathbf{I}_N \end{bmatrix}. \tag{A.82}$$

(**iii**)

$$\mathbf{I}_M \otimes \mathbf{I}_N = \mathbf{I}_{MN}. \tag{A.83}$$

(**iv**)

$$\mathbf{A} \otimes (\alpha\mathbf{B}) = (\alpha\mathbf{A}) \otimes \mathbf{B} = \alpha\left(\mathbf{A} \otimes \mathbf{B}\right). \tag{A.84}$$

(**v**)

$$\left(\mathbf{A} \otimes \mathbf{B}\right)^H = \mathbf{A}^H \otimes \mathbf{B}^H. \tag{A.85}$$

(**vi**)

$$\left(\mathbf{A} + \mathbf{B}\right) \otimes \mathbf{C} = \left(\mathbf{A} \otimes \mathbf{C}\right) + \left(\mathbf{B} \otimes \mathbf{C}\right). \tag{A.86}$$

(**vii**)

$$\mathbf{A} \otimes (\mathbf{B} + \mathbf{C}) = (\mathbf{A} \otimes \mathbf{B}) + (\mathbf{A} \otimes \mathbf{C}). \tag{A.87}$$

(**viii**)

$$\mathbf{A} \otimes (\mathbf{B} \otimes \mathbf{C}) = (\mathbf{A} \otimes \mathbf{B}) \otimes \mathbf{C}. \tag{A.88}$$

(**ix**)

$$(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = \mathbf{AC} \otimes \mathbf{BD}. \tag{A.89}$$

(**x**)

$$(\mathbf{A} \otimes \mathbf{B})^{-1} = \mathbf{A}^{-1} \otimes \mathbf{B}^{-1}. \tag{A.90}$$

(**xi**) If $\mathbf{A}$ is $M \times M$ and $\mathbf{B}$ is $N \times N$, then

$$\begin{aligned} (\mathbf{A} \otimes \mathbf{B}) &= (\mathbf{A} \otimes \mathbf{I}_N)(\mathbf{I}_M \otimes \mathbf{B}) \\ &= (\mathbf{I}_M \otimes \mathbf{B})(\mathbf{A} \otimes \mathbf{I}_N). \end{aligned} \tag{A.91}$$

(**xii**) If $\mathbf{A}_1, \mathbf{A}_2, \cdots \mathbf{A}_p$ are $M \times M$ and $\mathbf{B}_1, \mathbf{B}_2, \cdots \mathbf{B}_p$ are $N \times N$, then

$$(\mathbf{A}_1 \otimes \mathbf{B}_1)(\mathbf{A}_2 \otimes \mathbf{B}_2) \cdots (\mathbf{A}_p \otimes \mathbf{B}_p) = (\mathbf{A}_1\mathbf{A}_2 \cdots \mathbf{A}_p) \otimes (\mathbf{B}_1\mathbf{B}_2 \cdots \mathbf{B}_p). \tag{A.92}$$

(**xiii**) If $\mathbf{A}$ is $M \times M$ and $\mathbf{B}$ is $N \times N$, then

$$\det(\mathbf{A} \otimes \mathbf{B}) = (\det \mathbf{A})^M (\det \mathbf{B})^N. \tag{A.93}$$

(**xiv**) If $\mathbf{A}$ is $M \times M$ and $\mathbf{B}$ is $N \times N$

$$\operatorname{tr}(\mathbf{A} \otimes \mathbf{B}) = (\operatorname{tr} \mathbf{A})(\operatorname{tr} \mathbf{B}). \tag{A.94}$$

(**xv**) If $\mathbf{A}$ is $M \times M$ and $\mathbf{B}$ is $N \times N$

$$\operatorname{rank}(\mathbf{A} \otimes \mathbf{B}) = (\operatorname{rank} \mathbf{A})(\operatorname{rank} \mathbf{B}). \tag{A.95}$$

Khatri-Rao product

The **Khatri-Rao product** of an $N \times M$ matrix $\mathbf{A}$ and a $P \times M$ matrix $\mathbf{B}$ is defined as the $NP \times M$ matrix

$$\mathbf{A}\square\mathbf{B} \triangleq \left[\begin{array}{c:c:c:c} \mathbf{a}_1 \otimes \mathbf{b}_1 & \mathbf{a}_2 \otimes \mathbf{b}_2 & \cdots & \mathbf{a}_N \otimes \mathbf{b}_N \end{array} \right], \tag{A.96}$$

where $\mathbf{a}_j$ denotes the jth column of the matrix $\mathbf{A}$.

The first element is

$$\mathbf{a}_1 \otimes \mathbf{b}_1 = \begin{bmatrix} a_{11}\mathbf{b}_1 \\ \vdots \\ a_{N1}\mathbf{b}_1 \end{bmatrix} = \begin{bmatrix} a_{11}b_{11} \\ a_{11}b_{21} \\ \vdots \\ a_{N1}b_{P-1,1} \\ a_{N1}b_{P1} \end{bmatrix}. \tag{A.97}$$

The remaining elements have the same structure.

Several properties will be useful:

(i)

$$(\mathbf{A} \otimes \mathbf{B})(\mathbf{C}\square\mathbf{D}) = \mathbf{AC}\square\mathbf{BD}, \tag{A.98}$$

([RM71]).

(ii)

$$(\mathbf{A}\square\mathbf{B})^H(\mathbf{C}\square\mathbf{D}) = \mathbf{A}^H\mathbf{C} \odot \mathbf{B}^H\mathbf{D}, \tag{A.99}$$

([Kur89]),

(iii)

$$(\mathbf{A}\square\mathbf{B})^H(\mathbf{C} \otimes \mathbf{D})(\mathbf{E}\square\mathbf{F}) = (\mathbf{A}\square\mathbf{B})^H(\mathbf{CE}\square\mathbf{DF}) = \mathbf{A}^H\mathbf{CE} \odot \mathbf{B}^H\mathbf{DF}, \tag{A.100}$$

([Kur89]).

A.2.6 Matrix Inequalities

If $\mathbf{A}$ and $\mathbf{B}$ are $N \times N$ matrices, then the inequality,

$$\mathbf{A} > \mathbf{B} \tag{A.101}$$

means that $\mathbf{A} - \mathbf{B}$ is a positive definite matrix.

Similarly, the inequality

$$\mathbf{A} \geq \mathbf{B} \tag{A.102}$$

means that $\mathbf{A} - \mathbf{B}$ is a non-negative definite matrix.

A.3 Special Vectors and Matrices

In this section we define some special vectors and matrices that we encounter in our analyses.

A.3.1 Elementary Vectors and Matrices

The zero vector $\mathbf{0}_N$ is a $N \times 1$ vector whose elements are all zero. The vector $\mathbf{1}_N$ is a $N \times 1$ vector whose elements are all unity.

The vector $\mathbf{e}_j$ is a $N \times 1$ vector whose jth element is unity and whose remaining elements are zero. The dimension of $\mathbf{e}_j$ is inferred from its usage or is specified. The vector $\mathbf{e}_j$ is referred to as an **elementary** vector.

If a $N \times N$ matrix $\mathbf{A}$ is written as

$$\mathbf{A} = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_N \end{bmatrix}, \tag{A.103}$$

then

$$\mathbf{A}\mathbf{e}_i\mathbf{e}_i^T = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \cdots & \mathbf{a}_i & \cdots & \mathbf{0} \end{bmatrix}. \tag{A.104}$$

If a $N \times N$ matrix $\mathbf{B}$ is written as

$$\mathbf{B} = \begin{bmatrix} \mathbf{b}_1^T \\ \mathbf{b}_2^T \\ \vdots \\ \mathbf{b}_N^T \end{bmatrix}, \tag{A.105}$$

then

$$\mathbf{e}_i \mathbf{e}_i^T \mathbf{B} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \vdots \\ \mathbf{b}_N^T \\ \vdots \\ \mathbf{0} \end{bmatrix} . \tag{A.106}$$

The matrix $\mathbf{E}_{ij}$ is a matrix whose ijth element is 1 and all other elements are 0.

$$\mathbf{E}_{ij} = \mathbf{e}_i \mathbf{e}_j^T . \tag{A.107}$$

The $N \times N$ matrix $\mathbf{E}^{(ij)}$ with $i < j$ is defined as

$$\mathbf{E}^{(ij)} = \begin{array}{c} \\ \\ \\ (i) \\ \\ \\ \\ (j) \\ \\ \\ \\ \end{array} \begin{bmatrix} 1 & & & & & & & & & & \\ & \ddots & & & & & & & & & \\ & & 1 & & & & & & & & \\ & & & 0 & & \cdots & & 1 & & & \\ & & & & 1 & & & & & & \\ & & & \vdots & & \ddots & & \vdots & & & \\ & & & & & & 1 & & & & \\ & & & 1 & & \cdots & & 0 & & & \\ & & & & & & & & 1 & & \\ & & & & & & & & & \ddots & \\ & & & & & & & & & & 1 \end{bmatrix} . \tag{A.108}$$

Pre-multiplication by $\mathbf{E}^{(ij)}$ interchanges row i and row j. Post-multiplication by $\mathbf{E}^{(ij)}$ interchanges column i and column j.

The $N \times N$ matrix $\mathbf{E}^{(i)}$ is defined as

$$\mathbf{E}^{(i)} = \begin{array}{c} \\ \\ \\ (i) \\ \\ \\ \\ \end{array} \begin{bmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & & & \\ & & & k & & & \\ & & & & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{bmatrix} , \tag{A.109}$$

where k is a non-zero constant. Pre-multiplication by $\mathbf{E}^{(i)}$ multiplies the ith row by k. Post-multiplication by $\mathbf{E}^{(i)}$ multiplies the ith column by k.

The $N \times N$ matrix, $\mathbf{E}^{(i+kj)}$ is defined as:

$$\mathbf{E}^{(i+kj)} = \begin{array}{c} \\ \\ (i) \\ \\ \\ (j) \\ \\ \\ \end{array} \begin{bmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & \cdots & k & & \\ & & & 1 & \vdots & & \\ & & & \ddots & & & \\ & & & & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{bmatrix}, \tag{A.110}$$

for $i < j$, and

$$\mathbf{E}^{(i+kj)} = \begin{array}{c} \\ \\ (i) \\ \\ (j) \\ \\ \\ \end{array} \begin{bmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & & & \\ & & \vdots & \ddots & & & \\ & & k & \cdots & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{bmatrix}, \tag{A.111}$$

for $i > j$.

Pre-multiplication by $\mathbf{E}^{(i+kj)}$ adds k times the jth row to the ith row. Post-multiplication by $\mathbf{E}^{(i+kj)}$ adds k times the jth column to the ith column.

An $N \times N$ matrix $\mathbf{P}_{per}$ is a **permutation matrix** if it can be obtained from the identity matrix $\mathbf{I}_N$ by interchanges of rows or columns.

A.3.2 The $vec(\mathrm{A})$ matrix

Consider a $N \times M$ matrix $\mathbf{A}$ whose jth column is $\mathbf{a}_j$. The vec-function of $\mathbf{A}$ is written as $vec(\mathbf{A})$ and is obtained by stacking the columns to obtain an $NM \times 1$ vector:

$$vec(\mathbf{A}) = \begin{bmatrix} \mathbf{a}_1 \\ --- \\ \mathbf{a}_2 \\ --- \\ \vdots \\ \mathbf{a}_M \end{bmatrix}. \tag{A.112}$$

The function $vec(\mathbf{A})$ is closely related to the Kronecker product and will be useful when we study planar arrays.

Several properties will be useful:

(i)

$$vec\,(\alpha\mathbf{A}+\beta\mathbf{B}) = \alpha\, vec(\mathbf{A}) + \beta\, vec(\mathbf{B}), \tag{A.113}$$

where $\mathbf{A}$ and $\mathbf{B}$ are $N \times M$.

(ii) If $\mathbf{A}$ is $M \times M$, $\mathbf{B}$ is $M \times N$, and $\mathbf{C}$ is $N \times N$, then

$$vec\,(\mathbf{ABC}) = \left(\mathbf{C}^T \otimes \mathbf{A}\right) vec(\mathbf{B}). \tag{A.114}$$

(iii)

$$vec\,(\mathbf{AB}) = (\mathbf{I}_N \otimes \mathbf{A})\, vec(\mathbf{B}). \tag{A.115}$$

(iv)

$$vec\,(\mathbf{BC}) = \left(\mathbf{C}^T \otimes \mathbf{I}_M\right) vec(\mathbf{B}). \tag{A.116}$$

(v)

$$\operatorname{tr}[\mathbf{ABCD}] = vec^H \mathbf{B}^H \left(\mathbf{A}^T \odot \mathbf{C}\right) vec\,[\mathbf{D}]\,. \tag{A.117}$$

A.3.3 Diagonal Matrices

A **diagonal matrix** is a square $N \times N$ matrix with $a_{ij} = 0$ for $i \neq j$. All elements of the principal diagonal are zero. Thus, $\mathbf{A}$ can be written as

$$\mathbf{A} = \begin{bmatrix} a_{11} & 0 & \cdots & 0 \\ 0 & a_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{bmatrix}. \tag{A.118}$$

We also write

$$\mathbf{A} = \text{ diag } [a_{11}, a_{22}, \cdots, a_{nn}] . \tag{A.119}$$

The inverse is

$$\mathbf{A}^{-1} = \text{ diag } \left[a_{11}^{-1}, a_{22}^{-1}, \cdots, a_{nn}^{-1}\right] . \tag{A.120}$$

A **block diagonal matrix** is a square $N \times N$ block matrix,

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & 0 & \cdots & 0 \\ 0 & \mathbf{A}_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{A}_{mm} \end{bmatrix}, \tag{A.121}$$

where the submatrices $\mathbf{A}_{ii}$ are square and all other submatrices are zero. The dimensions of the $\mathbf{A}_{ii}$ need not be the same. We also write

$$\mathbf{A} = \text{ block diag } [\mathbf{A}_{11}, \mathbf{A}_{22}, \cdots, \mathbf{A}_{mm}] . \tag{A.122}$$

If the $\mathbf{A}_{ii}$ are non-singular, then

$$\mathbf{A}^{-1} = \text{ block diag } \left[\mathbf{A}_{11}^{-1}, \mathbf{A}_{22}^{-1}, \cdots, \mathbf{A}_{mm}^{-1}\right] . \tag{A.123}$$

The identity matrix defined in (A.40) is a special case of diagonal matrix whose diagonal elements equal one:

$$\mathbf{I} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} . \tag{A.124}$$

When it is necessary for clarity, we denote the order of the identity matrix by a subscript (e.g., $\mathbf{I}_N$).

A.3.4 Exchange Matrix and Conjugate Symmetric Vectors

The **exchange** or **reflection matrix J** is a square $N \times N$ matrix whose elements on the cross diagonal are unity and all other elements are zero. Thus,

$$\mathbf{J} = \begin{bmatrix} 0 & \cdots & 0 & 1 \\ 0 & \cdots & 1 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 1 & \cdots & 0 & 0 \end{bmatrix}. \tag{A.125}$$

When it is necessary for clarity we add the subscript N to denote the order of the exchange matrix (e.g., $\mathbf{J}_N$). We observe that $\mathbf{J}$ is symmetric, $\mathbf{J}^2 = \mathbf{I}$, and $\mathbf{J}^T = \mathbf{J}$.

Applying $\mathbf{J}$ to a vector exchanges the order of the elements,

$$\mathbf{J}\mathbf{x} = \begin{bmatrix} x_n \\ x_{n-1} \\ \vdots \\ x_2 \\ x_1 \end{bmatrix}. \tag{A.126}$$

We say that a vector $\mathbf{x}$ is **conjugate symmetric** if

$$\mathbf{x} = \mathbf{J}\mathbf{x}^*. \tag{A.127}$$

If $\mathbf{x}$ is $N \times 1$ and N is even, then we can denote the first $N/2$ elements as $\mathbf{x}_1$ and write

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{J}\mathbf{x}_1^* \end{bmatrix}. \tag{A.128}$$

If N is odd, the first $(N-1)/2$ elements are denoted by $\mathbf{x}_1$ and

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_1 \\ x_0 \\ \mathbf{J}\mathbf{x}_1^* \end{bmatrix}, \tag{A.129}$$

where x_0 is a real scalar.

A vector $\mathbf{x}$ is **conjugate asymmetric** if

$$\mathbf{x} = -\mathbf{J}\mathbf{x}^*. \tag{A.130}$$

When $\mathbf{J}$ is applied to a matrix, it reverses the rows or columns of the matrix. Pre-multiplying by $\mathbf{J}$ reverses the order of the rows,

$$\mathbf{J}_N\mathbf{A} = \begin{bmatrix} a_{N1} & a_{N2} & \cdots & a_{NM} \\ \vdots & \vdots & & \vdots \\ a_{21} & a_{22} & \cdots & a_{2M} \\ a_{11} & a_{12} & \cdots & a_{1M} \end{bmatrix}. \tag{A.131}$$

Post-multiplying by $\mathbf{J}$ reverses the order of the columns,

$$\mathbf{A}\mathbf{J}_M = \begin{bmatrix} a_{1M} & \cdots & a_{12} & a_{11} \\ a_{2M} & \cdots & a_{22} & a_{21} \\ \vdots & & \vdots & \vdots \\ a_{NM} & \cdots & a_{N2} & a_{N1} \end{bmatrix}. \tag{A.132}$$

A.3.5 Persymmetric and Centrohermitian Matrices

A **persymmetric matrix** is an $N \times N$ matrix that is symmetric about its cross diagonal. For $N = 4$,

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{13} \\ a_{31} & a_{32} & a_{22} & a_{12} \\ a_{41} & a_{31} & a_{21} & a_{11} \end{bmatrix}. \tag{A.133}$$

This implies

$$a_{ij} = a_{N-j+1,N-i+1}. \tag{A.134}$$

It follows that

$$\mathbf{A}^T = \mathbf{JAJ}, \tag{A.135}$$

and

$$\mathbf{A} = \mathbf{JA}^T\mathbf{J}. \tag{A.136}$$

A **centrohermitian matrix** is an $N \times N$ matrix with the property that

$$a_{ij} = a^*_{N-i+1,N-j+1}. \tag{A.137}$$

A centrohermitian matrix that is also Hermitian exhibits a double symmetry. It is Hermitian around the principal diagonal and persymmetric around the cross diagonal.

$$a_{ij} = a^*_{ji} = a_{N-j+1,N-i+1} = a^*_{N-i+1,N-j+1}. \tag{A.138}$$

For $N = 4$,

$$\mathbf{A} = \begin{bmatrix} a_{11} & a^*_{21} & a^*_{31} & a^*_{41} \\ a_{21} & a_{22} & a^*_{32} & a^*_{31} \\ a_{31} & a_{32} & a_{22} & a^*_{21} \\ a_{41} & a_{31} & a_{21} & a_{11} \end{bmatrix}. \tag{A.139}$$

For centrohermitian-persymmetric matrices,

$$\mathbf{A} = \mathbf{JA}^*\mathbf{J}. \tag{A.140}$$

If N is even, a centrohermitian-persymmetric matrix may be partitioned as

$$\mathbf{A} = \left[\begin{array}{c:c} \mathbf{B}_1 & \mathbf{B}_2 \\ --- & --- \\ \mathbf{JB}_2^*\mathbf{J} & \mathbf{JB}_1^*\mathbf{J} \end{array}\right], \tag{A.141}$$

where the partitioned matrices have dimensions $\frac{N}{2} \times \frac{N}{2}$.

If N is odd, a centrohermitian-persymmetric matrix may be partitioned as

$$\mathbf{A} = \begin{bmatrix} \mathbf{B}_1 & \mathbf{a} & \mathbf{B}_2 \\ \mathbf{a}^H & c & \mathbf{a}^H\mathbf{J} \\ \mathbf{JB}_2^*\mathbf{J} & \mathbf{Ja} & \mathbf{JB}_2^*\mathbf{J} \end{bmatrix}, \tag{A.142}$$

where $\mathbf{B}_1$ is a $(N-1)/2 \times (N-1)/2$ matrix, $\mathbf{B}_2$ is a $(N-1)/2 \times (N-1)/2$ matrix, $\mathbf{a}$ is a $(N-1)/2 \times 1$ vector, and c is a real constant.

These matrices will arise frequently in our study of linear arrays and their structure will lead to significant computational savings and analytic simplifications.

A.3.6 Toeplitz and Hankel Matrices

A **Toeplitz matrix** has the property that all of the elements along each diagonal are identical. Thus,

$$a_{ij} = a_{i-j}. \tag{A.143}$$

For example,

$$\mathbf{A} = \begin{bmatrix} a_0 & a_{-1} & a_{-2} & a_{-3} & a_{-4} \\ a_1 & a_0 & a_{-1} & a_{-2} & a_{-3} \\ a_2 & a_1 & a_0 & a_{-1} & a_{-2} \\ a_3 & a_2 & a_1 & a_0 & a_{-1} \end{bmatrix}. \tag{A.144}$$

If $\mathbf{A}$ is square, then it is a special case of a persymmetric matrix. If $\mathbf{A}$ is also Hermitian, then it is centrohermitian. The inverse of a Toeplitz matrix is persymmetric.

A **Hankel matrix** has the property that the elements along every cross diagonal are equal. Thus,

$$a_{ij} = a_{i+j-N-1}. \tag{A.145}$$

For example,

$$\mathbf{A} = \begin{bmatrix} a_{-3} & a_{-2} & a_{-1} & a_0 \\ a_{-2} & a_{-1} & a_0 & a_1 \\ a_{-1} & a_0 & a_1 & a_2 \\ a_0 & a_1 & a_2 & a_3 \\ a_1 & a_2 & a_3 & a_4 \end{bmatrix}. \tag{A.146}$$

Note that **JA** and **AJ** are Toeplitz matrices when **A** is a Hankel matrix.

A.3.7 Circulant Matrices

A **circulant matrix** is an $N \times N$ square matrix made up of N elements.

The elements of a right-circulant matrix obey the relationship

$$a_{R,ij} = \begin{cases} a_{j-i}, & j-i \geq 0, \\ a_{N-j+1}, & j-i < 0, \quad 1 \leq i, \quad j \leq N. \end{cases} \tag{A.147}$$

For example,

$$\mathbf{A}_R = \begin{bmatrix} a_0 & a_1 & a_2 & a_3 \\ a_3 & a_0 & a_1 & a_2 \\ a_2 & a_3 & a_0 & a_1 \\ a_1 & a_2 & a_3 & a_0 \end{bmatrix}. \tag{A.148}$$

Each row is obtained from the row above it by shifting each element right one column and bringing the last element on the right to the first column. A right-circulant matrix is a special case of a Toeplitz matrix.

The elements of a left-circulant matrix obey the relationship

$$a_{L,ij} = \begin{cases} a_{N+1-i-j}, & j+i \leq N+1, \\ a_{2N+1-i-j}, & j+i > N+1, \quad 1 \leq i, \quad j \leq N. \end{cases} \tag{A.149}$$

For example,

$$\mathbf{A}_L = \begin{bmatrix} a_3 & a_2 & a_1 & a_0 \\ a_2 & a_1 & a_0 & a_3 \\ a_1 & a_0 & a_3 & a_2 \\ a_0 & a_3 & a_2 & a_1 \end{bmatrix}. \tag{A.150}$$

Each row is obtained from the row above it by shifting each element left one column and bringing the last element on the left to the last column.

A left-circulant matrix is a special case of a Hankel matrix.

A.3.8 Triangular Matrices

A **lower triangular** square $N \times N$ matrix is defined as a matrix whose elements above the main diagonal are zero. Thus, we can write,

$$\mathbf{L} = \begin{bmatrix} l_{11} & 0 & \cdots & 0 \\ l_{21} & l_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ l_{n1} & l_{n2} & \cdots & l_{nn} \end{bmatrix}. \tag{A.151}$$

$\mathbf{L}^{-1}$ is also lower triangular. The determinant is

$$\det(\mathbf{L}) = \prod_{i=1}^{N} l_{ii}. \tag{A.152}$$

An **upper triangular** square $N \times N$ matrix is defined as a matrix whose elements below the main diagonal are zero. Thus, we can write,

$$\mathbf{U} = \begin{bmatrix} u_{11} & u_{12} & \cdots & u_{1n} \\ 0 & u_{22} & \cdots & u_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & u_{nn} \end{bmatrix}. \tag{A.153}$$

$\mathbf{U}^{-1}$ is also upper triangular. The determinant is

$$\det(\mathbf{U}) = \prod_{i=1}^{N} u_{ii}. \tag{A.154}$$

Clearly, $\mathbf{L}^T$ and $\mathbf{L}^H$ are upper triangular and $\mathbf{U}^T$ and $\mathbf{U}^H$ are lower triangular.

If $\mathbf{A}$ is a square Hermitian positive definite matrix, there are two factorizations of interest. The first factorization is called the $\mathbf{LDL}^H$ factorization. There is a unique factorization of $\mathbf{A}$,

$$\mathbf{A} = \mathbf{LDL}^H, \tag{A.155}$$

where $\mathbf{D}$ is a diagonal matrix with positive entries. (e.g., Golub and Van Loan [GVL89]).

The **Cholesky decomposition** is a unique factorization,

$$\mathbf{A} = \mathbf{G}\mathbf{G}^H, \tag{A.156}$$

where $\mathbf{G}$ is lower triangular,

$$\mathbf{G} = \mathbf{L}\ \text{diag}\left[\ \sqrt{d_1}, \quad \sqrt{d_2}, \quad \cdots \quad \sqrt{d_N}\ \right]. \tag{A.157}$$

The Cholesky decomposition plays a key role in many algorithms and analyses.

A.3.9 Unitary and Orthogonal Matrices

A square $N \times N$ matrix is **unitary** if

$$\mathbf{A}^{-1} = \mathbf{A}^H. \tag{A.158}$$

If $\mathbf{A}$ is unitary, then

$$\mathbf{A}^H\mathbf{A} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{A}\mathbf{A}^{-1} = \mathbf{A}\mathbf{A}^H = \mathbf{I}. \tag{A.159}$$

In order for $\mathbf{A}$ to be unitary, the columns must be orthonormal. Thus,

$$\mathbf{a}_i^H\mathbf{a}_j = \delta_{ij}. \tag{A.160}$$

A particular unitary matrix that is used in the text has columns that are conjugate symmetric and has a sparse structure.

For N even,

$$\mathbf{Q} = \frac{1}{\sqrt{2}}\begin{bmatrix} \mathbf{I} & j\mathbf{I} \\ \mathbf{J} & -j\mathbf{J} \end{bmatrix}, \tag{A.161}$$

where the $\mathbf{I}$ and $\mathbf{J}$ matrices have dimension $N/2$.

For N odd,

$$\mathbf{Q} = \frac{1}{\sqrt{2}}\begin{bmatrix} \mathbf{I} & \mathbf{0} & j\mathbf{I} \\ \mathbf{0}^T & \sqrt{2} & \mathbf{0}^T \\ \mathbf{I} & \mathbf{0} & -j\mathbf{J} \end{bmatrix}, \tag{A.162}$$

where the $\mathbf{I}$ and $\mathbf{J}$ matrices have dimension $(N-1)/2$ and $\mathbf{0}$ is a $(N-1)/2 \times 1$ vector whose elements are 0.

The $\mathbf{Q}$ matrices in (A.161) and (A.162) have a useful property. If $\mathbf{x}$ is a conjugate symmetric vector, then $\mathbf{Q}^H\mathbf{x}$ is real. For N even,

$$\begin{aligned} \mathbf{Q}^H\mathbf{x} &= \frac{1}{\sqrt{2}}\begin{bmatrix} \mathbf{I} & \mathbf{J} \\ -j\mathbf{I} & j\mathbf{J} \end{bmatrix}\begin{bmatrix} \mathbf{x}_1 \\ \mathbf{J}\mathbf{x}_1^* \end{bmatrix} = \frac{1}{\sqrt{2}}\begin{bmatrix} \mathbf{x}_1 + \mathbf{x}_2^* \\ -j(\mathbf{x}_1 - \mathbf{x}_2^*) \end{bmatrix} \\ &= \sqrt{2}\begin{bmatrix} Re(\mathbf{x}_1) \\ Im(\mathbf{x}_1) \end{bmatrix}. \end{aligned} \tag{A.163}$$

Other useful properties will be derived in Chapter 7.

An important consequence of the unitary property arises when we do a unitary transformation of a vector whose correlation matrix is the identity matrix,

$$\mathbf{y} = \mathbf{A}^H\mathbf{x}. \tag{A.164}$$

The correlation matrix of $\mathbf{y}$ is

$$E\left[\mathbf{y}\mathbf{y}^H\right] = \mathbf{A}^H E\left[\mathbf{x}\mathbf{x}^H\right]\mathbf{A} = \mathbf{A}^H\mathbf{I}\mathbf{A} = \mathbf{A}^H\mathbf{A} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}. \tag{A.165}$$

Thus, the correlation matrix of the output $\mathbf{y}$ is also an identity matrix.

A square $N \times N$ matrix is **orthogonal** if

$$\mathbf{A}^{-1} = \mathbf{A}^T. \tag{A.166}$$

In order for $\mathbf{A}$ to be orthogonal

$$\mathbf{a}_i^T\mathbf{a}_j = \delta_{ij}. \tag{A.167}$$

A.3.10 Vandermonde Matrices

A **Vandermonde matrix** is an $N \times M$ matrix in which the elements in the jth column can be expressed as powers of a parameter c_j. Thus,

$$a_{ij} = c_j^{i-1}, \quad 1 \le i \le N, \ 1 \le j \le M. \tag{A.168}$$

The matrix has the structure,

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ c_1 & c_2 & \cdots & c_M \\ c_1^2 & c_2^2 & \cdots & c_M^2 \\ \vdots & \vdots & & \vdots \\ c_1^{N-1} & c_2^{N-1} & \cdots & c_M^{N-1} \end{bmatrix}. \tag{A.169}$$

One can show that the determinant of a square $N \times N$ Vandermonde matrix is given by

$$\det \mathbf{A} = \prod_{1 \le i < k \le N} (c_k - c_i). \tag{A.170}$$

A.3.11 Projection Matrices

A matrix that will play a central role in many of our analyses is the projection matrix $\mathbf{P}$. An **idempotent matrix P** satisfies the relation

$$\mathbf{P}^2 = \mathbf{P}. \tag{A.171}$$

Another name for an idempotent matrix is a **projection matrix**. We restrict our discussion to projection matrices that are Hermitian,

$$\mathbf{P}^H = \mathbf{P}. \tag{A.172}$$

To motivate the construction of $\mathbf{P}$, we first consider the projection of a vector $\mathbf{b}$ onto a vector $\mathbf{a}$. We first construct a unit vector in the direction of $\mathbf{a}$,

$$\mathbf{u}_a = \frac{\mathbf{a}}{|\mathbf{a}|^2} = \left[\mathbf{a}^H\mathbf{a}\right]^{-1}\mathbf{a} = \mathbf{a}\left[\mathbf{a}^H\mathbf{a}\right]^{-1}. \tag{A.173}$$

Then, the projection of $\mathbf{b}$ onto $\mathbf{a}$ is

$$\mathbf{a}\left[\mathbf{u}_a^H\mathbf{b}\right] = \mathbf{a}\left[\mathbf{a}^H\mathbf{a}\right]^{-1}\mathbf{a}^H\mathbf{b}. \tag{A.174}$$

If we define the $N \times N$ matrix

$$\mathbf{P}_{\mathbf{a}} \triangleq \mathbf{a}\left[\mathbf{a}^H\mathbf{a}\right]^{-1}\mathbf{a}^H, \tag{A.175}$$

Then we can write the projection of $\mathbf{b}$ onto $\mathbf{a}$ as $\mathbf{P_a b}$. The matrix $\mathbf{P_a}$ is referred to as a projection matrix.

Now consider the complex $N \times M$ matrix, $\mathbf{V}$, and assume that the columns are linearly independent. Then, the projection of a complex $N \times 1$ vector $\mathbf{b}$ onto the M-dimensional subspace defined by the columns of $\mathbf{V}$ is,

$$\mathbf{b_V} = \mathbf{P_V b}, \tag{A.176}$$

where $\mathbf{P_V}$ is an $N \times N$-dimensional matrix defined by

$$\mathbf{P_V} = \mathbf{V}\left[\mathbf{V}^H\mathbf{V}\right]^{-1}\mathbf{V}^H. \tag{A.177}$$

We can also define a projection matrix that projects $\mathbf{b}$ onto a subspace orthogonal to the subspace defined by the columns of $\mathbf{V}$,

$$\mathbf{P_V^\perp} \triangleq \mathbf{I} - \mathbf{P_V}. \tag{A.178}$$

All of the vectors in this subspace are orthogonal to $\mathbf{V}$.

Note that, if the columns of $\mathbf{V}$ are orthonormal (i.e., it is a unitary matrix), then

$$\mathbf{P_V} = \mathbf{U}\mathbf{U}^H, \tag{A.179}$$

and

$$\mathbf{P_V^\perp} = \mathbf{I} - \mathbf{U}\mathbf{U}^H. \tag{A.180}$$

In many of our applications, we will find it useful to divide our N-dimensional "signal subspace," which contains both signal and'noise, and an $N - M$-dimensional "noise subspace," which contains only noise. Recall that we used this subspace technique in Chapter 2 of [VT68], [VT01a] when we solved the temporal detection problem.

A.3.12 Generalized Inverse

In this section we define a generalized inverse[2] for the $M \times N$ matrix $\mathbf{A}$.

[2] Our discussion follows Section 12.8 in [LT85].

A.3.12.1 Moore-Penrose pseudo-inverse

Assume that $\mathbf{A}$ is an $M \times N$ matrix. There exists a unique $N \times M$ matrix $\mathbf{B}$ for which

$$\mathbf{ABA} = \mathbf{A}, \tag{A.181}$$

$$\mathbf{BAB} = \mathbf{B}, \tag{A.182}$$

$$(\mathbf{AB})^H = \mathbf{AB}, \tag{A.183}$$

and

$$(\mathbf{BA})^H = \mathbf{BA}. \tag{A.184}$$

This matrix $\mathbf{B}$ is denoted by $\mathbf{A}^\dagger$, and

$$\mathbf{A}^\dagger = \mathbf{A}^H(\mathbf{A}\mathbf{A}^H)^{-1} \qquad M \le N, \tag{A.185}$$

$$\mathbf{A}^\dagger = \mathbf{A}^{-1} \qquad M = N, \tag{A.186}$$

$$\mathbf{A}^\dagger = (\mathbf{A}^H\mathbf{A})^{-1}\mathbf{A}^H \qquad M \ge N. \tag{A.187}$$

The matrix $\mathbf{A}^\dagger$ is referred to as the Moore-Penrose pseudo-inverse or the Moore-Penrose inverse in the literature ([Moo20] and [Pen55]). It is encountered in the solution of linear equations.

A.3.12.2 Application to the solution of $\mathbf{Ax} = \mathbf{b}$.

Consider the linear equation, $\mathbf{Ax} = \mathbf{b}$, where $\mathbf{A}$ is an $M \times N$ matrix, $\mathbf{x}$ is an $N \times 1$ vector, and $\mathbf{b}$ is an $M \times 1$ vector.

If $M < N$, there are multiple solutions. The solution,

$$\mathbf{x}_0 = \mathbf{A}^\dagger\mathbf{b}, \tag{A.188}$$

is one of the solutions,

$$\mathbf{A}\mathbf{x}_0 = \mathbf{A}\mathbf{A}^\dagger\mathbf{b} = \mathbf{b}. \tag{A.189}$$

It is the solution with the smallest Euclidean norm.

If $M = N$ and $\mathbf{A}$ is non-singular, $\mathbf{x}_0$ is the unique solution and $\mathbf{A}^\dagger = \mathbf{A}^{-1}$.

If $M > N$, there are, in general, no solutions to the equation. We define an approximate solution of the overdetermined equation $\mathbf{Ax} = \mathbf{b}$ to be a vector $\mathbf{x}_0$ given by

$$\mathbf{x}_0 = \arg\min_{\mathbf{x}} \|\mathbf{Ax} - \mathbf{b}\|, \tag{A.190}$$

where the norm is the Euclidean vector norm and the minimization is over all $N \times 1$ vectors. The solution to (A.190) is

$$\mathbf{x}_0 = \mathbf{A}^\dagger\mathbf{b}. \tag{A.191}$$

A.4 Eigensystems

In Chapters 2 and 3 of DEMT I [VT68], [VT01a] and Chapter 2 of DEMT III [VT71], [VT01b], we saw the advantages of an eigenfunction decomposition for solving detection and estimation problems. We will find similar advantages in the array processing area.

In Section A.4.1, we develop the concept of eigenvectors and eigenvalues and discuss properties. In Section A.4.2, we consider several matrices whose eigenvectors have special properties.

A.4.1 Eigendecomposition

Consider a square Hermitian $N \times N$ matrix $\mathbf{A}$ (symmetric if $\mathbf{A}$ is real). The **eigenvectors** of $\mathbf{A}$ are denoted by $\boldsymbol{\phi}$. They satisfies the equation,

$$\mathbf{A}\boldsymbol{\phi} = \lambda\boldsymbol{\phi}. \tag{A.192}$$

Thus when $\boldsymbol{\phi}$ is operated on by $\mathbf{A}$ the output is $\boldsymbol{\phi}$ multiplied by a scalar λ. The scalar λ is called the **eigenvalue** of $\mathbf{A}$. The relation in (A.192) can also be written as

$$[\lambda\mathbf{I} - \mathbf{A}]\,\boldsymbol{\phi} = \mathbf{0}. \tag{A.193}$$

In order to have a solution for the homogeneous equation in (A.193), we must have

$$\det[\lambda\mathbf{I} - \mathbf{A}] = 0. \tag{A.194}$$

This is an Nth order polynomial in λ, which is called the **characteristic polynomial**,

$$f(\lambda) \triangleq \det[\lambda\mathbf{I} - \mathbf{A}]. \tag{A.195}$$

The N roots of $f(\lambda)$ are the eigenvalues of $\mathbf{A}$:

$$f(\lambda) = \prod_{i=1}^{N} (\lambda_i - \lambda) = \sum_{k=0}^{N} f_k \lambda^k. \tag{A.196}$$

There is an eigenvector (there may be more than one) associated with each λ. The eigenvectors associated with unequal eigenvalues are linearly independent. For the distinct eigenvalues, we choose the eigenvectors to be orthonormal vectors. For eigenvalues of multiplicity M, we can construct orthonormal eigenvectors.

The eigenvalues of a Hermitian matrix are all real. We order the eigenvalues in decreasing size,

$$\lambda_{max} = \lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \cdots \geq \lambda_N = \lambda_{min}. \tag{A.197}$$

The eigenvectors associated with larger eigenvalues are referred to as **principal eigenvectors**.

Several properties follow:

(**i**) The highest coefficient in (A.196) is

$$f_N = (-1)^N. \tag{A.198}$$

(**ii**) The lowest coefficient in (A.196) is

$$f_0 = \det(\mathbf{A}) = \prod_{i=1}^{N} \lambda_i. \tag{A.199}$$

(**iii**) The sum of eigenvalues equals $\mathrm{tr}(\mathbf{A})$,

$$\mathrm{tr}\;(\mathbf{A}) = \sum_{i=1}^{N} \lambda_i. \tag{A.200}$$

(**iv**) If β is a scalar, then the eigenvalues of $\beta\mathbf{A}$ are $\beta\lambda_i; i = 1, \cdots, N$.

(**v**) The eigenvalues of the matrix $\mathbf{A}^m$ for m, a positive integer, are λ_i^m.

(**vi**) If $\mathbf{A}$ is non-singular, then the eigenvalues of $\mathbf{A}^{-1}$ are λ_i^{-1} and the eigenvectors are the eigenvectors of $\mathbf{A}$.

(**vii**) If $|\lambda_{\max}(\mathbf{A})| < 1$, then

$$(\mathbf{I} + \mathbf{A})^{-1} = \mathbf{I} - \mathbf{A} + \mathbf{A}^2 - \mathbf{A}^3 + \cdots. \tag{A.201}$$

(**viii**) The identity matrix $\mathbf{I}$ has N eigenvalues equal to unity and any set of N orthonormal vectors can be used as eigenvectors.

(**ix**) The eigenvalues of $\mathbf{A} + \sigma^2\mathbf{I}$ are $\lambda_i + \sigma^2$ and the eigenvectors are the eigenvectors of $\mathbf{A}$.

(**x**) If we define the $N \times N$ matrix

$$\mathbf{U_A} = \left[\begin{array}{c:c:c:c} \boldsymbol{\phi}_1 & \boldsymbol{\phi}_2 & \cdots & \boldsymbol{\phi}_N \end{array} \right], \tag{A.202}$$

where the $\boldsymbol{\phi}_i, i = 1, 2, \cdots, N$ are the orthonormal eigenvectors and the $N \times N$ diagonal matrix,

$$\boldsymbol{\Lambda} = \text{ diag } [\lambda_1, \lambda_2, \cdots, \lambda_N], \tag{A.203}$$

then we can write $\mathbf{A}$ as

$$\mathbf{A} = \mathbf{U_A} \boldsymbol{\Lambda} \mathbf{U_A^H}, \tag{A.204}$$

(**xi**) If $\mathbf{A}$ is non-singular, we can write

$$\mathbf{A}^{-1} = \mathbf{U_A} \boldsymbol{\Lambda}^{-1} \mathbf{U_A}. \tag{A.205}$$

(**xii**) We can also expand $\mathbf{A}$ as,

$$\mathbf{A} = \sum_{i=1}^{N} \lambda_i \boldsymbol{\phi}_i \boldsymbol{\phi}_i^H. \tag{A.206}$$

(**xiii**) If $\mathbf{A}$ is non-singular,

$$\mathbf{A}^{-1} = \sum_{i=1}^{N} \frac{1}{\lambda_i} \boldsymbol{\phi}_i \boldsymbol{\phi}_i^H. \tag{A.207}$$

(**xiv**) If $\mathbf{A}$ is positive definite, then all eigenvalues are positive.

(**xv**) If $\mathbf{A}$ is positive semi-definite, the number of positive eigenvalues equals rank $(\mathbf{A})$ and the remaining eigenvalues are zero.

(**xvi**) If the $N \times N$ matrix $\mathbf{A}$ consists of a weighted sum of D outer products,

$$\mathbf{A} = \sum_{i=1}^{D} \sigma_i^2 \mathbf{v}_i \mathbf{v}_i^H, \tag{A.208}$$

then rank $(\mathbf{A}) = D$ and the first D eigenvectors are linear combinations of the $\mathbf{v}_i$ and the eigenvalues are denoted by λ_i. There are $N - D$ eigenvalues equal to zero and the corresponding eigenvectors can be chosen to be any orthonormal set that are also orthogonal to the $\mathbf{v}_i$, $i = 1, \cdots, D$.

(**xvii**) If $\mathbf{A}$ is an $N \times N$ matrix that can be written as

$$\mathbf{A} = \sum_{i=1}^{D} \sigma_i^2 \mathbf{v}_i \mathbf{v}_i^H + \sigma_w^2 \mathbf{I}, \tag{A.209}$$

then the eigenvectors in (xvi) are still valid and

$$\lambda_i(\mathbf{A}) = \begin{cases} \lambda_i + \sigma_w^2, & i = 1, \cdots, D, \\ \\ \sigma_w^2, & i = D+1, \cdots, N. \end{cases} \tag{A.210}$$

In this case, we often divide the N-dimensional space into two orthogonal subspaces, a signal subspace and a noise subspace. Then,

$$\mathbf{A} = \mathbf{U}_S \mathbf{\Lambda}_S \mathbf{U}_S^H + \mathbf{U}_N \mathbf{\Lambda}_N \mathbf{U}_N^H, \tag{A.211}$$

where $\mathbf{U}_S$ is a $N \times D$ matrix of eigenvectors,

$$\mathbf{U}_S = \left[\; \boldsymbol{\phi}_1 \; \vdots \; \boldsymbol{\phi}_2 \; \vdots \; \cdots \; \vdots \; \boldsymbol{\phi}_D \; \right], \tag{A.212}$$

and $\mathbf{\Lambda}_S$ is a diagonal matrix,

$$\mathbf{\Lambda}_S = \text{ diag } \left[\lambda_1 + \sigma_w^2, \lambda_2 + \sigma_w^2, \cdots, \lambda_D + \sigma_w^2 \right]. \tag{A.213}$$

The matrix $\mathbf{U}_N$ is a $N \times (N-D)$ matrix of orthonormal vectors,

$$\mathbf{U}_N = \left[\; \boldsymbol{\phi}_{D+1} \; \vdots \; \boldsymbol{\phi}_{D+2} \; \vdots \; \cdots \; \vdots \; \boldsymbol{\phi}_N \; \right], \tag{A.214}$$

where

$$\boldsymbol{\phi}_i^H \boldsymbol{\phi}_j = 0, \qquad i = D+1, \cdots, N, \;\; j = 1, \cdots, D, \tag{A.215}$$

and

$$\mathbf{\Lambda}_N = \sigma_w^2 \mathbf{I}_{(N-D)}. \tag{A.216}$$

(**xviii**) In many cases of interest, $\mathbf{A}$ is a correlation matrix, $\mathbf{R_x}$,

$$\mathbf{R_x} = E\left[\mathbf{x}\mathbf{x}^H\right], \tag{A.217}$$

where $\mathbf{x}$ is a zero-mean random vector. Then we can expand

$$\mathbf{x} = \sum_{i=1}^{N} x_i \boldsymbol{\phi}_i, \tag{A.218}$$

where the $\boldsymbol{\phi}_i$ are the eigenvectors of $\mathbf{R_x}$, and

$$x_i = \mathbf{x}^H \boldsymbol{\phi}_i = \boldsymbol{\phi}_i^H \mathbf{x}. \tag{A.219}$$

The x_i are uncorrelated,

$$E\left[x_i x_j\right] = \lambda_i \delta_{ij}. \tag{A.220}$$

(**xix**) If $\mathbf{x}$ is a Gaussian random vector, then the x_i are statistically independent zero-mean Gaussian random variables with variance λ_i.

A.4.2 Special Matrices

In this section, we discuss several matrices whose eigendecompositions have special properties. The sections discuss:

(**i**) Separable kernels,

(**ii**) Centrohermitian matrices,

(**iii**) Circulant matrices,

(**iv**) Toeplitz matrices,

(**v**) Kronecker products.

A.4.2.1 Separable kernels

We use the term separable kernel to describe matrices of the form[3]

$$\mathbf{A} = \mathbf{V}\mathbf{S}\mathbf{V}^H, \tag{A.221}$$

where $\mathbf{V}$ is a $N \times K$ matrix whose columns are the array manifold vector $\mathbf{v}$ and $\mathbf{S}$ is a $K \times K$ Hermitian matrix. This corresponds to the model for K correlated signals impinging on a N-element array. We assume $K < N$. This matrix structure allows us to reduce the eigenanalysis to a K-dimensional problem instead of a N-dimensional problem.

The eigenequation is

$$\lambda \boldsymbol{\phi} = \mathbf{V}\mathbf{S}\mathbf{V}^H \boldsymbol{\phi}. \tag{A.222}$$

Since the rank of $\mathbf{A}$ is K, there will be K non-zero eigenvalues and corresponding eigenvectors. The vectors $\mathbf{v}_i, i = 1, 2, \cdots, K$ define a K-dimensional subspace. The $\boldsymbol{\phi}$ vectors are the orthogonal basis for that K-dimensional subspace. Therefore, each $\boldsymbol{\phi}_i$ can be obtained from $\mathbf{V}$ by a linear transformation.

$$\boldsymbol{\phi}_i = \mathbf{V}\mathbf{c}_i, \qquad i = 1, 2, \cdots, K, \tag{A.223}$$

where $\mathbf{c}_i$ is a $K \times 1$ vector.

Using (A.223) in (A.222) gives

$$\lambda \mathbf{V}\mathbf{c}_i = \mathbf{V}\mathbf{S}\mathbf{V}^H \mathbf{V}\mathbf{c}_i. \tag{A.224}$$

Equation (A.224) can be written as

$$\mathbf{V}\left\{\left[\lambda_i \mathbf{I} - \mathbf{S}\mathbf{V}^H\mathbf{V}\right]\mathbf{c}_i\right\} = \mathbf{0}. \tag{A.225}$$

In order for (A.225) to be satisfied, we require

$$\left[\lambda_i \mathbf{I} - \mathbf{S}\mathbf{V}^H\mathbf{V}\right]\mathbf{c}_i = \mathbf{0}. \tag{A.226}$$

Thus, the $\mathbf{c}_i$ are the eigenvectors of the $K \times K$ matrix $\mathbf{S}\mathbf{V}^H\mathbf{V}$.

This is a homogeneous equation that has a solution for K values of $\lambda_i, i = 1, 2, \cdots, K$. In order for (A.226) to have a solution,

$$\det\left[\lambda \mathbf{I} - \mathbf{S}\mathbf{V}^H\mathbf{V}\right] = 0, \tag{A.227}$$

[3] The use of the separable kernel descriptor follows from our separable kernel discussions in Chapter 3 of [VT68], [VT01a] and Chapters 3 and 4 of [VT71], [VT01b].

which will have K solutions: $\lambda_i, i = 1, 2, \cdots, K$.

For each λ_i we can find the corresponding $\mathbf{c}_i$, which determines the corresponding $\boldsymbol{\phi}_i, i = 1, 2, \cdots, K$.

In many cases of interest, $K \ll N$, so that we have achieved a significant reduction in the complexity of the problem.

Example A.4.1

Consider a single case in which $K = 2$ and $\mathbf{S}$ is diagonal matrix $\sigma_s^2\mathbf{I}$. Then,

$$\mathbf{V}^H\mathbf{S}\mathbf{V} = N\sigma_s^2 \begin{bmatrix} 1 & \frac{\mathbf{v}_1^H\mathbf{v}_2}{N} \\ \frac{\mathbf{v}_2^H\mathbf{v}_1}{N} & 1 \end{bmatrix}. \tag{A.228}$$

The determinant of (A.228) is a quadratic equation in λ,

$$\lambda^2 - b\lambda + c = 0, \tag{A.229}$$

where

$$b = 2N\sigma_s^2, \tag{A.230}$$

and

$$c = N^2\sigma_s^4\left[1 - |\beta_{12}|^2\right], \tag{A.231}$$

where $\beta_{12} \triangleq \frac{\mathbf{v}_1^H\mathbf{v}_2}{N}$. Then,

$$\lambda = N\sigma_s^2\left[1 \pm \sqrt{1 - \frac{4c}{b^2}}\right], \tag{A.232}$$

or

$$\lambda_{1(2)} = N\sigma_s^2\left[1 \pm |\beta_{12}|\right]. \tag{A.233}$$

Then $\mathbf{c}_1$ and $\mathbf{c}_2$ are obtained by substituting λ_1 and λ_2 from (A.233) into (A.226) to obtain,

$$\mathbf{c}_1 = \frac{1}{\sqrt{2N\left[1 + |\beta_{12}|\right]}}\begin{bmatrix} 1 \\ e^{-j\phi(\beta_{12})} \end{bmatrix}, \tag{A.234}$$

$$\mathbf{c}_1 = \frac{1}{\sqrt{2N\left[1 - |\beta_{12}|\right]}}\begin{bmatrix} 1 \\ -e^{-j\phi(\beta_{12})} \end{bmatrix}, \tag{A.235}$$

where $\phi(\beta_{12})$ denotes the phase of β_{12}. Substituting (A.234) and (A.235) into (A.223) gives,

$$\phi_1 = \frac{\left[\mathbf{v}_1 + e^{-j\phi(\beta_{12})}\mathbf{v}_2\right]}{\sqrt{2N(1+|\beta_{12}|)}} \tag{A.236}$$

and

$$\phi_2 = \frac{\left[\mathbf{v}_1 - e^{-j\phi(\beta_{12})}\mathbf{v}_2\right]}{\sqrt{2N(1-|\beta_{12}|)}}, \tag{A.237}$$

Plots of the λ_i and $\mathbf{c}_i$ are shown in Chapter 5 for typical array geometries.

A.4.2.2 Centrohermitian matrices

Centrohermitian-persymmetric matrices have the property that their eigenvectors are either conjugate symmetric or conjugate asymmetric [CB76b], [CB76a] (see, e.g., [Mar87], [Mak81] for further discussion).[4]

If the dimension N of the matrix $\mathbf{C}$ is even, then it has $\frac{N}{2}$ conjugate symmetric orthogonal eigenvectors ϕ_i $\left(i = 1, \cdots, \frac{N}{2}\right)$ of the form

$$\phi_i = \frac{1}{\sqrt{2}} \begin{bmatrix} \mathbf{y}_i \\ \mathbf{J}\mathbf{y}_i \end{bmatrix}, \tag{A.238}$$

where the $\mathbf{y}_i$ are the orthonormal eigenvectors of the matrix $\mathbf{B}_1 + \mathbf{B}_2\mathbf{J}$, where $\mathbf{B}_1$ and $\mathbf{B}_2$ are the submatrices defined in (A.141). Thus,

$$[\mathbf{B}_1 + \mathbf{B}_2\mathbf{J}]\,\mathbf{y}_i = \lambda_i \mathbf{y}_i. \tag{A.239}$$

It also has $\frac{N}{2}$ conjugate asymmetric orthonormal eigenvectors ϕ_i $\left(i = 1, \cdots, \frac{N}{2}\right)$ of the form

$$\phi_i = \frac{1}{\sqrt{2}} \begin{bmatrix} \mathbf{z}_i \\ -\mathbf{J}\mathbf{z}_i \end{bmatrix}, \tag{A.240}$$

where the $\mathbf{z}_i$ are the eigenvectors of the matrix $\mathbf{B}_1 - \mathbf{B}_2\mathbf{J}$. Thus,

$$[\mathbf{B}_1 - \mathbf{B}_2\mathbf{J}]\,\mathbf{z}_i = \lambda_i \mathbf{z}_i. \tag{A.241}$$

Note that we have reduced the dimension of the eigendecomposition problem by factor of 2.

If N is odd, then $\mathbf{C}$ will have $\frac{N+1}{2}$ conjugate symmetric orthonormal eigenvectors ϕ_i $\left(i = 1, 2, \cdots, \frac{(N+1)}{2}\right)$ of the form

[4] The results in this section are due to [CB76a], [CB76b].

$$\boldsymbol{\phi}_i = \frac{1}{\sqrt{2}} \begin{bmatrix} \mathbf{y}_i \\ 2\beta_i \\ \mathbf{J}\mathbf{y}_i \end{bmatrix}, \tag{A.242}$$

where the vector $\mathbf{y}_i$ and the real scalar β_i are the orthonormal eigenvectors of the matrix.

$$\begin{bmatrix} \mathbf{B}_1 + \mathbf{B}_2\mathbf{J} & \sqrt{2}\mathbf{a} \\ \sqrt{2}\mathbf{a}^H & c \end{bmatrix} \begin{bmatrix} \mathbf{y}_i \\ \beta_i \end{bmatrix} = \lambda_i \begin{bmatrix} \mathbf{y}_i \\ \beta_i \end{bmatrix}, \tag{A.243}$$

where $\mathbf{a}$ and c are defined in (A.142). $\mathbf{C}$ will also have $\frac{(N-1)}{2}$ conjugate asymmetric orthonormal eigenvectors of the form,

$$\boldsymbol{\phi}_i = \frac{1}{\sqrt{2}} \begin{bmatrix} \mathbf{z}_i \\ 0 \\ -\mathbf{J}\mathbf{z}_i \end{bmatrix}, \tag{A.244}$$

where the $\mathbf{z}_i$ are the orthonormal eigenvectors of the matrix $\mathbf{B}_1 - \mathbf{B}_2\mathbf{J}$,

$$\left(\mathbf{B}_1 - \mathbf{B}_2\mathbf{J}\right)\mathbf{z}_i = \lambda_i \mathbf{z}_i. \tag{A.245}$$

A.4.2.3 Toeplitz matrices

Toeplitz matrices are a special case of centrohermitian matrices so their eigendecomposition has the same properties. Makhoul [Mak81] analyzed their eigendecomposition and found that their additional structure did not lead to any additional properties that will be useful in our analyses.

The reader is referred to [Mak81]or [Mar87] for further discussion.

A.4.2.4 Kronecker products

Any important motivation for using Kronecker products is the relationship between the eigenvalues of $\mathbf{A}$ and $\mathbf{B}$ and the eigenvalues of $\mathbf{A} \otimes \mathbf{B}$. The following result is from Chapter 12 of [LT85], which they attribute to Stephanos [Ste00]. A good discussion of Kronecker products and their application is given in Graham [Gra81].

Let $\lambda_1, \cdots, \lambda_M$ be the eigenvalues of the $M \times M$ matrix $\mathbf{A}$. Let $\mu_1, \cdots, \mu_N$ be the eigenvalues of the $N \times N$ matrix $\mathbf{B}$.

Then the eigenvalues of $\mathbf{A} \otimes \mathbf{B}$ are the MN numbers, $\lambda_r \mu_s, r = 1, \cdots, M, s = 1, \cdots, N$.

A consequence of this result is that, if $\mathbf{A}$ and $\mathbf{B}$ are positive definite, then $\mathbf{A} \otimes \mathbf{B}$ is positive definite.

The eigenvalues of $(\mathbf{I}_N \otimes \mathbf{A}) + (\mathbf{B} \otimes \mathbf{I}_M)$ are the MN numbers,

$$\lambda_r + \mu_s, \qquad r = 1, \cdots, M, s = 1, \cdots, N. \tag{A.246}$$

The above matrix is called the Kronecker sum of $\mathbf{A}$ and $\mathbf{B}$.

A.5 Singular Value Decomposition

A significant number of applications in the text deal with the spatial spectral matrix of the received waveform at the sensors, $\mathbf{S}_{\mathbf{x}}$.

If we assume that we are dealing with a single frequency as a result of a DFT operation on the incoming vector, then

$$\mathbf{S}_{\mathbf{x}}(\omega_o) = E\left\{\mathbf{X}(\omega_o)\mathbf{X}^H(\omega_o)\right\}, \tag{A.247}$$

where $\mathbf{X}(\omega_o)$ is an $N \times 1$ vector corresponding to DFT of the incoming signal.

In practice, the output of each sensor is sampled, sectioned, and windowed. The DFT of each section is evaluated and a vector $\mathbf{X}_k(\omega_o)$ across the array is formed for each section. Assuming K sections are used, we form

$$\hat{\mathbf{S}}_{\mathbf{x}} = \frac{1}{K}\sum_{k=1}^{K} \mathbf{X}\left[\omega_o, k\right]\mathbf{X}^H\left[\omega_o, k\right], \tag{A.248}$$

and use this as an estimate of the spatial spectral matrix $\mathbf{S}_{\mathbf{x}}(\omega_o)$. For simplicity in the notation, we have suppressed the ω_o dependence in $\hat{\mathbf{S}}_x$ and will use $\mathbf{X}(k)$ to denote the kth sample. We restrict our attention to a ULA with interelement spacing d. We can write $\hat{\mathbf{S}}_{\mathbf{x}}$ in expanded form as

$$\hat{\mathbf{S}}_{\mathbf{x}} = \begin{bmatrix} \hat{s}_{xx}(0,0) & \hat{s}_{xx}(1,0) & \cdots & \hat{s}_{xx}(N-1,0) \\ \hat{s}_{xx}(0,1) & \hat{s}_{xx}(1,1) & \cdots & \hat{s}_{xx}(N-1,1) \\ \vdots & \vdots & \vdots & \vdots \\ \hat{s}_{xx}(0,N-1) & \hat{s}_{xx}(1,N-1) & \cdots & \hat{s}_{xx}(N-1,N-1) \end{bmatrix}, \tag{A.249}$$

where

$$s_{xx}(i,j) = \frac{1}{K}\sum_{k=1}^{K} \mathbf{X}_i[k]\mathbf{X}_j^*[k]. \tag{A.250}$$

$\hat{\mathbf{S}}_{\mathbf{x}}$ has several properties that we will use in the text.

The spatial spectral matrix is Hermitian,

$$\hat{\mathbf{S}}_{\mathbf{x}}^H = \hat{\mathbf{S}}_{\mathbf{x}}. \tag{A.251}$$

This follows directly from (A.249).

The spatial spectral matrix is non-negative definite,

$$\mathbf{y}^H \hat{\mathbf{S}}_{\mathbf{x}} \mathbf{y} \geq 0, \tag{A.252}$$

for any $N \times 1$ vector $\mathbf{y}$. This follows directly from (A.248),

$$\mathbf{y}^H \hat{\mathbf{S}}_{\mathbf{x}} \mathbf{y} = \frac{1}{K}\sum_{k=1}^{K} \left|\mathbf{y}^H \mathbf{X}(k)\right|^2 \geq 0. \tag{A.253}$$

If (A.252) is satisfied with an inequality sign, then $\hat{\mathbf{S}}_{\mathbf{x}}$ is non-singular and its inverse $\hat{\mathbf{S}}_{\mathbf{x}}^{-1}$ exists.

The eigenvalues of $\hat{\mathbf{S}}_{\mathbf{x}}$ are real and nonnegative. The spectral matrix $\hat{\mathbf{S}}_{\mathbf{x}}$ is, in general, not Toeplitz because, as can be seen from (A.249), the elements on the main diagonal (and other diagonals) have different values.

We can write the sample spectral matrix as the product of two rectangular matrices,

$$\hat{\mathbf{S}}_{\mathbf{x}} = \frac{1}{K}\left[\begin{array}{c;{2pt/2pt}c;{2pt/2pt}c;{2pt/2pt}c} \mathbf{X}(1) & \mathbf{X}(2) & \cdots & \mathbf{X}(K) \end{array}\right] \begin{bmatrix} \mathbf{X}^H(1) \\ \mathbf{X}^H(2) \\ \vdots \\ \mathbf{X}^H(K) \end{bmatrix}. \tag{A.254}$$

We now define the data matrix $\tilde{\mathbf{X}}$,

$$\tilde{\mathbf{X}} = \frac{1}{\sqrt{K}}\left[\begin{array}{c;{2pt/2pt}c;{2pt/2pt}c;{2pt/2pt}c} \mathbf{X}(1) & \mathbf{X}(2) & \cdots & \mathbf{X}(K) \end{array}\right], \tag{A.255}$$

or

$$\tilde{\mathbf{X}} = \frac{1}{\sqrt{K}} \left[\begin{array}{c:c:c:c} X_0(1) & X_0(2) & & X_0(K) \\ X_1(1) & X_1(2) & & \\ \vdots & & \cdots & \\ X_{N-1}(1) & X_{N-1}(2) & & X_{N-1}(K) \end{array} \right], \tag{A.256}$$

and

$$\hat{\mathbf{S}}_{\mathbf{x}} = \tilde{\mathbf{X}}\tilde{\mathbf{X}}^H. \tag{A.257}$$

If we use FB averaging, we have an $N \times 2K$ data matrix

$$\tilde{\mathbf{X}}_{fb} = \frac{1}{\sqrt{2}} \left[\begin{array}{c:c} \tilde{\mathbf{X}} & \mathbf{J}\tilde{\mathbf{X}}^* \end{array} \right]. \tag{A.258}$$

Writing (A.258) out gives an $N \times 2K$ data matrix

$$\tilde{\mathbf{X}}_{fb} = \frac{1}{\sqrt{2}} \left[\begin{array}{cccc:cccc} X_0(1) & X_0(2) & \cdots & X_0(K) & X_{N-1}^*(1) & X_{N-1}^*(2) & \cdots & X_{N-1}^*(K) \\ X_1(1) & X_1(2) & \cdots & X_1(K) & X_{N-2}^*(1) & X_{N-2}^*(2) & \cdots & X_{N-2}^*(K) \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ X_{N-1}(1) & X_{N-1}(2) & \cdots & X_{N-1}(K) & X_0^*(1) & X_0^*(2) & \cdots & X_0^*(K) \end{array} \right], \tag{A.259}$$

and

$$\hat{\mathbf{S}}_{\mathbf{x},fb} = \tilde{\mathbf{X}}_{fb}\tilde{\mathbf{X}}_{fb}^H. \tag{A.260}$$

We now discuss how we find the eigenvectors and eigenvalues of $\hat{\mathbf{S}}_{\mathbf{x}}$ and $\hat{\mathbf{S}}_{\mathbf{x},fb}$ directly from the data matrix.

Given the data matrix $\tilde{\mathbf{X}}$, we want to show that there are two unitary matrices, $\mathbf{U}$ and $\mathbf{V}$, such that,

$$\mathbf{U}^H\tilde{\mathbf{X}}^H\mathbf{V} = \begin{bmatrix} \mathbf{\Sigma} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}, \tag{A.261}$$

where $\mathbf{\Sigma}$ is a diagonal matrix,

$$\mathbf{\Sigma} = \operatorname{diag}\left(\sigma_1, \sigma_2, \cdots, \sigma_W\right), \tag{A.262}$$

and the σ's are ordered in a decreasing manner.

The result in (A.261) is called the **singular value decomposition theorem**. The matrix $\mathbf{U}^H$ is $K \times K$. The matrix $\mathbf{V}$ is $N \times N$. The subscript W is the rank of the matrix $\tilde{\mathbf{X}}$

$$W = \text{ rank } \left(\tilde{\mathbf{X}}\right) \leq \min\left[K, N\right]. \tag{A.263}$$

For $K > N$, the system in (A.260) is overdetermined. For $K < N$, the system is underdetermined. The overdetermined case is of most interest is our applications. (We need more snapshots (K) than the number of sensors (N) to get a stable estimate of $\mathbf{S}_{\mathbf{x}}$.) We derive the overdetermined case in the text and refer the reader to Haykin (p. 520 of [Hay96]) for the underdetermined case.

We form the $N \times N$ matrix, $\tilde{\mathbf{X}}\tilde{\mathbf{X}}^H$. The matrix $\tilde{\mathbf{X}}\tilde{\mathbf{X}}^H$ is Hermitian and non-negative definite, so its eigenvalues are real and non-negative. We denote the eigenvalues as $\lambda_1, \lambda_2, \cdots, \lambda_N$ where

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_W > 0, \tag{A.264}$$

and

$$\lambda_{W+1} = \lambda_{W+2} = \cdots = \lambda_N = 0. \tag{A.265}$$

The λ's and σ's are related by

$$\sigma_i = \lambda_i^{\frac{1}{2}}, \qquad i = 1, 2, \cdots, N. \tag{A.266}$$

We denote the orthonormal eigenvectors of the matrix $\tilde{\mathbf{X}}\tilde{\mathbf{X}}^H$ by $\mathbf{v}_k, k = 1, 2, \cdots, N$, and the $N \times N$ unitary matrix whose columns are the eigenvectors of $\tilde{\mathbf{X}}\tilde{\mathbf{X}}^H$ by $\mathbf{V}$. Then

$$\mathbf{V}^H\tilde{\mathbf{X}}\tilde{\mathbf{X}}^H\mathbf{V} = \begin{bmatrix} \mathbf{\Sigma}^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}. \tag{A.267}$$

We partition $\mathbf{V}$ as,

$$\mathbf{V} = \left[\mathbf{V}_1 \, \mathbf{V}_2\right], \tag{A.268}$$

where $\mathbf{V}_1$ is an $N \times W$ matrix,[5]

[5] In most of our applications, there is a white noise component present, so that $W = N$ and the $\mathbf{V}_2$ is not present. The modification for that case is straightforward.

$$\mathbf{V}_1 = [\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_W], \tag{A.269}$$

$$\mathbf{V}_2 = [\mathbf{v}_{W+1}, \mathbf{v}_{W+2}, \cdots, \mathbf{v}_M]. \tag{A.270}$$

We observe that

$$\mathbf{V}_1^H \mathbf{V}_2 = \mathbf{0}. \tag{A.271}$$

From (A.267), we have for $\mathbf{V}_1$,

$$\mathbf{V}_1^H \tilde{\mathbf{X}} \tilde{\mathbf{X}}^H \mathbf{V}_1 = \mathbf{\Sigma}^2, \tag{A.272}$$

and therefore,

$$\mathbf{\Sigma}^{-1} \mathbf{V}_1^H \tilde{\mathbf{X}} \tilde{\mathbf{X}}^H \mathbf{V}_1 \mathbf{\Sigma}^{-1} = \mathbf{I}. \tag{A.273}$$

For $\mathbf{V}_2$,

$$\mathbf{V}_2^H \tilde{\mathbf{X}} \tilde{\mathbf{X}}^H \mathbf{V}_2 = \mathbf{0}, \tag{A.274}$$

and therefore,

$$\tilde{\mathbf{X}}^H \mathbf{V}_2 = \mathbf{0}. \tag{A.275}$$

We next define a $K \times W$ matrix, $\mathbf{U}_1$,

$$\mathbf{U}_1 = \tilde{\mathbf{X}}^H \mathbf{V}_1 \mathbf{\Sigma}^{-1}. \tag{A.276}$$

Then, from (A.273),

$$\mathbf{U}_1^H \mathbf{U}_1 = \mathbf{I}. \tag{A.277}$$

This implies the columns of $\mathbf{U}_1$ are orthogonal.

Finally, we define a $K \times K - W$ matrix $\mathbf{U}_2$ so that

$$\mathbf{U} = [\mathbf{U}_1 \, \mathbf{U}_2], \tag{A.278}$$

is a unitary matrix.

This implies that

$$\mathbf{U}_1^H \mathbf{U}_2 = \mathbf{0}. \tag{A.279}$$

With these definitions we can write

$$
\begin{aligned}
\mathbf{U}^H\tilde{\mathbf{X}}^H\mathbf{V} &= \begin{bmatrix} \mathbf{U}_1^H \\ \mathbf{U}_2^H \end{bmatrix} \tilde{\mathbf{X}}^H [\mathbf{V}_1, \mathbf{V}_2] \\
&= \begin{bmatrix} \mathbf{U}_1^H\tilde{\mathbf{X}}^H\mathbf{V}_1 & \mathbf{U}_1^H\tilde{\mathbf{X}}^H\mathbf{V}_2 \\ \mathbf{U}_2^H\tilde{\mathbf{X}}^H\mathbf{V}_1 & \mathbf{U}_2^H\tilde{\mathbf{X}}^H\mathbf{V}_2 \end{bmatrix} \\
&= \begin{bmatrix} (\mathbf{\Sigma}^{-1}\mathbf{V}_1^H\tilde{\mathbf{X}})\tilde{\mathbf{X}}^H\mathbf{V}_1 & \mathbf{U}_1^H\mathbf{0} \\ \mathbf{U}_2^H(\mathbf{U}_1\mathbf{\Sigma}) & \mathbf{U}_2^H\mathbf{0} \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{\Sigma} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix},
\end{aligned} \tag{A.280}
$$

which is the relation in (A.261).

We refer to elements of $\mathbf{\Sigma}$ as the singular values of $\tilde{\mathbf{X}}^H$. The columns of the unitary matrix $\mathbf{V}$ are the right singular vectors of $\tilde{\mathbf{X}}^H$, and the columns of the unitary matrix $\mathbf{U}$ are the left singular vectors of $\tilde{\mathbf{X}}^H$.

The right singular vectors are the eigenvectors of the matrix $\tilde{\mathbf{X}}\tilde{\mathbf{X}}^H$ that corresponds to the sample covariance matrix, $\hat{\mathbf{S}}_{\mathbf{x}}$ (see (A.257)). The non-zero eigenvalues of the sample covariance matrix are

$$
[\mathbf{\Lambda}]_{ii} = \left[\mathbf{\Sigma}^2\right]_{ii}, \qquad i = 1, \cdots, W. \tag{A.281}
$$

Our primary usage of SVD will be in the context of the eigenvalues and eigenvectors of the sample correlation matrix.

There are important computation reasons for working directly with the data matrix $\tilde{\mathbf{X}}$ rather than the sample covariance matrix. The dynamic range required to deal with $\tilde{\mathbf{X}}\tilde{\mathbf{X}}^H$ is doubled. Thus, for a specified numerical accuracy, the required word length is doubled.

There are several efficient computational schemes for computing the SVD. Various algorithms are discussed in detail in Sections 11.7–11.12 of [Hay91] and Chapters 11 and 12 of [Hay96]. All of the various computational programs, such as LINPACK, EISPACK, and MATLAB, have SVD algorithms included. SVD is widely used in a number of signal processing applications, and there is extensive literature. We utilize it when we study adaptive beamforming and parameter estimation.

A.6 QR Decomposition

A.6.1 Introduction

The techniques developed in this section are motivated by the least squares estimation problem. The least squares equation has the form

$$\mathbf{\Phi}^*(K)\hat{\mathbf{w}}_{lse}(K) = \mathbf{\Phi}^*_{\mathbf{x}d^*}(K), \tag{A.282}$$

where

$$\mathbf{\Phi}(K) = \sum_{k=1}^{K} \mathbf{X}(k)\mu^{\frac{K-k}{2}}\mu^{\frac{K-k}{2}}\mathbf{X}^H(k), \tag{A.283}$$

is an exponentially weighted sample correlation matrix and

$$\mathbf{\Phi}_{\mathbf{x}d^*}(K) = \sum_{k=1}^{K} \mathbf{X}(k)\mu^{\frac{K-k}{2}}\mu^{\frac{K-k}{2}}D^*(k), \tag{A.284}$$

is an exponentially weighted sample cross-correlation matrix.

We define a $K \times N$ exponentially weighted data matrix,

$$\mathbf{A}_\mu(K) = \boldsymbol{\mu}(K)\begin{bmatrix} \mathbf{X}^T(1) \\ \hdashline \mathbf{X}^T(2) \\ \hdashline \vdots \\ \hdashline \mathbf{X}^T(K) \end{bmatrix}, \tag{A.285}$$

where

$$\boldsymbol{\mu}(K) = \text{diag}\left[\begin{array}{cccc} \mu^{\frac{K-1}{2}}, & \mu^{\frac{K-2}{2}}, & \cdots, & 1 \end{array}\right]. \tag{A.286}$$

Then,

$$\mathbf{\Phi}(K) = \mathbf{A}^T_\mu(K)\,\mathbf{A}^*_\mu(K), \tag{A.287}$$

and

$$\mathbf{\Phi}_{\mathbf{x}d^*}(K) = \mathbf{A}^T_\mu(K)\,\mathbf{d}^*_\mu, \tag{A.288}$$

where

$$\mathbf{d}_\mu = \boldsymbol{\mu}(K)\left[\begin{array}{ccc} D(1), & \cdots, & D(K) \end{array}\right]^T. \tag{A.289}$$

Then, (A.282) can be written as

$$\mathbf{A}^H_\mu(K)\,\mathbf{A}_\mu(K)\,\mathbf{w}^* = \mathbf{A}^H_\mu\,\mathbf{d}_\mu. \tag{A.290}$$

One approach to solving (A.282) is to factor $\mathbf{\Phi}$ using a Cholesky decomposition. Then, we can solve (A.282) using back-substitution (e.g., [GVL89]).

A problem that one encounters with this approach is the requirement for high numerical precision. In this section, we discuss techniques that utilize the data matrix rather than the sample covariance matrix.

All of these techniques are classical and are discussed in a number of references and textbooks. Books that discuss solutions to least squares estimation problems include Lawson and Hanson [LH74], Wilkinson [Wil65], Stewart [Ste73], and Golub and Van Loan [GVL89]. Books that include discussion of these techniques as background to adaptive filtering or adaptive beamforming include Haykin ([Hay96], [Hay91]), Haykin and Steinkardt [HS92], and Proakis et al. [PRLN92]. Our discussion is similar to Section 5.3 of [PRLN92].

In Section A.6.2, we discuss the QR decomposition and its application to the solution of (A.282). In Sections A.6.3 and A.6.4, we develop two techniques for accomplishing the QR decomposition: the Givens rotation and the Householder transformation. A third technique, the modified Gram-Schmidt algorithm, is discussed in [PRLN92] and will not be included.

All of our discussion in this section assumes K is fixed. In Chapter 7, we discussed recursive techniques that are the basis of our adaptive beamforming algorithm.

A.6.2 QR Decomposition

In this section, we show how to solve (A.282) in an efficient manner using the data matrix $\mathbf{A}_\mu$. By avoiding the computation of $\mathbf{\Phi}$, we can significantly improve our numerical accuracy. The key is the decomposition of the $K \times N$ matrix $\mathbf{A}_\mu$ into a $K \times N$ matrix that is partitioned into a $N \times N$ upper triangular matrix $\tilde{\mathbf{R}}$ and a $(K-N) \times N$ null matrix. There exists a $K \times K$ unitary matrix $\mathbf{Q}$ such that

$$\mathbf{Q}\mathbf{A}_\mu = \begin{bmatrix} \tilde{\mathbf{R}} \\ \mathbf{0} \end{bmatrix}, \tag{A.291}$$

where $\tilde{\mathbf{R}}$ has the form

$$\tilde{\mathbf{R}} = \begin{bmatrix} \tilde{r}_{11} & \tilde{r}_{12} & \cdots & \tilde{r}_{1N} \\ 0 & \tilde{r}_{22} & \cdots & \tilde{r}_{2N} \\ \cdots & \cdots & \ddots & \cdots \\ 0 & 0 & \cdots & \tilde{r}_{NN} \end{bmatrix}, \tag{A.292}$$

where we have suppressed the K dependence, and $\mathbf{0}$ is a $(K-N)\times N$ matrix with zero elements.

Note that

$$\mathbf{\Phi} = \mathbf{A}_\mu^H \mathbf{A}_\mu = \mathbf{A}_\mu^H \mathbf{Q}^H \mathbf{Q} \mathbf{A}_\mu = \tilde{\mathbf{R}}^H \tilde{\mathbf{R}}, \tag{A.293}$$

so that $\tilde{\mathbf{R}}$ is the Cholesky factor of the weighted sample covariance matrix.

The decomposition has several other properties that we will utilize.[6]

(i) Each row of the triangular matrix $\tilde{\mathbf{R}}$ is unique up to a complex scale factor with unit magnitude. Hence if $\tilde{\mathbf{R}}_1$ and $\tilde{\mathbf{R}}_2$ are two different $\tilde{\mathbf{R}}$-factors in a QR decomposition of the $\mathbf{A}_\mu$, we can always find a complex-valued diagonal matrix β whose complex-valued diagonal elements have unit magnitude, so that $\tilde{\mathbf{R}}_1 = \beta\tilde{\mathbf{R}}_2$.

(ii) For any matrix $\mathbf{A}_\mu$ there exists a unique $\tilde{\mathbf{R}}$ whose diagonal elements are all real and nonnegative.

(iii) If $K > N$, the unitary matrix $\mathbf{Q}$ in a QR decompositions of $\mathbf{A}_\mu$ is not unique for the same $\tilde{\mathbf{R}}$-factor.

(iv) Since the condition number of the unitary matrix $\mathbf{Q}$ is unity, it follows that the condition number of $\tilde{\mathbf{R}}$ is equal to the condition number of $\tilde{\mathbf{X}}_\mu$.

To solve (A.293), we substitute (A.291) into (A.293) and obtain

$$\begin{bmatrix} \tilde{\mathbf{R}} \\ \mathbf{0} \end{bmatrix}^H \mathbf{Q}\mathbf{Q}^H \begin{bmatrix} \tilde{\mathbf{R}} \\ \mathbf{0} \end{bmatrix} \mathbf{w}^* = \begin{bmatrix} \tilde{\mathbf{R}} \\ \mathbf{0} \end{bmatrix}^H \mathbf{Q}\mathbf{d}_\mu. \tag{A.294}$$

Now partition $\mathbf{d}_\mu$ as

$$\mathbf{Q}\mathbf{d}_\mu = \begin{bmatrix} \mathbf{p} \\ \mathbf{v} \end{bmatrix}, \tag{A.295}$$

where $\mathbf{p}$ is an $N \times 1$ vector.

Using (A.295) in (A.294), we obtain

$$\tilde{\mathbf{R}}^H \tilde{\mathbf{R}} \mathbf{w}^* = \tilde{\mathbf{R}}^H \mathbf{p}, \tag{A.296}$$

or

$$\tilde{\mathbf{R}} \mathbf{w}^* = \mathbf{p}. \tag{A.297}$$

It is now straightforward to solve (A.297) because $\tilde{\mathbf{R}}$ is upper triangular. Two observations are important:

[6]The list of properties is taken directly from p. 291 of [PRLN92].

(i) We do not need an explicit expression for $\mathbf{Q}$.

(ii) The condition number of $\tilde{\mathbf{R}}$ is the square root of the condition number of $\mathbf{\Phi}$, so our numerical accuracy has improved.

A.6.3 Givens Rotation

The Givens rotation is a method for implementing the QR decomposition by a sequence of plane rotations. The technique is due to Givens [Giv58].

To introduce the technique, we first consider a 2×1 complex vector $\mathbf{v}$,

$$\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}. \tag{A.298}$$

The Givens matrix $\mathbf{G}$ is a 2×2 unitary matrix,

$$\mathbf{G} = \begin{bmatrix} c^* & s \\ -s^* & c \end{bmatrix}. \tag{A.299}$$

We select the elements of $\mathbf{G}$ such that

$$\mathbf{G}\mathbf{v} = \begin{bmatrix} c^* v_1 + s v_2 \\ -s^* v_1 + c v_2 \end{bmatrix} \triangleq \begin{bmatrix} v_1' \\ 0 \end{bmatrix} = \mathbf{v}'. \tag{A.300}$$

The zero term in (A.300) requires,

$$s^* v_1 = c v_2, \tag{A.301}$$

and the unitary condition requires

$$|c|^2 + |s|^2 = 1. \tag{A.302}$$

Therefore,

$$c = \frac{v_1}{\sqrt{|v_1|^2 + |v_2|^2}}, \tag{A.303}$$

$$s = \frac{v_2^*}{\sqrt{|v_1|^2 + |v_2|^2}}, \tag{A.304}$$

and

$$v_1' = \sqrt{|v_1|^2 + |v_2|^2}. \tag{A.305}$$

Thus, $\mathbf{v}'$ has the same length as $\mathbf{v}$, but has only one non-zero component.

In a similar manner, if we have a row vector $\mathbf{u}$,

$$\mathbf{u} = \begin{bmatrix} u_1 & u_2 \end{bmatrix}. \tag{A.306}$$

We can write

$$\mathbf{u}\mathbf{G}^H = \begin{bmatrix} u_1' & 0 \end{bmatrix}, \tag{A.307}$$

where $\mathbf{G}$ is given by (A.303) and (A.304), with v_1 and v_2 replaced by u_1 and u_2. This result follows directly from (A.300) by letting $\mathbf{u} = \mathbf{v}^H$.

The operation described by (A.299) is called a plane rotation because, if v_1 and v_2 are real,

$$c = \cos\phi, \tag{A.308}$$

and

$$s = \sin\phi, \tag{A.309}$$

where ϕ is the angle that $\mathbf{v}$ is rotated. The adjective "plane" denotes that $\mathbf{v}$ stays in the same plane.

Now consider a $K \times 1$ complex vector $\mathbf{v}$,

$$\mathbf{v} = \begin{bmatrix} v_1, & \cdots, & v_m, & \cdots, & v_n, & \cdots, & v_K \end{bmatrix}^T. \tag{A.310}$$

We operate on $\mathbf{v}$ with a Givens rotation $\mathbf{G}$ and leave all of the elements of $\mathbf{v}$ unchanged except for v_m and v_n. This requires a $K \times K$ unitary matrix of the form,

$$\mathbf{G}(m,n) = \begin{bmatrix} \mathbf{I} & & & & & & 0 \\ & \ddots & & & & & \\ & & c_{mm} & & s_{mn} & & \\ & & & \ddots & & & \\ & & & \mathbf{I} & & & \\ & & & & \ddots & & \\ & & s_{nm} & & c_{nn} & & \\ & & & & & \ddots & \\ 0 & & & & & & \mathbf{I} \end{bmatrix}. \tag{A.311}$$

$\mathbf{G}(m,n)$ is an identity matrix except for four elements. The two c elements are on the diagonal,

$$c = c_{mm}^* = c_{nn} = \frac{v_m}{\sqrt{|v_m|^2 + |v_n|^2}}, \tag{A.312}$$

and the two s elements are off the diagonal, with

$$s = s_{mn} = \frac{v_n^*}{\sqrt{|v_m|^2 + |v_n|^2}}, \tag{A.313}$$

and

$$s_{nm} = s_{mn}^* = -s^*. \tag{A.314}$$

Then,

$$\mathbf{G}(m,n)\mathbf{v} = \begin{bmatrix} v_1, & \cdots, & \sqrt{|v_m|^2 + |v_n|^2}, & \cdots, & 0, & \cdots, & v_K \end{bmatrix}^T. \tag{A.315}$$

We have made the nth element zero and rotated it into the mth element.

We now demonstrate how to apply a sequence of Givens rotations to accomplish a QR decomposition of $\mathbf{A}_\mu$.

Example A.6.1

Consider the case when $N = 2$ and $K = 3$. Then the exponentially weighted data matrix can be written as

$$\mathbf{A}_\mu = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix}. \tag{A.316}$$

We utilize three successive Givens rotations to obtain the desired QR decomposition. We eliminate a_{31} first. From (A.311),

$$\mathbf{G}_1 = \begin{bmatrix} c_1^* & 0 & s_1 \\ 0 & 1 & 0 \\ -s_1^* & 0 & c_1 \end{bmatrix}, \tag{A.317}$$

where we have chosen to rotate a_{31} into a_{11}. Then

$$c_1 = \frac{a_{11}}{r_1}, \tag{A.318}$$

$$s_1 = \frac{a_{31}^*}{r_1}, \tag{A.319}$$

and

$$r_1 \triangleq \sqrt{|a_{11}|^2 + |a_{31}|^2}. \tag{A.320}$$

Then,

$$\mathbf{G}_1\mathbf{A}_\mu = \begin{bmatrix} r_1 & c_1^* a_{12} + s_1 a_{32} \\ a_{21} & a_{22} \\ 0 & -s_1^* a_{12} + c_1 a_{32} \end{bmatrix} \triangleq \begin{bmatrix} r_1 & a_{12}' \\ a_{21} & a_{22} \\ 0 & a_{32}' \end{bmatrix}. \tag{A.321}$$

Note that only the first and third rows were affected by $\mathbf{G}_1$.

The next step is to eliminate a_{21},

$$\mathbf{G}_2 = \begin{bmatrix} c_2^* & s_2 & 0 \\ -s_2^* & c_2 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{A.322}$$

where

$$c_2 = \frac{r_1}{\tilde{r}_{11}}, \tag{A.323}$$

$$s_2 = \frac{a_{21}^*}{\tilde{r}_{11}}, \tag{A.324}$$

and

$$\tilde{r}_{11} \triangleq \sqrt{|r_1|^2 + |a'_{12}|^2}. \tag{A.325}$$

Then,

$$\mathbf{G}_2\mathbf{G}_1\mathbf{A}_\mu = \begin{bmatrix} \tilde{r}_{11} & c_2^* a'_{12} + s_2 a_{22} \\ 0 & -s_2^* a'_{12} + c_2 a_{22} \\ 0 & a'_{32} \end{bmatrix} \triangleq \mathbf{G}_2\mathbf{G}_1\mathbf{A}_\mu = \begin{bmatrix} \tilde{r}_{11} & \tilde{r}_{12} \\ 0 & a'_{22} \\ 0 & a'_{32} \end{bmatrix}. \tag{A.326}$$

The notation is chosen because the first row will not be affected by the next step so that $\tilde{r}_{11}$ and $\tilde{r}_{12}$ are the elements in $\tilde{\mathbf{R}}$.

The last step is to eliminate a'_{32}. We do this without disturbing the first row.

$$\mathbf{G}_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & c_3^* & s_3 \\ 0 & -s_3^* & c_3 \end{bmatrix}, \tag{A.327}$$

where

$$c_3 = \frac{a'_{22}}{\tilde{r}_{22}}, \tag{A.328}$$

$$s_3 = \frac{(a'_{32})^*}{\tilde{r}_{22}}, \tag{A.329}$$

and

$$\tilde{r}_{22} \triangleq \sqrt{|a'_{22}|^2 + |a'_{32}|^2}. \tag{A.330}$$

Then,

$$\mathbf{G}_3\mathbf{G}_2\mathbf{G}_1\mathbf{A}_\mu = \begin{bmatrix} \tilde{r}_{11} & \tilde{r}_{12} \\ 0 & \tilde{r}_{22} \\ 0 & 0 \end{bmatrix}, \tag{A.331}$$

which is the desired result.

Note that $\tilde{r}_{11}$ and $\tilde{r}_{22}$ are real (see (A.325) and (A.330)). In each step, it is straightforward to locate the c and s parameters. The $-s^*$ parameter is in the same position as the element that we are eliminating and the c^* parameter is in the same position as the element we are rotating into. We should also observe that the elimination order is not unique (see Problem A.6.1), but $\tilde{\mathbf{R}}$ is. The $\mathbf{Q}$ matrix is

$$\mathbf{Q} = \mathbf{G}_3\mathbf{G}_2\mathbf{G}_1, \tag{A.332}$$

but is not explicitly used in the algorithm.

The extension of the above example to an arbitrary $K \times N$ $(K \geq N)$ data matrix is straightforward. We apply $K-1$ successive Givens rotations to eliminate all of the elements in the first column except for the elements in the first row. We then apply $K-2$ Givens rotations, without disturbing the first row, to eliminate the bottom $K-2$ elements in the second column. We process each successive column in a similar manner. This requires,

$$N_G = (K-1)(K-2)\cdots(K-N) = KN - \frac{N(N-1)}{2}, \tag{A.333}$$

rotations. We use the same sequence of rotations on $\mathbf{d}_\mu$ to obtain $\mathbf{p}$.

A.6.4 Householder Transformation

The Givens rotation technique obtains a triangular $\tilde{\mathbf{R}}$ by eliminating one element with each rotation. The Householder transformation eliminates all except one element in a column in each step. The transformation is due to Householder ([Hou58], [Hou64]). A tutorial paper by Steinhardt [Ste88] gives an excellent discussion of the transformation and its application to signal processing.

The Householder transformation can be written as

$$\mathbf{H} = \mathbf{I} - \frac{2\mathbf{u}\mathbf{u}^H}{\|\mathbf{u}\|^2} = \mathbf{I} - 2\mathbf{u}\left[\mathbf{u}^H\mathbf{u}\right]^{-1}\mathbf{u}^H, \tag{A.334}$$

where $\mathbf{u}$ is an $N \times 1$ vector whose norm is $\|\mathbf{u}\|$. We can also write it as

$$\mathbf{H} = \mathbf{I} - 2\mathbf{P}_{\mathbf{u}}, \tag{A.335}$$

where

$$\mathbf{P}_{\mathbf{u}} = \mathbf{u}\left[\mathbf{u}^H\mathbf{u}\right]^{-1}\mathbf{u}^H, \tag{A.336}$$

is the projection matrix onto the $\mathbf{u}$ subspace.

If we pre-multiply the $N \times 1$ complex vector $\mathbf{v}$ by $\mathbf{H}$, we have

$$\mathbf{H}\mathbf{v} = (\mathbf{I} - 2\mathbf{P}_{\mathbf{u}})\,\mathbf{v} = \mathbf{v} - 2\mathbf{P}_{\mathbf{u}}\mathbf{v}. \tag{A.337}$$

From (A.334), we observe that $\mathbf{H}$ is a Hermitian unitary matrix,

$$\mathbf{H}^{-1} = \mathbf{H}^H = \mathbf{H}, \tag{A.338}$$

so that the transformation preserves length

$$\|\mathbf{H}\mathbf{v}\| = \|\mathbf{v}\|. \tag{A.339}$$

Now consider the $N \times 1$ complex vector $\mathbf{v}$,

$$\mathbf{v} = \begin{bmatrix} v_1, & v_2, & \cdots, & v_i, & \cdots, & v_N \end{bmatrix}^T. \tag{A.340}$$

We use $\mathbf{H}$ to eliminate all of the elements except $\mathbf{v}_i$. We define

$$v_{in} = \frac{v_i}{|v_i|}, \tag{A.341}$$

and

$$\|\mathbf{v}\|^2 = \mathbf{v}^H\mathbf{v}. \tag{A.342}$$

We use (A.334) with

$$\mathbf{u} = \mathbf{v} + v_{in}\|\mathbf{v}\|\mathbf{e}_i, \tag{A.343}$$

where

$$\mathbf{e}_i^T = \begin{bmatrix} 0, & \cdots, & 0 & 1, & 0 & \cdots & 0 \end{bmatrix}^T, \tag{A.344}$$

has one as the ith element and zero elsewhere.

Using (A.341)–(A.344), we find

$$\|\mathbf{u}\|^2 = \mathbf{u}^H\mathbf{u} = 2\|\mathbf{v}\|\left(\|\mathbf{v}\| + |v_i|\right). \tag{A.345}$$

Then, we observe that

$$\mathbf{u}^H\mathbf{v} = \left(\mathbf{v}^H + v_{in}^*\|\mathbf{v}\|\mathbf{e}_i^T\right)\mathbf{v} = \|\mathbf{v}\|^2 + |v_i|\|\mathbf{v}\|. \tag{A.346}$$

Therefore,

$$\begin{aligned} \mathbf{Hv} &= \left(\mathbf{I} - \frac{2\mathbf{u}\mathbf{u}^H}{\|\mathbf{u}\|^2}\right)\mathbf{v} \\ &= \mathbf{v} - \mathbf{u} \\ &= -v_{in}\|\mathbf{v}\|\mathbf{e}_i, \end{aligned} \tag{A.347}$$

which is the desired result.

The procedure for using the Householder transformation to do a QR decomposition is analogous to the Givens rotation technique. We illustrate it with a similar model to the model in Example A.6.1.

Example A.6.2

Consider the case when $K = 4$ and $N = 2$. In view of our results in Example A.6.1, we define an $K \times (N+1)$ augmented data matrix

$$\begin{aligned} \mathbf{A}_{aug} &\triangleq \begin{bmatrix} \mathbf{A}_\mu & \mathbf{d}_\mu \end{bmatrix} \\ &\triangleq \begin{bmatrix} a_{11} & a_{12} & d_1 \\ a_{21} & a_{22} & d_2 \\ a_{31} & a_{32} & d_3 \\ a_{41} & a_{42} & d_4 \end{bmatrix} \\ &\triangleq \begin{bmatrix} \mathbf{a}_1 \vdots \mathbf{a}_2 \vdots \mathbf{d}_\mu \end{bmatrix}. \end{aligned} \tag{A.348}$$

The first step is to use a Householder transformation to eliminate a_{21}, a_{31}, and a_{41}:

$$\mathbf{H}_1 = \mathbf{I} - \frac{2\mathbf{u}_1\mathbf{u}_1^H}{\|\mathbf{u}_1\|^2}, \tag{A.349}$$

where

$$\mathbf{u}_1 = \mathbf{a}_1 + \frac{a_{11}}{|a_{11}|}\|\mathbf{a}_1\|\mathbf{e}_1. \tag{A.350}$$

Then,

$$\mathbf{H}_1\mathbf{A}_{aug} = \left[\begin{array}{c:c:c} -\frac{a_{11}}{|a_{11}|}\|\mathbf{a}_1\| & a'_{12} & d'_1 \\ 0 & a'_{22} & d'_2 \\ 0 & a'_{32} & d'_3 \\ 0 & a'_{42} & d'_4 \end{array}\right]. \tag{A.351}$$

In the next step we want to eliminate a'_{32} and a'_{42} without disturbing the first row. We define

$$\mathbf{a}_2^{(-)} = \left[\begin{array}{cccc} 0 & a'_{22} & a'_{32} & a'_{42} \end{array}\right]. \tag{A.352}$$

Then,

$$\mathbf{H}_2 = \mathbf{I} - \frac{2\mathbf{u}_2\mathbf{u}_2^H}{\|\mathbf{u}_2\|^2}, \tag{A.353}$$

where

$$\mathbf{u}_2 = \mathbf{a}_2^{(-)} + \frac{a'_{22}}{|a'_{22}|}\|\mathbf{a}_2^{(-)}\|\mathbf{e}_2. \tag{A.354}$$

Applying $\mathbf{H}_2$ to (A.351) gives

$$\mathbf{H}_2\mathbf{H}_1\mathbf{A}_{aug} = \left[\begin{array}{c:c:c} -\frac{a_{11}}{|a_{11}|}\|\mathbf{a}_1\| & a'_{12} & d''_1 \\ 0 & -\frac{a'_{22}}{|a'_{22}|}\|\mathbf{a}_2^{(-)}\| & d''_2 \\ 0 & 0 & d''_3 \\ 0 & 0 & d''_4 \end{array}\right]. \tag{A.355}$$

The last step is to eliminate d''_4. Define

$$\mathbf{a}_3^{(-)} = \left[\begin{array}{cccc} 0 & 0 & d''_3 & d''_4 \end{array}\right]. \tag{A.356}$$

Then

$$\mathbf{H}_3 = \mathbf{I} - \frac{2\mathbf{u}_3\mathbf{u}_3^H}{\|\mathbf{u}_3\|^2}, \tag{A.357}$$

where

$$\mathbf{u}_3 = \mathbf{a}_3^{(-)} + \frac{d''_3}{|d''_3|}\|\mathbf{a}_3^{(-)}\|\mathbf{e}_3. \tag{A.358}$$

Then

$$\mathbf{H}_3\mathbf{H}_2\mathbf{H}_1\mathbf{A}_{aug} = \begin{bmatrix} r_{11} & r_{12} & \hat{d}_1 \\ 0 & r_{22} & \hat{d}_2 \\ 0 & 0 & e_3 \\ 0 & 0 & 0 \end{bmatrix}. \tag{A.359}$$

The top two rows give the required QR decomposition.

Note that the diagonal elements are not necessarily real or positive. As a final step we multiply each row by a complex number with unity magnitude. The resulting $\tilde{\mathbf{R}}$ matrix is identical to the one obtained by the Givens rotation. We denote this diagonal matrix as

$$\boldsymbol{\beta} = \text{diag}\left[\begin{array}{cccc} \beta_1, & \beta_2, & \cdots, & \beta_N \end{array}\right]. \tag{A.360}$$

The extension to arbitrary K and N, $(K \geq N)$, is straightforward.

$$\mathbf{Q} = \mathbf{H}_{N+1}\mathbf{H}_N\mathbf{H}_{N-1}\cdots\mathbf{H}_1. \tag{A.361}$$

A.7 Derivative Operations

In many of our array processing applications we encounter a non-negative cost function that we want to maximize (or minimize) with respect to a vector or matrix. To accomplish this optimization we need to find a vector or matrix derivative.

Another application of derivatives that we will encounter is the calculation of the CRB. We recall from Chapter 2 of DEMT I [VT68], [VT01a] that the CRB is a fundamental tool in the parameter estimation.

In Section A.7.1, we develop the derivative of a scalar with respect to a vector.[7] In Section A.7.2, we develop the derivative of a scalar with respect to a matrix. In Section A.7.3, we develop derivatives with respect to a parameter. In Section A.7.4, we extend these results to complex vectors.

A.7.1 Derivative of Scalar with Respect to Vector

First consider a scalar $a(\boldsymbol{\theta})$ that is a function of the $M \times 1$ real vector parameter $\boldsymbol{\theta}$. The derivative of a with respect to $\boldsymbol{\theta}$ is an $M \times 1$ vector

$$\frac{\partial}{\partial \boldsymbol{\theta}} a(\boldsymbol{\theta}) = \begin{bmatrix} \frac{\partial a(\boldsymbol{\theta})}{\partial \theta_1} \\ \frac{\partial a(\boldsymbol{\theta})}{\partial \theta_2} \\ \vdots \\ \frac{\partial a(\boldsymbol{\theta})}{\partial \theta_M} \end{bmatrix}. \tag{A.362}$$

Similarly, if we consider a $1 \times N$ row vector,

$$\mathbf{a}^T(\boldsymbol{\theta}) = \left[\begin{array}{c:c:c:c} a_1(\boldsymbol{\theta}) & a_2(\boldsymbol{\theta}) & \cdots & a_N(\boldsymbol{\theta}) \end{array}\right], \tag{A.363}$$

the gradient is a $M \times N$ matrix,

[7] The results in Sections A.7.1 and A.7.2 appear in many places. An early engineering reference is [AS65]. Scharf [Sch91] has a concise summary in his appendix to Chapter 6. Most of the results are derived in Graham [Gra81].

$$\frac{\partial}{\partial \boldsymbol{\theta}}\mathbf{a}^T = \left[\begin{array}{ccc} \frac{\partial}{\partial \boldsymbol{\theta}}a_1(\boldsymbol{\theta}) & \cdots & \frac{\partial}{\partial \boldsymbol{\theta}}a_N(\boldsymbol{\theta}) \end{array}\right] \tag{A.364}$$

$$= \begin{bmatrix} \frac{\partial a_1(\boldsymbol{\theta})}{\partial \theta_1} & \cdots & \frac{\partial a_N(\boldsymbol{\theta})}{\partial \theta_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial a_1(\boldsymbol{\theta})}{\partial \theta_M} & \cdots & \frac{\partial a_N(\boldsymbol{\theta})}{\partial \theta_M} \end{bmatrix}. \tag{A.365}$$

Similarly,

$$\frac{\partial}{\partial \boldsymbol{\theta}}\mathbf{a} = \left[\frac{\partial}{\partial \boldsymbol{\theta}}\mathbf{a}^T\right]^T, \tag{A.366}$$

is a $N \times M$ matrix where each row is the gradient of the scalar a_i with respect to $\boldsymbol{\theta}$.

Some cases that we will encounter are:

Products

(i)

$$\frac{\partial}{\partial \boldsymbol{\theta}}\boldsymbol{\theta}^T = \mathbf{I}. \tag{A.367}$$

(ii)

$$\frac{\partial}{\partial \boldsymbol{\theta}}\mathbf{b}^T\boldsymbol{\theta} = \frac{\partial}{\partial \boldsymbol{\theta}}\boldsymbol{\theta}^T\mathbf{b} = \mathbf{b}. \tag{A.368}$$

(iii)

$$\frac{\partial}{\partial \boldsymbol{\theta}}\mathbf{a}^T(\boldsymbol{\theta})\mathbf{b}(\boldsymbol{\theta}) = \left(\frac{\partial}{\partial \boldsymbol{\theta}}\mathbf{a}^T(\boldsymbol{\theta})\right)\mathbf{b}(\boldsymbol{\theta}) + \left(\frac{\partial}{\partial \boldsymbol{\theta}}\mathbf{b}^T(\boldsymbol{\theta})\right)\mathbf{a}(\boldsymbol{\theta}). \tag{A.369}$$

Quadratic Forms

In the next set of equations, $\mathbf{Q}$ is not a function of $\boldsymbol{\theta}$

(iv)

$$\frac{\partial}{\partial \boldsymbol{\theta}}\boldsymbol{\theta}^T\mathbf{Q}\boldsymbol{\theta} = 2\mathbf{Q}\boldsymbol{\theta}. \tag{A.370}$$

(v)

$$\frac{\partial}{\partial \boldsymbol{\theta}} \mathbf{m}^T \mathbf{Q} \mathbf{m} = 2 \left(\frac{\partial}{\partial \boldsymbol{\theta}} \mathbf{m}^T \right) \mathbf{Q} \mathbf{m}. \tag{A.371}$$

(vi)

$$\frac{\partial}{\partial \boldsymbol{\theta}} \exp\left\{ -\frac{1}{2} \boldsymbol{\theta}^T \mathbf{Q}^{-1} \boldsymbol{\theta} \right\} = -\exp\left\{ -\frac{1}{2} \boldsymbol{\theta}^T \mathbf{Q}^{-1} \boldsymbol{\theta} \right\} \mathbf{Q}^{-1} \boldsymbol{\theta}. \tag{A.372}$$

(vii)

$$\frac{\partial}{\partial \boldsymbol{\theta}} \ln \left(\boldsymbol{\theta}^T \mathbf{Q} \boldsymbol{\theta} \right) = 2 \left(\boldsymbol{\theta}^T \mathbf{Q} \boldsymbol{\theta} \right)^{-1} \mathbf{Q} \boldsymbol{\theta}. \tag{A.373}$$

A.7.2 Derivative of Scalar with Respect to Matrix

Consider a scalar function of a $M \times N$ matrix $\mathbf{X}$, which denotes by $a(\mathbf{X})$. The derivative of $a(\mathbf{X})$ with respect to $\mathbf{X}$ is the $M \times N$ matrix,

$$\frac{\partial}{\partial \mathbf{X}} a(\mathbf{X}) = \begin{bmatrix} \frac{\partial a(\mathbf{X})}{\partial x_{11}} & \frac{\partial a(\mathbf{X})}{\partial x_{12}} & \cdots & \frac{\partial a(\mathbf{X})}{\partial x_{1N}} \\ \frac{\partial a(\mathbf{X})}{\partial x_{21}} & & & \vdots \\ \vdots & & & \\ \frac{\partial a(\mathbf{X})}{\partial x_{M1}} & \cdots & & \frac{\partial a(\mathbf{X})}{\partial x_{MN}} \end{bmatrix}. \tag{A.374}$$

Each column is the vector derivative of the scalar $a(\mathbf{X})$ with respect to a column of $\mathbf{X}$:

$$\frac{\partial}{\partial \mathbf{X}} a(\mathbf{X}) = \left(\frac{\partial}{\partial \mathbf{r}_1} a(\mathbf{X}) \,\vdots\, \cdots \,\vdots\, \frac{\partial}{\partial \mathbf{r}_N} a(\mathbf{X}) \right). \tag{A.375}$$

$$\mathbf{X} = \left(\mathbf{x}_1 \quad \cdots \quad \mathbf{x}_N \right). \tag{A.376}$$

$$\mathbf{x}_i = \begin{bmatrix} x_{1i} \\ x_{2i} \\ \vdots \\ x_{Mi} \end{bmatrix}. \tag{A.377}$$

One application where we will utilize matrix gradients is in the computation of the CRB.

Typical functions are:

1.

$$\frac{\partial}{\partial \mathbf{X}} \operatorname{tr} \mathbf{X} = \mathbf{I}. \tag{A.378}$$

In (2)–(5), $\mathbf{A}$ and $\mathbf{B}$ are not functions of $\mathbf{X}$.

2.

$$\frac{\partial}{\partial \mathbf{X}} \operatorname{tr} \left[\mathbf{A}\mathbf{X}\right] = \frac{\partial}{\partial \mathbf{X}^T} \operatorname{tr} \left[\mathbf{A}\mathbf{X}^T\right] = \left(\frac{\partial}{\partial \mathbf{X}} \operatorname{tr} \left[\mathbf{A}\mathbf{X}^T\right]\right)^T = \mathbf{A}^T. \tag{A.379}$$

3.

$$\frac{\partial}{\partial \mathbf{X}} \operatorname{tr} \left[\mathbf{A}\mathbf{X}^T\right] = \frac{\partial}{\partial \mathbf{X}^T} \operatorname{tr} \left[\mathbf{A}\mathbf{X}\right] = \mathbf{A}. \tag{A.380}$$

4.

$$\frac{\partial}{\partial \mathbf{X}} \operatorname{tr} \left[\mathbf{A}\mathbf{X}\mathbf{B}\right] = \frac{\partial}{\partial \mathbf{X}^T} \operatorname{tr} \left[\mathbf{A}\mathbf{X}^T\mathbf{B}\right] = \mathbf{A}^T\mathbf{B}^T. \tag{A.381}$$

5.

$$\frac{\partial}{\partial \mathbf{X}} \operatorname{tr} \left[\mathbf{A}\mathbf{X}^T\mathbf{B}\right] = \frac{\partial}{\partial \mathbf{X}^T} \operatorname{tr} \left[\mathbf{A}\mathbf{X}\mathbf{B}\right] = \mathbf{B}\mathbf{A}. \tag{A.382}$$

6.

$$\frac{\partial}{\partial \mathbf{X}} \operatorname{tr} \left[\mathbf{A}\mathbf{X}^{-1}\right] = \frac{\partial}{\partial \mathbf{X}} \operatorname{tr} \left[\mathbf{X}^{-1}\mathbf{A}\right] = \left(-\mathbf{X}^{-1}\mathbf{A}\mathbf{X}^{-1}\right)^T. \tag{A.383}$$

7.

$$\frac{\partial}{\partial \mathbf{X}} \operatorname{tr} \mathbf{X}^n = n\left(\mathbf{X}^{n-1}\right)^T. \tag{A.384}$$

8.

$$\frac{\partial}{\partial \mathbf{X}} \operatorname{tr} \left[\exp \mathbf{X}\right] = \exp \mathbf{X}. \tag{A.385}$$

9.

$$\frac{\partial}{\partial \mathbf{X}} \det \mathbf{X} = \det \mathbf{X} \left(\mathbf{X}^{-1}\right)^T. \tag{A.386}$$

10.

$$\frac{\partial}{\partial \mathbf{X}} \ln \det \mathbf{X} = \left(\mathbf{X}^{-1}\right)^T. \tag{A.387}$$

11.

$$\frac{\partial}{\partial \mathbf{X}} \det \mathbf{X}^n = n\left(\det \mathbf{X}\right)^n \left(\mathbf{X}^{-1}\right)^T. \tag{A.388}$$

12. Kronecker products:

$$\begin{aligned}\frac{\partial(\mathbf{X}\otimes\mathbf{Z})}{\partial\mathbf{Y}} &= \frac{\partial\mathbf{X}}{\partial\mathbf{Y}}\otimes\mathbf{Z}\\ &+(\mathbf{I}_s\otimes\mathbf{U}_{p\times r})\left(\frac{\partial\mathbf{Z}}{\partial\mathbf{Y}}\otimes\mathbf{X}\right)(\mathbf{I}_t\otimes\mathbf{U}_{l\times q}),\end{aligned}\tag{A.389}$$

where $\mathbf{U}$ is the permutation matrix and the dimensions of $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{Z}$ are $p\times q$, $s\times t$, and $r\times l$, respectively.

A.7.3 Derivatives with Respect to Parameter

If $\mathbf{X}$ is a function of a real parameter, $\boldsymbol{\theta}$, we can use the chain rule to obtain

1.

$$\begin{aligned}\frac{\partial\ln\det\mathbf{X}}{\partial\theta_i} &= \operatorname{tr}\left[\left[\frac{\partial\ln\det\mathbf{X}}{\partial\mathbf{X}}\right]^T\frac{\partial\mathbf{X}}{\partial\theta_i}\right]\\ &= \operatorname{tr}\left[\mathbf{X}^{-1}\frac{\partial\mathbf{X}}{\partial\theta_i}\right].\end{aligned}\tag{A.390}$$

2.

$$\begin{aligned}\frac{\partial\operatorname{tr}\left[\mathbf{X}^{-1}\mathbf{A}\right]}{\partial\theta_i} &= \operatorname{tr}\left[\left[\frac{\partial\operatorname{tr}\left[\mathbf{X}^{-1}\mathbf{A}\right]}{\partial\mathbf{X}}\right]^T\frac{\partial\mathbf{X}}{\partial\theta_i}\right]\\ &= \operatorname{tr}\left[\left[\left[-\mathbf{X}^{-1}\mathbf{A}\mathbf{X}^{-1}\right]^T\right]^T\frac{\partial\mathbf{X}}{\partial\theta_i}\right]\\ &= -\operatorname{tr}\left[\left[\mathbf{X}^{-1}\mathbf{A}\mathbf{X}^{-1}\right]\frac{\partial\mathbf{X}}{\partial\theta_i}\right].\end{aligned}\tag{A.391}$$

3.

$$\frac{\partial\mathbf{X}^{-1}}{\partial\theta_i} = -\mathbf{X}^{-1}\frac{\partial\mathbf{X}}{\partial\theta_i}\mathbf{X}^{-1}.\tag{A.392}$$

4.

$$\frac{\partial^2\ln\det\mathbf{X}}{\partial\theta_i\partial\theta_j} = \operatorname{tr}\left\{-\mathbf{X}^{-1}\frac{\partial\mathbf{X}}{\partial\theta_i}\mathbf{X}^{-1}\frac{\partial\mathbf{X}}{\partial\theta_j}+\mathbf{X}^{-1}\frac{\partial^2\mathbf{X}}{\partial\theta_i\partial\theta_j}\right\}.\tag{A.393}$$

5. The derivative of the projection matrix is,

$$\frac{\partial}{\partial\theta_i}\mathbf{P}_{\mathbf{V}}^{\perp} = -\frac{\partial}{\partial\theta_i}\mathbf{P}_{\mathbf{V}} = -\mathbf{P}_{\mathbf{V}}^{\perp}\mathbf{V}_i\mathbf{V}^H - \left(\mathbf{P}_{\mathbf{V}}^{\perp}\mathbf{V}_i\mathbf{V}^H\right)^H.\tag{A.394}$$

A.7.4 Complex Gradients

In many applications we have a real-valued function of a complex vector $\mathbf{z}$ and want to find either the minimum or the maximum. One approach is to write

$$\mathbf{z} = \mathbf{x} + j\,\mathbf{y} \tag{A.395}$$

and take derivatives with respect to $\mathbf{x}$ and $\mathbf{y}$. However, the resulting expressions are cumbersome. For the functions of interest in array applications, we can work directly with the vector $\mathbf{z}$ and obtain simpler expressions. However, we must be careful with the differentiation. Brandwood [Bra83] has developed a suitable approach, which we summarize.[8]

First consider the case when z is a complex scalar and the function of interest is

$$f(z) = f(x, y). \tag{A.396}$$

Define a function $g(z, z^*)$ that is analytic with respect to z and z^* independently, and

$$g(z, z^*) = f(x, y). \tag{A.397}$$

Brandwood proves that the partial of $g(z, z^*)$ with respect to z (treating z^* as a constant in g) gives the result,

$$\frac{\partial g(z, z^*)}{\partial z}\Big|_{z=x+jy} = \frac{1}{2}\left(\frac{\partial f(x,y)}{\partial x} - j\frac{\partial f(x,y)}{\partial y}\right). \tag{A.398}$$

Similarly,

$$\frac{\partial g(z, z^*)}{\partial z^*}\Big|_{z^*=x-jy} = \frac{1}{2}\left(\frac{\partial f(x,y)}{\partial x} + j\frac{\partial f(x,y)}{\partial y}\right). \tag{A.399}$$

Brandwood then shows that a necessary and sufficient condition for $f(z)$ to have a stationary point is either

$$\frac{\partial g(z, z^*)}{\partial z} = 0, \tag{A.400}$$

where z^* is treated as a constant in the partial derivative, or

$$\frac{\partial g(z, z^*)}{\partial z^*} = 0, \tag{A.401}$$

where z is treated as a constant in the partial derivative.

[8] Our discussion follows [Bra83].

For the case when $\mathbf{z}$ is a vector, we define the complex gradient operator as

$$\nabla_{\mathbf{z}} = \left[\begin{array}{cccc} \frac{\partial}{\partial z_1}, & \frac{\partial}{\partial z_2}, & \cdots, & \frac{\partial}{\partial z_N} \end{array} \right]^T, \tag{A.402}$$

where

$$\frac{\partial}{\partial z_n} \triangleq \frac{\partial}{\partial x_n} - j\frac{\partial}{\partial y_n}, \quad n = 1, \cdots, N. \tag{A.403}$$

Similarly,

$$\nabla_{\mathbf{z}^H} = \left[\begin{array}{cccc} \frac{\partial}{\partial z_1^*}, & \frac{\partial}{\partial z_2^*}, & \cdots, & \frac{\partial}{\partial z_N^*} \end{array} \right], \tag{A.404}$$

where

$$\frac{\partial}{\partial z_n^*} \triangleq \frac{\partial}{\partial x_n} + j\frac{\partial}{\partial y_n}, \quad n = 1, \cdots, N. \tag{A.405}$$

Let $f(\mathbf{z}) = f(\mathbf{x}, \mathbf{y}) = g(\mathbf{z}, \mathbf{z}^H)$ where $g(\mathbf{z}, \mathbf{z}^H)$ is a real-valued function of $\mathbf{z}$ and $\mathbf{z}^H$, which is analytic with respect to $\mathbf{z}$ and $\mathbf{z}^H$ independently. Then, either

$$\nabla_{\mathbf{z}} g(\mathbf{z}, \mathbf{z}^H) = \mathbf{0}, \tag{A.406}$$

where $\mathbf{z}^H$ is treated as a constant, or

$$\nabla_{\mathbf{z}^H} g(\mathbf{z}, \mathbf{z}^H) = \mathbf{0}, \tag{A.407}$$

where $\mathbf{z}$ is treated as a constant, are necessary and sufficient to determine a stationary point of $f(\mathbf{z})$. We will normally use (A.407) in our applications.

We consider two applications that are encountered in the text. The first application is to minimize $\mathbf{w}^H\mathbf{R}\mathbf{w}$ subject to the constraint $\mathbf{w}^H\mathbf{c} = a$. Here $\mathbf{R}$ is an $N \times N$ Hermitian matrix and $\mathbf{w}$ is an $N \times 1$ complex vector.

We define a real cost function,

$$\begin{aligned} g(\mathbf{w}, \mathbf{w}^H) &= \mathbf{w}^H\mathbf{R}\mathbf{w} + 2Re\left[\lambda(\mathbf{w}^H\mathbf{c} - a)\right] &\text{(A.408)}\\ &= \mathbf{w}^H\mathbf{R}\mathbf{w} + \lambda(\mathbf{w}^H\mathbf{c} - a) + \lambda^*(\mathbf{w}^H\mathbf{c} - a^*), &\text{(A.409)} \end{aligned}$$

where λ is a Lagrange multiplier. Taking the gradient with respect to $\mathbf{w}^H$ and setting the result equal to zero gives

$$\nabla_{\mathbf{w}^H} g(\mathbf{w}, \mathbf{w}^H) = \mathbf{R}\mathbf{w}_o + \lambda\mathbf{c} = \mathbf{0}, \tag{A.410}$$

where the subscript denotes optimum. Then

$$\mathbf{w}_o = -\lambda\mathbf{R}^{-1}\mathbf{c}. \tag{A.411}$$

The constraint equation is used to find λ,

$$\mathbf{c}^H\mathbf{w}_o = -\lambda\mathbf{c}^H\mathbf{R}^{-1}\mathbf{c} = a^*, \tag{A.412}$$

or

$$\lambda = -(\mathbf{c}^H\mathbf{R}^{-1}\mathbf{c})^{-1}a^*. \tag{A.413}$$

Using (A.413) in (A.411) gives

$$\mathbf{w}_o = \frac{a^*\mathbf{R}^{-1}\mathbf{c}}{\mathbf{c}^H\mathbf{R}^{-1}\mathbf{c}}. \tag{A.414}$$

As a second example, we maximize the same function, $\mathbf{w}^H\mathbf{R}\mathbf{w}$, subject to the constraint

$$\mathbf{w}^H\mathbf{w} = a, \tag{A.415}$$

where a is real.

The cost function is

$$g(\mathbf{w}, \mathbf{w}^H) = \mathbf{w}^H\mathbf{R}\mathbf{w} + \lambda(\mathbf{w}^H\mathbf{w} - a), \tag{A.416}$$

where λ is a real Lagrange multiplier. Then,

$$\nabla_{\mathbf{w}^H} g(\mathbf{w}, \mathbf{w}^H) = \mathbf{R}\mathbf{w}_o + \lambda\mathbf{w}_o = \mathbf{0}, \tag{A.417}$$

or

$$\mathbf{R}\mathbf{w}_o = -\lambda\mathbf{w}_o. \tag{A.418}$$

The result in (A.418) is familiar as the eigenvector equation (A.188). Thus, $\mathbf{w}_o$ is the eigenvector of $\mathbf{R}$ corresponding to the largest eigenvalue.

We use the complex gradient in a number of derivations in the text.

Bibliography

[AS65] M. Athans and F. C. Schweppe. Gradient matrices and matrix calculations. Technical Report, MIT Lincoln Laboratory, Lexington, Massachusetts, November 1965.

[Bel72] R. Bellmann. *Introduction to Matrix Analysis.* McGraw-Hill, New York, 2nd edition, 1972.

[Bra83] D. H. Brandwood. A complex gradient operator and its application in adaptive array theory. *Proc. IEE,* Special issue on adaptive arrays, vol.130, P. F, pp. 11–17, February 1983.

[CB76a] A. Cantoni and P. Butler. Eigenvalues and eigenvectors of symmetric centrosymmetric matrices. *Linear Algebra Appl.,* vol.13 pp. 275–288, 1976.

[CB76b] A. Cantoni and P. Butler. Properties of the eigenvectors of persymmetric matrices with applications to communication theory. *IEEE Trans. Commun.*, vol. COM-24 pp. 804–809, August 1976.

[Giv58] W. Givens. Computation of plane unitary rotations transforming a general matrix to triangular form. *J. Soc. Ind. Appl. Math.*, 6:26–50, 1958.

[Gra81] A. Graham. *Kronecker Products and Matrix Calculus: With Applications.* Wiley, New York, 1981.

[GS58] O. Grenander and G. Szego. *Toeplitz Forms and Their Applications.* University of California Press, Berkeley, 1958.

[GVL89] G. H. Golub and C. F. Van Loan. *Matrix Computations.* The Johns Hopkins University Press, Baltimore, Maryland, 1989.

[Hay91] S. Haykin. *Adaptive Filter Theory* Prentice-Hall, Englewood Cliffs, New Jersey, 2nd edition 1991.

[Hay96] S. Haykin. *Adaptive Filter Theory.* Prentice-Hall, Upper Saddle River, New Jersey, 3rd edition, 1996.

[Hoh73] F. E. Hohn. *Elementary Matrix Algebra.* Macmillan, New York, 1973.

[Hou58] A. S. Householder. Unitary triangularization of a non-symmetric matrix. *J. Assoc. Comput. Math.*, vol. 5 pp. 204–243, 1958.

[Hou64] A. S. Householder. *The Theory of Matrices in Numerical Analysis.* Blaisdell, Waltham, Massachusetts, 1964.

[HS92] S. Haykin and A. Steinhardt, editors. *Adaptive Radar Detection and Estimation.* Wiley, New York, 1992.

[Hud81] J.E. Hudson. *Adaptive Array Principles.* Peter Peregrinus, New York and London, 1981.

[Kay88] S. M. Kay. *Modern Spectral Estimation: Theory and Application.* Prentice-Hall, Englewood Cliffs, New Jersey, 1988.

[Kur89] A. R. Kuruc. Lower bounds on multiple-source direction finding in the presence of direction-dependent antenna-array-calibration errors. Technical Report, MIT Lincoln Laboratory, Lexington, Massachusetts, 1989.

[LH74] C. L. Lawson and R. J. Hanson. *Solving Least Squares Problems.* Prentice-Hall, Englewood Cliffs, New Jersey, 1974.

[LT85] P. Lancaster and M. Tismenetsky. *The Theory of Matrices with Applications.* Academic Press, San Diego, California 1985.

[Mak81] J. Makhoul. On the eigenvectors of symmetric Toeplitz matrices. *IEEE Trans. on Acoust., Speech, Signal Process.*, ASSP-29 pp. 868–872, August 1981.

[Mar60] M. Marcus. Basic theorems in matrix theory. Technical Report Applied Math. Ser. 57, National Bureau of Standards, Washington D.C., January 1960.

[Mar87] S. L. Marple, Jr. *Digital Spectral Analysis.* Prentice-Hall, Englewood Cliffs, New Jersey, 1987.

[MP73] J. H. McClellan and T. W. Parks. A unified approach to the design of optimum FIR linear phase digital filters. *IEEE Trans. Circuit Theory*, vol. CT-20 pp. 697–701, 1973.

[Moo20] E. H. Moore. *Bull. Am. Math. Soc.*, vol. 26, pp. 394–395, 1920

[ND77] B. Noble and J. W. Daniel. *Applied Linear Algebra.* Prentice-Hall, Englewood Cliffs, New Jersey, 2nd edition, 1977.

[Pen55] R. Penrose. *Proc. Cambridge Philos. Soc.*, vol. 1, pp. 406–413, 1955.

[PRLN92] J. G. Proakis, C. M. Rader, F. Ling, and C. L. Nikias. *Advanced Digital Signal Processing.* Macmillan, New York, 1992.

[RM71] C. R. Rao and S. K. Mitra. *Generalized Inverse of Matrices and Its Applications.* Wiley, New York, 1971.

[Sch91] L. L. Scharf. *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis.* Addison-Wesley, Reading, Massachusetts, 1991.

[SN89] P. Stoica and A. Nehorai. MUSIC, maximum likelihood, and Cramér-Rao bound. *IEEE Trans. Acoust., Speech, Signal Process.*, vol.ASSP-37, pp. 720–741, May 1989.

[Ste00] C. Stephanos. Pures applications. *J. Math.*, vol. 6 pp. 73–128, 1900.

[Ste73] G. W. Stewart. *Introduction to Matrix Computations.* Academic Press, New York, 1973.

[Ste88] A. O. Steinhardt. Householder transforms in signal processing. *IEEE Acoust. Speech, Signal Process. Mag*, pp. 4–12, July 1988.

[VT68] H. L. Van Trees. *Detection, Estimation, and Modulation Theory, Part I.* Wiley, New York, 1968.

[VT71] H. L. Van Trees. *Detection, Estimation, and Modulation Theory, Part III.* Wiley, New York, 1971.

[VT01a] H. L. Van Trees. *Detection, Estimation, and Modulation Theory, Part I.* Wiley Interscience, New York, 2001.

[VT01b] H. L. Van Trees. *Detection, Estimation, and Modulation Theory, Part III.* Wiley Interscience, New York, 2001.

[Wil65] J. H. Wilkinson. *The Algebraic Eigenvalue Problem.* Oxford University Press, New York, 1965.