Appendix A

Review of Important Concepts

A.1 Matrix Fundamentals

A.1.1 Definitions

A matrix $\mathbf{A}$ is a rectangular array whose elements a_{ij} are arranged in m rows and n columns, referred to as an $m \times n$ matrix,

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}. \tag{A.1}$$

The matrix is said to be *square* when $m = n$, a *row matrix* when $m = 1$, and a *column matrix* when $n = 1$. The column matrix has particular significance and is referred to as a *vector*.

It may be useful to *partition* a matrix into smaller *submatrices*. For example, an $m \times n$ matrix may be written as

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}, \tag{A.2}$$

where $\mathbf{A}_{11}$, $\mathbf{A}_{12}$, $\mathbf{A}_{21}$, and $\mathbf{A}_{22}$ are $m_1 \times n_1$, $m_1 \times n_2$, $m_2 \times n_1$, and $m_2 \times n_2$ matrices, respectively, such that $m_1 + m_2 = m$ and $n_1 + n_2 = n$. By denoting the i^{th} column of $\mathbf{A}$ as $\mathbf{a}_i$,

$$\mathbf{a}_i = \begin{bmatrix} a_{i1} \\ a_{i2} \\ \vdots \\ a_{in} \end{bmatrix}, \tag{A.3}$$

the matrix $\mathbf{A}$ can be written as a partitioned matrix

$$\mathbf{A} = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \end{bmatrix}. \tag{A.4}$$

The *transpose* of a matrix $\mathbf{A}$ is another matrix $\mathbf{A}^T$ whose rows and columns have been interchanged, and thus, if $\mathbf{A}$ is $m \times n$ then $\mathbf{A}^T$ is $n \times m$. The *Hermitian* transpose of a complex matrix $\mathbf{A}$ is another matrix $\mathbf{A}^H$ which has been transposed and with all elements complex-conjugated. The *inverse* of a square matrix $\mathbf{A}$ is a square matrix $\mathbf{A}^{-1}$ for which

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}, \tag{A.5}$$

where $\mathbf{I}$ is the *identity matrix* with ones in the principal diagonal and elsewhere zeros,

$$\mathbf{I} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}. \tag{A.6}$$

The *rank* of a matrix is defined as the number of linearly independent rows or columns, whichever is less.

The *reversal* of a vector

$$\mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} \tag{A.7}$$

is defined by

$$\tilde{\mathbf{a}} = \begin{bmatrix} a_n \\ a_{n-1} \\ \vdots \\ a_1 \end{bmatrix}, \tag{A.8}$$

where the tilde denotes the reversal operation. For the matrix $\mathbf{A}$ in (A.1), its reversal is given by

$$\tilde{\mathbf{A}} = \begin{bmatrix} a_{mn} & a_{m(n-1)} & \cdots & a_{m1} \\ a_{(m-1)n} & a_{(m-1)(n-1)} & \cdots & a_{(m-1)2} \\ \vdots & \vdots & & \vdots \\ a_{1n} & a_{1(n-1)} & \cdots & a_{11} \end{bmatrix}, \tag{A.9}$$

that is, the matrix $\mathbf{A}$ is reversed about both its vertical axis and its horizontal axis.

The inverse of a product of two matrices $\mathbf{A}$ and $\mathbf{B}$ is the product of the inverse matrices in reversed order,

$$(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}. \tag{A.10}$$

Similarly, the transpose of the product of two matrices is given by

$$(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T. \tag{A.11}$$

A matrix is said to be *symmetric* if

$$\mathbf{A}^T = \mathbf{A}, \tag{A.12}$$

and *orthogonal* if

$$\mathbf{A}^T\mathbf{A} = \mathbf{I}, \tag{A.13}$$

or, equivalently,

$$\mathbf{A}^T = \mathbf{A}^{-1}. \tag{A.14}$$

A square matrix is called *diagonal* if all elements off the principal diagonal are zero,

$$\operatorname{diag}(a_{11}, a_{22}, \ldots, a_{nn}) = \begin{bmatrix} a_{11} & 0 & 0 & \cdots & 0 \\ 0 & a_{22} & 0 & \cdots & 0 \\ 0 & 0 & a_{33} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & a_{nn} \end{bmatrix}. \tag{A.15}$$

A *lower triangular matrix* has all of its elements above the principal diagonal equal to zero,

$$\begin{bmatrix} a_{11} & 0 & 0 & \cdots & 0 \\ a_{21} & a_{22} & 0 & \cdots & 0 \\ a_{31} & a_{32} & a_{33} & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nn} \end{bmatrix}. \tag{A.16}$$

An *upper triangular matrix* is defined as the transpose of a lower triangular matrix. A matrix is said to be *Toeplitz* if all of its elements along each

diagonal have the same value,

$$\begin{bmatrix} a_0 & a_1 & a_2 & \cdots & a_{n-2} & a_{n-1} \\ a_{-1} & a_0 & a_1 & \cdots & a_{n-3} & a_{n-2} \\ a_{-2} & a_{-1} & a_0 & \cdots & a_{n-4} & a_{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{-n+2} & a_{-n+3} & a_{-n+4} & \cdots & a_0 & a_1 \\ a_{-n+1} & a_{-n+2} & a_{-n+3} & \cdots & a_{-1} & a_0 \end{bmatrix}. \tag{A.17}$$

The matrix is *symmetric Toeplitz* if $a_{-k} = a_k$.

The product of a transposed vector and another vector of the same dimension is a scalar a and is referred to as the *inner product* or the *scalar product*

$$a = \mathbf{x}^T\mathbf{y} = \mathbf{y}^T\mathbf{x}. \tag{A.18}$$

The vectors are orthogonal if their scalar product is equal to zero,

$$\mathbf{x}^T\mathbf{y} = 0. \tag{A.19}$$

The product of a vector and a transposed vector is a matrix $\mathbf{A}$ and is referred to as the *outer product*

$$\mathbf{A} = \mathbf{x}\mathbf{y}^T. \tag{A.20}$$

Note that $\mathbf{x}\mathbf{y}^T \neq \mathbf{y}\mathbf{x}^T$ in general.

For a symmetric $n \times n$ matrix $\mathbf{A}$, a *quadratic form* Q is defined as

$$Q = \mathbf{x}^T\mathbf{A}\mathbf{x} = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_i x_j. \tag{A.21}$$

A symmetric matrix $\mathbf{A}$ is said to be *positive definite* if

$$\mathbf{x}^T\mathbf{A}\mathbf{x} > 0 \tag{A.22}$$

for all $\mathbf{x} \neq \mathbf{0}$; the matrix $\mathbf{A}$ is *positive semidefinite* if the quadratic form Q is greater or equal to zero.

The sum of the diagonal elements of a square matrix $\mathbf{A}$ is called the *trace* of $\mathbf{A}$ and is

$$\mathrm{tr}(\mathbf{A}) = \sum_{i=1}^{n} a_{ii}. \tag{A.23}$$

The trace of the sum of matrices is equal to the sum of the traces

$$\mathrm{tr}(\mathbf{A} + \mathbf{B}) = \mathrm{tr}(\mathbf{A}) + \mathrm{tr}(\mathbf{B}), \tag{A.24}$$

and the trace of a matrix product is

$$\mathrm{tr}(\mathbf{AB}) = \mathrm{tr}(\mathbf{BA}). \tag{A.25}$$

The determinant for an $n \times n$ matrix $\mathbf{A}$ multiplied with a scalar a is

$$\det(a\mathbf{B}) = a^n \det(\mathbf{A}), \tag{A.26}$$

and for a matrix product it is

$$\det(\mathbf{AB}) = \det(\mathbf{BA}). \tag{A.27}$$

The determinant of a diagonal matrix $\mathbf{A}$ is

$$\det(\mathbf{A}) = \prod_{i=1}^{n} a_{ii}. \tag{A.28}$$

The determinant of an inverse matrix is

$$\det(\mathbf{A}^{-1}) = (\det(\mathbf{A}))^{-1}. \tag{A.29}$$

A useful relationship for the trace of the outer product of the vectors $\mathbf{x}$ and $\mathbf{y}$ is

$$\mathrm{tr}(\mathbf{x}\mathbf{y}^T) = \mathbf{y}^T\mathbf{x}. \tag{A.30}$$

The *matrix inversion lemma* states that

$$(\mathbf{A} + \mathbf{BCD})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}(\mathbf{DA}^{-1}\mathbf{B} + \mathbf{C}^{-1})^{-1}\mathbf{DA}^{-1}, \tag{A.31}$$

where $\mathbf{A}$ is $n \times n$, $\mathbf{B}$ is $n \times m$, $\mathbf{C}$ is $m \times m$, and $\mathbf{D}$ is $m \times n$, with the assumption that the included inverse matrices exist.

The norm of a vector is a number that characterizes the magnitude of the vector. The *Euclidean norm* is one such measure which is defined by

$$\|\mathbf{x}\|_2 = \left(\sum_{i=1}^{n} |x_i|^2\right)^{\frac{1}{2}} = \left(\mathbf{x}^T\mathbf{x}\right)^{\frac{1}{2}}. \tag{A.32}$$

Similarly, the norm of a matrix is a number that characterizes the magnitude of the matrix. Several matrix norm definitions exist, of which one of the most used is the *Frobenius norm*,

$$\|\mathbf{A}\|_F = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} |a_{ij}|^2} = \left(\mathrm{tr}\left(\mathbf{AA}^T\right)\right)^{1/2}. \tag{A.33}$$

A.1.2 Matrix Decomposition

A square $n \times n$ matrix $\mathbf{A}$ has an *eigenvector* $\mathbf{v}$ which satisfies

$$\mathbf{A}\mathbf{v} = \lambda\mathbf{v} \tag{A.34}$$

for some scalar λ, also referred to as the *eigenvalue* corresponding $\mathbf{v}$. The eigenvectors are normalized to have unit length, i.e., $\mathbf{v}^T\mathbf{v} = 1$. The n eigenvalues of (A.34) are obtained as the roots of the *characteristic equation*

$$\det(\mathbf{A} - \lambda\mathbf{I}) = \mathbf{0}.$$

With these eigenvalues, the corresponding eigenvectors can be determined from

$$(\mathbf{A} - \lambda\mathbf{I})\mathbf{v} = \mathbf{0}.$$

When $\mathbf{A}$ is symmetric and positive definite (semidefinite), all eigenvalues are real-valued and positive (non-negative). The corresponding eigenvectors are orthonormal,

$$\mathbf{v}_i^T\mathbf{v}_j = \begin{cases} 1, & i = j; \\ 0, & i \neq j. \end{cases} \tag{A.35}$$

The relation in (A.34) can be expressed to include all n eigenvalues and eigenvectors

$$\mathbf{A}\begin{bmatrix}\mathbf{v}_1 & \mathbf{v}_1 & \cdots & \mathbf{v}_n\end{bmatrix} = \begin{bmatrix}\lambda_1\mathbf{v}_1 & \lambda_2\mathbf{v}_2 & \cdots & \lambda_n\mathbf{v}_n\end{bmatrix},$$

or

$$\mathbf{A}\mathbf{V} = \mathbf{V}\Lambda, \tag{A.36}$$

where $\mathbf{V} = \begin{bmatrix}\mathbf{v}_1 & \mathbf{v}_1 & \cdots & \mathbf{v}_n\end{bmatrix}$ and $\Lambda = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$. Since the columns of $\mathbf{V}$ are orthonormal, i.e., $\mathbf{V}^T = \mathbf{V}^{-1}$, the matrix $\mathbf{A}$ can be decomposed into a weighted sum of rank-one matrices $\mathbf{v}_i\mathbf{v}_i^T$,

$$\mathbf{A} = \mathbf{V}\Lambda\mathbf{V}^T = \sum_{i=1}^{n} \lambda_i\mathbf{v}_i\mathbf{v}_i^T. \tag{A.37}$$

Thus, for a matrix defined by an outer product, i.e., $\mathbf{A} = \mathbf{a}\mathbf{a}^T$, it is evident from (A.37) that only one eigenvalue can be nonzero. The corresponding eigenvector $\mathbf{v}_1$ is, apart from a normalization factor, identical to the vector $\mathbf{a}$. The remaining eigenvectors must be selected such that these are orthogonal to $\mathbf{v}_1$.

Based on the expansion in (A.37), the determinant and the trace of a symmetric matrix $\mathbf{A}$ can be related to its eigenvalues,

$$\begin{aligned}\det(\mathbf{A}) &= \det(\mathbf{V}\Lambda\mathbf{V}^T) = \det(\mathbf{V}^T\mathbf{V}\Lambda) \\ &= \prod_{i=1}^{n} \lambda_i,\end{aligned} \tag{A.38}$$

and

$$\begin{aligned}\mathrm{tr}(\mathbf{A}) &= \mathrm{tr}(\mathbf{V}\Lambda\mathbf{V}^T) = \mathrm{tr}(\mathbf{V}^T\mathbf{V}\Lambda) \\ &= \sum_{i=1}^{n} \lambda_i,\end{aligned} \tag{A.39}$$

respectively.

Another, more general type of matrix decomposition is the *singular value decomposition* (SVD) by which an $m \times n$ matrix $\mathbf{A}$ can be decomposed into two orthogonal matrices, an $m \times m$ matrix $\mathbf{U}$ and an $n \times n$ matrix $\mathbf{V}$ such that

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T, \tag{A.40}$$

where $\mathbf{\Sigma}$ is an $m \times n$ non-negative diagonal matrix defined by

$$\mathbf{\Sigma} = \begin{bmatrix} \mathbf{S} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}, \tag{A.41}$$

and

$$\mathbf{S} = \mathrm{diag}(\sigma_1, \sigma_2, \ldots, \sigma_l), \tag{A.42}$$

where $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_l > 0$ and $\sigma_{l+1} = \cdots = \sigma_p = 0$ denote the singular values of $\mathbf{A}$; the number of nonzero singular values l does not exceed $p = \min(m, n)$. Similar to eigendecomposition of a symmetric matrix in (A.37), the matrix $\mathbf{A}$ can be decomposed into a weighted sum of rank-one matrices $\mathbf{u}_i\mathbf{v}_i^T$,

$$\mathbf{A} = \sum_{i=1}^{l} \sigma_i \mathbf{u}_i \mathbf{v}_i^T. \tag{A.43}$$

The nonzero singular values of $\mathbf{A}$ are equal to the square-root of the positive eigenvalues of the matrices $\mathbf{A}^T\mathbf{A}$ and $\mathbf{A}\mathbf{A}^T$.

A.1.3 Matrix Optimization

When optimizing matrix equations the following vector differentiation rules are useful:

$$\nabla_{\mathbf{x}}(\mathbf{x}^T\mathbf{y}) = \mathbf{y}, \tag{A.44}$$

and

$$\nabla_{\mathbf{x}}(\mathbf{y}^T\mathbf{x}) = \mathbf{y}, \tag{A.45}$$

where the gradient $\nabla_{\mathbf{x}}$ of a vector function $f(\mathbf{x})$ is defined as

$$\nabla_{\mathbf{x}} f(\mathbf{x}) \stackrel{\text{def}}{=} \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix}, \tag{A.46}$$

where $x_1, \ldots, x_n$ are the elements of $\mathbf{x}$. The gradient of a quadratic form is given by

$$\nabla_{\mathbf{x}}(\mathbf{x}^T\mathbf{A}\mathbf{x}) = 2\mathbf{A}\mathbf{x}, \tag{A.47}$$

where $\mathbf{A}$ is symmetric.

It may be necessary to optimize a vector function $f(\mathbf{x})$ subject to a number of constraints. For a set of L different constraints $g_l(\mathbf{x}) = 0$ with $l = 1, \ldots, L$, optimization with respect to $\mathbf{x}$ can be done in the following way.

1. First, define the *Lagrangian* function $\mathcal{L}(\mathbf{x}, \mu_1, \ldots, \mu_L)$,

$$\mathcal{L}(\mathbf{x}, \mu_1, \ldots, \mu_L) = f(\mathbf{x}) + \sum_{l=1}^{L} \mu_l g_l(\mathbf{x}), \tag{A.48}$$

   where $\mu_1, \ldots, \mu_l$ are called Lagrange multipliers.

2. Then, solve the equation system

$$\begin{cases} \nabla_{\mathbf{x}}\mathcal{L}(\mathbf{x}, \mu_1, \ldots, \mu_L) = \nabla_{\mathbf{x}}\left(f(\mathbf{x}) + \sum_{l=1}^{L} \mu_l g_l(\mathbf{x})\right) = \mathbf{0}, \\ \dfrac{\partial \mathcal{L}(\mathbf{x}, \mu_1, \ldots, \mu_L)}{\partial \mu_1} = g_1(\mathbf{x}) = 0, \\ \vdots \\ \dfrac{\partial \mathcal{L}(\mathbf{x}, \mu_1, \ldots, \mu_L)}{\partial \mu_L} = g_L(\mathbf{x}) = 0, \end{cases} \tag{A.49}$$

where the solutions to the L different equations

$$\frac{\partial \mathcal{L}(\mathbf{x}, \mu_1, \ldots, \mu_L)}{\partial \mu_l} = 0$$

are the specified constraints.

A.1.4 Linear Equations

The solution to a set of n linear equations in the m unknowns x_i, $i = 1, \ldots, m$,

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1m}x_m &= b_1, \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2m}x_m &= b_2, \\ &\vdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nm}x_m &= b_n, \end{aligned}$$

or, equivalently,

$$\mathbf{A}\mathbf{x} = \mathbf{b}, \tag{A.50}$$

depend on various factors such as the relative size of m and n and the rank of $\mathbf{A}$. When $\mathbf{A}$ is a square matrix, i.e., $m = n$, the solution is defined by

$$\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}, \tag{A.51}$$

provided that the inverse matrix $\mathbf{A}^{-1}$ exists.

In many signal processing problems, the number of unknowns x_i is less than the number of linear equations, i.e., $m < n$, and the solution is said to be overdetermined. In such cases, the *least-squares solution* is usually considered and results from minimizing the error norm

$$\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 \tag{A.52}$$

with respect to $\mathbf{x}$. When the matrix $\mathbf{A}$ has full rank, the least-squares solution is [1–3]

$$\mathbf{x} = \left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{A}^T\mathbf{b}, \tag{A.53}$$

where the matrix $\left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{A}^T$ is referred to as the pseudo-inverse of $\mathbf{A}$ for the overdetermined problem.

A.2 Discrete-Time Stochastic Processes

A stochastic process represents an ensemble of possible realizations of a process. Each realization of the stochastic process is called a time series, e.g., $x(0), \ldots, x(N-1)$, which in vector form can be written

$$\mathbf{x} = \begin{bmatrix} x(0) \\ x(1) \\ \vdots \\ x(N-1) \end{bmatrix}. \tag{A.54}$$

It is assumed that $x(n)$ is real-valued, stochastic process.

A.2.1 Definitions

A stochastic process can be described using expected values or ensemble averages which are averages over all realizations. The mean or expected value of a function $f(x)$ of a random variable X, characterized by its PDF $p_X(x)$, is defined as

$$E[f(x)] = \int_{-\infty}^{\infty} f(x) p_X(x) dx. \tag{A.55}$$

The mean value function of a stochastic process is defined by

$$m_x(n) = E[x(n)] \tag{A.56}$$

and contains the averages of all possible outcomes for each individual sample. In vector representation (A.56) can be written

$$\mathbf{m}_x = E[\mathbf{x}]. \tag{A.57}$$

The variance function contains for each sample the ensemble average of the squared deviation from the mean value for that sample. The variance function is given by

$$\sigma_x^2(n) = E\left[|x(n) - m_x(n)|^2\right], \tag{A.58}$$

or in vector form

$$\sigma_x^2 = E\left[|\mathbf{x} - \mathbf{m}_x|^2\right]. \tag{A.59}$$

The covariance function (also called the autocovariance function) describes the average joint deviation from the mean value for two samples n_1 and n_2 and is defined by

$$c_x(n_1, n_2) = E[(x(n_1) - m_x(n_1))(x(n_2) - m_x(n_2))]. \tag{A.60}$$

A positive covariance value indicates that the deviations from the mean value for these two samples, in average, have the same sign, while a negative value indicates that the deviations tend to have opposite sign. The covariance matrix $\mathbf{C}_x$ is defined by

$$\begin{aligned}\mathbf{C}_x &= E\left[(\mathbf{x}-\mathbf{m}_x)(\mathbf{x}-\mathbf{m}_x)^T\right] \\ &= \begin{bmatrix} c_x(0,0) & c_x(0,1) & \cdots & c_x(0,N-1) \\ c_x(1,0) & c_x(1,1) & \cdots & c_x(1,N-1) \\ \vdots & \vdots & & \vdots \\ c_x(N-1,0) & c_x(N-1,1) & \cdots & c_x(N-1,N-1) \end{bmatrix},\end{aligned} \tag{A.61}$$

which is symmetric. The correlation function (also called the autocorrelation function) is defined by

$$r_x(n_1,n_2) = E[x(n_1)x(n_2)], \tag{A.62}$$

which has an interpretation similar to that of the covariance function, although it does not reflect deviations from the mean value. The correlation matrix is defined by

$$\begin{aligned}\mathbf{R}_x &= E\left[\mathbf{x}\mathbf{x}^T\right] \\ &= \begin{bmatrix} r_x(0,0) & r_x(0,1) & \cdots & r_x(0,N-1) \\ r_x(1,0) & r_x(1,1) & \cdots & r_x(1,N-1) \\ \vdots & \vdots & & \vdots \\ r_x(N-1,0) & r_x(N-1,1) & \cdots & r_x(N-1,N-1) \end{bmatrix}.\end{aligned} \tag{A.63}$$

A close relation exists between the covariance and the correlation matrices since

$$\mathbf{C}_x = \mathbf{R}_x - \mathbf{m}_x\mathbf{m}_x^T, \tag{A.64}$$

and thus a zero-mean process has identical covariance and correlation matrices.

A cross-correlation function can be defined which describes the correlation properties between two different stochastic processes $x(n)$ and $y(n)$

$$r_{xy}(n_1,n_2) = E[x(n_1)y(n_2)]. \tag{A.65}$$

The corresponding cross-correlation matrix is defined by

$$\mathbf{R}_{xy} = E\left[\mathbf{x}\mathbf{y}^T\right]. \tag{A.66}$$

A.2.2 Stationarity

In the previous subsection the stochastic process was characterized by its first and second moments. It should be noted, however, that the first two moments do not provide a complete statistical description of a stochastic process. On the other hand, it is seldom possible to exactly determine the probability density function, and the process may be reasonably well-described by its first two moments.

A stochastic process is said to be strictly stationary if all of its moments are time-invariant. A less strict assumption is that its first two moments are time-invariant, and the process is called *wide-sense stationary.* A process is wide-sense stationary if the mean value function is a constant,

$$m_x(n) = m_x, \tag{A.67}$$

and the covariance and correlation functions depend only on the lag between the two samples,

$$c_x(n, n-k) = c_x(k) \tag{A.68}$$

and

$$r_x(n, n-k) = r_x(k). \tag{A.69}$$

The correlation function of a wide-sense stationary process is symmetric,

$$r_x(k) = r_x(-k). \tag{A.70}$$

For a zero lag, the correlation function is non-negative and equals the mean-square value of the process,

$$r_x(0) = E\left[|x(n)|^2\right] \geq 0. \tag{A.71}$$

Furthermore, the correlation function is bounded by the mean-square value,

$$|r_x(k)| \leq r_x(0). \tag{A.72}$$

The correlation matrix $\mathbf{R}_x$ of a wide-sense stationary process $x(n)$ is symmetric and Toeplitz,

$$\mathbf{R}_x = E[\mathbf{x}\mathbf{x}^T] = \begin{bmatrix} r_x(0) & r_x(-1) & \cdots & r_x(-N+1) \\ r_x(1) & r_x(0) & \cdots & r_x(-N+2) \\ \vdots & \vdots & & \vdots \\ r_x(N-1) & r_x(N-2) & \cdots & r_x(0) \end{bmatrix}, \tag{A.73}$$

and is positive semidefinite with non-negative, real-valued eigenvalues.

A.2.3 Ergodicity

Often different realizations of a signal are unavailable, and the mean value and the correlation function can be estimated using time averages instead of ensemble averages. If a large number of samples of a wide-sense stationary process are available, the mean value of a stochastic process can be estimated by using

$$\hat{m}_x(N) = \frac{1}{N} \sum_{n=0}^{N-1} x(n). \tag{A.74}$$

If

$$\lim_{N \to \infty} E\left[|\hat{m}_x(N) - m_x|^2\right] = 0, \tag{A.75}$$

the process $x(n)$ is said to be *ergodic in the mean.* Similarly, the correlation function can be estimated by

$$\hat{r}_x(k, N) = \frac{1}{N} \sum_{n=k}^{N-1} x(n)x(n-k). \tag{A.76}$$

If

$$\lim_{N \to \infty} E\left[|\hat{r}_x(k, N) - r_x(k)|^2\right] = 0, \tag{A.77}$$

the process is said to be *correlation ergodic.*

A.2.4 Bias and Consistency

An estimate $\hat{\theta}_N$ of the unknown parameter θ is referred to as being *unbiased* if

$$E\left[\hat{\theta}_N\right] = \theta, \tag{A.78}$$

or, otherwise, the difference

$$b(\theta) = E\left[\hat{\theta}_N\right] - \theta$$

is referred to as the *bias*; the parameter N denotes the number of observations used to compute an estimate of θ. The estimate is *asymptotically unbiased* if the bias approaches zero for an increasing number of observations,

$$\lim_{N \to \infty} E\left[\hat{\theta}_N\right] = \theta. \tag{A.79}$$

The estimate is said to be *consistent* if it is asymptotically unbiased and has a variance that approaches zero as the number of observations goes to infinity.

A.2.5 Power Spectrum

The discrete-time Fourier transform of the correlation function is called *power spectrum* or *power spectral density*

$$S_x(e^{\jmath\omega}) = \sum_{k=-\infty}^{\infty} r_x(k)e^{-\jmath\omega k}. \tag{A.80}$$

Inversely, the correlation function can be calculated from the power spectrum by

$$r_x(k) = \frac{1}{2\pi}\int_{-\pi}^{\pi} S_x(e^{\jmath\omega})e^{\jmath\omega k}d\omega. \tag{A.81}$$

The z-transform can be used instead of the discrete-time Fourier transform, and the power spectrum can be written

$$S_x(z) = \sum_{k=-\infty}^{\infty} r_x(k)z^{-k}. \tag{A.82}$$

For a wide-sense stationary process the power spectrum is symmetric,

$$S_x(e^{\jmath\omega}) = S_x(e^{-\jmath\omega}), \tag{A.83}$$

and non-negative,

$$S_x(e^{\jmath\omega}) \geq 0. \tag{A.84}$$

The average power of a zero-mean, wide-sense stationary process can be written

$$E\left[|x(n)|^2\right] = \frac{1}{2\pi}\int_{-\pi}^{\pi} S_x(e^{\jmath\omega})d\omega. \tag{A.85}$$

A.2.6 White Noise

A zero-mean process that has constant power spectrum is called a *white noise* process. A stationary, white noise process $v(n)$ is completely described by its second-order moment,

$$r_v(k) = \sigma_v^2\delta(k), \tag{A.86}$$

where σ_v^2 is the variance of the process. The correlation function shows that white noise is a sequence of uncorrelated random variables. If the process is a sequence of Gaussian random variables it is called white, Gaussian noise. The power spectrum of white noise is

$$S_v(e^{\jmath\omega}) = \sigma_v^2. \tag{A.87}$$

A.2.7 Filtering of Stochastic Processes

Stochastic processes are often inputs to linear, time-invariant filters. The first and second-order moments of the output process $y(n)$ of a filter with impulse response $h(n)$ then relate to the input process $x(n)$ as follows:

$$m_y = E[y(n)] = m_x \sum_{k=-\infty}^{\infty} h(k), \tag{A.88}$$

and

$$r_y(k) = r_x(k) * h(k) * h(-k). \tag{A.89}$$

In the frequency domain these relations can be written

$$m_y = m_x H(e^{j0}), \tag{A.90}$$

and

$$S_y(e^{j\omega}) = |H(e^{j\omega})|^2 S_x(e^{j\omega}). \tag{A.91}$$

Similarly, using the z-transform,

$$S_y(z) = H(z)H(z^{-1})S_x(z). \tag{A.92}$$

Filtering of white noise with variance σ_x^2 can then be expressed as

$$S_y(z) = H(z)H(z^{-1})\sigma_x^2. \tag{A.93}$$

A white noise process $v(n)$ filtered with a filter $H(z)$ having a rational transfer function with q zeros and p poles of the form

$$H(z) = \frac{B_q(z)}{A_p(z)} = \frac{\sum_{k=0}^{q} b_q(k) z^{-k}}{1 + \sum_{k=1}^{p} a_q(k) z^{-k}} \tag{A.94}$$

is called an *autoregressive moving average* (ARMA) process of order (p, q). The power spectrum of an ARMA process $x(n)$ can then be written

$$S_x(e^{j\omega}) = \frac{|B_q(e^{j\omega})|^2}{|A_p(e^{j\omega})|^2} \sigma_v^2. \tag{A.95}$$

There are two special cases of an ARMA process and that is when $q = 0$ and when $p = 0$, respectively. When $q = 0$ the process is referred to as

an *autoregressive* (AR) process of order p. The transfer function and power spectrum are given by

$$H(z) = \frac{b(0)}{1 + \sum_{k=1}^{p} a_q(k) z^{-k}} \tag{A.96}$$

and

$$S_x(e^{\jmath\omega}) = \frac{|b(0)|^2}{|A_p(e^{\jmath\omega})|^2} \sigma_v^2, \tag{A.97}$$

respectively. Similarly, when $p = 0$ the process is referred to as a *moving average* (MA) process of order q for which the transfer function and power spectrum are

$$H(z) = \sum_{k=0}^{q} b_q(k) z^{-k} \tag{A.98}$$

and

$$S_x(e^{\jmath\omega}) = |B_q(e^{\jmath\omega})|^2 \sigma_v^2, \tag{A.99}$$

respectively.

Bibliography

[1] G. Strang, *Linear Algebra and Its Applications.* Philadelphia: Saunders, 3rd ed., 1988.

[2] G. H. Golub and C. F. van Loan, *Matrix Computations.* Baltimore: The Johns Hopkins University Press, 2nd ed., 1989.

[3] M. Hayes, *Statistical Digital Signal Proccessing and Modeling.* New York: John Wiley & Sons, 1996.