1

MATHEMATICAL PRELIMINARIES

1.1 INTRODUCTION

Free space optical systems engineering addresses how light energy is created, manipulated, transferred, changed, processed, or any combination of these entities, for use in atmospheric and space remote sensing and communications applications. To understand the material in this book, some basic mathematical concepts and relationships are needed. Because our audience is envisioned to be junior and senior undergraduates, it is not possible to require the readers to have had exposure to these topics at this stage of their education. Normally, they will have a working knowledge of algebra, geometry, and differential and integral calculus by now, but little else.

In this chapter, we provide a concise, but informative, summary of the additional mathematical concepts and relationships needed to perform optical systems engineering. The vision is to form a strong foundation for understanding what follows in subsequent chapters.

The intent is to establish the lexicon and mathematical basis of the various topics, and what they really mean in very simple, straightforward means, establishing that envisioned foundation.

1.2 LINEAR ALGEBRA

Imaging sensors create pictures stored as two-dimensional, discrete element arrays, that is, matrices. It is often convenient to convert these image arrays into a vector form

Free Space Optical Systems Engineering: Design and Analysis, First Edition. Larry B. Stotts.

Companion website: www.wiley.com\go\stotts\free_space_optical_systems_engineering

by column (or row) scanning the matrix, and then stringing the elements together in a long vector; that is, a lexicographic form [1]. This is called "vectorization." This means that the optical engineers need to be adept in linear algebra to take on problems in optical signal processing and target detection. This section reviews the notational conventions and basics of linear algebra following Bar-Shalom and Fortmann, Appendix A [2]. A more extensive review can be found in books on linear algebra.

1.2.1 Matrices and Vectors

A *matrix*, $\boldsymbol{A}$, is a two-dimensional array that can be mathematically written as

$$\boldsymbol{A} = [a_{ij}] = \begin{bmatrix} a_{11} & \cdots & a_{1J} \\ \vdots & \ddots & \vdots \\ a_{I1} & \cdots & a_{IJ} \end{bmatrix}. \tag{1.1}$$

The first index in the matrix element indicates the row number and the second index, the column number. The dimensions of the matrix are $I \times J$. The *transpose* of the above matrix is written as

$$\boldsymbol{A}^T = \begin{bmatrix} a_{11} & \cdots & a_{1I} \\ \vdots & \ddots & \vdots \\ a_{J1} & \cdots & a_{JI} \end{bmatrix}. \tag{1.2}$$

A *square matrix* is said to be symmetric, which means that

$$\boldsymbol{A} = \boldsymbol{A}^T, \tag{1.3}$$

which means that $a_{ij} = a_{ji} \ \forall i, j$.

A *vector* is a one-dimensional matrix array, which is written as

$$\boldsymbol{a} = \text{column}(a_i) = \begin{bmatrix} a_1 \\ \vdots \\ a_I \end{bmatrix}. \tag{1.4}$$

The column vector has dimension I in this case. By convention, we assume all vectors are column vectors. The transpose of a column vector is a row vector and the transpose of Eq. (1.4) can be written as

$$\boldsymbol{a}^T = \text{row}(a_i) = \begin{bmatrix} a_1 & \cdots & a_I \end{bmatrix}. \tag{1.5}$$

Comparing Eqs. (1.4) and (1.5), it is clear that

$$\boldsymbol{a} = \begin{bmatrix} a_1 & \cdots & a_I \end{bmatrix}^T. \tag{1.6}$$

1.2.2 Linear Operations

The *addition* of matrices and *multiplication* of a matrix by a scalar are given by the following equation:

$$\boldsymbol{C} = r\boldsymbol{A} + s\boldsymbol{B}, \tag{1.7}$$

where

$$c_{ij} = ra_{ij} + sb_{ij} \tag{1.8}$$

for $\{i = 1, \ldots, I; j = 1, \ldots, J\}$. Obviously, all three matrices have the same dimensions.

The *product* of two matrices is written in general as

$$\boldsymbol{C} = \boldsymbol{AB}, \tag{1.9}$$

where

$$c_{ip} = \sum_{j=1}^{J} a_{ij} b_{jp} \tag{1.10}$$

for $\{i = 1, \ldots, I; p = 1, \ldots, P\}$. Here, $\boldsymbol{A}$ is a $I \times J$ matrix, $\boldsymbol{B}$ is a $J \times P$ matrix, and $\boldsymbol{C}$ is a $I \times P$ matrix. In general, matrix products are not commutative, that is, $\boldsymbol{AB} \neq \boldsymbol{BA}$.

The transpose of a product is

$$\boldsymbol{C}^T = (\boldsymbol{AB})^T = \boldsymbol{B}^T \boldsymbol{A}^T. \tag{1.11}$$

Equation (1.10) implies that if the matrix–vector product is written as

$$\boldsymbol{Ab} = \boldsymbol{c}, \tag{1.12}$$

where $\boldsymbol{A}$ is a $I \times J$ matrix, $\boldsymbol{b}$ is a $J \times 1$ vector, and $\boldsymbol{c}$ is a $I \times 1$ vector, then its transpose is equal to

$$\boldsymbol{c}^T = \boldsymbol{b}^T \boldsymbol{A} \tag{1.13}$$

with $\boldsymbol{c}^T$ being a $1 \times I$ (row) vector, $\boldsymbol{b}^T$ being a $1 \times J$ vector, and $\boldsymbol{A}$ still being a $I \times J$ matrix.

1.2.3 Traces, Determinants, and Inverses

The *trace* of a $I \times I$ matrix $\boldsymbol{A}$ is defined as

$$\text{tr}(\boldsymbol{A}) = \sum_{i=1}^{I} a_{ii} = \text{tr}(\boldsymbol{A}^T), \tag{1.14}$$

which implies that

$$\text{tr}(\boldsymbol{AB}) = \text{tr}(\boldsymbol{BA}). \tag{1.15}$$

The *determinant* of a $I \times I$ matrix $\boldsymbol{A}$ is defined as

$$|\boldsymbol{A}| = a_{11}c_{11} + a_{12}c_{12} + \cdots + a_{1I}c_{1I} = |\boldsymbol{A}^T|, \tag{1.16}$$

where

$$c_{ij} = (-1)^{i+j} |\boldsymbol{A}^{ij}|; \quad i, j = 1, \ldots, I. \tag{1.17}$$

The parameter set $\{c_{ij}\}$ are called the *cofactors* of $\boldsymbol{A}$ and $\boldsymbol{A}^{ij}$ is the $(I-1)\times(I-1)$ matrix formed by deleting the ith row and jth column from $\boldsymbol{A}$. The determinant of a scalar is defined as the scalar itself since essentially is a 1×1 matrix. This implies that the determinant of a matrix multiplied by a scalar is given by

$$|r\boldsymbol{A}| = r^I|\boldsymbol{A}|, \tag{1.18}$$

and the determinant of a product of two matrices is written as

$$|\boldsymbol{AB}| = |\boldsymbol{BA}|. \tag{1.19}$$

Example 1.1

(a) The determinant of a 2×2 matrix is given by

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{21}a_{12}. \tag{1.20}$$

(b) The determinant of a 3×3 matrix is given by

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11}\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12}\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13}\begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} \tag{1.21}$$

$$= a_{11}(a_{22}a_{33} - a_{23}a_{32}) - a_{12}(a_{21}a_{33} - a_{23}a_{31}) + a_{13}(a_{21}a_{32} - a_{22}a_{31}). \tag{1.22}$$

Example 1.2 Let us now look at the solution for Eq. (1.12), where matrix $\boldsymbol{A}$ is a 3×3 matrix. Multiplying Eq. (1.12) out, we have three simultaneous equations:

$$\begin{aligned} a_{11}x + a_{12}y + a_{13}z &= u, \\ a_{21}x + a_{22}y + a_{23}z &= v, \\ a_{31}x + a_{32}y + a_{33}z &= w. \end{aligned} \tag{1.23}$$

The solutions to these equations are:

$$x = \frac{\begin{vmatrix} u & a_{12} & a_{13} \\ v & a_{22} & a_{23} \\ w & a_{32} & a_{33} \end{vmatrix}}{\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}}, \tag{1.24}$$

$$y = \frac{\begin{vmatrix} a_{11} & u & a_{13} \\ a_{21} & v & a_{23} \\ a_{31} & w & a_{33} \end{vmatrix}}{\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}}, \tag{1.25}$$

and

$$z = \frac{\begin{vmatrix} a_{11} & a_{12} & u \\ a_{21} & a_{22} & v \\ a_{31} & a_{32} & w \end{vmatrix}}{\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}}, \tag{1.26}$$

assuming the determinant of matrix $\boldsymbol{A}$ is not zero.

The *inverse* $\boldsymbol{A}^{-1}$ of a $I \times I$ matrix $\boldsymbol{A}$ (if it exists) can be expressed as

$$\boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{A}\boldsymbol{A}^{-1} = \begin{bmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{bmatrix} = \boldsymbol{I}. \tag{1.27}$$

In Eq. (1.20), the $I \times I$ matrix $\boldsymbol{I}$ is called the *identity matrix*, which has 1's down the diagonal and 0's everywhere else. The inverse is given by the equation

$$\boldsymbol{A}^{-1} = \frac{1}{|\boldsymbol{A}|}\boldsymbol{C}^T, \tag{1.28}$$

where $\boldsymbol{C}$ are the cofactors of $\boldsymbol{A}$. The matrix $\boldsymbol{C}^T$ is called the *adjugate* of $\boldsymbol{A}$. A matrix is considered *invertible* or *nonsingular* if and only if its determinant is nonzero; otherwise, it is said to be singular. Let us discuss these points a little more.

The inverse of a matrix exists if and only if the columns of the matrix (or its rows) are linearly independent. This means that

$$\sum_{i=1}^{m} r_i \boldsymbol{a}_i = \boldsymbol{0} \rightarrow r_i = 0 \quad \text{for} \quad i = 1, \ldots, m, \tag{1.29}$$

where $\boldsymbol{0}$ is the zero vector.

A general $m \times m$ matrix can be inverted using methods such as the Cayley–Hamilton (CH) method, Gauss–Jordan elimination, Gaussian elimination, or LU decomposition.

Example 1.3 The cofactor equation given in Eq. (1.17) gives the following expression for the inverse of a 2×2 matrix:

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}^{-1} = \frac{1}{\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}} \begin{bmatrix} a_{22} & -a_{21} \\ -a_{12} & a_{11} \end{bmatrix}. \tag{1.30}$$

The CH method gives the solution

$$\boldsymbol{A}^{-1} = \frac{1}{|\boldsymbol{A}|}[(\mathrm{tr}(\boldsymbol{A})\,\boldsymbol{I} - \boldsymbol{A})]. \tag{1.31}$$

Example 1.4 The inverse of a 3×3 matrix is given by

$$\boldsymbol{A}^{-1} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}^{-1} = \frac{1}{|\boldsymbol{A}|}\begin{bmatrix} d_{11} & d_{12} & d_{13} \\ d_{21} & d_{22} & d_{23} \\ d_{31} & d_{32} & d_{33} \end{bmatrix}^{T} = \frac{1}{|\boldsymbol{A}|}\begin{bmatrix} d_{11} & d_{21} & d_{31} \\ d_{12} & d_{22} & d_{32} \\ d_{13} & d_{23} & d_{33} \end{bmatrix}, \tag{1.32}$$

where

$$\begin{array}{lll} d_{11} = a_{22}a_{33} - a_{23}a_{32} & d_{12} = -(a_{21}a_{33} - a_{23}a_{31}) & d_{13} = (a_{21}a_{32} - a_{22}a_{31}) \\ d_{21} = -(a_{12}a_{33} - a_{13}a_{32}) & d_{22} = -(a_{11}a_{33} - a_{13}a_{31}) & d_{23} = -(a_{11}a_{32} - a_{12}a_{31}). \\ d_{31} = a_{12}a_{23} - a_{13}a_{22} & d_{32} = -(a_{11}a_{23} - a_{13}a_{21}) & d_{33} = (a_{21}a_{22} - a_{12}a_{21}) \end{array}$$

The CH method gives the solution

$$\boldsymbol{A}^{-1} = \frac{1}{|\boldsymbol{A}|}\left[\frac{1}{2}[(\mathrm{tr}(\boldsymbol{A}))^2 - \mathrm{tr}(\boldsymbol{A}^2)]\,\boldsymbol{I} - \boldsymbol{A}\,\mathrm{tr}(\boldsymbol{A}) + \boldsymbol{A}^2\right]. \tag{1.33}$$

Example 1.5 With increasing dimensions, expressions for $\boldsymbol{A}^{-1}$ becomes complicated. However, for $m = 4$, the CH method yields

$$\boldsymbol{A}^{-1} = \frac{1}{|\boldsymbol{A}|}\begin{bmatrix} \frac{1}{6}[(\mathrm{tr}(\boldsymbol{A}))^3 - 3\mathrm{tr}(\boldsymbol{A})\mathrm{tr}(\boldsymbol{A}^2) + \mathrm{tr}(\boldsymbol{A}^3)] \\ \boldsymbol{I} - \frac{1}{2}\boldsymbol{A}[(\mathrm{tr}(\boldsymbol{A}))^2 - \mathrm{tr}(\boldsymbol{A}^2)] - \boldsymbol{A}^2\,\mathrm{tr}(\boldsymbol{A}) - \boldsymbol{A}^3 \end{bmatrix}. \tag{1.34}$$

Example 1.6 The inverse of a (nonsingular) partitioned $I \times I$ matrix also can be shown to be given by

$$\begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{bmatrix}^{-1} = \begin{bmatrix} \boldsymbol{E} & \boldsymbol{F} \\ \boldsymbol{G} & \boldsymbol{M} \end{bmatrix}, \tag{1.35}$$

where $\boldsymbol{A}$ is a $I_1 \times I_1$ matrix, $\boldsymbol{B}$ is a $I_1 \times I_2$ matrix, $\boldsymbol{C}$ is a $I_2 \times I_1$ matrix, $\boldsymbol{D}$ is a $I_2 \times I_2$ matrix, and $I_1 + I_2 = I$. In the above,

$$\boldsymbol{E} = \boldsymbol{A}^{-1} + \boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{M}\boldsymbol{C}\boldsymbol{A}^{-1} = (\boldsymbol{A} - \boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{C})^{-1} \tag{1.36}$$

$$\boldsymbol{F} = \boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{M} = -\boldsymbol{E}\boldsymbol{B}\boldsymbol{D}^{-1} \tag{1.37}$$

$$\boldsymbol{G} = -\boldsymbol{M}\boldsymbol{C}\boldsymbol{A}^{-1} = -\boldsymbol{D}^{-1}\boldsymbol{C}\boldsymbol{E} \tag{1.38}$$

and

$$\boldsymbol{M} = (\boldsymbol{D} - \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B})^{-1} = \boldsymbol{D}^{-1} + \boldsymbol{D}^{-1}\boldsymbol{C}\boldsymbol{E}\boldsymbol{B}\boldsymbol{D}^{-1}. \tag{1.39}$$

If $\boldsymbol{R} = -\boldsymbol{A}$, $\boldsymbol{P} = \boldsymbol{D}^{-1}$, and $\boldsymbol{H} = \boldsymbol{B} = \boldsymbol{C}^T$, then the following matrix equation

$$(\boldsymbol{P}^{-1} + \boldsymbol{H}^T\boldsymbol{R}^{-1}\boldsymbol{H})^T = \boldsymbol{P} - \boldsymbol{P}\boldsymbol{H}^T(\boldsymbol{H}\boldsymbol{P}\boldsymbol{H}^T + \boldsymbol{R})^{-1}\boldsymbol{H}\boldsymbol{P} \tag{1.40}$$

can be rewritten as

$$(\boldsymbol{R} + \boldsymbol{H}^T\boldsymbol{P}\boldsymbol{H})^T = \boldsymbol{R}^{-1} - \boldsymbol{R}^{-1}\boldsymbol{H}(\boldsymbol{P}^{-1} + \boldsymbol{H}^T\boldsymbol{R}^{-1}\boldsymbol{H} + \boldsymbol{R})^{-1}\boldsymbol{H}^T\boldsymbol{R}^{-1}. \tag{1.41}$$

The above is known as *matrix inversion formula*.

It is easy to show that

$$(\boldsymbol{A}\boldsymbol{B})^{-1} = \boldsymbol{B}^{-1}\boldsymbol{A}^{-1}. \tag{1.42}$$

1.2.4 Inner Products, Norms, and Orthogonality

The inner product of two arbitrary vectors of the same dimension is given by

$$\boldsymbol{a}^T\boldsymbol{b} = \sum_{i=1}^{I} a_i b_i. \tag{1.43}$$

If $\boldsymbol{a} = \boldsymbol{b}$, then we write

$$\boldsymbol{a}^T\boldsymbol{a} = \boldsymbol{a}^2 = \sum_{i=1}^{I} a_i^2. \tag{1.44}$$

Equation (1.44) is called the squared *norm* of the vector $\boldsymbol{a}$. The *Schwartz inequality* states that

$$|\boldsymbol{a}^T\boldsymbol{b}| \leq \boldsymbol{a}\boldsymbol{b}. \tag{1.45}$$

Two vectors are defined to be orthogonal ($\boldsymbol{a} \perp \boldsymbol{b}$) if

$$\boldsymbol{a}^T\boldsymbol{b} = \boldsymbol{0}. \tag{1.46}$$

The orthogonal projection of the vector $\boldsymbol{a}$ onto $\boldsymbol{b}$ is

$$\Pi_b(\boldsymbol{a}) = \frac{\boldsymbol{a}^T\boldsymbol{b}}{\boldsymbol{b}^2}\,\boldsymbol{b}, \tag{1.47}$$

and the difference between it and the original vector $\boldsymbol{a}$ is orthogonal to $\boldsymbol{b}$. That is, we have

$$[\boldsymbol{a} - \Pi_b(\boldsymbol{a})] \perp \boldsymbol{b}. \tag{1.48}$$

Finally, the outer product of the vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ is the matrix $\boldsymbol{C}$, that is,

$$\boldsymbol{a}\boldsymbol{b}^T = \boldsymbol{C}. \tag{1.49}$$

1.2.5 Eigenvalues, Eigenvectors, and Rank

The *eigenvalues* of a square matrix are the scalars λ_i such that

$$\boldsymbol{A}\boldsymbol{x}_i = \lambda_i \boldsymbol{x}_i, \tag{1.50}$$

where the vectors $\boldsymbol{x}_i$ are the corresponding *eigenvectors*. Equation (1.50) has the following properties:

- A matrix $\boldsymbol{A}$ is nonsingular if and only if all its eigenvalues are nonzero.
- The ***rank*** of the matrix $\boldsymbol{A}$ is equal to the number of its nonzero eigenvalues. A nonsingular matrix is said to be of full rank.
- The eigenvalues of a real matrix can be either real or complex, but a symmetric matrix only has eigenvalues that are real.
- The trace of a matrix $\boldsymbol{A}$ is equal to the sum of its eigenvalues.
- The determinant of a matrix $\boldsymbol{A}$ is equal to the product of its eigenvalues.

1.2.6 Quadratic Forms and Positive Definite Matrices

The scalar equation

$$q(\boldsymbol{x}) = \boldsymbol{x}\boldsymbol{A}\boldsymbol{x} \tag{1.51}$$

is called a *quadratic form*. This relationship is true when the matrix $\boldsymbol{A}$ is symmetric. The equation of this form is called *positive definite* if

$$q(\boldsymbol{x}) = \boldsymbol{x}\boldsymbol{A}\boldsymbol{x} > \boldsymbol{0} \quad \forall\, \boldsymbol{x} \neq \boldsymbol{0}. \tag{1.52}$$

The matrix $\boldsymbol{A}$ also is called positive definite, which we denote by $\boldsymbol{A} > \boldsymbol{0}$. It goes without saying that a matrix is positive definite if and only if its eigenvalues are positive. If the inequality in Eq. (1.52) is nonnegative rather than positive, then the matrix is referred to as positive semidefinite or nonnegative definite.

The quadratic equation given in Eq. (1.51) is the squared weighted norm of the vector $\boldsymbol{x}$ where the weighting is in accordance with the matrix $\boldsymbol{A}$.

1.2.7 Gradients, Jacobians, and Hessians

The *gradient operator* is written as

$$\nabla_x = \begin{bmatrix} \dfrac{\partial}{\partial x_1} \\ \vdots \\ \dfrac{\partial}{\partial x_I} \end{bmatrix}, \tag{1.53}$$

with the properties

$$\nabla_x \boldsymbol{x}^T = \boldsymbol{I} \tag{1.54}$$

and

$$\nabla_x(\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}) = 2\boldsymbol{A}\boldsymbol{x}. \tag{1.55}$$

The gradient of a vector-valued function, say $\boldsymbol{f}(\boldsymbol{x})$, is equal to

$$\nabla_x \boldsymbol{f}^T(\boldsymbol{x}) = \begin{bmatrix} \dfrac{\partial}{\partial x_1} \\ \vdots \\ \dfrac{\partial}{\partial x_I} \end{bmatrix} \left\{ f_1(\boldsymbol{x}) \ \cdots \ f_I(\boldsymbol{x}) \right\} = \begin{bmatrix} \dfrac{\partial f_1(\boldsymbol{x})}{\partial x_1} & \cdots & \dfrac{\partial f_I(\boldsymbol{x})}{\partial x_1} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial f_1(\boldsymbol{x})}{\partial x_I} & \cdots & \dfrac{\partial f_I(\boldsymbol{x})}{\partial x_I} \end{bmatrix}. \tag{1.56}$$

The transpose of Eq. (1.56) is called the *Jacobian* and is defined as

$$\boldsymbol{f}_x(\boldsymbol{x}) = \frac{\partial \boldsymbol{f}(\boldsymbol{x})}{\partial \boldsymbol{x}} = [\nabla_x \boldsymbol{f}^T(\boldsymbol{x})]^T. \tag{1.57}$$

The Hessian of the scalar function $f(\boldsymbol{x})$ is defined to be given by

$$\boldsymbol{f}_{xx}(\boldsymbol{x}) = \frac{\partial^2 \boldsymbol{f}(\boldsymbol{x})}{\partial \boldsymbol{x}^2} = \nabla_x \, \nabla_x{}^T f(\boldsymbol{x}) = \begin{bmatrix} \dfrac{\partial^2 f_1(\boldsymbol{x})}{\partial x_1 \partial x_1} & \cdots & \dfrac{\partial^2 f_I(\boldsymbol{x})}{\partial x_1 \partial x_I} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial^2 f_1(\boldsymbol{x})}{\partial x_I \partial x_1} & \cdots & \dfrac{\partial^2 f_I(\boldsymbol{x})}{\partial x_I \partial x_I} \end{bmatrix}, \tag{1.58}$$

which is a symmetric matrix.

1.3 FOURIER SERIES

One of the main tools used in physics and engineering are Fourier series and Fourier integrals, which were published by Jean Baptiste Fourier in 1822, almost 200 years ago. This section reviews the basics of these important entities [3–5] as they are the foundation for a lot of the mathematical concepts used in optics.

1.3.1 Real Fourier Series

A *Fourier series* is used to represent a periodic function $f(x)$. The definition of a periodic function is

$$f(x) = f(x + md); \quad \text{where } |m| = 0, 1, 2, 3, \ldots \tag{1.59}$$

where d is the length of the period (Figure 1.1) and $\nu = \frac{1}{d}$ is the fundamental spatial frequency in one dimension. The following expression is the usual way for writing a Fourier series representing the periodic function $f(x)$:

$$f(x) = \sum_{n=0}^{\infty} [a_n \cos(2\pi n \nu x) + b_n \sin(2\pi n \nu x)]. \tag{1.60}$$

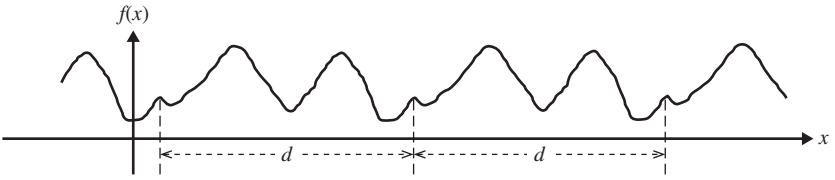


FIGURE 1.1 Example of a periodic function $f(x)$.

In short, a Fourier series is an expansion of a periodic function $f(x)$ in terms of sines and cosines. The coefficients for this series are given by:

$$a_0 = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\, dx, \tag{1.61}$$

$$a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\ \cos(nx)dx, \tag{1.62}$$

and

$$b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\ \sin(nx)dx \tag{1.63}$$

for $n = 1, 2, \ldots, \infty$. It is noted that this basis set is orthonormal, as seen by the following the integral identities:

$$\int_{-\pi}^{\pi} \sin(mx)\ \sin(nx)dx = \pi\delta_{mn}, \tag{1.64}$$

$$\int_{-\pi}^{\pi} \cos(mx)\ \cos(nx)dx = \pi\delta_{mn}, \tag{1.65}$$

$$\int_{-\pi}^{\pi} \sin(mx)\ \cos(nx)dx = 0, \tag{1.66}$$

$$\int_{-\pi}^{\pi} \sin(mx)\, dx = 0, \tag{1.67}$$

and

$$\int_{-\pi}^{\pi} \cos(mx)\, dx = 0 \tag{1.68}$$

for $m, n \neq 0$. In the first two equations, δ_{mn} denotes the Kronecker delta function.

1.3.2 Complex Fourier Series

The natural extension of the above is to express a Fourier series in terms of complex coefficients. Consider a real-valued function $f(x)$. In this case, we have

$$f(x) = \sum_{n=-\infty}^{\infty} C_n\, e^{+inx}, \tag{1.69}$$

where

$$C_n = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)\, e^{-inx} dx. \tag{1.70}$$

The above coefficients can be expressed in terms of those in the Fourier series given in Eq. (1.60), namely,

$$C_n = \begin{cases} \dfrac{1}{2\pi}\displaystyle\int_{-\pi}^{\pi} f(x)[\cos(nx) + i\, \sin(nx)]dx & \text{for } n < 0 \\ \dfrac{1}{2\pi}\displaystyle\int_{-\pi}^{\pi} f(x)\, dx & \text{for } n < 0 \\ \dfrac{1}{2\pi}\displaystyle\int_{-\pi}^{\pi} f(x)\, [\cos(nx) - i\, \sin(nx)]dx & \text{for } n > 0 \end{cases} \tag{1.71}$$

$$= \begin{cases} \dfrac{1}{2\pi}(a_n + i b_n) & \text{for } n < 0 \\ \dfrac{1}{2\pi}\, a_0 & \text{for } n = 0\ . \\ \dfrac{1}{2\pi}(a_n - i b_n) & \text{for } n < 0 \end{cases} \tag{1.72}$$

For a function that is periodic in $\left\{-\frac{d}{2}, \frac{d}{2}\right\}$, this complex series become

$$f(x) = \sum_{n=-\infty}^{\infty} C_n e^{+2\pi i n \nu x}, \tag{1.73}$$

where

$$C_n = \frac{1}{d}\int_{-\frac{d}{2}}^{\frac{d}{2}} f(x) e^{-2\pi i n \nu x} dx \tag{1.74}$$

and $\nu = \frac{1}{d}$. These equations are the basis for the extremely important Fourier transform, which is obtained by transforming C_n from a discrete variable to a continuous one as the length $d \to \infty$. We will discuss Fourier Transforms shortly.

1.3.3 Effects of Finite Fourier Series Use

Sometimes, it is necessary not to have an infinite series, but a finite series to represent the periodic function $f(x)$. In this case, we write

$$f(x) = \sum_{n=0}^{N} [a_n \cos(2\pi n \nu x) + b_n \sin(2\pi n \nu x)] = S_N(x). \tag{1.75}$$

What is the best choice of coefficients, $\{a_n\}$ and $\{b_n\}$? Let us define an error function

$$E_N(x) = f(x) - S_N(x). \tag{1.76}$$

Rewriting Eq. (1.75), we find that

$$S_N(x) = \sum_{n=0}^{\infty} [a_n \cos(2\pi n \nu x) + b_n \sin(2\pi n \nu x)]$$
$$= a_0 + \sum_{n=1}^{\infty} \left[\frac{(a_n - i b_n)}{2}\right] e^{+2\pi n \nu x} + \sum_{n=1}^{\infty} \left[\frac{(a_n + i b_n)}{2}\right] e^{-2\pi n \nu x}. \quad (1.77)$$

Manipulating the terms in Eq. (1.77), we find that

$$S_N(x) = \sum_{n=-N}^{N} C_n e^{+2\pi i n \nu x}; \quad C_n = f(x) \cdot e^{-2\pi i n \nu x}. \quad (1.78)$$

Now let us return to the questions: "How good is our approximation?" or "Does $\sigma_N^2 \to 0$ as $N \to \infty$?" Using our complex notation, the error σ_N^2 can be rewritten as

$$\sigma_N^2 = \left(f - \sum_{n=-N}^{N} C_n e^{+2\pi i n \nu x}\right) \cdot \left(f^* - \sum_{n=-N}^{N} C_n^* e^{-2\pi i n \nu x}\right) \quad (1.79)$$
$$= f \cdot f^* - \sum_{m=-N}^{N} |C_m|^2. \quad (1.80)$$

This equation implies the following desirable feature, $\sigma_{N+1}^2 - \sigma_N^2 = -2|C_{N+1}|^2 \leq 0$. This means that the error σ_N^2 can only decrease if more terms are added to the Fourier series as in $S_N \to S_{N+1}$. If $\sigma_N^2 \to 0$, then

$$\sum_{m=-N}^{N} |C_m|^2 \overset{\text{yields}}{\to} f \cdot f^*.$$

The ultimate result, which holds for all healthy functions (and also for some strange ones too), is the *completeness relationship*,

$$f \cdot f^* = \sum_{m=-\infty}^{\infty} |C_m|^2. \quad (1.81)$$

However, this relationship is not completely satisfactory because it always cannot easily be checked and $f(x)$ can be quite different from $\sum_{n=-\infty}^{\infty} C_n e^{+2\pi n \nu x}$ for a finite number of points.

Example 1.7 Let

$$f(x) = \begin{cases} 1 & \text{for } 0 \leq x \leq \frac{d}{2} \\ -1 & \text{for } -\frac{d}{2} < x < 0 \end{cases}, \quad (1.82)$$

which implies that

$$C_n = \begin{cases} 0 & \text{if } n \text{ is even} \\ \frac{2}{\pi i n} & \text{if } n \text{ is odd} \end{cases}. \tag{1.83}$$

Equation (1.82) is depicted in Figure 1.2a.

From Eq. (1.83), we see that the Fourier series for the function in Eq. (1.70) can be written as

$$\sum_{n=-\infty}^{\infty} C_n e^{+2\pi i n \nu x} = \sum_{n=1,3,5,\ldots} \frac{4}{\pi n} \sin(2\pi n \nu x) = S_{\infty(x)}. \tag{1.84}$$

It is clear that the series in Eq. (1.84) certainly is zero at $x = 0$, while $f(x)$ is $+1$ for $x = 0$ (Figure 1.2b).

For functions with a finite number of discontinuities, Dirichlet has shown that at discontinuous point of $f(x)$ the infinite Fourier series assumes the arithmetic means of the right- and left-hand limits (Figure 1.2):

$$\sum_{m=-\infty}^{\infty} C_n e^{+2\pi i n \nu x} = \frac{1}{2} \lim_{\mathcal{E} \to 0} [f(x + \mathcal{E}) + f(x - \mathcal{E})]. \tag{1.85}$$

Figure 1.2b shows how a square wave is perfectly represented by the infinite series, except at points of discontinuity, where the series assumes the mean value.

Another effect occurs when one tries to approximate a discontinuous function $f(x)$ by the *finite* series Fourier series $S_N(x)$. Figure 1.3 shows a rectangular function $f(x)$ and examples of the series, $S_N(x)$, for $N = 9$, 15, and 25, for the interval $0 \le x \le \frac{d}{2}$. These graphs clearly show there is overshoot and ringing occurring in the S_N curves for $N > 1$. This is the so-called Gibbs effect. Looking closely at these three plots, we see that as you add more sinusoids to the series, the width of the overshoot decreases, but the amplitude of the overshoot peak remains about the same, $\frac{4}{\pi}$. These characteristics continue as $N \to \infty$. What is most interesting is that the overshoot is still present with an infinite number of sinusoids, but it has zero width; hence, no energy. Exactly

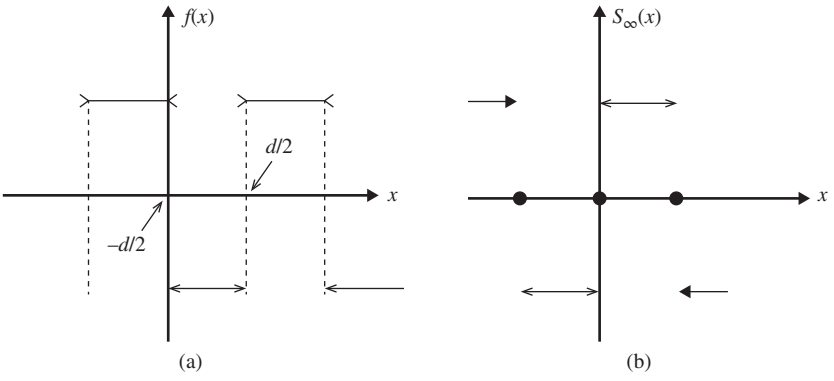


FIGURE 1.2 Dirichlet examples of (a) $f(x)$ and (b) $S_\infty(x)$.

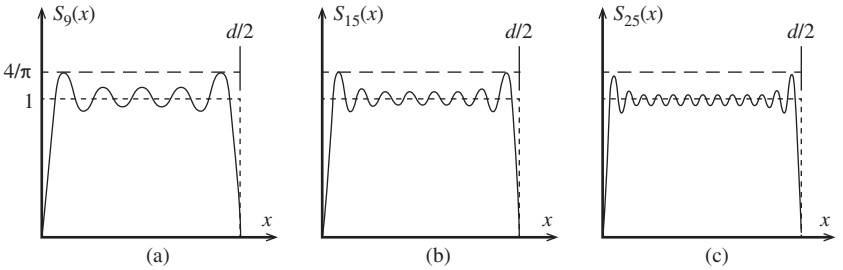


FIGURE 1.3 Plots of (a) $S_9(x)$, (b) $S_{15}(x)$, and (c) $S_{25}(x)$ as a function of x.

at the discontinuity the value of the reconstructed signal converges to the midpoint of the step. The result is the rectangle of amplitude 1. In other words, the summation converges to the signal in the sense that the error between the two has zero energy (Gibbs).

1.3.4 Some Useful Properties of Fourier Series

This section provides some of the interesting properties for Fourier series. If the function $f(x)$ has the Fourier coefficients $\{A_n\}$, $g(x)$ have the coefficients $\{B_n\}$ and both $f(x)$ and $g(x)$ are periodic in $|x| \leq \frac{d}{2}$, then the following properties in both Fourier domains are equivalent:

$$[f(x) + g(x)] \leftrightarrow [A_n + B_n] \tag{1.86}$$

$$f(x) = a\,g(x) \leftrightarrow A_n = aB_n \tag{1.87}$$

$$f(x) = g(Mx) \leftrightarrow A_{Mn} = B_n \ (M \text{ is a fixed integer} > 0). \tag{1.88}$$

The following two properties, which are quite important, are called the "shift theorem":

$$f(x) = g(x + c) \leftrightarrow A_n = B_n e^{+2\pi i n \nu c} \tag{1.89}$$

$$f(x) = g(x)e^{+2\pi i M \nu c} (M, \text{ an integer}) \leftrightarrow A_n = B_{n-M}. \tag{1.90}$$

The next two properties are sometimes called "reality symmetry" or Hermitian:

$$f(x) = g^*(x) \leftrightarrow A_n = B_n^* \tag{1.91}$$

$$f(x) = f^*(x) \leftrightarrow A_n = A_n^* \tag{1.92}$$

$$f(x) = -g^*(x) \leftrightarrow A_n = -B_n^* \tag{1.93}$$

$$f(x) = g_1(x)\,g_2(x) \leftrightarrow A_n = \sum_m B_m^{(1)} B_{n-m}^{(2)} \tag{1.94}$$

$$f(x) = g(x)g^*(x) \leftrightarrow A_n = \sum_m B_m B_{m-n}^* \tag{1.95}$$

$$f(x) = \frac{dg(x)}{dx} \leftrightarrow A_n = (2\pi i n \nu) B_n \tag{1.96}$$

$$f(x) = \int_{-x}^{x} g(x')dx' \leftrightarrow A_n = \frac{(B_n + B_{-n})}{(2\pi i n \nu)} \quad [B_0 = 0 \text{ assumed}] \tag{1.97}$$

$$\underbrace{f(x) = \left(\frac{1}{d}\right) \int_{-d/2}^{d/2} g_1(x') g_2(x' - x) dx' \leftrightarrow A_n = B_n^{(1)} B_{-n}^{(2)}}_{\text{[Cross-Correlation]}} \tag{1.98}$$

$$\underbrace{f(x) = \left(\frac{1}{d}\right) \int_{-d/2}^{d/2} g_1(x') g_2(x - x') dx' \leftrightarrow A_n = B_n^{(1)} B_n^{(2)}}_{\text{[Convolution]}} \tag{1.99}$$

$$\underbrace{f(x) = \left(\frac{1}{d}\right) \int_{-d/2}^{d/2} g(x') g^*(x - x') dx' \leftrightarrow A_n = |B_n|^2}_{\text{[Autocorrelation]}} \tag{1.100}$$

$$\underbrace{f(x) = \left(\frac{1}{d}\right) \int_{-d/2}^{d/2} g(x') g^*(x - x') dx' \leftrightarrow A_n = {B_n}^2.}_{\text{[Autoconvolution]}} \tag{1.101}$$

1.4 FOURIER TRANSFORMS

As noted earlier, the Fourier transform is a generalization of the complex Fourier series in the limit as $d \to \infty$. We first replace the discrete C_n with the continuous $\widehat{f}(\nu)d\nu$ while letting $\frac{n}{d} \to \nu$, and then change the sum to an integral. The result is that Fourier series representation for $f(x)$ becomes

$$f(x) = \int_{-\infty}^{\infty} \widehat{f}(\nu) e^{2\pi i \nu x} dx. \tag{1.102}$$

Its counterpart, the inverse Fourier transform, is given by

$$\widehat{f}(k) = \int_{-\infty}^{\infty} f(x) e^{-2\pi i \nu x} dx. \tag{1.103}$$

1.4.1 Some General Properties

Since we generated the Fourier-integral transform as an extension of the Fourier-series transform for periodic functions, we can expect some of the same properties. In particular the Fourier-integral representation is optimized in the Gaussian sense:

$$\int \left| f(x) - \int_{-\infty}^{\infty} \widehat{f}(\nu) e^{2\pi i \nu x} dx \right|^2 dx \to \text{Minimum.} \tag{1.104}$$

Hence, it is plausible that the Fourier integral also exhibits the Dirichlet effect. Here are some other properties:

$$f(x) + g(x) \leftrightarrow \widehat{f}(\nu) + \widehat{g}(\nu) \tag{1.105}$$

$$f(x) = ag(x) \leftrightarrow \widehat{f}(\nu) = a\widehat{g}(\nu) \tag{1.106}$$

$$f(x) = g(mx) \leftrightarrow \widehat{f}(\nu) = \frac{\widehat{g}\left(\frac{\nu}{m}\right)}{|m|} \tag{1.107}$$

$$f(x) = g(x + c) \leftrightarrow \widehat{f}(\nu) = \widehat{g}(\nu)e^{2\pi i \nu c} \quad \text{[Shift Theorem]} \tag{1.108}$$

$$f(x) = g(x)e^{2\pi i \nu_0 c} \leftrightarrow \widehat{f}(\nu) = \widehat{g}(\nu - \nu_0) \tag{1.109}$$

$$f(x) = g(-x) \leftrightarrow \widehat{f}(\nu) = \widehat{g}(-\nu) \tag{1.110}$$

$$f(x) = -g(-x) \leftrightarrow \widehat{f}(\nu) = \widehat{g}(\nu) \tag{1.111}$$

$$f(x) = f^*(x) \leftrightarrow \widehat{f}(\nu) = \widehat{f}^*(-\nu) \quad \text{[Reality Symmetry]} \tag{1.112}$$

$$f(x) = -f^*(x) \leftrightarrow \widehat{f}(\nu) = -\widehat{f}^*(-\nu) \tag{1.113}$$

$$f(x) = g_1(x)g_2(x) \leftrightarrow \widehat{f}(\nu) = \int \widehat{g}_1(\mu)\widehat{g}_2(\nu - \mu)d\mu \quad \text{[Convolution]} \tag{1.114}$$

$$\underbrace{f(x) = g(x)g^*(x)}_{\text{Real nonnegative}} \leftrightarrow \widehat{f}(\nu) = \int \widehat{g}(\mu)\widehat{g}^*(\nu - \mu)d\mu \quad \text{[Auto-correlation]} \tag{1.115}$$

$$f(x) = \frac{dg(x)}{dx} \leftrightarrow \widehat{f}(\nu) = 2\pi i \nu \widehat{g}(\nu) \tag{1.116}$$

$$f(x) = \int_{-x}^{x} g(x')dx' \leftrightarrow \widehat{f}(\nu) = \frac{\widehat{g}(\nu) + \widehat{g}(-\nu)}{2\pi i \nu} \tag{1.117}$$

$$f(x) = \int g_1(x')g_2(x' - x)dx' \leftrightarrow \widehat{f}(\nu) = \widehat{g}_1(\nu)\widehat{g}_2(-\nu) \tag{1.118}$$

$$f(x) = \int g_1(x')g_2(x - x')dx' \leftrightarrow \widehat{f}(\nu) = \widehat{g}_1(\nu)\widehat{g}_2(\nu) \tag{1.119}$$

$$f(x) = \int g(x')g^*(x' - x)dx' \leftrightarrow \widehat{f}(\nu) = |\widehat{g}(\nu)|^2 \tag{1.120}$$

$$f(x) = \int g(x')g^*(x - x')dx' \leftrightarrow \widehat{f}(\nu) = \widehat{g}^2(\nu) \tag{1.121}$$

$$f(x, y) = g(x + x_0, y + y_0) \leftrightarrow \widehat{f}(\nu, \mu) = \widehat{g}(\nu, \mu)e^{2\pi i(\nu x_0 + \mu y_0)} \tag{1.122}$$

$$f(x, y) = g(-x, -y) \leftrightarrow \widehat{f}(\nu, \mu) = (-\nu, -\mu) \tag{1.123}$$

$$\underbrace{f(x, y) = g(-x, -y)}_{\text{Inversion around the } y\text{-axis}} \leftrightarrow \widehat{f}(\nu, \mu) = \widehat{g}(-\nu, +\mu) \tag{1.124}$$

$$f(x,y) = f^*(x,y) \leftrightarrow \widehat{f}, \mu = \widehat{f}^*(-\nu,-\mu) \tag{1.125}$$

$$f(x,y) = \iint g_1(x',y')g_2(x'-x,y'-y)dx'dy' \leftrightarrow \widehat{f}(\nu,\mu) = \widehat{g}_1(\nu,\mu)\widehat{g}_2^*(\nu,\mu). \tag{1.126}$$

One of the most important properties of Fourier Transform is the following:

$$\iint f(x',y')g^*(x',y')dx'dy' = \iint \widehat{f}(\nu,\mu)\widehat{g}^*(\nu,\mu)d\nu d\mu \tag{1.127}$$

or

$$f \cdot g^* = \widehat{f} \cdot \widehat{g}^*, \tag{1.128}$$

that many other formulae can be derived from it. For example, if $g(x',y') = f(x',y')$, then we have

$$\iint |f(x',y')|^2 dx'dy' = \iint |\widehat{f}(\nu,\mu)|^2 d\nu d\,\mu. \tag{1.129}$$

Equation (1.129) comes from *Parseval's Theorem*. In addition, the *Wiener–Khinchin Formula*

$$\iint |f(x',y')|^2 e^{-2\pi i(\nu' x' + \mu' y')} dx'dy' = \iint \widehat{f}(\nu,\mu)\widehat{f}^*(\nu-\nu',\mu-\mu')d\nu d\,\mu \tag{1.130}$$

also can be created via Eq. (1.127) by setting $g(x',y') = f(x',y')\,e^{2\pi i(\nu' x' + \mu' y')}$ and $\widehat{g}^*(\nu,\mu) = \widehat{f}^*(\nu-\nu',\mu-\mu')$, using previously stated properties.

Before leaving this topic, let us review the Fourier series in polar coordinates. Mathematically, we can write the series as

$$f(r,\varphi) = \sum f_m(r)\, e^{im\varphi}, \tag{1.131}$$

where the coefficients can be written as

$$f_m(r) = \frac{1}{2\pi}\int_0^{2\pi} f(r,\varphi)\, e^{im\varphi}\, d\varphi \tag{1.132}$$

or

$$f_m(r) = \int_{-\infty}^{\infty} \widehat{f}_m(\vartheta)\, e^{2\pi i\, r\vartheta} d\vartheta. \tag{1.133}$$

This last equation implies that Eq. (1.131) can be rewritten as

$$f(r,\varphi) = \sum e^{im\varphi} \int_{-\infty}^{\infty} \widehat{f}_m(\vartheta) e^{2\pi i\, r\vartheta} d\vartheta. \tag{1.134}$$

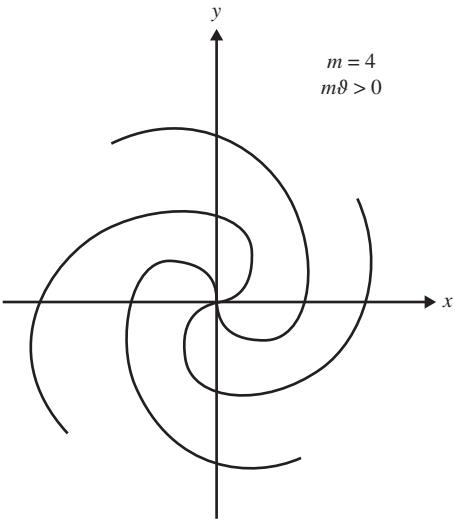


FIGURE 1.4 Plots of polar spirals.

The real part of $e^{im\varphi+2\pi ir\vartheta}$ is $\cos\{m\varphi + 2\pi r\vartheta\}$ and has maximum curves that form as set of m spirals that obey the equation

$$m\varphi + 2\pi r\vartheta = W, \quad \text{for } W = 0, \pm 1, \pm 2, \ldots \tag{1.135}$$

This suggests that Eq. (1.134) represents a superposition of spirals following this expression. The parameter $\frac{2\pi\vartheta}{m}$ specifies the extent of the spirals as illustrated in Figure 1.4.

An alternative to the above is the two-dimensional version of the Fourier transform in polar coordinates. Expressing the Cartesian coordinates in terms of polar coordinates, we have

$$\begin{aligned} x = r\cos\{\varphi\} \quad & y = r\sin\{\varphi\}, \\ \nu = \rho\cos\{\theta\} \quad & \mu = \rho\sin\{\theta\}, \\ x\nu = r\rho\cos\{\varphi\}\cos\{\theta\} \quad & y\mu = r\rho\sin\{\varphi\}\sin\{\theta\}, \end{aligned}$$

and

$$\cos\{\varphi\}\cos\{\theta\} + \sin\{\varphi\}\sin\{\theta\} = \cos\{\varphi - \theta\}. \tag{1.136}$$

This implies that

$$f(x,y) = \int\int_{-\infty}^{\infty} \widehat{f}(\mu,\nu)e^{-2\pi i(\nu x+\mu y)}\,d\nu d\,\mu \tag{1.137a}$$

$$= \int_0^{2\pi}\int_0^{\infty} \widehat{f}_{\text{pol}}(\varrho,\theta)e^{2\pi ir\varrho\cos\{\varphi-\theta\}}\varrho d\varrho d\theta = f_{\text{pol}}(r,\varphi). \tag{1.137b}$$

Comparing Eqs. (1.137a) and (1.137b), we see that the transformation to polar coordinates changes the functions $f(x, y)$ and $\widehat{f}(\mu, \nu)$ to $f_{\text{pol}}(r, \varphi)$ and $\widehat{f}_{\text{pol}}(\varrho, \theta)$, respectively, even though they represent the same spatial pattern. This is important for the reader to recognize and not assume it is just a change of variables in the cited functions.

Rewriting Eq. (1.137b) into the following form:

$$\begin{aligned} f_{\text{pol}}(r, \varphi) &= \int_0^{2\pi} \int_0^{\infty} \widehat{f}_{\text{pol}}(\varrho, \theta) e^{2\pi i r \varrho \cos\{\varphi - \theta\}} \varrho d\varrho d\theta \\ &= \int_0^{2\pi} \int_0^{\infty} \widehat{f}_{\text{pol}}(\varrho, \theta) e^{2\pi i r \varrho \sin\left\{\frac{\pi}{2} + \varphi - \theta\right\}} \varrho d\varrho d\theta, \end{aligned} \tag{1.138}$$

one can use the following Bessel Function Series

$$e^{A \sin(\alpha)} = \sum_{n=-\infty}^{\infty} J_n(A) e^{in\alpha} \tag{1.139}$$

to yield

$$f_{\text{pol}}(r, \varphi) = \sum_{n=-\infty}^{\infty} \int_0^{\infty} J_n(2\pi r \rho) \int_0^{2\pi} \widehat{f}_{\text{pol}}(\langle), \theta) e^{in\left\{\frac{\pi}{2} + \varphi - \theta\right\}} \rho d\rho d\theta. \tag{1.140}$$

Expressing $\widehat{f}_{\text{pol}}(\rho, \theta)$ as a Fourier series, we obtain

$$\begin{aligned} f_{\text{pol}}(r, \varphi) &= \sum_{n=-\infty}^{\infty} \int_0^{\infty} J_n(2\pi r \rho) \sum_{m=-\infty}^{\infty} \widehat{f}_m^{\text{pol}}(\rho) \int_0^{2\pi} e^{i(m-n)\theta + in\left\{\frac{\pi}{2} + \varphi\right\}} \rho d\rho d\theta \\ &= \sum_{n=-\infty}^{\infty} \int_0^{\infty} J_n(2\pi r \rho) \sum_{m=-\infty}^{\infty} \widehat{f}_m^{\text{pol}}(\rho) e^{in\left\{\frac{\pi}{2} + \varphi\right\}} \int_0^{2\pi} e^{i(m-n)\theta} \rho d\rho d\theta. \qquad (1.141) \\ &= \sum_{n=-\infty}^{\infty} \int_0^{\infty} J_n(2\pi r \rho) \sum_{n=-\infty}^{\infty} \widehat{f}_m^{\text{pol}}(\rho) e^{in\left\{\frac{\pi}{2} + \varphi\right\}} (2\pi \delta_{mn}) \rho d\rho. \\ &= 2\pi \sum_{n=-\infty}^{\infty} \int_0^{\infty} J_n(2\pi r \rho) \widehat{f}_n^{\text{pol}}(\rho) e^{in\left\{\frac{\pi}{2} + \varphi\right\}} \rho d\rho. \qquad (1.142) \end{aligned}$$

Example 1.8 If $f_{\text{pol}}(r, \varphi)$ does not depend on φ, then all the coefficients $\{\widehat{f}_n^{\text{pol}}(\rho)\}$ are zero, except for $\widehat{f}_0^{\text{pol}}(\rho)$. If we set $2\pi \widehat{f}_n^{\text{pol}}(\rho) = \widehat{f}_{\text{pol}}(\rho)$, we obtain the Bessel Transformations given in the following equations:

$$f_{\text{pol}}(r) = \int_0^{\infty} J_0(2\pi r \rho) \widehat{f}_{\text{pol}}(\rho) \rho d\rho \tag{1.143}$$

and

$$\widehat{f}_{\text{pol}}(\rho) = \int_0^\infty J_0(2\pi r\rho) f_{\text{pol}}(r) r dr. \tag{1.144}$$

1.5 DIRAC DELTA FUNCTION

The companion to the Kronecker Delta Function introduced in the previous section is the Dirac Delta function. It is used with continuous functions and has application throughout science, engineering, and other disciplines. Consequently, it warrants a discussion to ensure the reader understands its function and properties.

The *Dirac Delta Function* $\delta(x)$ often is described as infinitely high spike with an infinitesimal width such that the "area" essentially is 1. This is not too satisfying as a mathematical function definition and a better way is to put in terms of limiting function, say like the following:

$$\delta(x) = \lim_{A\to\infty} \begin{cases} A & \text{if } |x| \le \frac{1}{2A} \\ 0 & \text{otherwise} \end{cases} \tag{1.145a}$$

$$= \lim_{A\to\infty} A \text{ rect}\left[\frac{x}{\frac{1}{A}}\right], \tag{1.145b}$$

where

$$\text{rect}[x] = \begin{cases} 1 & \text{if } |x| \le \frac{1}{2A} \\ 0 & \text{if } |x| > \frac{1}{2A} \end{cases} \tag{1.146}$$

is the *rectangle function*. This form of the delta function satisfies the implicit definition by keeping the area constant while its height gets larger and its width gets narrower as A moves to infinity. This implies that

$$\int_{-\infty}^{\infty} g(x)\delta(x-c)dx = \lim_{A\to\infty} \int_{-\infty}^{\infty} g(x)\, A \text{ rect}\left[\frac{x-c}{\frac{1}{A}}\right] dx = g(c). \tag{1.147}$$

There are other forms for the delta function than the above that satisfy the implicit definition and normalize to one. However, not all of them mean that $\delta(x-c) = 0$ whenever $x \neq c$. For example, the function

$$A \,\text{sinc}[A(x-c)] = \frac{\text{sinc}[\pi A(x-c)]}{\pi A(x-c)}. \tag{1.148}$$

Equation (1.148) defines the *sinc function*. This means that we hypothesize that

$$\int_{-\infty}^{\infty} g(x)\delta(x-c)dx = \lim_{A\to\infty} \int_{-\infty}^{\infty} g(x)\, A \,\text{sinc}\left[\frac{x-c}{A}\right] dx = g(c). \tag{1.149}$$

Because of the periodic nature of the sine function, there will be points where this sinc function will equal 1 when $x \neq c$. Fortunately, when A is large, it will oscillate

rapidly, resulting in the $g(x)$ contributions outside of $x = c$, averaging to zeros because of that fact over the long integration interval. This implies that

$$\delta(x) = \lim_{A\to\infty} A \operatorname{sinc}\left[\frac{x}{A}\right] \tag{1.150}$$

is a legitimate representation of the Dirac Delta Function.

One of the properties of this function is that it is the derivative of the *Unit Step Function*. In particular, we have

$$\frac{d}{dx}U(x-c) = \delta(x-c), \tag{1.151}$$

where

$$U(x) = \begin{cases} 1 & \text{if } x > 0 \\ \frac{1}{2} & \text{if } x = 0 \\ 0 & \text{if } x < 0 \end{cases} \tag{1.152}$$

is the Heaviside (or Unit) Step Function.

1.6 PROBABILITY THEORY

Optical detection and estimation theory is based on the concepts derived from probability theory. This theory is fundamental to understanding the performance of any optical remote sensing or communications system. This section provides a brief review of probability theory following Bar-Shalom and Fortmann, Appendix B [2] and Helstrom [6]. It recognizes that this theory is designed to understand the possible outcomes of *chance experiments*. These outcomes are unpredictable, but known with certain knowledge of their relative frequency of occurrence, countable or not. The theory provides a means for creating a strategy for "predicting" the various outcomes, which hopefully maximizes possible gains and minimizes losses. More information can be found in books on probability, for example, by Venkatesh [7] and Helstrom [6].

1.6.1 Axioms of Probability

Let $\boldsymbol{A}$ specify an *event* in a chance experiment whose results come from a specific set of random outcome. This could be "heads" if one was flipping a coin or the number "3" in throwing a die. Let $\boldsymbol{S}$ represent an assured event in the experiment; this would mean "a number between 1 and 6 coming up after throwing a die." The *probability* of an event, denoted by $P\{\boldsymbol{A}\}$, occurring obeys these following axioms:

Axiom 1

It is nonnegative, that is,

$$P\{\boldsymbol{A}\} \geq 0. \tag{1.153}$$

Axiom 2

It is unity for assured event, that is,

$$P\{\boldsymbol{S}\} = 1. \tag{1.154}$$

Axiom 3

It is additive over the union of mutually exclusive events, that is, if the events $\boldsymbol{A}$ and $\boldsymbol{B}$ have no common elements, then

$$\boldsymbol{A} \cap \boldsymbol{B} \equiv \{\boldsymbol{A} \text{ and } \boldsymbol{B}\} = \{\boldsymbol{A}, \boldsymbol{B}\} = \varnothing, \tag{1.155}$$

and

$$P\{\boldsymbol{A} \cup \boldsymbol{B}\} = P\{\boldsymbol{A} \text{ or } \boldsymbol{B}\} = P\{\boldsymbol{A} + \boldsymbol{B}\} = P\{\boldsymbol{A}\} + P\{\boldsymbol{B}\}. \tag{1.156}$$

Axiom 4

If $\boldsymbol{A}_i \cap \boldsymbol{A}_j = \varnothing$ for all i and j, $i \neq j$, then

$$P\{\cup_{i=1}^{\infty} \boldsymbol{A}_i\} = \sum_{i=1}^{\infty} P(\boldsymbol{A}_i). \tag{1.157}$$

An assignment of probabilities to the events consistent with these axioms is known as a *probability measure*.

Corollaries to the above are

Corollary 1

$$P(\varnothing) = 0. \tag{1.158}$$

Corollary 2

$$0 \leq P\{\boldsymbol{A}\} \leq 1. \tag{1.159}$$

Corollary 3

With $\overline{\boldsymbol{A}}$ being the complement of the event $\boldsymbol{A}$, then

$$P\{\overline{\boldsymbol{A}}\} = 1 - P\{\boldsymbol{A}\}, \tag{1.160}$$

using Eqs. (1.154) and (1.157).

Corollary 4

If the n events $\boldsymbol{A}_i$, $\{1 \leq i \leq n\}$, are mutually exclusive, for example, $\boldsymbol{A}_i \cap \boldsymbol{A}_j = \varnothing$ for all i and j, $i \neq j$, then

$$P\{\cup_{i=1}^{n} \boldsymbol{A}_i\} = \sum_{i=1}^{n} P(\boldsymbol{A}_i). \tag{1.161}$$

Corollary 5

$$P\{\boldsymbol{A} \cup \boldsymbol{B}\} = P\{\boldsymbol{A}\} + P\{\boldsymbol{B}\} - P\{\boldsymbol{A} \cap \boldsymbol{B}\}. \tag{1.162}$$

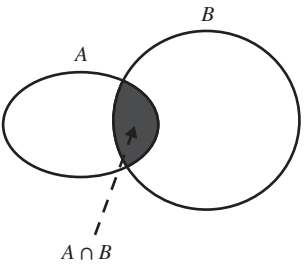


FIGURE 1.5 Venn diagram of events $\boldsymbol{A}$ and $\boldsymbol{B}$.

It should be noted that probability theory is essentially computing the new probabilities from an initial probability measure set over the events of the chance experiment. It is a Bookkeeping drill that requires that all the unit quantities of probability be accounted for, no probability is negative, and no excess probability emerges.

1.6.2 Conditional Probabilities

Sometimes the outcomes from two events are common. These type of outcomes form the intersection $\boldsymbol{A} \cap \boldsymbol{B}$, as depicted in Figure 1.5. The conditional probability of the event $\boldsymbol{A}$, given event $\boldsymbol{B}$, is defined as

$$P\{\boldsymbol{A} \mid \boldsymbol{B}\} = \frac{P\{\boldsymbol{A} \cap \boldsymbol{B}\}}{P\{\boldsymbol{B}\}} \tag{1.163}$$

in terms of the probability measure assigned to the event. The probability $P\{\boldsymbol{A} \mid \boldsymbol{B}\}$ can be interpreted as the likelihood of event $\boldsymbol{A}$ occurring when it is known that event $\boldsymbol{B}$ has occurred.

Example 1.9 A coin is tossed three times in succession. The possible outcomes from this chance experiment are

ϑ_1	ϑ_2	ϑ_3	ϑ_4	ϑ_5	ϑ_6	ϑ_7	ϑ_8
$\{HHH\}$	$\{HHT\}$	$\{HTH\}$	$\{HTT\}$	$\{THH\}$	$\{THT\}$	$\{TTH\}$	$\{TTT\}$

It is clear that there are eight elements to the *universal set* of outcomes of this experiment. For illustrative purposes, let us assume that the coin is unbalanced and the individual probabilities of occurrence for the above universal set after flipping the coin many times come out nonuniform and have the probabilities given in Table 1.1.

TABLE 1.1 Probabilities for Possible Outcomes from Flipping a Coin

ϑ_1	ϑ_2	ϑ_3	ϑ_4	ϑ_5	ϑ_6	ϑ_7	ϑ_8
$\{HHH\}$	$\{HHT\}$	$\{HTH\}$	$\{HTT\}$	$\{THH\}$	$\{THT\}$	$\{TTH\}$	$\{TTT\}$
0.07	0.31	0.17	0.05	0.29	0.01	0.06	0.04

(a) What is the probability that three heads will appear, given the coin's first toss is heads? From our table, the possible outcomes beginning with heads are $\boldsymbol{B} = \{\vartheta_1, \vartheta_2, \vartheta_3, \vartheta_4\}$. This conditioning event has probability

$$P\{\boldsymbol{B}\} = 0.07 + 0.31 + 0.17 + 0.05 = 0.60.$$

The probability that three heads come up is

$$P\{HHH\} = 0.07.$$

Using Eq. (1.163), we have

$$P\{HHH\,|\boldsymbol{B}\} = \frac{P\{HHH\}}{P\{\boldsymbol{B}\}} = \frac{0.07}{0.6} \approx 0.11.$$

(b) What is the probability that three heads will appear, given the coin was heads on the first two tosses? In this case, the possible outcomes are $\boldsymbol{B} = \{\vartheta_1, \vartheta_2\}$. This implies that

$$P\{\boldsymbol{B}\} = 0.07 + 0.31 = 0.38,$$

and we have

$$P\{HHH\,|\boldsymbol{B}\} = \frac{0.07}{0.38} \approx 0.18.$$

(c) What is the probability that tails appeared on the second and third tosses, given the number of tails is odd? In this case, the events are $\boldsymbol{A} = \{\vartheta_4, \vartheta_8\}$ and $\boldsymbol{B} = \{\vartheta_2, \vartheta_3, \vartheta_5, \vartheta_8\}$. The intersection of $\boldsymbol{A}$ and $\boldsymbol{B}$ is $\boldsymbol{A} \cap \boldsymbol{B} = \{\vartheta_8\}$ and we find that

$$P\{\boldsymbol{A}\,|\boldsymbol{B}\} = \frac{P\{\boldsymbol{A} \cap \boldsymbol{B}\}}{P\{\boldsymbol{B}\}} = \frac{0.04}{0.31 + 0.17 + 0.29 + 0.04} = \frac{0.04}{0.81} \approx 0.5.$$

Example 1.10 The reliability of a rectifier is such that the probability of it lasting at least t hours is given by $e^{-\alpha t^2}$. What is the probability that the rectifier fails between times t_1 and t_2, given it is still operating after time τ, where $\tau < t_1 < t_2$.

The universal set of time possible goes from zero to infinity. The conditioning event $\boldsymbol{B}$ is that the failure happens in the interval $\tau < t < \infty$. The associated probability of that happening is $e^{-\alpha\tau^2}$, by the hypothesis. Event $\boldsymbol{A}$ is that the rectifier fails within the interval $t_1 < t < t_2$. This implies that

$$P\{\text{failure in } t_1 < t < \infty\} = P\{\text{failure in } t_1 < t < t_2\} + P\{\text{failure in } t_2 < t < \infty\}$$

allows us to write

$$\begin{aligned} P\{\text{failure in } t_1 < t < t_2\} &= P\{\text{failure in } t_2 < t < \infty\} - P\{\text{failure in } t_1 < t < \infty\} \\ &= e^{-\alpha {t_2}^2} - e^{-\alpha {t_1}^2} \end{aligned}$$

using Eq. (1.144). From the above description, we have

$$P\{\boldsymbol{A}\,|\boldsymbol{B}\} = \frac{P\{\boldsymbol{A} \cap \boldsymbol{B}\}}{P\{\boldsymbol{B}\}} = \frac{e^{-\alpha {t_2}^2} - e^{-\alpha {t_1}^2}}{e^{-\alpha\tau^2}}.$$

If one cross multiplies the above conditional probability and use the fact that

$$\boldsymbol{A} \cap \boldsymbol{B} = \boldsymbol{B} \cap \boldsymbol{A},$$

then the following theorem can be derived:

$$P\{\boldsymbol{A} \cap \boldsymbol{B}\} = P\{\boldsymbol{B}\}\, P\{\boldsymbol{A} \mid \boldsymbol{B}\}.$$

This results in the following corollary:

Corollary 6

For any set of events $\{\boldsymbol{A}_1, \boldsymbol{A}_2, \ldots, \boldsymbol{A}_m\}$, we have

$$\begin{aligned} &P\{\boldsymbol{A}_1 \cap \boldsymbol{A}_2, \cap \cdots \cap \boldsymbol{A}_m\} \\ &\quad = P\{\boldsymbol{A}_1\} P\{\boldsymbol{A}_2 \mid \boldsymbol{A}_2\} P\{\boldsymbol{A}_3 \mid \boldsymbol{A}_2 \cap \boldsymbol{A}_2\} \cdots P\{\boldsymbol{A}_m \mid \boldsymbol{A}_1 \cap \boldsymbol{A}_2, \cap \cdots \cap \boldsymbol{A}_{m-1}\} \end{aligned}$$

The above theorem is known as the *multiplication theorem*.

Example 1.11 A class contains 12 boys and 4 girls. If three students are chosen at random from the class, what is the probability that they are all boys?

The probability that the first student chosen is a boy is $\frac{12}{16}$ since there are 12 boys in a class of 16 students. If the first selection is a boy, then the probability that the next student chosen is a boy is $\frac{11}{15}$. Finally, If the first two selections are boys, then the probability that the final selected student is a boy is $\frac{10}{14}$. Then by the multiplication theorem, the probability that all three students chosen are boys is

$$\left(\frac{12}{16}\right)\left(\frac{11}{15}\right)\left(\frac{10}{14}\right) = \left(\frac{1}{4}\right)\left(\frac{11}{1}\right)\left(\frac{1}{7}\right) = \frac{11}{28}.$$

1.6.3 Probability and Cumulative Density Functions

Nearly all applications of probability to science and engineering are derived from the outcomes from chance experiments, which are associated with specific numbers, for example, voltages, currents, and power. Thus, events are sets of outcomes from chance experiments. A scalar *random variable* is a real-valued parameter that labels an outcome of a chance experiment for a given event with a probability measure assigned to it. Its value is known as its *realization*. It follows that the outcome of experiment can be a single number, a pair or more of numbers, and even a function of the parameter. We now turn to defining the two key functions for characterizing all of the above. They are the *probability density function* (*PDF*) and *cumulative distribution function* (*CDF*).

The PDF $p_x(x)$ of a continuous-valued random variable x at $x = \vartheta$ is defined as

$$p_x(\vartheta)d\vartheta = P\{\vartheta \leq \boldsymbol{x} \leq \vartheta + d\vartheta\} \geq 0. \tag{1.164}$$

For simplicity, the PDF can be shorted to $p(\vartheta)$, where the argument of the function defines it. We will do so for the remainder of this section.

Example 1.12 Let us assume we have a device for counting the number of electrons radiated by some sort of emissive surface during the period of τ seconds. We also assume we have a constant time average rate of occurrence for said electron emissions equal to λ. A stochastic process

$$Z(t) = C\sum_{m=1}^{k}\delta(t - t_m), \tag{1.165}$$

composed of a sequence of electron emissions (impulses) occurring at times t_m in a time period $-\frac{\tau}{2}$ to $\frac{\tau}{2}$ and multiplied by a constant C, is a Poisson process if the probability that the number, X, of impulses occurring in the time period τ is an integer with probability

$$P\{X(t) = k\} = \frac{(\lambda t)^k}{k!}e^{-\lambda t}. \tag{1.166}$$

[8]. In Eq. (1.165), an impulse is represented by $\delta(t)$, the Dirac delta function defined previously.

The PDF for $Z(t)$ is not Poisson, but rather equals

$$P\{Z(t)\} = P\{t_1, t_2, \ldots, t_k \,|k\}P\{X(t) = k\}. \tag{1.167}$$

If the times of the impulse occurrences are independent of one another and independent of the total number of pulses, then Eq. (1.167) can be written as

$$P\{Z(t)\} = \frac{1}{\tau^k}\,P\{X(t) = k\}. \tag{1.168}$$

Using Eq. (1.164) and Corollary 4, we can write

$$P\{a \le \boldsymbol{x} \le b\} = \int_a^b p(x)dx. \tag{1.169}$$

The function

$$P\{-\infty \le \boldsymbol{x} \le b\} \equiv P_x(b) = P\{\boldsymbol{x} \le b\} = \int_{-\infty}^{b} p(x)dx \tag{1.170}$$

is called the CDF of $\boldsymbol{x}$ at b. Since the event $S \equiv (\boldsymbol{x} \le \infty)$ is a sure thing, then

$$P\{\boldsymbol{x} \le \infty\} = \int_{-\infty}^{\infty} p(x)dx = 1. \tag{1.171}$$

Equation (1.171) is known as the normalization integral for the PDF $p(x)$. This equation must be true for $p(x)$ to be a proper PDF.

The relationship between the PDF and CDF can be derived from Eq. (1.170), namely, we see that

$$p(x) = \frac{d}{d\vartheta}P_{\boldsymbol{x}}(\vartheta)|_{\vartheta=x}, \tag{1.172}$$

if the derivative exists.

1.6.4 Probability Mass Function

Let us assume that we have discrete random variable $\boldsymbol{x}$ comprising one set of possible countable values $\{\vartheta_1, \vartheta, \ldots, \vartheta_m\}$. (Note: the set could extend to infinity and still be countable.) As a result, their probabilities are written as

$$P\{\boldsymbol{x} = \vartheta_i\} = \mu_{\boldsymbol{x}}(\vartheta_i) = \mu_i, \quad \text{for } i = 1, 2, \ldots, m. \tag{1.173}$$

The parameter μ_i is the point mass associated with outcome ϑ_i. Similar to Eq. (1.171), we have

$$\sum_{i=1}^{m} \mu_i = 1. \tag{1.174}$$

Using the Dirac Delta Function, we can write the PDF for the above random variable as

$$p(x) = \sum_{i=1}^{m} \mu_i \, \delta(x - \vartheta_i). \tag{1.175}$$

This PDF normalizes properly, so it is a true PDF. The distribution expressed in Eq. (1.175) has jumps in it at values of ϑ_i. Its CDF is a "stair step" or "staircase" function whose derivative is zero everywhere but at the jumps, where it is an impulse.

As a final note, a random variable $\boldsymbol{x}$ can take on values that are continuous, discrete, or a combination of both. The PDF for a mixed random variable has the form

$$p(x) = p_c(x) + \sum_{i=1}^{m} \mu_i \, \delta(x - \vartheta_i). \tag{1.176}$$

where $p_c(x)$ represents the continuous part of the PDF. As one would expect,

$$\begin{aligned} \int p(x)dx &= \int_{x \in X} p_c(x)dx + \sum_{i=1}^{m} \mu_i \int \delta(x - \vartheta_i)dx \\ &= \int_{x \in X} p_c(x)dx + \sum_{i=1}^{m} \mu_i = 1. \end{aligned} \tag{1.177}$$

Example 1.13 Let us again assume that we have discrete random variable $\boldsymbol{x}$ that only takes on the following values $\{\vartheta_1, \vartheta, \ldots, \vartheta_m\}$. Their associated probabilities are $P\{\boldsymbol{x} = \vartheta_i\}; i = 1, \ldots, m$, which normalize properly. Its CDF will have steps of height $P\{\boldsymbol{x} = \vartheta_i\}$ at all points ϑ_i and can be written as

$$P_{\boldsymbol{x}}(x) = \sum_{i=1}^{m} P\{\boldsymbol{x} = \vartheta_i\} \; U(x - \vartheta_i). \tag{1.178}$$

Using Eq. (1.172) for the PDF, we find that

$$p_{\boldsymbol{x}}(x) = \sum_{i=1}^{m} P\{\boldsymbol{x} = \vartheta_i\} \; \delta(x - \vartheta_i) \tag{1.179a}$$

$$= \sum_{i=1}^{m} \mu_i \, \delta(x - \vartheta_i). \tag{1.179b}$$

Equation (1.179b) is the same one we used in Eq. (1.173) for characterizing the PDF in terms of the point masses.

1.6.5 Expectation and Moments of a Scalar Random Variable

The expected value, or *mean* or first moment, of a random variable is defined as

$$\bar{\boldsymbol{x}} \equiv \mathcal{E}\{\boldsymbol{x}\} = \int_{-\infty}^{\infty} x \, p(x) dx. \tag{1.180}$$

The expected value of any function of the random variable therefore is

$$\mathcal{E}\{g(\boldsymbol{x})\} = \int_{-\infty}^{\infty} g(x) \, p(x) dx. \tag{1.181}$$

The *variance*, or second moment, of a random variable is defined as

$$\text{var}(\boldsymbol{x}) = \sigma_{\boldsymbol{x}}^2 \equiv \mathcal{E}\{(\boldsymbol{x} - \bar{\boldsymbol{x}})^2\} = \mathcal{E}\{\boldsymbol{x}^2\} - (\bar{\boldsymbol{x}})^2 = \int_{-\infty}^{\infty} (x - \bar{\boldsymbol{x}})^2 \, p(x) dx. \tag{1.182}$$

The square root of the variance is the called the *standard deviation*. Looking at the forms of Eqs. (1.180) and (1.182) implies that the *n*th (noncentral) moment of the random variable is given by

$$\mathcal{E}\{\boldsymbol{x}^n\} = \int_{-\infty}^{\infty} x^n \, p(x) dx. \tag{1.183}$$

Example 1.14 Let us assume a random variable $\boldsymbol{x}$ that is uniformly distributed. Its PDF is given by

$$p(x) = \begin{cases} \frac{1}{x_2 - x_1} & \text{for } x_1 \le x \le x_2 \\ 0 & \text{otherwise} \end{cases}. \tag{1.184}$$

The mean for this random variable equals

$$\bar{\boldsymbol{x}} \equiv \mathcal{E}\{\boldsymbol{x}\} = \int_{x_1}^{x_2} \left(\frac{x}{x_2 - x_1} \right) dx = \frac{1}{2} \left(\frac{x^2}{x_2 - x_1} \right) \Bigg|_{x_1}^{x_2} = \frac{1}{2} \left(\frac{{x_2}^2 - {x_1}^2}{x_2 - x_1} \right) = \frac{(x_2 + x_1)}{2}.$$

The covariance equals

$$\begin{aligned} \text{var}(\boldsymbol{x}) &= \int_{x_1}^{x_2} \left(\frac{x^2}{x_2 - x_1} \right) dx - (\bar{x})^2 = \frac{1}{3} \left(\frac{x^3}{x_2 - x_1} \right) \Bigg|_{x_1}^{x_2} - \left(\frac{(x_2 + x_1)}{2} \right)^2 \\ &= \frac{(x_2 - x_1)^2}{12}. \end{aligned}$$

1.6.6 Joint PDF and CDF of Two Random Variables

Given two random variables $\boldsymbol{x}$ and $\boldsymbol{y}$, the joint PDF is defined as the probability of the joint event, which is denoted by the set intersection symbol, and is written as

$$p_{xy}\{\vartheta,\zeta\}d\vartheta\,d\zeta \equiv P_{xy}\{[\vartheta < x < \vartheta + d\vartheta] \cap [\zeta < y < \zeta + d\zeta]\}. \tag{1.185}$$

The function $p_{xy}\{x, y\}$ is called the *join* or *bivariate PDF*. It follows that

$$p_{xy}(x, y) = \frac{d}{d\vartheta\, d\zeta} P_{\boldsymbol{x}}(\vartheta,\zeta)|_{\vartheta=x,\zeta=y} \tag{1.186}$$

defines the relationship between the bivariate PDF and bivariate CDF, provided that $P_x(x, y)$ is continuous and differentiable. Integrating Eq. (1.185) over one of the variables yields the PDF of the other variable, that is,

$$\int_{-\infty}^{\infty} p_{xy}(\vartheta,\zeta)d\zeta = p_x(\vartheta) \tag{1.187}$$

or

$$\int_{-\infty}^{\infty} p(\vartheta,\zeta)d\zeta = p(\vartheta) \tag{1.188}$$

in reduced notation. The resulting PDF is called the *marginal PDF* or *marginal density*.

The *covariance* of two scalar random variables ϑ_1 and ϑ_2 with means $\overline{\vartheta}_1$ and $\overline{\vartheta}_2$, respectively, is equal to

$$\mathcal{E}\{(\vartheta_1 - \overline{\vartheta}_1)(\vartheta_2 - \overline{\vartheta}_2)\} = \int_{-\infty}^{\infty} (\vartheta_1 - \overline{\vartheta}_1)(\vartheta_2 - \overline{\vartheta}_2)p(\vartheta_1, \vartheta_2)d\vartheta_1 d\vartheta_2 \tag{1.189}$$

$$= \sigma^2_{\vartheta_1\vartheta_2}. \tag{1.190}$$

The correlation coefficient of these two random variables can be written as

$$\rho_{12} = \frac{\sigma^2_{\vartheta_1\vartheta_2}}{\sigma_{\vartheta_1}\sigma_{\vartheta_2}}, \tag{1.191}$$

where σ_{ϑ_i} is the standard deviation of the variable ϑ_i, $i = 1, 2$. The correlation coefficient of any two random variables must obey the following inequality:

$$|\rho_{12}| \leq 1. \tag{1.192}$$

1.6.7 Independent Random Variables

If the conditional PDF $p_x(x|y)$ does not depend on y, then we have

$$p_x(x|y) = p_x(x), \tag{1.193}$$

which says that any observation of the random variables $\boldsymbol{x}$ provides no information about the variable $\boldsymbol{y}$. This situation establishes that the two random variables $\boldsymbol{x}$ and $\boldsymbol{y}$ are *statistically independent*. Another way to say this is that two events are independent if the probability of their joint event equals the product of their marginal probabilities, or

$$P\{\boldsymbol{A} \cap \boldsymbol{B}\} = P\{\boldsymbol{A}, \boldsymbol{B}\} = P\{\boldsymbol{A}\}P\{\boldsymbol{B}\}. \tag{1.194}$$

The conclusion from this is that

$$p_{xy}(x, y) = p_x(x)p_y(y) \tag{1.195}$$

and

$$P_{xy}(x, y) = P_x(x)P_y(y), \tag{1.196}$$

both in terms of marginal probabilities.

1.6.8 Vector-Valued Random Variables

The PDF of the vector-valued random variable

$$\boldsymbol{x} = [x_1 \cdots x_m]^T \tag{1.197}$$

is defined as the joint density of its components

$$p_{x_1 \cdots x_m}(\vartheta_1 \cdots \vartheta_m)d\vartheta_1 \cdots d\vartheta_m \equiv p_x(\boldsymbol{\vartheta})d\boldsymbol{\vartheta} = P_x(\cap_{i=1}^m \{\vartheta_i < x_i \leq \vartheta_i + d\vartheta_i\}), \tag{1.198}$$

where the set intersection symbol is used to denote the joint event

$$\boldsymbol{A} \cap \boldsymbol{B} = \{\boldsymbol{A} \text{ and } \boldsymbol{B}\} = \{\boldsymbol{A}, \boldsymbol{B}\}. \tag{1.199}$$

The *mean* of $\boldsymbol{x}$ is the result of the m-fold integration

$$\overline{\boldsymbol{x}} = \mathcal{E}\{\boldsymbol{x}\} = \int \boldsymbol{x} p_x(\boldsymbol{x})d\boldsymbol{x}. \tag{1.200}$$

The *covariance matrix* of $\boldsymbol{x}$ is given by

$$\boldsymbol{\Gamma}_{xx} = \mathcal{E}\left\{(\boldsymbol{x} - \overline{\boldsymbol{x}})(\boldsymbol{x} - \overline{\boldsymbol{x}})^T\right\} = \int (\boldsymbol{x} - \overline{\boldsymbol{x}})(\boldsymbol{x} - \overline{\boldsymbol{x}})^T \, p_x(\boldsymbol{x})d\boldsymbol{x}. \tag{1.201}$$

The covariance matrix is a positive definite (or semidefinite) matrix, whose diagonal elements are the variances of the random variable $\boldsymbol{x}$. The off-diagonal elements are the covariances between various components of $\boldsymbol{x}$.

The *characteristic* or *moment generating function* of a vector random variable is defined as

$$M_x(\boldsymbol{S}) = \mathcal{E}\left\{e^{\boldsymbol{S}^T\boldsymbol{x}}\right\} = \int e^{\boldsymbol{S}^T\boldsymbol{x}} p_x(\boldsymbol{x})d\boldsymbol{x}, \tag{1.202}$$

which is the Fourier transform of the PDF. The first moment of $\boldsymbol{x}$ is related to the characteristic function via the relation

$$\mathcal{E}\{\boldsymbol{x}\} = \nabla_{\boldsymbol{S}} M_x(\boldsymbol{S})|_{\boldsymbol{S}=\boldsymbol{0}}, \tag{1.203}$$

where $\nabla_{\boldsymbol{S}}$ is the (column) gradient operator defined in Section 1.2.7. Similarly, we have

$$\mathcal{E}\left\{\boldsymbol{x}\boldsymbol{x}^T\right\} = \nabla_{\boldsymbol{S}} \nabla_{\boldsymbol{S}}^T M_x(\boldsymbol{S})|_{\boldsymbol{S}=\boldsymbol{0}}. \tag{1.204}$$

1.6.9 Gaussian Random Variables

The PDF of a *Gaussian or normally distributed* random variable is

$$p(x) = N(x, \bar{x}, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\bar{x})^2}{2\sigma^2}\right\}, \tag{1.205}$$

where $N(x, \bar{x}, \sigma^2)$ denotes the normal PDF, $\bar{x}$ is the mean, and σ^2 its variance. These first two moments totally characterize the Gaussian random variance and are known as its statistics.

A vector-valued Gaussian random variable has a PDF of the form

$$p(\boldsymbol{x}) = N(\boldsymbol{x}, \bar{\boldsymbol{x}}, \boldsymbol{\Gamma}) = \frac{1}{\sqrt{2\pi\boldsymbol{\Gamma}}} \exp\{-(\boldsymbol{x}-\bar{\boldsymbol{x}})^T \boldsymbol{\Gamma}^{-1} (\boldsymbol{x}-\bar{\boldsymbol{x}})\}, \tag{1.206}$$

where $\bar{\boldsymbol{x}}$ is the mean of the vector $\boldsymbol{x}$ and $\boldsymbol{\Gamma}$ is its covariance matrix. If $\boldsymbol{\Gamma}$ is a diagonal matrix, then the elements of vector $\boldsymbol{x}$ are uncorrelated and independent. Consequently, their joint PDF equals the product of their marginal PDFs.

Two vector random vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ are jointly Gaussian if the stacked vector

$$z = \begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{bmatrix}. \tag{1.207}$$

Its PDF is given by

$$p(\boldsymbol{x}, \boldsymbol{y}) = p(z) = N(z, \bar{z}, \boldsymbol{\Gamma}_{zz}). \tag{1.208}$$

The mean and covariance matrix of z in terms of those vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ are

$$\bar{z} = \begin{bmatrix} \bar{\boldsymbol{x}} \\ \bar{\boldsymbol{y}} \end{bmatrix} \tag{1.209}$$

and

$$\boldsymbol{\Gamma}_{zz} = \begin{bmatrix} \boldsymbol{\Gamma}_{xx} & \boldsymbol{\Gamma}_{xy} \\ \boldsymbol{\Gamma}_{yx} & \boldsymbol{\Gamma}_{yy} \end{bmatrix}, \tag{1.210}$$

where

$$\boldsymbol{\Gamma}_{xx} = \mathcal{E}\left\{(\boldsymbol{x}-\bar{\boldsymbol{x}})(\boldsymbol{x}-\bar{\boldsymbol{x}})^T\right\}, \tag{1.211}$$

$$\mathbf{\Gamma}_{xy} = \mathcal{E}\left\{(\boldsymbol{x} - \overline{\boldsymbol{x}})(\boldsymbol{y} - \overline{\boldsymbol{y}})^T\right\}, \tag{1.212}$$

$$\mathbf{\Gamma}_{yx} = \mathcal{E}\left\{(\boldsymbol{y} - \overline{\boldsymbol{y}})(\boldsymbol{x} - \overline{\boldsymbol{x}})^T\right\}, \tag{1.213}$$

and

$$\mathbf{\Gamma}_{yy} = \mathcal{E}\left\{(\boldsymbol{y} - \overline{\boldsymbol{y}})(\boldsymbol{y} - \overline{\boldsymbol{y}})^T\right\}. \tag{1.214}$$

The conditional probability for $\boldsymbol{x}$, given $\boldsymbol{y}$, is written as

$$p(\boldsymbol{x}|\boldsymbol{y}) = \frac{p(\boldsymbol{x}, \boldsymbol{y})}{p(\boldsymbol{y})}. \tag{1.215}$$

Let

$$\boldsymbol{\xi} = (\boldsymbol{x} - \overline{\boldsymbol{x}}) \quad \text{and} \quad \boldsymbol{\eta} = (\boldsymbol{y} - \overline{\boldsymbol{y}}).$$

Substituting Eqs. (1.206) and (1.208) into Eq. (1.215), we obtain for the argument in the resulting exponent the following equation:

$$q = \begin{bmatrix}\boldsymbol{\xi} \\ \boldsymbol{\eta}\end{bmatrix}^T \begin{bmatrix}\mathbf{\Gamma}_{xx} & \mathbf{\Gamma}_{xy} \\ \mathbf{\Gamma}_{yx} & \mathbf{\Gamma}_{yy}\end{bmatrix}^{-1} \begin{bmatrix}\boldsymbol{\xi} \\ \boldsymbol{\eta}\end{bmatrix} - \boldsymbol{\eta}^T \mathbf{\Gamma}_{yy}{}^{-1} \boldsymbol{\eta} \tag{1.216}$$

$$= \begin{bmatrix}\boldsymbol{\xi} \\ \boldsymbol{\eta}\end{bmatrix}^T \begin{bmatrix}\mathbf{\Upsilon}_{xx} & \mathbf{\Upsilon}_{xy} \\ \mathbf{\Upsilon}_{yx} & \mathbf{\Upsilon}_{yy}\end{bmatrix} \begin{bmatrix}\boldsymbol{\xi} \\ \boldsymbol{\eta}\end{bmatrix} - \boldsymbol{\eta}^T \mathbf{\Gamma}_{yy}{}^{-1} \boldsymbol{\eta}, \tag{1.217}$$

using the results in Section 1.3. In particular, we find for the partitions that

$$\mathbf{\Upsilon}_{xx}{}^{-1} = \mathbf{\Gamma}_{xx} - \mathbf{\Gamma}_{xy} \mathbf{\Gamma}_{yy}{}^{-1} \mathbf{\Gamma}_{yx}, \tag{1.218}$$

$$\mathbf{\Gamma}_{yy}{}^{-1} = \mathbf{\Upsilon}_{yy} - \mathbf{\Upsilon}_{yx} \mathbf{\Upsilon}_{xx}{}^{-1} \mathbf{\Upsilon}_{xy}, \tag{1.219}$$

and

$$\mathbf{\Upsilon}_{xx}{}^{-1} \mathbf{\Upsilon}_{xy} = -\mathbf{\Gamma}_{xy} \mathbf{\Gamma}_{yy}{}^{-1}, \tag{1.220}$$

which allows us to write Eq. (1.217) [after a multiplication of -2] as

$$\begin{aligned} q &= \boldsymbol{\xi}^T \mathbf{\Upsilon}_{xx} \boldsymbol{\xi} + \boldsymbol{\xi}^T \mathbf{\Upsilon}_{xy} \boldsymbol{\eta} + \boldsymbol{\eta}^T \mathbf{\Upsilon}_{yx} \boldsymbol{\xi} + \boldsymbol{\eta}^T \mathbf{\Upsilon}_{yy} \boldsymbol{\eta} - \boldsymbol{\eta}^T \mathbf{\Gamma}_{yy}{}^{-1} \boldsymbol{\eta} \\ &= (\boldsymbol{\xi} + \mathbf{\Upsilon}_{xx}{}^{-1} \mathbf{\Upsilon}_{xy} \boldsymbol{\eta})^T \mathbf{\Upsilon}_{xx} (\boldsymbol{\xi} + \mathbf{\Upsilon}_{xx}{}^{-1} \mathbf{\Upsilon}_{xy} \boldsymbol{\eta}) + \boldsymbol{\eta}^T (\mathbf{\Upsilon}_{yy} - \mathbf{\Upsilon}_{yx} \mathbf{\Upsilon}_{xx}{}^{-1} \mathbf{\Upsilon}_{xy}) \boldsymbol{\eta} - \boldsymbol{\eta}^T \Gamma_{yy}{}^{-1} \boldsymbol{\eta} \\ &= (\boldsymbol{\xi} + \mathbf{\Upsilon}_{xx}{}^{-1} \mathbf{\Upsilon}_{xy} \boldsymbol{\eta})^T \mathbf{\Upsilon}_{xx} (\boldsymbol{\xi} + \mathbf{\Upsilon}_{xx}{}^{-1} \mathbf{\Upsilon}_{xy} \boldsymbol{\eta}), \end{aligned} \tag{1.221}$$

using the partitioning results. It is clear that Eq. (1.221) has a quadratic form, which means that the conditional PDF of $\boldsymbol{x}$, given $\boldsymbol{y}$, also is Gaussian.

Using Eq. (1.220), we find that

$$\boldsymbol{\xi} + \mathbf{\Upsilon}_{xx}{}^{-1} \mathbf{\Upsilon}_{xy} \boldsymbol{\eta} = (\boldsymbol{x} - \overline{\boldsymbol{x}}) - \mathbf{\Gamma}_{xy} \mathbf{\Gamma}_{yy}{}^{-1} (\boldsymbol{y} - \overline{\boldsymbol{y}}). \tag{1.222}$$

The conditional mean of $\boldsymbol{x}$, given $\boldsymbol{y}$, is

$$\widehat{\boldsymbol{x}} = \mathcal{E}\{\boldsymbol{x}|\boldsymbol{y}\} = \overline{\boldsymbol{x}} - \mathbf{\Gamma}_{xy} \mathbf{\Gamma}_{yy}{}^{-1} (\boldsymbol{y} - \overline{\boldsymbol{y}}), \tag{1.223}$$

and its covariance matrix

$$\boldsymbol{\Gamma}_{xx|y} = \mathcal{E}\{(\boldsymbol{x} - \hat{\boldsymbol{x}})(\boldsymbol{x} - \hat{\boldsymbol{x}})^T|\boldsymbol{y}\} = \boldsymbol{\Upsilon}_{xx}{}^{-1} = \boldsymbol{\Gamma}_{xx} - \boldsymbol{\Gamma}_{xy}\boldsymbol{\Gamma}_{yy}{}^{-1}\boldsymbol{\Gamma}_{yx}, \tag{1.224}$$

using Eq. (1.218).

As a final comment, Gaussian random variables remain Gaussian even under linear or affine transformations.

1.6.10 Quadratic and Quartic Forms

The expected value of quadratic and quartic forms of Gaussian random variables can be derived as follows. Let us assume that

$$p(\boldsymbol{x}) = N(\boldsymbol{x}, \overline{\boldsymbol{x}}, \boldsymbol{\Gamma}).$$

Then the characteristic function for the above is

$$M_x(\boldsymbol{S}) = \mathcal{E}\left\{e^{\boldsymbol{S}^T\boldsymbol{x}}\right\} = e^{\frac{1}{2}\,\boldsymbol{S}^T\boldsymbol{\Gamma}\,\boldsymbol{S} + \boldsymbol{S}^T\overline{\boldsymbol{x}}}. \tag{1.225}$$

For convenience and without any loss of generality, let us assume $\overline{\boldsymbol{x}} = 0$.

Given the vector random variable $\boldsymbol{x}$, we can write the following general equation:

$$\mathcal{E}\left\{\boldsymbol{x}^T\boldsymbol{A}\,\boldsymbol{x}\right\} = \mathcal{E}\left\{\mathrm{tr}[\boldsymbol{A}\,\boldsymbol{x}\boldsymbol{x}^T]\right\} = \mathrm{tr}[\boldsymbol{A}\,\boldsymbol{\Gamma}] \tag{1.226}$$

for any arbitrary matrix $\boldsymbol{A}$. The same result can be obtained using the characteristic function; specifically, we have

$$\begin{aligned}\mathcal{E}\left\{\boldsymbol{x}^T\boldsymbol{A}\,\boldsymbol{x}\right\} &= \mathcal{E}\left\{\nabla_{\boldsymbol{S}}{}^T e^{\boldsymbol{S}^T\boldsymbol{x}}\boldsymbol{A}\,\boldsymbol{x}\right\}\Big|_{\boldsymbol{S}=\boldsymbol{0}} = \mathcal{E}\left\{\nabla_{\boldsymbol{S}}{}^T\boldsymbol{A}\,\boldsymbol{x}e^{\boldsymbol{S}^T\boldsymbol{x}}\right\}\Big|_{\boldsymbol{S}=\boldsymbol{0}} \\ &= \nabla_{\boldsymbol{S}}{}^T\boldsymbol{A}\,\mathcal{E}\left\{\boldsymbol{x}e^{\boldsymbol{S}^T\boldsymbol{x}}\right\}\Big|_{\boldsymbol{S}=\boldsymbol{0}} = \nabla_{\boldsymbol{S}}{}^T\boldsymbol{A}\,\mathcal{E}\left\{\nabla_{\boldsymbol{S}}e^{\boldsymbol{S}^T\boldsymbol{x}}\right\}\Big|_{\boldsymbol{S}=\boldsymbol{0}} \\ &= \nabla_{\boldsymbol{S}}{}^T\boldsymbol{A}\,\nabla_{\boldsymbol{S}}M_{\boldsymbol{x}}(\boldsymbol{S})|_{\boldsymbol{S}=\boldsymbol{0}}.\end{aligned} \tag{1.227}$$

Substituting Eq. (1.225) into Eq. (1.227), with $\overline{\boldsymbol{x}} = 0$, will give us Eq. (1.226). This same procedure can be applied to the quartic form.

Specifically, we find that

$$\mathcal{E}\{\boldsymbol{x}^T\boldsymbol{A}\,\boldsymbol{x}\boldsymbol{x}^T\boldsymbol{B}\,\boldsymbol{x}\} = \nabla_{\boldsymbol{S}}{}^T\boldsymbol{A}\,\nabla_{\boldsymbol{S}}\nabla_{\boldsymbol{S}}{}^T\boldsymbol{B}\,\nabla_{\boldsymbol{S}}M_x(\boldsymbol{S})|_{\boldsymbol{S}=\boldsymbol{0}} \tag{1.228}$$

$$= \mathrm{tr}[\boldsymbol{A}\,\boldsymbol{\Gamma}]\mathrm{tr}[\boldsymbol{B}\,\boldsymbol{\Gamma}] + 2\mathrm{tr}[\boldsymbol{A}\,\boldsymbol{\Gamma}\,\boldsymbol{B}\,\boldsymbol{\Gamma}]. \tag{1.229}$$

Example 1.15 If $\boldsymbol{A}$ and $\boldsymbol{B}$ equal the scalar 1 and $\boldsymbol{\Gamma}$ equals the scalar σ^2, then the scalar counterpart for Eq. (1.224) is the well-known expression for the fourth moment of a Gaussian random variable given by

$$\mathcal{E}\left\{x^4\right\} = 3\sigma^4. \tag{1.230}$$

Using Eqs. (1.226) and (1.229), the covariance of two quartic forms is equal to

$$\begin{aligned}
\mathcal{E}\left\{\left(\boldsymbol{x}^T\boldsymbol{A}\,\boldsymbol{x}-\mathcal{E}\left\{\boldsymbol{x}^T\boldsymbol{A}\,\boldsymbol{x}\right\}\right)\left(\boldsymbol{x}^T\boldsymbol{B}\,\boldsymbol{x}-\mathcal{E}\left\{\boldsymbol{x}^T\boldsymbol{B}\,\boldsymbol{x}\right\}\right)\right\} &= \mathcal{E}\left\{\boldsymbol{x}^T\boldsymbol{A}\,\boldsymbol{x}\boldsymbol{x}^T\boldsymbol{B}\,\boldsymbol{x}\right\} \\
&\quad -\mathcal{E}\left\{\boldsymbol{x}^T\boldsymbol{A}\,\boldsymbol{x}\right\}\mathcal{E}\left\{\boldsymbol{x}^T\boldsymbol{B}\,\boldsymbol{x}\right\} \\
&\quad -\mathcal{E}\left\{\boldsymbol{x}^T\boldsymbol{B}\,\boldsymbol{x}\right\}\mathcal{E}\left\{\boldsymbol{x}^T\boldsymbol{A}\,\boldsymbol{x}\right\} \\
&\quad +\mathcal{E}\left\{\boldsymbol{x}^T\boldsymbol{B}\,\boldsymbol{x}\right\}\mathcal{E}\left\{\boldsymbol{x}^T\boldsymbol{A}\,\boldsymbol{x}\right\} \\
&= \mathcal{E}\left\{\boldsymbol{x}^T\boldsymbol{A}\,\boldsymbol{x}\boldsymbol{x}^T\boldsymbol{B}\,\boldsymbol{x}\right\} \\
&\quad -\mathcal{E}\left\{\boldsymbol{x}^T\boldsymbol{A}\,\boldsymbol{x}\right\}\mathcal{E}\left\{\boldsymbol{x}^T\boldsymbol{B}\,\boldsymbol{x}\right\} \\
&= \mathrm{tr}\left[\boldsymbol{A}\,\boldsymbol{\Gamma}\right]\mathrm{tr}\left[\boldsymbol{B}\,\boldsymbol{\Gamma}\right]+2\mathrm{tr}\left[\boldsymbol{A}\,\boldsymbol{\Gamma}\,\boldsymbol{B}\,\boldsymbol{\Gamma}\right] \\
&\quad -\mathrm{tr}\left[\boldsymbol{A}\,\boldsymbol{\Gamma}\right]\mathrm{tr}\left[\boldsymbol{B}\,\boldsymbol{\Gamma}\right] \\
&\quad = 2\mathrm{tr}\left[\boldsymbol{A}\,\boldsymbol{\Gamma}\,\boldsymbol{B}\,\boldsymbol{\Gamma}\right]. \qquad (1.231)
\end{aligned}$$

1.6.11 Chi-Squared Distributed Random Variable

If an m-dimensional random variable $\boldsymbol{x}$ is Gaussian with mean $\overline{\boldsymbol{x}}$ and covariance matrix $\boldsymbol{\Gamma}$, then the scalar random variable of the form

$$q=(\boldsymbol{x}-\overline{\boldsymbol{x}})^T\boldsymbol{\Gamma}^{-1}(\boldsymbol{x}-\overline{\boldsymbol{x}}) \qquad (1.232)$$

is the sum of the squares of m independent, zero-mean, unity-variance Gaussian random variables and consequently, has a *chi-squared distribution* with m degrees of freedom. Let's see why this is true.

Define

$$\boldsymbol{u}=\boldsymbol{\Gamma}^{-1/2}\,(\boldsymbol{x}-\overline{\boldsymbol{x}}). \qquad (1.233)$$

Obviously, the form of the above implies that $\boldsymbol{u}$ is Gaussian, with

$$\mathcal{E}\{\boldsymbol{u}\}=0 \qquad (1.234)$$

and

$$\mathcal{E}\{\boldsymbol{u}\boldsymbol{u}^T\}=\boldsymbol{\Gamma}^{-1/2}\,\mathcal{E}\left\{(\boldsymbol{x}-\overline{\boldsymbol{x}})(\boldsymbol{x}-\overline{\boldsymbol{x}})^T\right\}\,\boldsymbol{\Gamma}^{-1/2}=\boldsymbol{\Gamma}^{-1/2}\boldsymbol{\Gamma}\,\boldsymbol{\Gamma}^{-1/2}=\boldsymbol{I}. \qquad (1.235)$$

Because the above covariance matrix is a diagonal matrix, its components are independent. This means that

$$q=\boldsymbol{u}^T\boldsymbol{u}=\sum_{i=1}^{m}u_i^2, \qquad (1.236)$$

where $u_i\sim N(0,1)$.

Given the above, the convention is to write

$$q\sim\chi_m^2. \qquad (1.237)$$

The mean for q is

$$\mathcal{E}\{q\} = \mathcal{E}\left\{\sum_{i=1}^{m} u_i^2\right\} = m, \tag{1.238}$$

and its variances equals

$$\mathcal{E}\left\{\left[\sum_{i=1}^{m}(u_i^2 - 1)\right]^2\right\} = \sum_{i=1}^{m}\mathcal{E}\{(u_i^2 - 1)^2\}$$

$$= \sum_{i=1}^{m}[\mathcal{E}\{u_i^4\} - 2\mathcal{E}\{u_i^2\} + 1] = 2m, \tag{1.239}$$

using Eq. (1.230) for the fourth moment of a Gaussian random variable as $\sigma^4 = \sigma^2 = 1$.

1.6.12 Binomial Distribution

A chance experiment $\mathbb{E}$ is repeated m times. The m trials effectively create an m -fold chance experiment, that is, $\mathbb{E}^m$ defines our new chance experiment. The outcomes of $\mathbb{E}^m$ are m -tuples of outcomes generated by the various outcome combinations from the m trials of $\mathbb{E}$. Let us look how to characterize these experiments statistically by an example.

The convention is to recognize that any event $\boldsymbol{B}$ in the experiment $\mathbb{E}$ has a probability p and its complement $\boldsymbol{B'}$ has probability $(1 - p) = q$. An occurrence of $\boldsymbol{B}$ is a "success" and the occurrence of its complement is a "failure." For simplicity discussion here, let us assume that our experiment involves a biased coin, and "heads" denote "success" and "tails" denote "failure. Clearly, these trials are statistically independent. The universal set has 2^m possible outcomes. To characterize the compound experiments, we seek the probability density for the number of "heads" and "tails" in m trials of chance experiment $\mathbb{E}$, which we interpret as a single experiment, $\mathbb{E}' = \mathbb{E}^m$. $\mathbb{E}'$ is often described as a succession of *Bernoulli trials*.

Let A_k denote the event where k "heads" and $(m - k)$ "tails" both occur after m coin tosses happen. The number of outcomes of $\mathbb{E}'$ in event A_k is equal to

$$\binom{m}{k} = \frac{m!}{r!(m-k)!}, \tag{1.240}$$

which is called the *binominal coefficient* [6, pp. 40–41]. Taking the sum of Eq. (1.240) gives

$$\sum_{k=0}^{m}\binom{m}{k} = \sum_{k=0}^{m}\frac{m!}{r!(m-k)!} = (1+1)^m = 2^m \tag{1.241}$$

using the binomial theorem,

$$(x+y)^n = \sum_{k=0}^{n}\frac{n!}{r!(n-r)!}\,x^{n-k}y^r. \tag{1.242}$$

Equation (1.241) confirms our earlier claim that we have 2^m possible outcomes in the universal set.

Let us now interpret our compound experiment as an atomic event (single outcome) in $\mathbb{E}'$ consisting of k "successes" and $(m-k)$ "failures" in a particular order. The probability of this event is $p^k q^{m-k}$ because each trial is independent of the others. The number of such atomic events in A_k is $\binom{m}{k}$ and the probability of A_k is

$$P\{A_k\} = \binom{m}{k} p^k q^{m-k} \tag{1.243}$$

from Corollary 4, Eq. (1.161). Equation (1.243) is known as *binomial distribution*. Taking the sum of Eq. (1.243) yields

$$\sum_{k=0}^{m} P\{A_k\} = \sum_{k=0}^{m} \binom{m}{k} p^k q^{m-k} = (p+q)^m = (1)^m = 1. \tag{1.244}$$

Example 1.16 Let us assume we have a pair of fair dice that we toss 10 times. What is the probability that the dice totals seven points exactly four times? The possible outcomes that total seven points are $(1,6), (2,5), (3,4), (4,3), (5,2),$ and $(6,1)$. The probability of success equals to $p = 6 \times 6^{-2} = \frac{1}{6}$. The probability of failure is $q = 1 - p = \frac{5}{6}$. The probability of the cited event is

$$\binom{10}{4} \left(\frac{1}{6}\right)^4 \left(\frac{5}{6}\right)^6 = (210)(0.000771605)(0.334897977) = 0.0543.$$

The probability that at most k successes occur in m trials is defined as

$$B(k, m; p) = \sum_{r=0}^{k} \binom{m}{r} p^r q^{m-r}, \tag{1.245}$$

and its results comprise what is known as the *cumulative binomial distribution*.

The mean of the binomial distribution can be calculated as follows:

$$\begin{aligned}
\bar{k} = \mathcal{E}\{k\} &= \sum_{k=0}^{m} k \left(\frac{m!}{k!(m-k)!}\right) p^k q^{m-k} = \sum_{k=1}^{m} \left(\frac{m!}{(k-1)!(m-k)!}\right) p^k q^{m-k} \\
&= \sum_{l=0}^{m-1} \left(\frac{m!}{l!(m-1-l)!}\right) p^{l+1} q^{m-1-l} = mp \sum_{l=0}^{m-1} \left(\frac{(m-1)!}{l!(m-1-l)!}\right) p^l q^{m-1-l} \\
&= mp \sum_{l=0}^{n} \left(\frac{n!}{l!(n-l)!}\right) p^l q^{n-l} = mp.
\end{aligned} \tag{1.246}$$

The variance of the binomial distribution can be derived as follows:

$$\sigma^2 = \mathcal{E}\{k^2\} - [\mathcal{E}\{k\}]^2 = \mathcal{E}\{k^2\} - \mathcal{E}\{k\} + \mathcal{E}\{k\} - [\mathcal{E}\{k\}]^2$$
$$= \mathcal{E}\{k(k-1)\} + \mathcal{E}\{k\} - [\mathcal{E}\{k\}]^2. \tag{1.247}$$

Now,

$$\mathcal{E}\{k(k-1)\} = \sum_{k=0}^{m} k(k-1) \left(\frac{m!}{k!(m-k)!}\right) p^k q^{m-k} = \sum_{k=2}^{m} \left(\frac{m!}{(k-2)!(m-k)!}\right) p^k q^{m-k}$$
$$= \sum_{l=0}^{m-2} \left(\frac{m!}{l!(m-2-l)!}\right) p^{l+2} q^{m-2-l}$$
$$= m(m-1)p^2 \sum_{l=0}^{m-1} \left(\frac{(m-2)!}{l!(m-2-l)!}\right) p^l q^{m-2-l}$$
$$= m(m-1)p^2 \sum_{l=0}^{n} \left(\frac{n!}{l!(n-l)!}\right) p^l q^{n-l} = m(m-1)p^2. \tag{1.248}$$

Substituting Eqs. (1.246) and (1.248) into Eq. (1.247) yields

$$\sigma^2 = \mathcal{E}\{k(k-1)\} + \mathcal{E}\{k\} - [\mathcal{E}\{k\}]^2 = m(m-1)p^2 + mp - (mp)^2$$
$$= m^2p^2 - mp^2 + mp - m^2p^2 = mp - mp^2 = mp(1-p). \tag{1.249}$$

Example 1.17 The probability that a certain diode fails before operating 1000 h is 0.15. The probability that among 20 such diodes, at least 5 diodes will fail before reaching 1000 h is

$$\sum_{r=5}^{20} \binom{20}{r} (0.15)^m (0.85)^{m-r} = 1 - B(4, 20; 0.15) \approx 0.17.$$

1.6.13 Poisson Distribution

When the number of Bernoulli trial is large and the probability p gets small, the binomial distribution approximates another probability distribution, namely, the Poisson distribution. We show that as follows:

Recall from the last section that the mean of the binomial distribution is mp, which we now will call λ. This implies that

$$p = \frac{\lambda}{m}. \tag{1.250}$$

This implies that

$$\left(\frac{m!}{k!(m-k)!}\right) p^k q^{m-k} = \left(\frac{m!}{k!(m-k)!}\right) \left(\frac{\lambda}{m}\right)^k \left(1 - \frac{\lambda}{m}\right)^{m-k}$$

$$= \left(\frac{m(m-1)\cdots(m-k+1)}{k!}\right)\left(\frac{\lambda}{m}\right)^k\left(1-\frac{\lambda}{m}\right)^{m-k}$$
$$= \left(\frac{m}{m}\right)\left(\frac{m-1}{m}\right)\cdots\left(\frac{(m-k+1)}{m}\right)\frac{\lambda^k}{k!}\left(1-\frac{\lambda}{m}\right)^{m-k}$$
$$= \left(\frac{m}{m}\right)\left(\frac{m-1}{m}\right)\cdots\left(\frac{(m-k+1)}{m}\right)$$
$$\frac{\lambda^k}{k!}\left(1-\frac{\lambda}{m}\right)^m\left(1-\frac{\lambda}{m}\right)^{-k}. \tag{1.251}$$

Now if we let m get very large, then all but the last three factors go to one. Looking at the second to last factor, if m is getting large while keeping k and λ fixed, then

$$\left(1-\frac{\lambda}{m}\right)^m \approx \left(\sum_{i=0}^{\infty}\frac{1}{i!}\left(-\frac{\lambda}{m}\right)^i\right)^m = \left(e^{-\frac{\lambda}{m}}\right)^m = e^{-\lambda}. \tag{1.252}$$

The last factor goes to 1 as m is getting large. The result is that

$$\lim_{m\to\infty} P\{A_k\} = \frac{\lambda^k}{k!}e^{-\lambda}. \tag{1.253}$$

The function on the right side of Eq. (1.253) is called *Poisson distribution*. The Poisson distribution has a mean equal to $\lambda = mp$ and its variance also is its mean.

Example 1.18 Suppose that the expected number of electrons counted in $(0, \tau)$ is $m = 4.5$. Assume the electron count follows a Poisson distribution. What is the probability that exactly two electrons are counted? The probability in this case is given by

$$p(2, 4.5) = \frac{(4.5)^2}{2!}e^{-4.5} = 0.1125.$$

What is the probability that at least six electrons are counted? The probability in this case equals

$$P\{k \geq 6\} = 1 - \sum_{k=0}^{5} p(k, 4.5) = \sum_{k=0}^{5}\frac{(4.5)^k}{k!}e^{-4.5} = 0.2971.$$

1.6.14 Random Processes

In the previous sections, we dealt with the random variable $\boldsymbol{x}$ that is a real number defined by the outcome ϑ of a chance experiment, that is,

$$\boldsymbol{x} = \boldsymbol{x}(\vartheta). \tag{1.254}$$

A *random or stochastic process* is function of time determined by the outcome of that chance experiment. That is, we rewrite Eq. (1.254) as

$$\boldsymbol{x}(t) = \boldsymbol{x}(t, \vartheta), \tag{1.255}$$

which now represents a family of functions in time, ever changing for each outcome ϑ. Let us now define its properties.

The mean of the random process is written as

$$\overline{\boldsymbol{x}}(t) = \mathcal{E}\{\boldsymbol{x}(t)\}, \tag{1.256}$$

while its *autocorrelation* is given as

$$\boldsymbol{R}(t_1, t_2) = \mathcal{E}\{\boldsymbol{x}(t_1)\boldsymbol{x}(t_2)\}. \tag{1.257}$$

The autocovariance of the random process is defined as

$$\boldsymbol{\Gamma}(t_1, t_2) = \mathcal{E}\{[\boldsymbol{x}(t_1) - \overline{\boldsymbol{x}}(t_1)][\boldsymbol{x}(t_2) - \overline{\boldsymbol{x}}(t_2)]\} = \boldsymbol{R}(t_1, t_2) - \overline{\boldsymbol{x}}(t_1)\overline{\boldsymbol{x}}(t_2). \tag{1.258}$$

Looking at the above two equations, it is clear that $\boldsymbol{R}(t_1, t_2)$ is a joint noncentral moment as its mean of the process has not been removed, while $\boldsymbol{\Gamma}(t_1, t_2)$ is joint central moment because it had its mean removed. A zero-mean random process makes them equivalent.

When a random process has a mean that is time independent, its autocorrelation depends on the length of the time interval between t_1 and t_2, not their specific values. In other words, we write

$$\boldsymbol{R}(t_1, t_2) = \boldsymbol{R}(\tau), \tag{1.259}$$

where

$$\tau = t_2 - t_1. \tag{1.260}$$

This type of random process is called *stationary*. (Its counterpart where the process depends on the specific times used is called *nonstationary*.)

Let us now define the power spectrum of a stationary random process. The Fourier transform of the autocovariance function is equal to

$$\int_{-\infty}^{\infty} \boldsymbol{\Gamma}(\xi) e^{-i\omega\xi} d\xi = \int_{-\infty}^{\infty} \boldsymbol{R}(\xi) e^{-i\omega\xi} d\xi + \int_{-\infty}^{\infty} \overline{\boldsymbol{x}}^2 e^{-i\omega\xi} d\xi \tag{1.261a}$$

$$= \int_{-\infty}^{\infty} \boldsymbol{R}(\xi) e^{-i\omega\xi} d\xi + \overline{\boldsymbol{x}}^2 \delta(\omega), \tag{1.261b}$$

where ω denotes radial (temporal) frequency. The second term on the right of Eq. (1.261b) is the "dc" component of the power spectrum ($\omega = 0$). It has no bandwidth, but has a magnitude equal to the square of the mean. On the other hand, the first term on the right of Eq. (1.261b) refers to the AC terms and has bandwidth. We define this term as the *power spectrum* of the random process, that is,

$$\boldsymbol{S}(\omega) = \int_{-\infty}^{\infty} \boldsymbol{R}(\xi) e^{-i\omega\xi} d\xi. \tag{1.262}$$

In other words, it is the Fourier transform of the autocorrelation function. This result is the Wiener–Khinchin Theorem. The integral of $\boldsymbol{S}(\omega)$ is related to the variance, which can be seen as follows:

$$\int_{-\infty}^{\infty} \boldsymbol{S}(\omega) e^{i\omega\xi} d\omega \Big|_{\tau=0} = \int_{-\infty}^{\infty} \boldsymbol{S}(\omega) d\omega = \boldsymbol{R}(0) \tag{1.263a}$$

$$= \mathcal{E}\{(\boldsymbol{x}(t_1))^2\} = \boldsymbol{\sigma}_x^2 + \bar{\boldsymbol{x}}^2. \tag{1.263b}$$

For a zero-mean random process, it is the variance.

An important random process found in both radio frequency (RF) and optical systems is *white noise*. This is where the random process has both a zero mean and zero autocorrelation. Mathematically, this is characterized by the equation

$$\boldsymbol{R}(t_1, t_2) = \boldsymbol{R}(\tau) = \delta(\tau). \tag{1.264}$$

This means that $\boldsymbol{S}(\omega) = 1$, which implies that we have a constant power spectrum. The term "white noise" comes from the analogy with white light possessing all wavelengths (frequencies) of light.

1.7 DECIBELS

The decibel, or dB, for short, is one of the most widely used mathematical tools for those involved with system engineering. It is related to the common (base 10) logarithm and has the following mathematical form:

$$X, \text{ measure in dB} = 10 \log_{10} X \text{dB}. \tag{1.265}$$

The decibel, named after Alexander Graham Bell, was created to measure the ratio of two power levels; the most typical use is the ratio of power out of systems over its input power. The parameter X typically is the ratio of two power levels P_1 and P_2, or

$$\frac{\text{Gain}}{\text{Loss (dB)}} = 10 \log_{10} \left(\frac{P_2}{P_1} \right) \text{ dB}. \tag{1.266}$$

One of the key reasons that the decibel is used in engineering is that it can transform a multiplicative equation into an additive equation. However, one must be careful in its use since there are various units involved. Let us look at an example to show how to do it properly.

Consider the range equation we will define in Chapter 4:

$$P_{rx} \approx \gamma_{tx} P_{tx} \frac{A_{tx} A_{rx}}{(\lambda R)^2}, \tag{1.267}$$

where P_{rx} is the received power, γ_{tx} the transmitter optics transmittance, P_{tx} the laser transmitter power, λ the wavelength of light, R the distance between the transmitter

and receiver, and A_{tx} and A_{rx} the areas of the transmitter and receiver apertures, respectively. Applying Eq. (1.265) to Eq. (1.267) yields

$$10 \log_{10}(P_{rx}) = 10 \log_{10}(\gamma_{tx}) + 10 \log_{10}(P_{tx}) + 10 \log_{10}\left(\frac{A_{tx}A_{rx}}{(\lambda R)^2}\right). \quad (1.268)$$

Equation (1.268) has a unit problem. The elements for the two power levels have unit of watts, while everything else is unitless. To remedy this problem, we can subtract $10 \log_{10}(1\ \text{W})$ to each side of Eq. (1.268) without affecting the correctness of the equations and obtain the following new equation:

$$10 \log_{10}\left(\frac{P_{rx}}{1\ \text{W}}\right) = 10 \log_{10}(\gamma_{tx}) + 10 \log_{10}\left(\frac{P_{tx}}{1\,\text{W}}\right) + 10 \log_{10}\left(\frac{A_{tx}A_{rx}}{(\lambda R)^2}\right). \quad (1.269)$$

This corrects the problem. If we define

$$P'_{tx}(\text{dB W}) = 10 \log_{10}\left(\frac{P_{rx}}{1\ \text{W}}\right) \quad (1.270)$$

and

$$P'_{rx}(\text{dB W}) = 10 \log_{10}\left(\frac{P_{rx}}{1\ \text{W}}\right), \quad (1.271)$$

then Eq. (1.268) becomes

$$P'_{rx}(\text{dB W}) = 10 \log_{10}(\gamma_{tx})dB + P'_{tx}(\text{dB W}) + 10 \log_{10}\left(\frac{A_{tx}A_{rx}}{(\lambda R)^2}\right) \text{dB}. \quad (1.272)$$

If we had specified 1 mW rather than 1 W in the above derivation, Eq. (1.272) would be written as

$$P'_{rx}(\text{dB mW}) = 10 \log_{10}(\gamma_{tx}) \text{ dB} + P'_{tx}(\text{dB mW}) + 10 \log_{10}\left(\frac{A_{tx}A_{rx}}{(\lambda R)^2}\right) \text{dB}. \quad (1.273)$$

Clearly, we could write the last term in Eq. (1.272) as

$$\begin{aligned} 10 \log_{10}\left(\frac{A_{tx}A_{rx}}{(\lambda R)^2}\right) &= \text{dB}10 \log_{10}\left(\frac{A_{tx}}{1\ \text{m}}\right) \text{dB} \\ &\quad + 10 \log_{10}\left(\frac{A_{rx}}{1\ \text{m}}\right) \text{dB} - 20 \log_{10}(\lambda R)\,\text{dB} \\ &= A'_{tx}(\text{dB m}) + A'_{rx}(\text{dB m}) - 20 \log_{10}(\lambda R) \text{ dB} \quad (1.274) \end{aligned}$$

$$= A'_{tx}(\text{dB m}) + A'_{rx}(\text{dB m}) - 2\lambda'(\text{dB m}) - 2R'(\text{dB m}) \quad (1.275)$$

TABLE 1.2 Equivalent Power Ratios

Decibel Value	Positive dB Power Ratios	Negative dB Power Ratios
0	1	1
1	1.3	0.79
2	1.6	0.63
3	2.0	0.50
4	2.5	0.40
5	3.2	0.32
6	4.0	0.25
7	5.0	0.20
8	6.3	0.16
9	7.9	0.13

following the same procedure. This implies that

$$P'_{rx}(\text{dB W}) = 10\log_{10}(\gamma_{tx})\ \text{dB} + P'_{tx}(\text{dB W}) + = A'_{tx}(\text{dB m}) + A'_{rx}(\text{dB m}) - 2\lambda'(\text{dB m}) - 2R'(\text{dB m}). \tag{1.276}$$

Now, the reader seeing Eq. (1.276) for the first time may think we again have a unit problem, mixing watts and meters since they have learned that quantities that have different units should not add. However, logarithmic units are different and entities with different logarithmic or decibel units do add. The reason is that all dB m element units cancel, leaving only the unitless $10\log_{10}(\gamma_{tx})$ dB and (dB W) terms left, agreeing with the right side of the equation. This may be hard to get used to at first, but the benefit will pay off when dealing with complex equations that must be having several parameters varied in a trade study. In addition, adding and subtracting numbers is easier to do in one's head than multiplying or dividing numbers; at least for most people. This is an easy bookkeeping approach for engineers.

Some common increments in decibels should be memorized by the reader so they are proficient in using decibels in link budgets. A factor of 1 equals 0 dB, a factor of 10 equals 10 dB, a factor of 100 is 20 dB, and so on. However, the workhorse engineering number to remember is 3 dB. Its +3 dB value is 1.995, which basically is 2, and its −3 dB number is 0.5012, which essentially is 0.5. Therefore, if one increases the signal power by 2, we say we have a 3 dB increase in power; if we cut the signal power by half, we decrease the power by −3 dB. Table 1.2 provides a list of the real number equivalents for decibels between 1 and 9.

1.8 PROBLEMS

Problem 1.1. Let

$$\boldsymbol{A} = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 4 \end{bmatrix} \text{ and } \boldsymbol{B} = \begin{bmatrix} 2 & 3 & 0 \\ -1 & 2 & 5 \end{bmatrix}.$$

(a) what is $\boldsymbol{A} + \boldsymbol{B}$?
(b) What is $\boldsymbol{A} - \boldsymbol{B}$?

Problem 1.2. Compute the determinant

$$|\boldsymbol{A}| = \begin{vmatrix} 1 & 2 & 10 \\ 2 & 3 & 9 \\ 4 & 5 & 11 \end{vmatrix}.$$

Problem 1.3. Compute the determinant

$$|\boldsymbol{A}| = \begin{vmatrix} +2 & +3 & -2 & +4 \\ +3 & -2 & +1 & +2 \\ +3 & +2 & +3 & +4 \\ -2 & +4 & 0 & +5 \end{vmatrix}.$$

Problem 1.4. Compute the determinant

$$|\boldsymbol{A}| = \begin{vmatrix} +2 & +3 & -2 & +4 \\ +7 & +4 & -3 & +10 \\ +3 & +2 & +3 & +4 \\ -2 & +4 & 0 & +5 \end{vmatrix}.$$

Problem 1.5. Show that the cofactor of each element of

$$\boldsymbol{A} = \begin{bmatrix} -\frac{1}{3} & -\frac{2}{3} & -\frac{2}{3} \\ +\frac{2}{3} & +\frac{1}{3} & -\frac{2}{3} \\ +\frac{2}{3} & -\frac{2}{3} & +\frac{1}{3} \end{bmatrix}$$

is that element.

Problem 1.6. Show that the cofactor of an element of any row of

$$\boldsymbol{A} = \begin{bmatrix} -4 & -3 & -3 \\ +1 & 0 & +1 \\ +4 & +3 & +3 \end{bmatrix}$$

is the corresponding element of the same numbered column.

Problem 1.7. Find the inverse of

$$\boldsymbol{A} = \begin{vmatrix} 2 & 3 \\ 1 & 4 \end{vmatrix}.$$

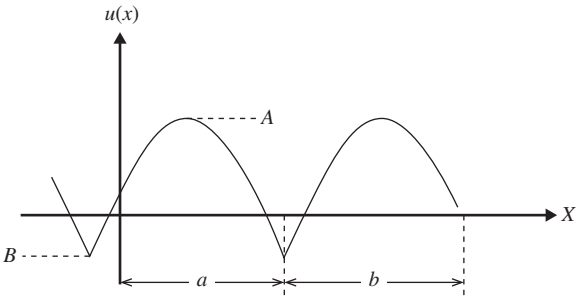
Problem 1.8. Find the inverse of

$$\boldsymbol{A} = \begin{vmatrix} 2 & 3 & 1 \\ 1 & 2 & 3 \\ 3 & 1 & 2 \end{vmatrix}.$$

Problem 1.9. Find the inverse of

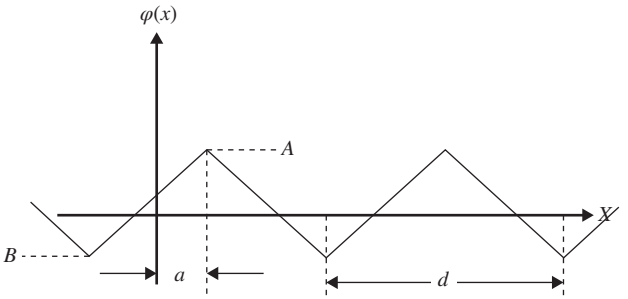
$$A = \begin{vmatrix} +2 & +4 & +3 & +2 \\ +3 & +6 & +5 & +2 \\ +2 & +5 & +2 & -3 \\ +4 & +5 & +14 & +14 \end{vmatrix}.$$

Problem 1.10. Calculate the Fourier series coefficients C_n for the following periodic function:

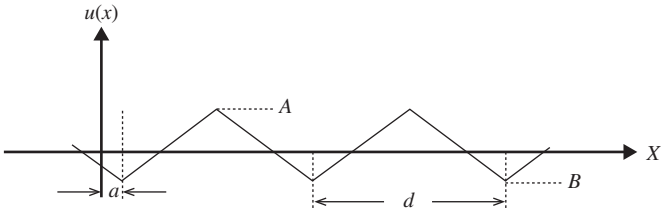


NOTE: The curve essentially follows the positive part of the sine function with period $2b$.

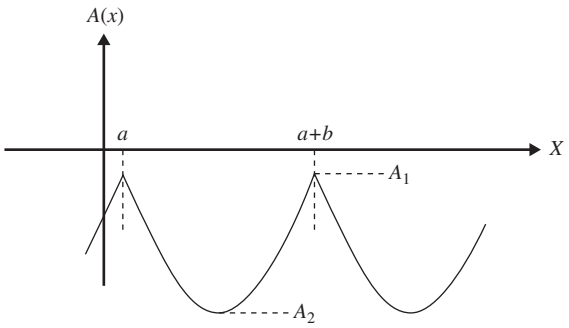
Problem 1.11. Calculate the Fourier series coefficients C_n for the function $u(x)$ plotted as follows, but do so exploiting some of the properties contained in Eqs. (1.4)–(1.24) to provide a solution derived from the coefficient calculation of a square wave. The slopes up and down are equally steep, and that B depicts a negative number in the graph.



Problem 1.12. Calculate the Fourier series coefficients C_n for the function $u(x)$ plotted as follows, but do so exploiting some of the properties contained in Eqs. (1.4)–(1.24) to provide a solution derived from the coefficient calculation of a square wave. The slopes up and down are equally steep, and that B depicts a negative number in the graph.



Problem 1.13. Calculate the Fourier series coefficients C_n for $u(x) = A(x)e^{i\varphi(x)}$. $\varphi(x) = \varphi_0 + 10\pi\frac{x}{b}$. The curve shown in the following figure is like the lower portion of the cosine function.



Problem 1.14. Let

$$f(x) = e^{-\left(\frac{x}{a}\right)^2}.$$

Prove that its Fourier transform is

$$\widehat{f}(\nu) = \sqrt{\pi}a\, e^{-(\pi\nu a)^2}.$$

Problem 1.15. Let

$$f(x) = \text{rect}\left(\frac{x}{a}\right) = \begin{cases} 1 & \text{if } |x| \leq \frac{a}{2} \\ 0 & \text{otherwise} \end{cases}.$$

Prove that its Fourier transform is

$$\widehat{f}(\nu) = \frac{\sin(a\nu)}{a\nu} \equiv \text{sinc}(a\nu).$$

Problem 1.16. Let

$$f(x, y) = \text{rect}\left(\frac{x}{a}\right)\text{rect}\left(\frac{y}{b}\right).$$

What is its two-dimensional Fourier transform?

Problem 1.17. Let

$$f(x,y) = \text{Circ}(r) \equiv \text{circ function}$$
$$= \begin{cases} 1 & |r| = \sqrt{x^2+y^2} \le 1 \\ 0 & \text{otherwise} \end{cases}.$$

What is its two-dimensional Fourier transform?

Problem 1.18. Assume a card selected out of an ordinary deck of 52 cards. Let

$$\boldsymbol{A} = \{\text{the card is a spade}\}$$

and

$$\boldsymbol{B} = \{\text{the card is a face card, that is, jack, queen, or king}\}.$$

Compute $P\{\boldsymbol{A}\}$, $P\{\boldsymbol{B}\}$, and $P\{\boldsymbol{A} \cap \boldsymbol{B}\}$.

Problem 1.19. Let two items be chosen out of a lot of 12 items where 4 of them are defective. Assume

$$\boldsymbol{A} = \{\text{both chosen items are defective}\}$$

and

$$\boldsymbol{B} = \{\text{both chosen items are not defective}\}.$$

Compute $P\{\boldsymbol{A}\}$ and $P\{\boldsymbol{B}\}$.

Problem 1.20. Given the problem laid out in Problem 1.19. Assume now that

$$\boldsymbol{C} = \{\text{At least one chosen item is defective}\}.$$

What is the probability that event $\boldsymbol{C}$ occurs?

Problem 1.21. Let a pair of fair dice be tossed. If the sum is 6, what is the probability that one of the dice is a 2? In other words, we have

$$\boldsymbol{A} = \{\text{sum is 6}\}$$

and

$$\boldsymbol{B} = \{\text{a 2 appears on at least one die}\}.$$

Find $P\{\boldsymbol{B}|\boldsymbol{A}\}$.

Problem 1.22. In a certain college, 25% of the students fail in mathematics, 15% of the students fail in chemistry, and 10% of the students fail both in mathematics and chemistry. A student is selected at random.

(a) If the student failed in chemistry, what is the probability that the student also failed in mathematics?

(b) If he failed in mathematics, what is the probability that he failed in chemistry too?

(c) What is the probability that he failed both in mathematics and chemistry?

Problem 1.23. Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be events with $P\{\boldsymbol{A}\} = 1/2$, $P\{\boldsymbol{B}\} = 1/3$, and $P\{\boldsymbol{A} \cap \boldsymbol{B}\} = 1/4$. Find (a) $P\{\boldsymbol{B}|\boldsymbol{A}\}$, (b) $P\{\boldsymbol{A}|\boldsymbol{B}\}$, and (c) $P\{\boldsymbol{A} \cup \boldsymbol{B}\}$, $P\{\boldsymbol{A}_c|\boldsymbol{B}_c\}$, and $P\{\boldsymbol{B}_c|\boldsymbol{A}_c\}$. Here, $\boldsymbol{A}_c$ and $\boldsymbol{B}_c$ are the complements of $\boldsymbol{A}$ and $\boldsymbol{B}$, respectively.

Problem 1.24. A lot contains 12 items of which 4 are defective. Three items are drawn at random from that lot one after another. Find the probability that all three are nondefective.

Problem 1.25. A card player is dealt 5 cards one right after another from an ordinary deck of 52 cards. What is the probability that they are all spades?

Problem 1.26. Let $\varphi(t)$ be the standard normal distribution (i.e., mean equals zero and variance equals to unity)? Find $\varphi(t)$ for (a) $t = 1.63$, (b) $t = -0.75$ and $t = -2.08$.

HINT: You may need a standard normal distribution table to solve this problem.

Problem 1.27. Let $\boldsymbol{x}$ be a random variable with a standard normal distribution $\varphi(t)$. Find

(a)
$$P\{0 \leq \boldsymbol{x} \leq 1.42\}.$$

(b)
$$P\{-0.73 \leq \boldsymbol{x} \leq 0\}.$$

(c)
$$P\{-1.37 \leq \boldsymbol{x} \leq 2.01\}.$$

(d)
$$P\{0.65 \leq \boldsymbol{x} \leq 1.26\}.$$

(e)
$$P\{-1.79 \leq \boldsymbol{x} \leq -0.54\}.$$

(f)
$$P\{\boldsymbol{x} \geq 1.13\}.$$

(g)
$$P\{|\boldsymbol{x}| \leq 0.5\}.$$

HINT: You may need a standard normal distribution table to solve this problem.

Problem 1.28. A fair die is tossed seven times. Let us assume that success occurs if a 5 or 6 appear. Let $n = 7$, $p = P\{5, 6\} = \frac{1}{3}$, and $q = 1 - p = \frac{2}{3}$.

(a) What is the probability that a 5 or a 6 occurs exactly three times (i.e., $k = 3$)?

(b) What is the probability that a 5 or a 6 occurs at least once?

Problem 1.29. A fair coin is tossed six times. Let us assume that success is a heads. Let $n = 6$ and $p = q = \frac{1}{2}$.

(a) What is the probability that exactly two heads occur (i.e., $k = 2$)?

(b) What is the probability of getting at least four heads (i.e., $k = 4,\ 5,$ and 6)?

(c) What is the probability that at least one head occurs?

Problem 1.30. For a Poisson distribution

$$p(k, \lambda) = \frac{\lambda^k}{k!} e^{-\lambda},$$

find (a) $p(2, 1)$, (b) $p\left(3, \frac{1}{2}\right)$, and (c) $p(2, 7)$.

Problem 1.31. Suppose 300 misprints are randomly distributed throughout a book of 500 pages. Find the probability that a given page contains (a) exactly 2 misprints, (b) 2 or more misprints.

HINT: You may want to consider the number of misprints on one page as the number of successes in a sequence of Bernoulli trials. Note that we are dealing with large numbers.

Problem 1.32. Suppose 2% of the items made by a factory are defective. Find the probability that there are 3 defective items in a sample of 100 items.

Problem 1.33. Given

$$X(\text{dB}) = 10 \log_{10} X,$$

derive an equation for X in terms of $X(\text{dB})$.

Problem 1.34. Given

$$X(\text{dB}) = 10 \log_{10} X,$$

find $X(\text{dB})$ for (a) $X = 63^2$, (b) $X = 4000$, and (c) $X = \frac{1}{2500}$.

REFERENCES

1. Pratt, W.K. (2007) *Digital Image Processing*, 4th edn, John Wiley and Sons, New York, pp. 121–123.
2. Bar-Shalom, Y. and Fortmann, T.E. (1988) *Tracking and Data Association*, Academic Press, New York, 353 pages.
3. Lohmann, A.W. (2006) in *Optical Information Processing* (ed. S. Sinzinger), Universitätsverlag, Ilmenau, Germany, ISBN 4-939474-00-6

4. Goodman, J.W. (2004) *Introduction to Fourier Optics*, 3rd edn, Roberts and Company, Englewood, CO.
5. Papoulis, A. (1968) *Systems and Transforms With Applications in Optics*, McGraw-Hill Series in Systems Science, Edition reprint, Robert Krieger Publishing Company, Malabar, FL, 474 pages, ISBN 0898743583, 9780898743586
6. Helstrom, C.W. (1991) *Probability and Stochastic Processes for Engineers*, 2nd edn, Macmillan Publishing Company, New York, 610 pages.
7. Venkatesh, S.S. (2013) *The Theory of Probability: Exploration and Applications*, Cambridge University Press, Cambridge, UK, 805 pages.
8. Fry, W. (1928) *Probability and Its Engineering Uses*, Von Nostrand, New York.