CHAPTER

2

Stationary Processes and Models

The term *stochastic process* or *random process* is used to describe the time evolution of a statistical phenomenon according to probabilistic laws. The time evolution of the phenomenon means that the stochastic process is a function of time, defined on some observation interval. The statistical nature of the phenomenon means that, before conducting an experiment, it is not possible to define exactly the way it evolves in time. Examples of a stochastic process include speech signals, television signals, radar signals, digital computer data, the output of a communication channel, seismological data, and noise.

The form of a stochastic process that is of interest to us is one that is defined at *discrete and uniformly spaced instants of time* (Box and Jenkins, 1976; Priestley, 1981). Such a restriction may arise naturally in practice, as in the case of radar signals or digital computer data. Alternatively, the stochastic process may be defined originally for a continuous range of real values of time; however, before processing, it is *sampled uniformly* in time, with the sampling rate chosen to be greater than twice the highest frequency component of the process (Haykin, 1994).

A stochastic process is *not* just a single function of time; rather, it represents, in theory, an infinite number of *different* realizations of the process. One particular realization of a discrete-time stochastic process is called a *discrete-time series* or simply *time series*. For convenience of notation, *we normalize time with respect to the sampling period.* For example, the sequence $u(n), u(n-1), \ldots, u(n-M)$ represents a time series that consists of the *present* observation $u(n)$ made at time n and M past observations of the process made at times $n-1, \ldots, n-M$.

We say that a stochastic process is *strictly stationary* if its statistical properties are *invariant* to a shift of time. Specifically, for a discrete-time stochastic process represented by the time series $u(n), u(n-1) \ldots, u(n-M)$ to be strictly stationary, the *joint probability density function* of these observations made at times $n, n-1, \ldots, n-M$ must remain the same no matter what values we assign to n for fixed M.

2.1 PARTIAL CHARACTERIZATION OF A DISCRETE-TIME STOCHASTIC PROCESS

In practice, we usually find that it is not possible to determine (by means of suitable measurements) the joint probability density function for an arbitrary set of observations made on a stochastic process. Accordingly, we must content ourselves with a partial characterization of the process by specifying its first and second moments.

Consider a discrete-time stochastic process represented by the time series $u(n)$, $u(n-1), \ldots, u(n-M)$, which may be complex valued. We define the *mean-value function* of the process as

$$\mu(n) = E[u(n)] \tag{2.1}$$

where E denotes the *statistical expectation operator*. We define the *autocorrelation function* of the process as

$$r(n, n-k) = E[u(n)u^*(n-k)], \quad k = 0, \pm 1, \pm 2, \ldots, \tag{2.2}$$

where the asterisk denotes *complex conjugation*. We define the *autocovariance function* of the process as

$$c(n, n-k) = E[(u(n) - \mu(n))(u(n-k) - \mu(n-k))^*], \quad k = 0, \pm 1, \pm 2, \ldots \tag{2.3}$$

From Eqs. (2.1) to (2.3), we see that the mean-value, autocorrelation and autocovariance functions of the process are related by

$$c(n, n-k) = r(n, n-k) - \mu(n)\mu^*(n-k) \tag{2.4}$$

For a partial characterization of the process, we therefore need to specify (1) the mean-value function $\mu(n)$ and (2) the autocorrelation function $r(n, n-k)$ or the autocovariance function $c(n, n-k)$ for various values of n and k that are of interest. Note also the autocorrelation and autocovariance functions have the same value when the mean $\mu(n)$ is zero for all n.

This form of partial characterization offers two important advantages:

1. It lends itself to practical measurements.
2. It is well suited to *linear* operations on stochastic processes.

For a discrete-time stochastic process that is strictly stationary, all three quantities defined in Eqs. (2.1) to (2.3) assume simpler forms. In particular, we find that the mean-value

function of the process is a constant μ (say), so we may write

$$\mu(n) = \mu \qquad \text{for all } n \tag{2.5}$$

We also find that both the autocorrelation and autocovariance functions depend only on the *difference* between the observation times n and $n - k$, that is, k, as shown by

$$r(n, n - k) = r(k) \tag{2.6}$$

and

$$c(n, n - k) = c(k) \tag{2.7}$$

Note that when $k = 0$, corresponding to a time difference or *lag* of zero, $r(0)$ equals the *mean-square value* of $u(n)$:

$$r(0) = E[|u(n)|^2] \tag{2.8}$$

and $c(0)$ equals the *variance* of $u(n)$:

$$c(0) = \sigma_u^2 \tag{2.9}$$

The conditio of Eqs. (2.5) to (2.7) are *not* sufficient to guarantee that the discrete-time stochastic process is strictly stationary. However, a discrete-time stochastic process that is not strictly stationary, but for which these conditions hold, is said to be *wide-sense stationary*, or *stationary to the second order*. A strictly stationary process $\{u(n)\}$, or $u(n)$ for short, is stationary in the wide sense if and only if (Doob, 1953)

$$E[|u(n)|^2] < \infty \qquad \text{for all } n$$

This condition is ordinarily satisfied by stochastic processes encountered in the physical sciences and engineering.

2.2 MEAN ERGODIC THEOREM

The *expectations* or *ensemble averages* of a stochastic process are averages "across the process." Clearly, we may also define *long-term sample averages* or *time averages* that are averages "along the process." Indeed, time averages may be used to build a *stochastic model* of a physical process by *estimating* unknown parameters of the model. For such an approach to be rigorous, however, we have to show that time averages converge to corresponding ensemble averages of the process in some statistical sense. A popular criterion for convergence is that of mean square-error, as described next.

To be specific, consider a discrete-time stochastic process $u(n)$ that is wide-sense stationary. Let a constant μ denote the mean of the process, and $c(k)$ denote its autocovariance function for lag k. For an estimate of the mean μ, we may use the time average

$$\hat{\mu}(N) = \frac{1}{N} \sum_{n=0}^{N-1} u(n) \tag{2.10}$$

where N is the total number of samples used in the estimation. Note that the estimate $\hat{\mu}(N)$ is a random variable with a mean and variance of its own. In particular, we readily find from Eq. (2.10) that the mean (expectation) of $\hat{\mu}(N)$ is

$$E[\hat{\mu}(N)] = \mu \qquad \text{for all } N \tag{2.11}$$

It is in the sense of Eq. (2.11) that we say the time average $\hat{\mu}(N)$ is an *unbiased* estimator of the ensemble average (mean) of the process.

Moreover, we say that the process $u(n)$ is *mean ergodic in the mean-square error sense* if the mean-square value of the error between the ensemble average μ and the time average $\hat{\mu}(N)$ approaches zero as the number of samples N approaches infinity; that is,

$$\lim_{N\to\infty} [(\mu - \hat{\mu}(N))^2] = 0$$

Using the time average formula of Eq. (2.10), we may write

$$\begin{aligned} E[|\mu - \hat{\mu}(N)|^2] &= E\left[\left|\mu - \frac{1}{N}\sum_{n=0}^{N-1} u(n)\right|^2\right] \\ &= \frac{1}{N^2} E\left[\left|\sum_{n=0}^{N-1} (u(n) - \mu)\right|^2\right] \\ &= \frac{1}{N^2} E\left[\sum_{n=0}^{N-1}\sum_{k=0}^{N-1} (u(n) - \mu)(u(k) - \mu)^*\right] \\ &= \frac{1}{N^2}\sum_{n=0}^{N-1}\sum_{k=0}^{N-1} E[(u(n) - \mu)(u(k) - \mu)^*] \\ &= \frac{1}{N^2}\sum_{n=0}^{N-1}\sum_{k=0}^{N-1} c(n-k) \end{aligned} \tag{2.12}$$

Let $l = n - k$. We may then simplify the double summation in Eq. (2.12) as follows:

$$E[|\mu - \hat{\mu}(N)|^2] = \frac{1}{N}\sum_{l=-N+1}^{N-1} \left(1 - \frac{|l|}{N}\right) c(l)$$

Accordingly, we may state that the necessary and sufficient condition for the process $u(n)$ to be mean ergodic in the mean-square error sense is that

$$\lim_{N\to\infty} \frac{1}{N}\sum_{l=-N+1}^{N-1} \left(1 - \frac{|l|}{N}\right) c(l) = 0 \tag{2.13}$$

In other words, if the process $u(n)$ is asymptotically uncorrelated in the sense of Eq. (2.13), then the time average $\hat{\mu}(N)$ of the process converges to the ensemble average μ in

the mean-square error sense. This is the statement of a particular form of the *mean ergodic theorem* (Gray and Davisson, 1986).

The use of the mean ergodic theorem may be extended to other time averages of the process. Consider, for example, the following time average used to estimate the autocorrelation function of a wide-sense stationary process:

$$\hat{r}(k,N) = \frac{1}{N}\sum_{n=0}^{N-1} u(n)u(n-k), \qquad 0 \le k \le N-1 \tag{2.14}$$

The process $u(n)$ is said to be *correlation ergodic* in the mean-square error sense if the mean-square value of the difference between the true value $r(k)$ and the estimate $\hat{r}(k, N)$ approaches zero as the number of samples N approaches infinity. Let $z(n, k)$ denote a new discrete-time stochastic process related to the original process $u(n)$ as follows:

$$z(n, k) = u(n)u(n - k) \tag{2.15}$$

Hence, by substituting $z(n, k)$ for $u(n)$, we may use the mean ergodic theorem to establish the conditions for $z(n, k)$ to be mean ergodic or, equivalently, for $u(n)$ to be correlation ergodic.

2.3 CORRELATION MATRIX

Let the M-by-1 *observation vector* $\mathbf{u}(n)$ represent the elements of the time series $u(n)$, $u(n - 1), \ldots, u(n - M + 1)$. To show the composition of the vector $\mathbf{u}(n)$ explicitly, we write

$$\mathbf{u}(n) = [u(n), u(n - 1), \ldots, u(n - M + 1)]^T \tag{2.16}$$

where the superscript T denotes *transposition*. We define the *correlation matrix* of a stationary discrete-time stochastic process represented by this time series as *the expectation of the outer product of the observation vector* $\mathbf{u}(n)$ *with itself*. Let $\mathbf{R}$ denote the M-by-M correlation matrix defined in this way. We thus write

$$\mathbf{R} = E[\mathbf{u}(n)\,\mathbf{u}^H(n)] \tag{2.17}$$

where the superscript H denotes *Hermitian transposition* (i.e., the operation of transposition combined with complex conjugation). By substituting Eq. (2.16) in (2.17) and using the condition of wide-sense stationarity, we may express the correlation matrix $\mathbf{R}$ in the expanded form:

$$\mathbf{R} = \begin{bmatrix} r(0) & r(1) & \cdots & r(M-1) \\ r(-1) & r(0) & \cdots & r(M-2) \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ r(-M+1) & r(-M+2) & \cdots & r(0) \end{bmatrix} \tag{2.18}$$

The element $r(0)$ on the main diagonal is always real valued. For complex-valued data, the remaining elements of $\mathbf{R}$ assume complex values.

Properties of the Correlation Matrix

The correlation matrix $\mathbf{R}$ plays a key role in the statistical analysis and design of discrete-time filters. It is therefore important that we understand its various properties and their implications. In particular, using the definition of Eq. (2.17), we find that the correlation matrix of a stationary discrete-time stochastic process has the following properties.

Property 1. *The correlation matrix of a stationary discrete-time stochastic process is Hermitian.*

We say that a *complex-valued* matrix is *Hermitian* if it is equal to its *conjugate transpose*. We may thus express the Hermitian property of the correlation matrix $\mathbf{R}$ by writing

$$\mathbf{R}^H = \mathbf{R} \tag{2.19}$$

This property follows directly from the definition of Eq. (2.17).

Another way of stating the Hermitian property of the correlation matrix $\mathbf{R}$ is to write

$$r(-k) = r^*(k) \tag{2.20}$$

where $r(k)$ is the autocorrelation function of the stochastic process $u(n)$ for a lag of k. Accordingly, for a wide-sense stationary process we only need M values of the autocorrelation function $r(k)$ for $k = 0, 1, \ldots, M-1$ in order to completely define the correlation matrix $\mathbf{R}$. We may thus rewrite Eq. (2.18) as follows:

$$\mathbf{R} = \begin{bmatrix} r(0) & r(1) & \cdots & r(M-1) \\ r^*(1) & r(0) & \cdots & r(M-2) \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ r^*(M-1) & r^*(M-2) & \cdots & r(0) \end{bmatrix} \tag{2.21}$$

From here on, we will use this representation for the expanded matrix form of the correlation matrix of a wide-sense stationary discrete-time stochastic process. Note that for the special case of *real-valued data*, the autocorrelation function $r(k)$ is real for all k, and the correlation matrix $\mathbf{R}$ is *symmetric*.

Property 2. *The correlation matrix of a stationary discrete-time stochastic process is Toeplitz.*

We say that a square matrix is *Toeplitz* if all the elements on its main diagonal are equal, and if the elements on any other diagonal parallel to the main diagonal are also equal. From the expanded form of the correlation matrix $\mathbf{R}$ given in Eq. (2.21), we see that all the elements on the main diagonal are equal to $r(0)$, all the elements on the first diagonal above the main diagonal are equal to $r(1)$, all the elements along the first diagonal

below the main diagonal are equal to $r^*(1)$, and so on for the other diagonals. We conclude therefore that the correlation matrix $\mathbf{R}$ is Toeplitz.

It is important to recognize, however, that the Toeplitz property of the correlation matrix $\mathbf{R}$ is a direct consequence of the assumption that the discrete-time stochastic process represented by the observation vector $\mathbf{u}(n)$ is wide-sense stationary. Indeed, we may state that if the discrete-time stochastic process is wide-sense stationary, then its correlation matrix $\mathbf{R}$ must be Toeplitz; and, conversely, if the correlation matrix $\mathbf{R}$ is Toeplitz, then the discrete-time stochastic process must be wide-sense stationary.

Property 3. *The correlation matrix of a discrete-time stochastic process is always nonnegative definite and almost always positive definite.*

Let $\mathbf{x}$ be an arbitrary (nonzero) M-by-1 complex-valued vector. Define the scalar random variable y as the *inner product* of $\mathbf{x}$ and the observation vector $\mathbf{u}(n)$, as shown by

$$y = \mathbf{x}^H\mathbf{u}(n)$$

Taking the Hermitian transpose of both sides and recognizing that y is a scalar, we get

$$y^* = \mathbf{u}^H(n)\mathbf{x}$$

where the asterisk denotes *complex conjugation*. The mean-square value of the random variable y equals

$$\begin{aligned} E[|y|^2] &= E[yy^*] \\ &= E[\mathbf{x}^H\mathbf{u}(n)\mathbf{u}^H(n)\mathbf{x}] \\ &= \mathbf{x}^H E[\mathbf{u}(n)\mathbf{u}^H(n)]\mathbf{x} \\ &= \mathbf{x}^H\mathbf{R}\mathbf{x} \end{aligned}$$

where $\mathbf{R}$ is the correlation matrix defined in Eq. (2.17). The expression $\mathbf{x}^H\mathbf{R}\mathbf{x}$ is called a Hermitian form. Since

$$E[|y|^2] \geq 0$$

it follows that

$$\mathbf{x}^H\mathbf{R}\mathbf{x} \geq 0 \tag{2.22}$$

A Hermitian form that satisfies this condition for every nonzero $\mathbf{x}$ is said to be *nonnegative definite* or *positive semidefinite*. Accordingly, we may state that the correlation matrix of a wide-sense stationary process is always nonnegative definite.

If the Hermitian form $\mathbf{x}^H\mathbf{R}\mathbf{x}$ satisfies the condition

$$\mathbf{x}^H\mathbf{R}\mathbf{x} > 0$$

for every nonzero $\mathbf{x}$, we say that the correlation matrix $\mathbf{R}$ is *positive definite*. This condition is satisfied for a wide-sense stationary process unless there are linear dependencies between the random variables that constitute the M elements of the observation vector

$\mathbf{u}(n)$. Such a situation arises essentially only when the process $u(n)$ consists of the sum of K sinusoids with $K \leq M$; see Section 2.4 for more details. In practice, we find that this idealized situation is so rare in occurrence that the correlation matrix $\mathbf{R}$ is almost always positive definite.

The positive definiteness of a correlation matrix implies that its determinant and all principal minors are greater than zero. For example, for $M = 2$, we must have

$$\begin{vmatrix} r(0) & r(1) \\ r^*(1) & r(0) \end{vmatrix} > 0$$

Similarly, for $M = 3$, we must have

$$\begin{vmatrix} r(0) & r(1) \\ r^*(1) & r(0) \end{vmatrix} > 0$$

$$\begin{vmatrix} r(0) & r(2) \\ r^*(2) & r(0) \end{vmatrix} > 0$$

$$\begin{vmatrix} r(0) & r(1) & r(2) \\ r^*(1) & r(0) & r(1) \\ r^*(2) & r^*(1) & r(0) \end{vmatrix} > 0$$

and so on for higher values of M. These conditions, in turn, imply that the correlation matrix is nonsingular. We say that a matrix is *nonsingular* if its inverse exists; otherwise, it is singular. Accordingly, we may state that a correlation matrix is almost always nonsingular.

Property 4. *When the elements that constitute the observation vector of a stationary discrete-time stochastic process are rearranged backward, the effect is equivalent to the transposition of the correlation matrix of the process.*

Let $\mathbf{u}^B(n)$ denote the M-by-1 vector obtained by rearranging the elements that constitute the observation vector $\mathbf{u}(n)$ *backward*. We illustrate this operation by writing

$$\mathbf{u}^{BT}(n) = [u(n - M + 1), u(n - M + 2), \ldots, u(n)] \tag{2.23}$$

where the superscript B denotes the backward rearrangement of a vector. The correlation matrix of the vector $\mathbf{u}^B(n)$ equals, by definition,

$$E[\mathbf{u}^B(n)\mathbf{u}^{BH}(n)] = \begin{bmatrix} r(0) & r^*(1) & \cdots & r^*(M-1) \\ r(1) & r(0) & \cdots & r^*(M-2) \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ r(M-1) & r(M-2) & \cdots & r(0) \end{bmatrix} \tag{2.24}$$

Hence, comparing the expanded correlation matrix of Eq. (2.24) with that of Eq. (2.21), we see that

$$E[\mathbf{u}^B(n)\,\mathbf{u}^{BH}(n)] = \mathbf{R}^T \tag{2.25}$$

which is the desired result.

Property 5. *The correlation matrices $\mathbf{R}_M$ and $\mathbf{R}_{M+1}$ of a stationary discrete-time stochastic process, pertaining to M and M + 1 observations of the process, respectively, are related by*

$$\mathbf{R}_{M+1} = \begin{bmatrix} r(0) & \mathbf{r}^H \\ \mathbf{r} & \mathbf{R}_M \end{bmatrix} \tag{2.26}$$

or equivalently,

$$\mathbf{R}_{M+1} = \begin{bmatrix} \mathbf{R}_M & \mathbf{r}^{B*} \\ \mathbf{r}^{BT} & r(0) \end{bmatrix} \tag{2.27}$$

where $r(0)$ is the autocorrelation of the process for a lag of zero, and

$$\mathbf{r}^H = [r(1), r(2), \ldots, r(M)] \tag{2.28}$$

and

$$\mathbf{r}^{BT} = [r(-M), r(-M+1), \ldots, r(-1)] \tag{2.29}$$

Note that in describing Property 5 we have added a subscript, M or $M + 1$, to the symbol for the correlation matrix in order to display dependence on the number of observations used to define this matrix. We follow such a practice (in the context of the correlation matrix and other vector quantities) *only* when the issue at hand involves dependence on the number of observations or dimensions of the matrix.

To prove the relation of Eq. (2.26), we express the correlation matrix $\mathbf{R}_{M+1}$ in its expanded form, partitioned as follows:

$$\mathbf{R}_{M+1} = \left[\begin{array}{c|ccccc} r(0) & r(1) & r(2) & \cdots & r(M) \\ \hline r^*(1) & r(0) & r(1) & \cdots & r(M-1) \\ r^*(2) & r^*(1) & r(0) & \cdots & r(M-2) \\ \vdots & \vdots & \vdots & & \vdots \\ r^*(M) & r^*(M-1) & r^*(M-2) & \cdots & r(0) \end{array}\right] \tag{2.30}$$

Using Eqs. (2.18), (2.20), and (2.28) in (2.30), we get the result given in Eq. (2.26). Note that according to this relation, the observation vector $\mathbf{u}_{M+1}(n)$ is *partitioned* in the form

$$\mathbf{u}_{M+1}(n) = \begin{bmatrix} u(n) \\ \cdots\cdots\cdots \\ u(n-1) \\ u(n-2) \\ \vdots \\ u(n-M) \end{bmatrix} = \begin{bmatrix} u(n) \\ \cdots\cdots\cdots \\ \mathbf{u}_M(n-1) \end{bmatrix} \tag{2.31}$$

where the subscript $M + 1$ is intended to denote the fact that the vector $\mathbf{u}_{M+1}(n)$ has $M + 1$ elements, and likewise for $\mathbf{u}_M(n)$.

To prove the relation of Eq. (2.27), we express the correlation matrix $\mathbf{R}_{M+1}$ in its expanded form, partitioned in the alternative form

$$\mathbf{R}_{M+1} = \left[\begin{array}{cccc:c} r(0) & r(1) & \cdots & r(M-1) & r(M) \\ r^*(1) & r(0) & \cdots & r(M-2) & r(M-1) \\ \vdots & \vdots & \ddots & \vdots & \\ r^*(M-1) & r^*(M-2) & \cdots & r(0) & r(1) \\ \hdashline r^*(M) & r^*(M-1) & \cdots & r^*(1) & r(0) \end{array}\right] \tag{2.32}$$

Here again, using Eqs. (2.18), (2.20), and (2.29) in (2.32), we get the result given in Eq. (2.27). Note that according to this second relation the observation vector $\mathbf{u}_{M+1}(n)$ is partitioned in the alternative form

$$\mathbf{u}_{M+1}(n) = \begin{bmatrix} u(n) \\ u(n-1) \\ \vdots \\ u(n-M+1) \\ \cdots\cdots\cdots \\ u(n-M) \end{bmatrix} = \begin{bmatrix} \mathbf{u}_M(n) \\ \cdots\cdots\cdots \\ u(n-M) \end{bmatrix} \tag{2.33}$$

2.4 CORRELATION MATRIX OF SINE WAVE PLUS NOISE

A time series of special interest is one that consists of a *complex sinusoid corrupted by additive noise*. Such a time series is representative of several important signal-processing applications. In the *temporal context*, for example, this time series represents the composite signal at the input of a receiver, with the complex sinusoid representing a *target signal* and the noise representing thermal noise generated at the front end of the receiver. In the *spatial context*, it represents the received signal in a linear array of sensors, with the complex sinusoid representing a *plane wave* produced by a remote source (emitter) and the noise representing *sensor noise*.

Let α denote the amplitude of the complex sinusoid, and ω denote its angular frequency. Let $v(n)$ denote a sample of the noise, assumed to have zero mean. We may then write a corresponding sample of the time series that consists of the complex sinusoid plus noise as follows:

$$u(n) = \alpha \exp(j\omega n) + v(n), \qquad n = 0, 1, \ldots, N-1 \tag{2.34}$$

The sources of the complex sinusoid and the noise are independent of each other. Since the noise component $v(n)$ has zero mean, by assumption, we see from Eq. (2.34) that the mean of $u(n)$ is equal to $\alpha \exp(j\omega n)$.

To calculate the autocorrelation function of the process $u(n)$, we clearly need to know the autocorrelation function of the noise process $v(n)$. To proceed then, we assume a special form of noise characterized by the autocorrelation function

$$E[v(n)v^*(n-k)] = \begin{cases} \sigma_v^2, & k = 0 \\ 0, & k \neq 0 \end{cases} \tag{2.35}$$

Such a form of noise is commonly referred to as *white noise*; more will be said about it in Chapter 3. Since the sources responsible for the generation of the complex sinusoid and the noise are independent and, therefore, uncorrelated, it follows that the autocorrelation function of the process $u(n)$ equals the sum of the autocorrelation functions of its two individual components. Accordingly, using Eqs. (2.34) and (2.35), we find that the autocorrelation function of the process $u(n)$ for a lag k is given by

$$\begin{aligned} r(k) &= E[u(n)u^*(n-k)] \\ &= \begin{cases} |\alpha|^2 + \sigma_v^2, & k = 0 \\ |\alpha|^2 \exp(j\omega k), & k \neq 0 \end{cases} \end{aligned} \tag{2.36}$$

where $|\alpha|$ is the magnitude of the complex amplitude α. Note that for a lag $k \neq 0$, the autocorrelation function $r(k)$ varies with k in the same sinusoidal fashion as the sample $u(n)$ varies with n, except for a change in amplitude. Given the series of samples $u(n)$, $u(n-1)$, $\ldots$, $u(n-M+1)$, we may thus express the correlation matrix of $u(n)$ as

$$\mathbf{R} = |\alpha|^2 \begin{bmatrix} 1+\frac{1}{\rho} & \exp(j\omega) & \cdots & \exp(j\omega(M-1)) \\ \exp(-j\omega) & 1+\frac{1}{\rho} & \cdots & \exp(j\omega(M-2)) \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \exp(-j\omega(M-1)) & \exp(-j\omega(M-2)) & \cdots & 1+\frac{1}{\rho} \end{bmatrix} \tag{2.37}$$

where ρ is the *signal-to-noise ratio*, defined by

$$\rho = \frac{|\alpha|^2}{\sigma_v^2} \tag{2.38}$$

The correlation matrix $\mathbf{R}$ of Eq. (2.37) has all of the properties described in Section 2.3; the reader is invited to verify them.

Equation (2.36) provides the mathematical basis of a two-step practical procedure for estimating the parameters of a complex sinusoid in the presence of additive noise:

1. Measure the mean-square value $r(0)$ of the process $u(n)$. Hence, given the noise variance σ_v^2, determine the magnitude $|\alpha|$.
2. Measure the autocorrelation function $r(k)$ of the process $u(n)$ for a lag $k \neq 0$. Hence, given $|\alpha|^2$ from step 1, determine the angular frequency ω.

Note that this estimation procedure is *invariant to the phase of* α, which is a direct consequence of the definition of the autocorrelation function $r(k)$.

Example 1

Consider the idealized case of a noiseless sinusoid of angular frequency ω. For the purpose of illustration, we assume that the time series of interest consists of three uniformly spaced samples drawn from this sinusoid. Hence, setting the signal-to-noise ratio $\rho = \infty$ and the number of samples $M = 3$, we find from Eq. (2.37) that the correlation matrix of the time series so obtained has the following value:

$$\mathbf{R} = |\alpha|^2 \begin{bmatrix} 1 & \exp(j\omega) & \exp(j2\omega) \\ \exp(-j\omega) & 1 & \exp(j\omega) \\ \exp(-j2\omega) & \exp(-j\omega) & 1 \end{bmatrix}$$

From this expression we readily see that the determinant of $\mathbf{R}$ and all principal minors are identically zero. Hence, this correlation matrix is singular.

We may generalize the result of this example by stating that when a process $u(n)$ consists of M samples drawn from the sum of K sinusoids with $K < M$ and there is *no* additive noise, then the correlation matrix of that process is singular.

2.5 STOCHASTIC MODELS

The term *model* is used for any hypothesis that may be applied to explain or describe the hidden laws that are supposed to govern or constrain the generation of physical data of interest. The representation of a stochastic process by a model dates back to an idea by Yule (1927). The idea is that a time series $u(n)$, consisting of highly correlated observations, may be generated by applying a series of statistically independent "shocks" to a linear filter, as in Fig. 2.1. The shocks are random variables drawn from a fixed distribution that is usually assumed to be *Gaussian* with zero mean and constant variance. Such a series of random variables constitutes a purely random process, commonly referred to as *white Gaussian noise*. Specifically, we may describe the input $v(n)$ in Figure 2.1 in statistical terms as follows:

$$E[v(n)] = 0 \qquad \text{for all } n \tag{2.39}$$

and

$$E[v(n)v^*(k)] = \begin{cases} \sigma_v^2, & k = n \\ 0, & \text{otherwise} \end{cases} \tag{2.40}$$

where σ_v^2 is the noise variance. Equation (2.39) follows from the zero-mean assumption, and Eq. (2.40) follows from the white assumption. The implication of the Gaussian assumption is discussed in Section 2.11.

In general, the time-domain description of the input–output relation for the stochastic model of Fig. 2.1 may be described as follows:

$$\begin{pmatrix}\text{present value}\\ \text{of model output}\end{pmatrix} + \begin{pmatrix}\text{linear combination}\\ \text{of past values}\\ \text{of model output}\end{pmatrix} = \begin{pmatrix}\text{linear combination of}\\ \text{present and past values}\\ \text{of model input}\end{pmatrix} \tag{2.41}$$

A stochastic process so described is referred to as a *linear process*.

The structure of the linear filter in Fig. 2.1 is determined by the manner in which the two linear combinations indicated in Eq. (2.41) are formulated. We may thus identify three popular types of linear stochastic models:

1. Autoregressive models, in which no past values of the model input are used.
2. Moving average models, in which no past values of the model output are used.
3. Mixed autoregressive-moving average models, in which the description of Eq. (2.41) applies in its entire form. Hence, this class of stochastic models includes autoregressive and moving average models as special cases.

Figure 2.1 Stochastic model.

These models are described next, in that order.

Autoregressive Models

We say that the time series $u(n), u(n-1), \ldots, u(n-M)$ represents the realization of an *autoregressive process (AR) of order M* if it satisfies the difference equation

$$u(n) + a_1^* u(n-1) + \cdots + a_M^* u(n-M) = v(n) \tag{2.42}$$

where $a_1, a_2, \ldots, a_M$ are constants called the AR *parameters*, and $v(n)$ is a white-noise process. The term $a_k^* u(n-k)$ is the scalar version of *inner product* of a_k and $u(n-k)$, where $k = 1, \ldots, M$.

To explain the reason for the term "autoregressive," we rewrite Eq. (2.42) in the form

$$u(n) = w_1^* u(n-1) + w_2^* u(n-2) + \cdots + w_M^* u(n-M) + v(n) \tag{2.43}$$

where $w_k = -a_k$. We thus see that the present value of the process, that is, $u(n)$, equals a *finite linear combination of past values* of the process, $u(n-1), \ldots, u(n-M)$, plus an *error term* $v(n)$. We now see the reason for the term "autoregressive." Specifically, a linear model

$$y = \sum_{k=1}^{M} w_k^* x_k + v$$

relating a *dependent* variable y to a set of *independent* variables $x_1, x_2, \ldots, x_M$ plus an error term v is often referred to as a *regression model*, and y is said to be "regressed" on $x_1, x_2, \ldots, x_M$. In Eq. (2.43), the variable $u(n)$ is *regressed* on previous values of *itself*; hence the term "autoregressive."

The left-hand side of Eq. (2.42) represents the *convolution* of the input sequence $u(n)$ and the sequence of parameters a_n^*. To highlight this point, we rewrite Eq. (2.42) in the form of a convolution sum:

$$\sum_{k=0}^{M} a_k^* u(n-k) = v(n) \tag{2.44}$$

where $a_0 = 1$. By taking the *z-transform* of both sides of Eq. (2.44), we transform the convolution sum on the left-hand side of the equation into a multiplication of the z-transforms of the two sequences $u(n)$ and a_n^*. Let $H_A(z)$ denote the z-transform of the sequence a_n^*:

$$H_A(z) = \sum_{n=0}^{M} a_n^* z^{-n} \tag{2.45}$$

Let $U(z)$ denote the z-transform of the input sequence $u(n)$:

$$U(z) = \sum_{n=0}^{\infty} u(n) z^{-n} \tag{2.46}$$

where z is a *complex variable*. We may thus transform the difference equation (2.42) into the equivalent form

$$H_A(z)U(z) = V(z) \tag{2.47}$$

where

$$V(z) = \sum_{n=0}^{\infty} v(n)z^{-n} \tag{2.48}$$

The z-transform of Eq. (2.47) offers two interpretations, depending on whether the AR process $u(n)$ is viewed as the input or output of interest:

1. Given the AR process $u(n)$, we may use the filter shown in Fig. 2.2(a) to produce the white noise process $v(n)$ as output. The parameters of this filter bear a one-to-one correspondence with those of the AR process $u(n)$. Accordingly, this filter represents a *process analyzer* with discrete transfer function $H_A(z) = V(z)/U(z)$. The impulse response of the AR process analyzer, that is, the inverse z-transform of $H_A(z)$, has *finite duration*.
2. With the white noise $v(n)$ acting as input, we may use the filter shown in Fig. 2.2(b) to produce the AR process $u(n)$ as output. Accordingly, this second filter represents a *process generator*, whose transfer function equals

$$\begin{aligned} H_G(z) &= \frac{U(z)}{V(z)} \\ &= \frac{1}{H_A(z)} \\ &= \frac{1}{\sum_{n=0}^{M} a_n^* z^{-n}} \end{aligned} \tag{2.49}$$

   The impulse response of the AR process generator, that is, the inverse z-transform of $H_G(z)$, has *infinite duration*.

The AR process analyzer of Fig. 2.2(a) is an *all-zero filter*. It is so called because its transfer function $H_A(z)$ is completely defined by specifying the locations of its *zeros*. This filter is inherently stable.

The AR process generator of Fig. 2.2(b) is an *all-pole filter*. It is so called because its transfer function $H_G(z)$ is completely defined by specifying the locations of its *poles*, as shown by

$$H_G(z) = \frac{1}{(1 - p_1 z^{-1})(1 - p_2 z^{-1}) \cdots (1 - p_M z^{-1})} \tag{2.50}$$

The parameters $p_1, p_2, \ldots, p_M$ are *poles* of $H_G(z)$; they are defined by the roots of the *characteristic equation*

$$1 + a_1^* z^{-1} + a_2^* z^{-2} + \cdots + a_M^* z^{-M} = 0 \tag{2.51}$$

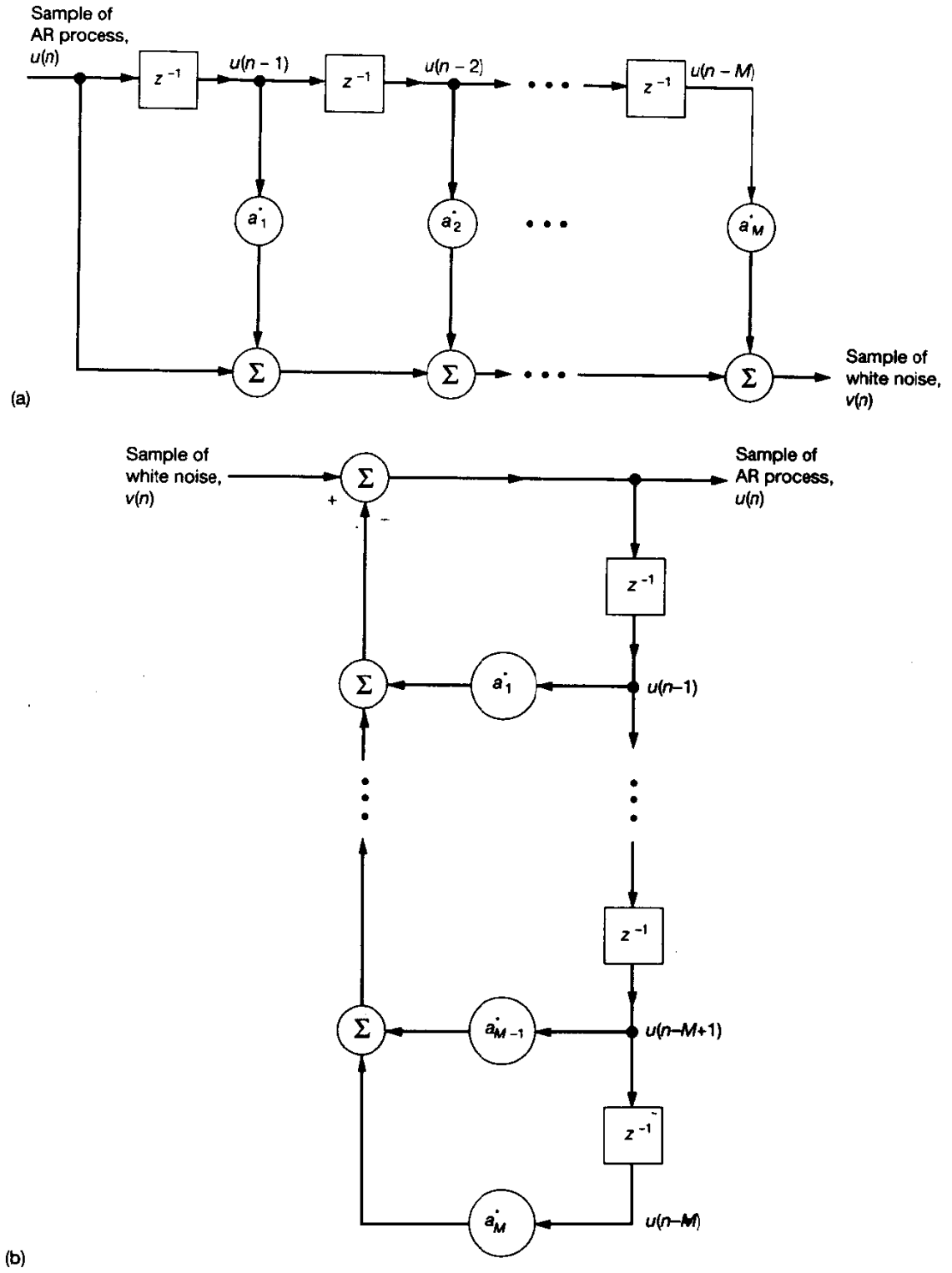

Figure 2.2 (a) AR process analyzer; (b) AR process generator.

For the all-pole AR process generator of Fig. 2.2(b) to be stable, the roots of the characteristic equation (2.51) must all lie inside the unit circle in the z-plane. This is also a necessary and sufficient condition for wide-sense stationarity of the AR process produced by the model of Fig. 2.2(b). We have more to say on the issue of stationarity in Section 2.7.

Moving Average Models

In a *moving average (MA) model*, the discrete-time linear filter of Fig. 2.1 consists of an *all-zero filter* driven by white noise. The resulting process $u(n)$, produced at the filter output, is described by the difference equation:

$$u(n) = v(n) + b_1^* v(n-1) + \cdots + b_K^* v(n-K) \tag{2.52}$$

where $b_1, \ldots, b_K$ are constants called the *MA parameters*, and $v(n)$ is a white-noise process of zero mean and variance σ_v^2. Except for $v(n)$, each term on the right-hand side of Eq. (2.52) represents the scalar version of an inner product. The *order* of the MA process equals K. The term moving average is a rather quaint one; nevertheless, its use is firmly established in the literature. Its usage arose in the following way: If we are given a complete temporal realization of the white-noise process $v(n)$, we may compute $u(n)$ by constructing a *weighted average* of the sample values $v(n), v(n-1), \ldots, v(n-K)$.

From Eq. (2.52), we readily obtain the MA model (i.e., process-generator) depicted in Fig. 2.3. Specifically, we start with a white-noise process $v(n)$ at the model input and generate an MA process $u(n)$ of order K at the model output. To proceed in the reverse manner, that is, to produce the white-noise process $v(n)$, given the MA process $u(n)$, we require the use of an *all-pole filter*. In other words, the filters used in the generation and analysis of an MA process are the *opposite* of those used in the case of an AR process.

Autoregressive–Moving Average Models

To generate a mixed *autoregressive–moving average (ARMA) process* $u(n)$, we use a discrete-time linear filter in Fig. 2.1 with a transfer function that contains *both poles and zeros*. Accordingly, given a white-noise process $v(n)$ as the filter input, the ARMA process $u(n)$ produced at the filter output is described by the difference equation

$$u(n) + a_1^* u(n-1) + \cdots + a_M^* u(n-M) = v(n) + b_1^* v(n-1) + \cdots + b_K^* v(n-K) \tag{2.53}$$

where $a_1, \ldots, a_M$ and $b_1, \ldots, b_K$ are called the *ARMA parameters*. Except for $u(n)$ on the left-hand side and $v(n)$ on the right-hand side of Eq. (2.53), all of the terms represent scalar versions of inner products. The *order* of the ARMA process equals (M, K).

From Eq. (2.53), we readily deduce the ARMA model (i.e., process generator) depicted in Fig. 2.4. Comparing this figure with Figs. 2.2(b) and 2.3, we clearly see that AR and MA models are indeed special cases of an ARMA model.

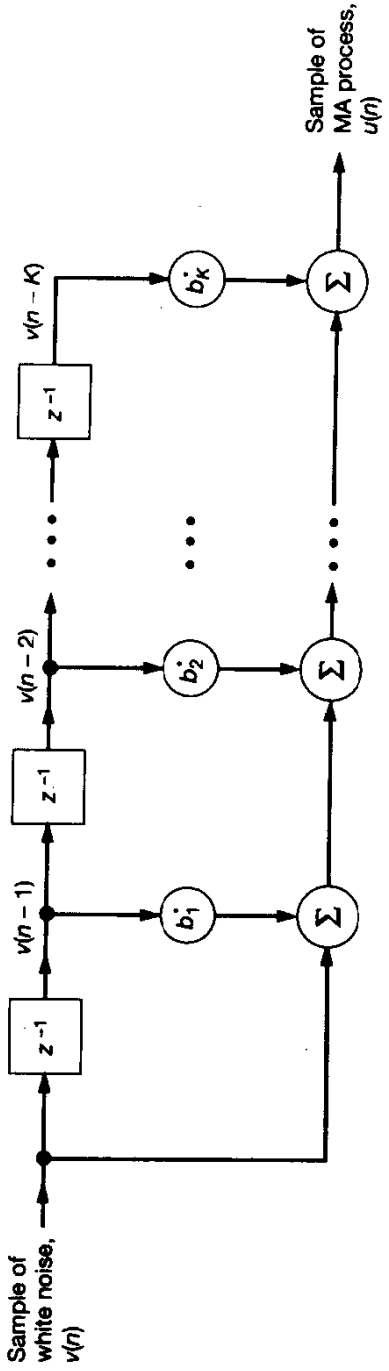

Figure 2.3 Moving average model (process generator).

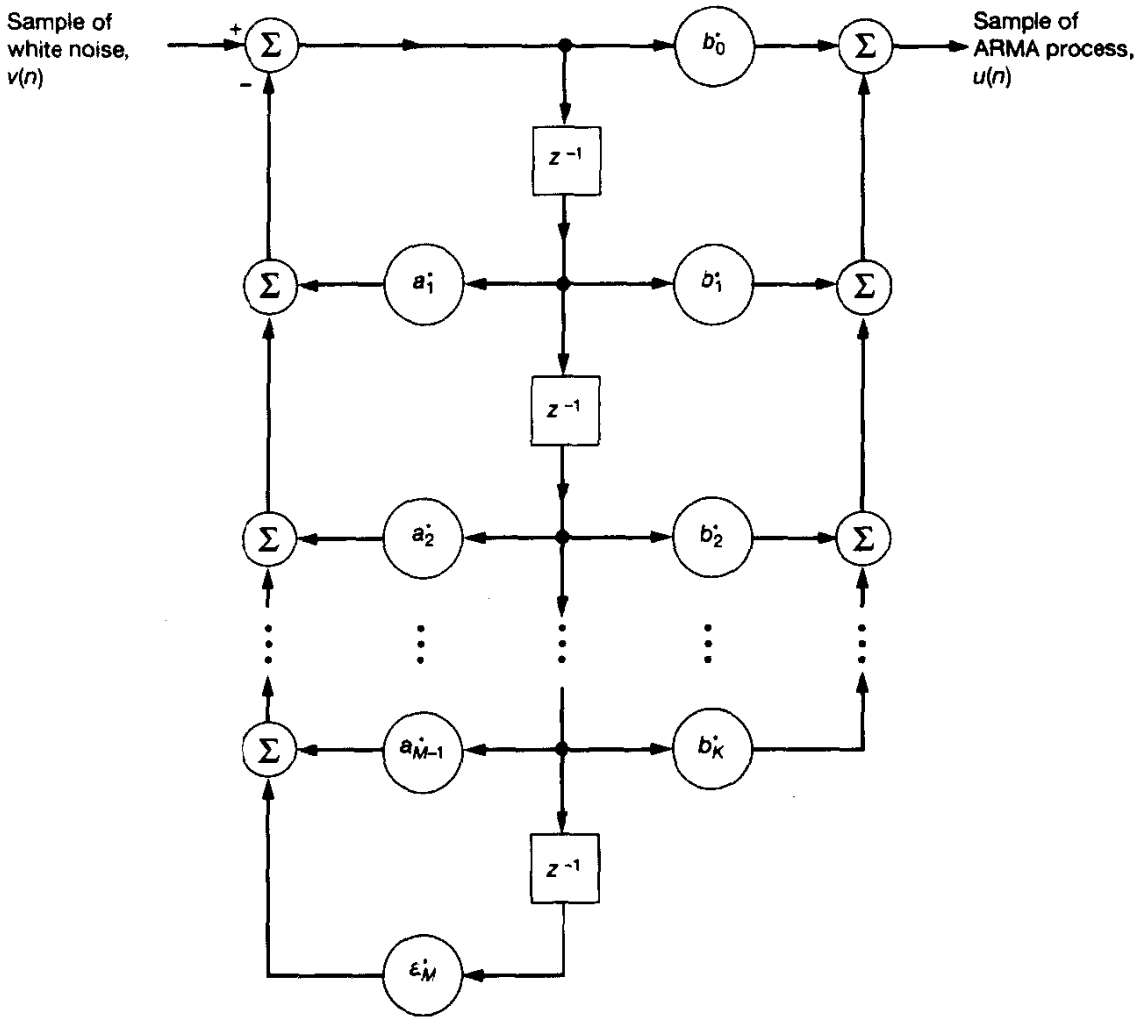

Figure 2.4 ARMA model (process generator) of order (M, K), assuming that $M > K$.

The transfer function of the ARMA process generator in Fig. 2.4 has both poles and zeros. Similarly, the ARMA analyzer used to generate a white-noise process $v(n)$, given an ARMA process $u(n)$, is characterized by a transfer function containing both poles and zeros.

From a computational viewpoint, the AR model has an advantage over the MA and ARMA models. Specifically, the computation of the AR coefficients in the model of Fig. 2.2(a) involves a system of *linear equations* known as the Yule–Walker equations, details of which are given in Section 2.8. On the other hand, the computation of the MA coefficients in the model of Fig. 2.3 and the computation of the ARMA coefficients in the model of Fig. 2.4 are much more complicated. Both of these computations require solving systems of *nonlinear equations*. It is for this reason that, in practice, we find that the use of AR models is more popular than MA and ARMA models. The wide application of AR

models may also be justified by virtue of a fundamental theorem of time series analysis, which is discussed next.

2.6 WOLD DECOMPOSITION

Wold (1938) proved a fundamental theorem, which states that any stationary discrete-time stochastic process may be decomposed into the sum of a *general linear process* and a *predictable process*, with these two processes being uncorrelated with each other. More precisely, Wold proved the following result:

Any stationary discrete-time stochastic process $x(n)$ may be expressed in the form

$$x(n) = u(n) + s(n) \tag{2.54}$$

where

1. $u(n)$ and $s(n)$ are uncorrelated processes,
2. $u(n)$ is a general linear process represented by the MA model:

$$u(n) = \sum_{k=0}^{\infty} b_k^* v(n-k) \tag{2.55}$$

with $b_0 = 1$, and

$$\sum_{k=0}^{\infty} |b_k|^2 < \infty,$$

and where $v(n)$ is a white-noise process uncorrelated with $s(n)$; that is,

$$E[v(n)s^*(k)] = 0 \qquad \text{for all } (n, k)$$

3. $s(n)$ is a predictable process; that is, the process can be predicted from its own past with zero prediction variance.

This result is known as *Wold's decomposition theorem*. A proof of this theorem is given in Priestley (1981).

According to Eq. (2.55), the general linear process $u(n)$ may be generated by feeding an *all-zero filter* with the white-noise process $v(n)$ as in Fig. 2.5(a). The zeros of the transfer function of this filter equal the roots of the equation:

$$B(z) = \sum_{n=0}^{\infty} b_n^* z^{-n} = 0$$

A solution of particular interest is an all-zero filter that is *minimum phase*, which means that all the zeros of the polynomial $B(z)$ lie inside the unit circle. In such a case, we may replace the all-zero filter with an *equivalent* all-pole filter that has the same impulse response $h_n = b_n^*$, as in Fig. 2.5(b). This means that except for a predictable component, a stationary discrete-time stochastic process may also be represented as an AR process of the appropriate order, subject to the above-mentioned restriction on $B(z)$. The basic difference between the MA and AR models is that $B(z)$ operates on the input $v(n)$ in the MA model, whereas the inverse $B^{-1}(z)$ operates on the output $u(n)$ in the AR model.

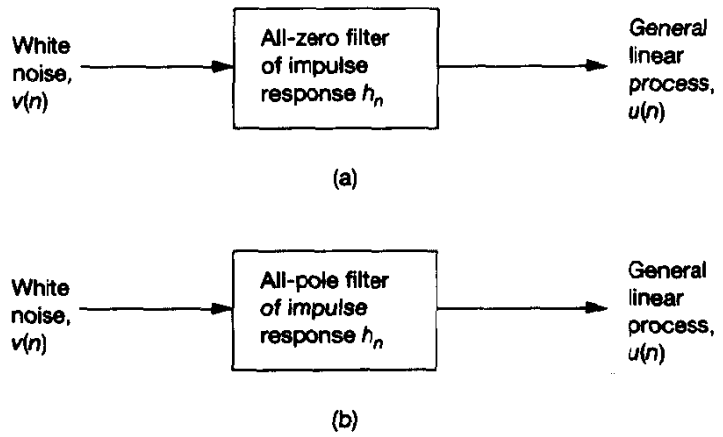

Figure 2.5 (a) Model, based on all-zero filter, for generating the linear process $u(n)$; (b) model, based on all-pole filter, for generating the general linear process $u(n)$. Both filters have exactly the same impulse response.

2.7 ASYMPTOTIC STATIONARITY OF AN AUTOREGRESSIVE PROCESS

Equation (2.42) represents a *linear, constant coefficient, difference equation of order M*, in which $v(n)$ plays the role of *input* or *driving function* and $u(n)$ that of *output* or *solution*. By using the *classical method*[1] for solving such an equation, we may formally express the solution $u(n)$ as the sum of a *complementary function*, $u_c(n)$, and a *particular solution*, $u_p(n)$, as follows:

$$u(n) = u_c(n) + u_p(n) \tag{2.56}$$

The evaluation of the solution $u(n)$ may thus proceed in two stages:

1. The complementary function $u_c(n)$ is the solution of the *homogeneous equation*

$$u(n) + a_1^* u(n-1) + a_2^* u(n-2) + \cdots + a_M^* u(n-M) = 0$$

In general, the complementary function $u_c(n)$ will therefore be of the form

$$u_c(n) = B_1 p_1^n + B_2 p_2^n + \cdots + B_M p_M^n \tag{2.57}$$

where $B_1, B_2, \ldots, B_M$ are arbitrary constants, and $p_1, p_2, \ldots, p_M$ are roots of the characteristic equation (2.51).

2. The particular solution $u_p(n)$ is defined by

$$u_p(n) = H_G(D)[v(n)] \tag{2.58}$$

[1]We may also use the z-transform method to solve the difference equation (2.42). However, for the discussion presented here, we find it more informative to use the classical method.

where D is the *unit-delay operator*, and the operator $H_G(D)$ is obtained by substituting D for z^{-1} in the discrete-transfer function of Eq. (2.49). The unit-delay operator D has the property

$$D^k[u(n)] = u(n-k), \qquad k = 0, 1, 2, \ldots \tag{2.59}$$

The constants $B_1, B_2, \ldots, B_M$ are determined by the choice of *initial conditions* that equal M in number. It is customary to set

$$\begin{aligned} u(0) &= 0 \\ u(-1) &= 0 \\ &\vdots \\ u(-M+1) &= 0 \end{aligned} \tag{2.60}$$

This is equivalent to setting the output of the model in Fig. 2.2(b) as well as the succeeding $(M-1)$ tap inputs equal to zero at time $n = 0$. Thus, by substituting these initial conditions into Eqs. (2.56) – (2.58), we obtain a set of M simultaneous equations that can be solved for the constants $B_1, B_2, \ldots, B_M$.

The result of imposing the initial conditions of Eq. (2.60) on the solution $u(n)$ is to make the discrete-time stochastic process represented by this solution nonstationary. On reflection, it is clear that this must be so, since we have given a "special status" to the time point $n = 0$, and the property of *invariance under a shift of time origin* cannot hold, even for second-order moments. If, however, the solution $u(n)$ is able to "forget" its initial conditions, the resulting process is asymptotically stationary in the sense that it settles down to a stationary behavior as n approaches infinity (Priestley, 1981). This requirement may be achieved by choosing the parameters of the AR model in Fig. 2.2(b) such that the complementary function $u_c(n)$ decays to zero as n approaches infinity. From Eq. (2.57) we see that, for arbitrary constants in the equation, this requirement can be met if and only if

$$|p_k| < 1 \qquad \text{for all } k$$

Hence, *for asymptotic stationarity of the discrete-time stochastic process represented by the solution $u(n)$, we require that all the poles of the filter in the AR model lie inside the unit circle in the z-plane.* This is intuitively satisfying.

Correlation Function of an Asymptotically Stationary AR Process

Assuming that the condition for asymptotic stationarity is satisfied, we may derive an important recursive relation for the autocorrelation function of the resulting AR process $u(n)$ as follows. We first multiply both sides of Eq. (2.42) by $u^*(n-l)$ and then apply the expectation operator, thereby obtaining

$$E\left[\sum_{k=0}^{M} a_k^* u(n-k)u^*(n-l)\right] = E[v(n)u^*(n-l)] \tag{2.61}$$

Next, we simplify the left-hand side of Eq. (2.61) by interchanging the expectation and summation, and recognizing that the expectation $E[u(n-k)u^*(n-l)]$ equals the autocorrelation function of the AR process for a lag of $l-k$. We simplify the right-hand side by observing that the expectation $E[v(n)u^*(n-l)]$ is zero for $l > 0$, since $u(n-l)$ only involves samples of the white-noise process at the filter input in Fig. 2.2(b) up to time $n-l$, which are uncorrelated with the white-noise sample $v(n)$. Accordingly, we simplify Eq. (2.61) as follows:

$$\sum_{k=0}^{M} a_k^* r(l-k) = 0, \qquad l > 0 \tag{2.62}$$

where $a_0 = 1$. We thus see that the autocorrelation function of the AR process satisfies the difference equation

$$r(l) = w_1^* r(l-1) + w_2^* r(l-2) + \cdots + w_M^* r(l-M), \qquad l > 0 \tag{2.63}$$

where $w_k = -a_k$, $k = 1, 2, \ldots, M$. Note that Eq. (2.63) is analogous to the difference equation satisfied by the AR process $u(n)$ itself.

We may express the general solution of Eq. (2.63) as follows:

$$r(m) = \sum_{k=1}^{M} C_k p_k^m \tag{2.64}$$

where $C_1, C_2, \ldots, C_M$ are constants, and $p_1, p_2, \ldots, p_M$ are roots of the characteristic equation (2.51). Note that when the AR model of Fig. 2.2(b) satisfies the condition for asymptotic stationarity, $|p_k| < 1$ for all k, in which case the autocorrelation function $r(m)$ approaches zero as the lag m approaches infinity.

The exact form of the contribution made by a pole p_k in Eq. (2.64) depends on whether the pole is real or complex. When p_k is real, the corresponding contribution decays geometrically to zero as the lag m increases. We refer to such a contribution as a *damped exponential*. On the other hand, complex poles occur in conjugate pairs, and the contribution of a complex-conjugate pair of poles is in the form of *a damped sine wave*. We thus find that, in general, the autocorrelation function of an asymptotically stationary AR process consists of a mixture of damped exponentials and damped sine waves.

2.8 YULE–WALKER EQUATIONS

In order to uniquely define the AR model of order M, depicted in Fig. 2.2(b), we need to specify two sets of model parameters:

1. The AR coefficients $a_1, a_2, \ldots, a_M$
2. The variance σ_v^2 of the white noise $v(n)$ used as excitation.

We now address these two issues in turn.

First, writing Eq. (2.63) for $l = 1, 2, \ldots, M$, we get a set of M simultaneous equations with the values $r(0), r(1), \ldots, r(M)$ of the autocorrelation function of the AR process as the known quantities and the AR parameters $a_1, a_2, \ldots, a_M$ as the unknowns. This set of equations may be expressed in the expanded matrix form

$$\begin{bmatrix} r(0) & r(1) & \cdots & r(M-1) \\ r^*(1) & r(0) & \cdots & r(M-1) \\ \vdots & \vdots & \ddots & \vdots \\ r^*(M-1) & r^*(M-2) & \cdots & r(0) \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_M \end{bmatrix} = \begin{bmatrix} r^*(1) \\ r^*(2) \\ \vdots \\ r^*(M) \end{bmatrix} \tag{2.65}$$

where we have $w_k = -a_k$. The set of equations (2.65) is called the *Yule–Walker equations* (Yule, 1927; Walker, 1931).

We may express the Yule–Walker equations in the compact matrix form

$$\mathbf{Rw} = \mathbf{r} \tag{2.66}$$

and its solution as (assuming that the correlation matrix $\mathbf{R}$ is nonsingular)

$$\mathbf{w} = \mathbf{R}^{-1}\mathbf{r} \tag{2.67}$$

where $\mathbf{R}^{-1}$ is the inverse of matrix $\mathbf{R}$, and the vector $\mathbf{w}$ is defined by

$$\mathbf{w} = [w_1, w_2, \ldots, w_M]^T$$

The correlation matrix $\mathbf{R}$ is defined by Eq. (2.21), and vector $\mathbf{r}$ is defined by Eq. (2.28). From these two equations, we see that we may uniquely determine both the matrix $\mathbf{R}$ and the vector $\mathbf{r}$, given the autocorrelation sequence $r(0), r(1), \ldots, r(M)$. Hence, using Eq. (2.67) we may compute the coefficient vector $\mathbf{w}$ and, therefore, the AR coefficients $a_k = -w_k$, $k = 1, 2, \ldots, M$. In other words, there is a unique relationship between the coefficients $a_1, a_2, \ldots, a_M$ of the AR model and the *normalized* correlation coefficients $\rho_1, \rho_2, \ldots, \rho_M$ of the AR process $u(n)$, as shown by

$$\{a_1, a_2, \ldots, a_M\} \rightleftharpoons \{\rho_1, \rho_2, \ldots, \rho_M\} \tag{2.68}$$

where the *correlation coefficient* ρ_k is defined by

$$\rho_k = \frac{r(k)}{r(0)}, \qquad k = 1, 2, \ldots, M \tag{2.69}$$

Variance of the White Noise

For $l = 0$, we find that the expectation on the right-hand side of Eq. (2.61) assumes the special form

$$\begin{aligned} E[v(n)u^*(n)] &= E[v(n)v^*(n)] \\ &= \sigma_v^2 \end{aligned} \tag{2.70}$$

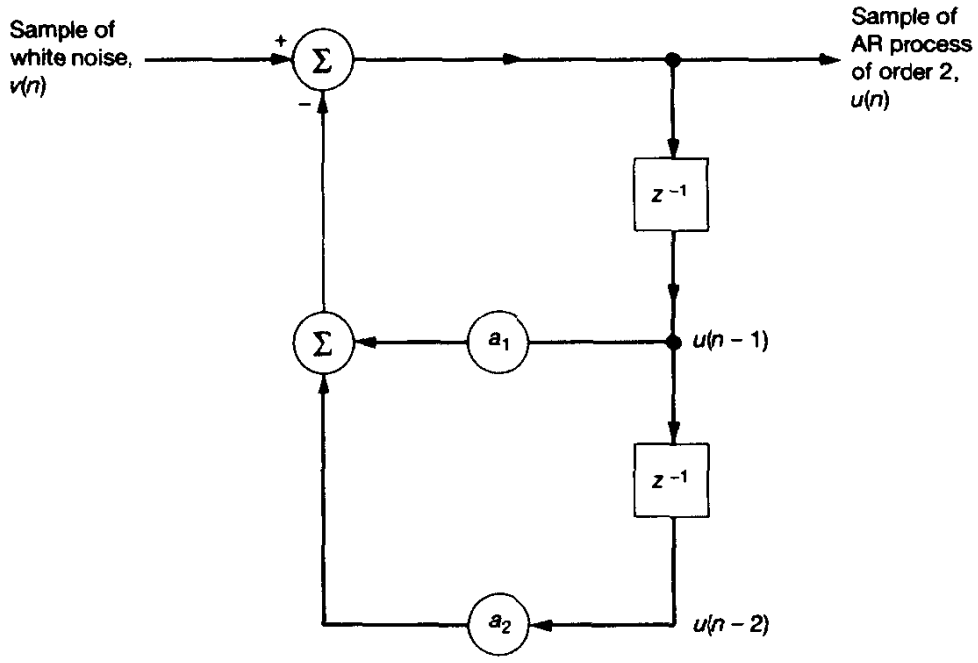

Figure 2.6 Model of (real-valued) autoregressive process of order 2.

where σ_v^2 is the variance of the zero-mean white noise $v(n)$. Accordingly, setting $l = 0$ in Eq. (2.61) and complex-conjugating both sides, we get the following formula for the variance of the white-noise process:

$$\sigma_v^2 = \sum_{k=0}^{M} a_k r(k) \tag{2.71}$$

where $a_0 = 1$. Hence, given the autocorrelations $r(0), r(1), \ldots, r(M)$, we may determine the white-noise variance σ_v^2.

2.9 COMPUTER EXPERIMENT: AUTOREGRESSIVE PROCESS OF ORDER 2

To illustrate the theory developed above for the modeling of an AR process, we consider the example of a second-order AR process that is real valued.[2] Figure 2.6 shows the block diagram of the model used to generate this process. Its time-domain description is governed by the second-order difference equation

$$u(n) + a_1 u(n-1) + a_2 u(n-2) = v(n) \tag{2.72}$$

[2]In this example, we follow the approach described by Box and Jenkins (1976).

where $v(n)$ is drawn from a white-noise process of zero mean and variance σ_v^2. Figure 2.7(a) shows one realization of this white-noise process. The variance σ_v^2 is chosen to make the variance of $u(n)$ equal unity.

Conditions for Asymptotic Stationarity

The second-order AR process $u(n)$ has the characteristic equation

$$1 + a_1 z^{-1} + a_2 z^{-2} = 0 \tag{2.73}$$

Let p_1 and p_2 denote the two roots of this equation:

$$p_1,\ p_2 = \tfrac{1}{2}(-a_1 \pm \sqrt{a_1^2 - 4a_2}) \tag{2.74}$$

To ensure the asymptotic stationarity of the AR process $u(n)$, we require that these two roots lie inside the unit circle in the z-plane. That is, both p_1 and p_2 must have a magnitude less than 1. This, in turn, requires that the AR parameters a_1 and a_2 lie in the triangular region defined by

$$\begin{aligned} -1 &\le a_2 + a_1 \\ -1 &\le a_2 - a_1 \\ -1 &\le a_2 \le 1 \end{aligned} \tag{2.75}$$

as shown in Fig. 2.8.

Autocorrelation Function

The autocorrelation function $r(m)$ of an asymptotically stationary AR process for lag m satisfies the difference equation (2.62). Hence, using this equation, we obtain the following second-order difference equation for the autocorrelation function of a second-order AR process:

$$r(m) + a_1 r(m-1) + a_2 r(m-2) = 0, \quad m > 0 \tag{2.76}$$

For the initial values, we have (as will be explained later)

$$\begin{aligned} r(0) &= \sigma_u^2 \\ r(1) &= \frac{-a_1}{1 + a_2}\sigma_u^2 \end{aligned} \tag{2.77}$$

Thus, solving Eq. (2.76) for $r(m)$, we get (for $m > 0$)

$$r(m) = \sigma_u^2 \left[\frac{p_1(p_2^2 - 1)}{(p_2 - p_1)(p_1 p_2 + 1)} p_1^m - \frac{p_2(p_1^2 - 1)}{(p_2 - p_1)(p_1 p_2 + 1)} p_2^m \right] \tag{2.78}$$

where p_1 and p_2 are defined by Eq. (2.74).

There are two specific cases to be considered, depending on whether the roots p_1 and p_2 are real or complex valued, as described next.

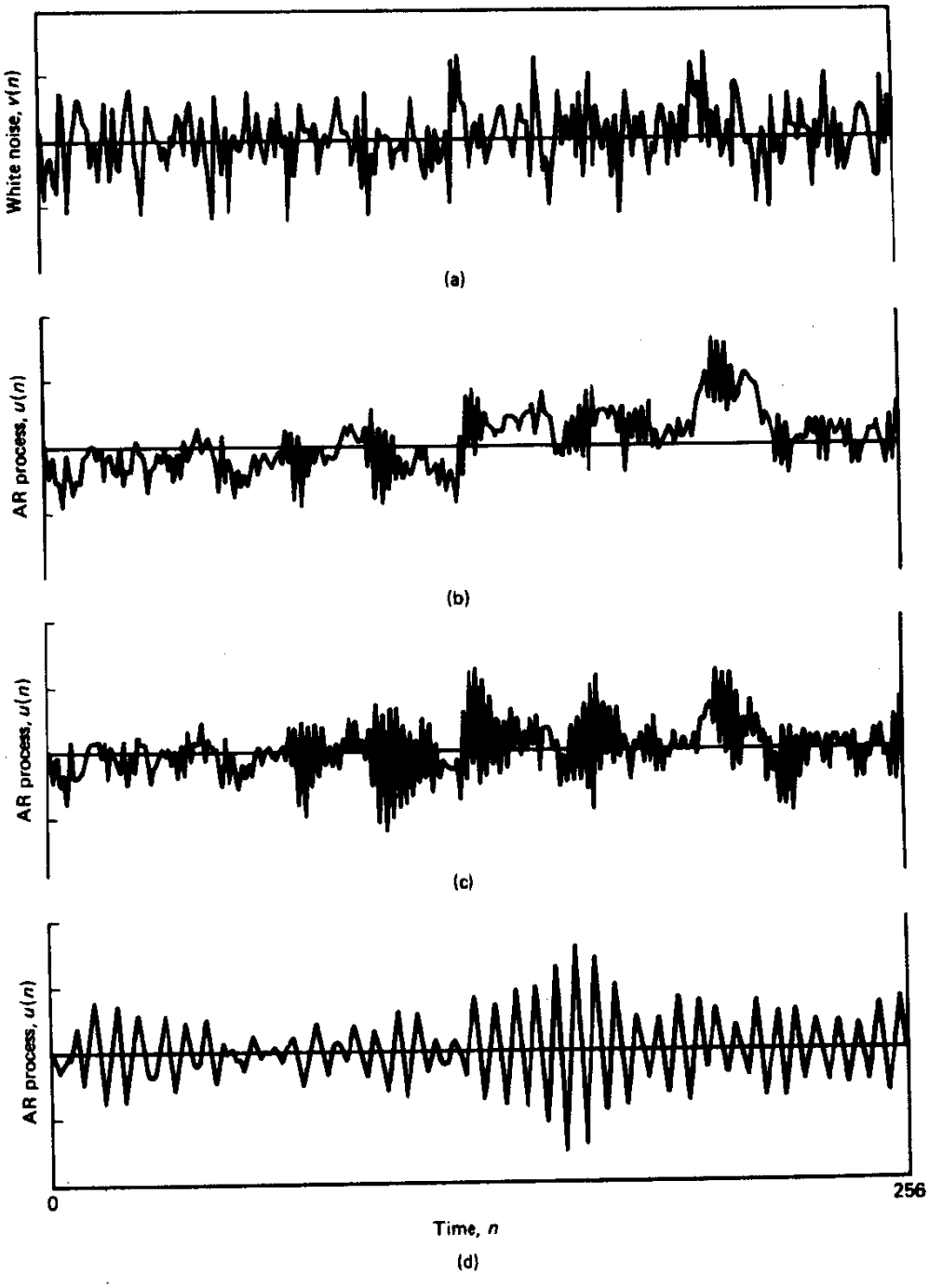

Figure 2.7 (a) One realization of white-noise input; (b), (c), (d) corresponding outputs of AR model of order 2 for parameters of Eqs. (2.79), (2.80), and (2.81), respectively.

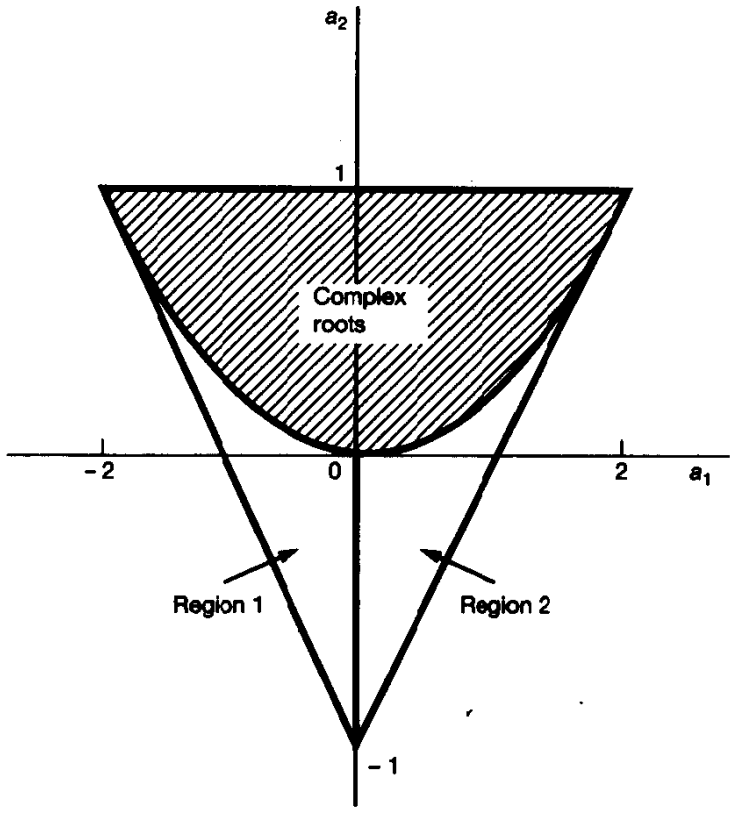

Figure 2.8 Permissible region for the AR parameters a_1 and a_2.

Case 1: *Real Roots.* This case occurs when

$$a_1^2 - 4a_2 > 0$$

which corresponds to regions 1 and 2 below the parabolic boundary in Fig. 2.8. In region 1, the autocorrelation function remains positive as it damps out, corresponding to a positive dominant root. This situation is illustrated in Fig. 2.9(a) for the AR parameters

$$\begin{aligned} a_1 &= -0.10 \\ a_2 &= -0.8 \end{aligned} \tag{2.79}$$

In Fig. 2.7(b), we show the time variation of the output of the model in Fig. 2.6 [with a_1 and a_2 assigned the values given in Eq. (2.79)]. This output is produced by the white-noise input shown in Fig. 2.7(a).

In region 2 of Fig. 2.8, the autocorrelation function alternates in sign as it damps out, corresponding to a negative dominant root. This situation is illustrated in Fig. 2.9(b) for the AR parameters

$$\begin{aligned} a_1 &= 0.1 \\ a_2 &= -0.8 \end{aligned} \tag{2.80}$$

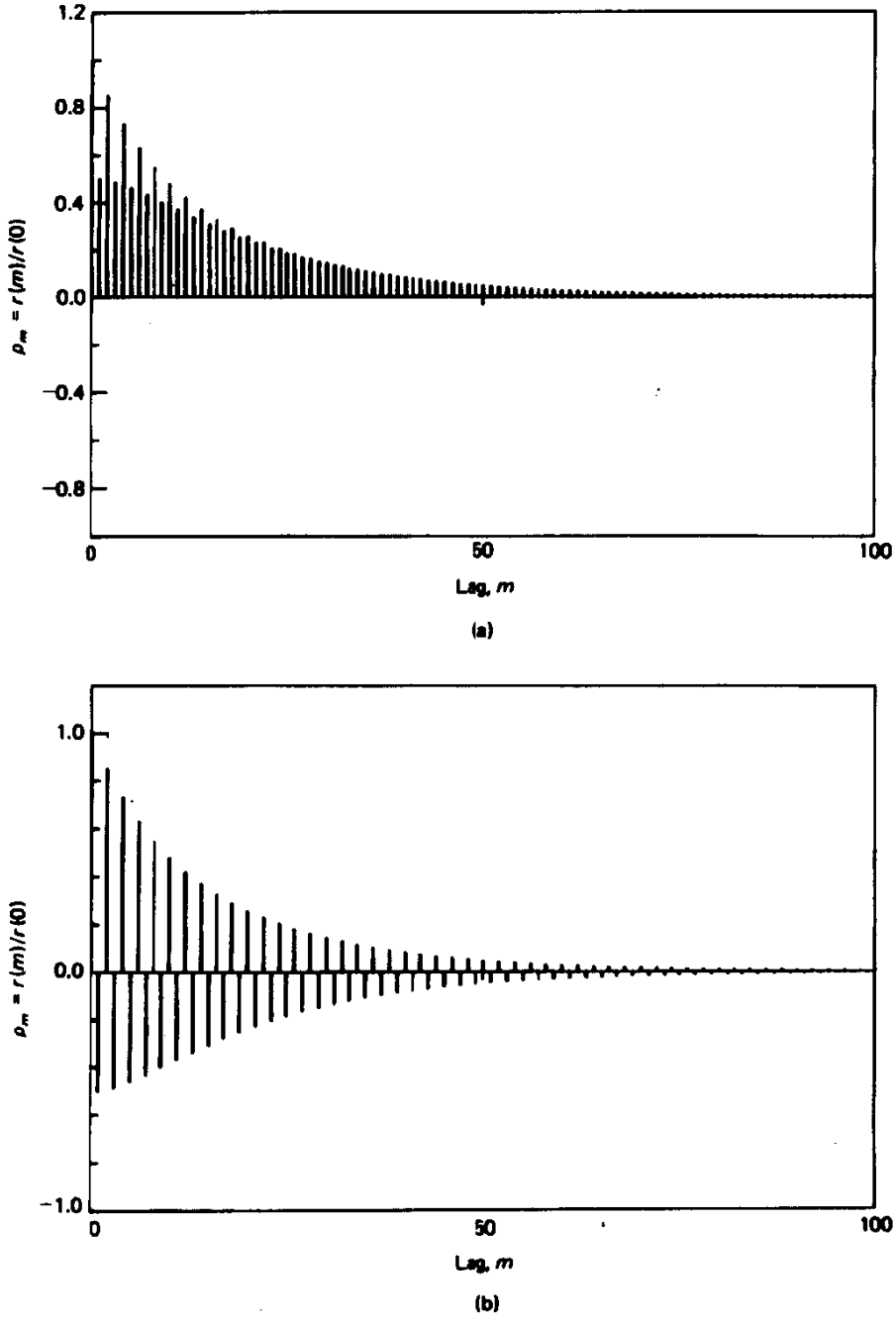

Figure 2.9 Plots of normalized autocorrelation function of real-valued AR(2) process; (a) $r(1) > 0$; (b) $r(1) < 0$; (c) conjugate roots.

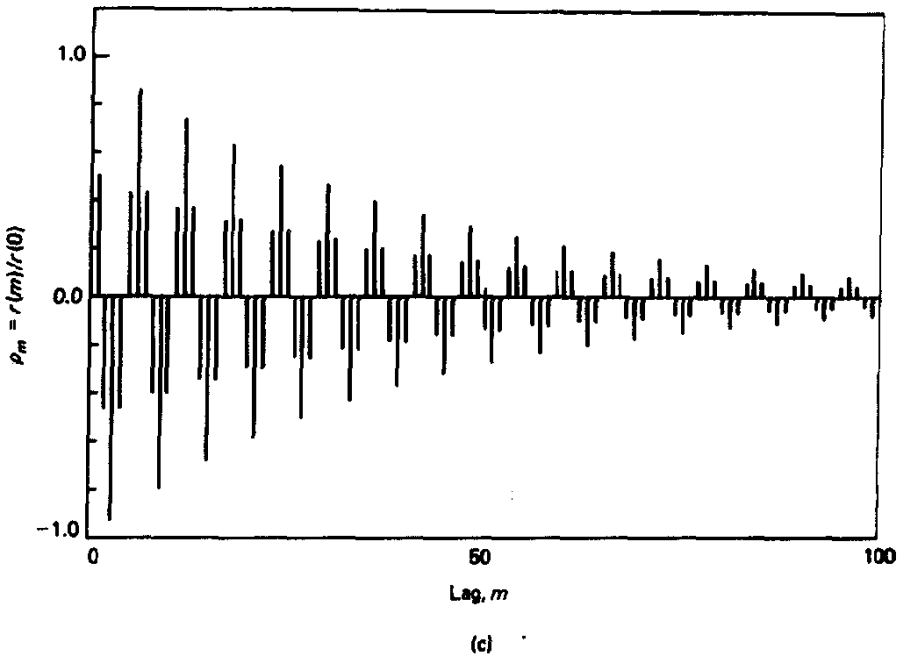

Figure 2.9 *(continued)*

In Fig. 2.7(c) we show the time variation of the output of the model in Fig. 2.6 [with a_1 and a_2 assigned the values given in Eq. (2.80)]. This output is also produced by the white-noise input shown in Fig. 2.7(a).

Case 2: *Complex-Conjugate Roots.* This occurs when

$$a_1^2 - 4a_2 < 0$$

which corresponds to the shaded region shown in Fig. 2.8 above the parabolic boundary. In this case, the autocorrelation function displays a pseudoperiodic behavior, as illustrated in Fig. 2.9(c) for the AR parameters

$$\begin{aligned} a_1 &= -0.975 \\ a_2 &= 0.95 \end{aligned} \tag{2.81}$$

In Fig. 2.7(d) we show the time variation of the output of the model in Fig. 2.6 [with a_1 and a_2 assigned the values given in Eq. (2.81)], which is produced by the white-noise input shown in Fig. 2.7(a).

Yule–Walker Equations

Substituting the value $M = 2$ for the AR model order in Eq. (2.65), we get the following Yule–Walker equations for the second-order AR process:

$$\begin{bmatrix} r(0) & r(1) \\ r(1) & r(0) \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} r(1) \\ r(2) \end{bmatrix} \tag{2.82}$$

where we have used the fact that $r(-1) = r(1)$ for a real-valued process. Solving Eq. (2.82) for w_1 and w_2, we get

$$\begin{aligned} w_1 &= -a_1 = \frac{r(1)[r(0) - r(2)]}{r^2(0) - r^2(1)} \\ w_2 &= -a_2 = \frac{r(0)r(2) - r^2(1)}{r^2(0) - r^2(1)} \end{aligned} \tag{2.83}$$

We may also use Eq. (2.82) to express $r(1)$ and $r(2)$ in terms of the AR parameters a_1 and a_2 as follows:

$$\begin{aligned} r(1) &= \frac{-a_1}{1 + a_2}\sigma_u^2 \\ r(2) &= \left(-a_2 + \frac{a_1^2}{1 + a_2}\right)\sigma_u^2 \end{aligned} \tag{2.84}$$

where $\sigma_u^2 = r(0)$. This solution explains the initial values for $r(0)$ and $r(1)$ that were quoted in Eq. (2.77).

The conditions for asymptotic stationarity of the second-order AR process are given in terms of the AR parameters a_1 and a_2 in Eq. (2.75). Using the expressions for $r(1)$ and $r(2)$ in terms of a_1 and a_2, given in Eq. (2.84), we may reformulate the conditions for asymptotic stationarity as follows:

$$\begin{aligned} -1 < \rho_1 < 1 \\ -1 < \rho_2 < 1 \\ \rho_1^2 < \frac{1}{2}(1 + \rho_2) \end{aligned} \tag{2.85}$$

where ρ_1 and ρ_2 are the normalized *correlation coefficients* defined by

$$\left.\begin{aligned} \rho_1 &= \frac{r(1)}{r(0)} \\ \rho_2 &= \frac{r(2)}{r(0)} \end{aligned}\right\} \tag{2.86}$$

and

Figure 2.10 shows the admissible region for ρ_1 and ρ_2.

Variance of the White-Noise Process

Putting $M = 2$ in Eq. (2.71), we may express the variance of the white-noise process $v(n)$ as

$$\sigma_v^2 = r(0) + a_1 r(1) + a_2 r(2) \tag{2.87}$$

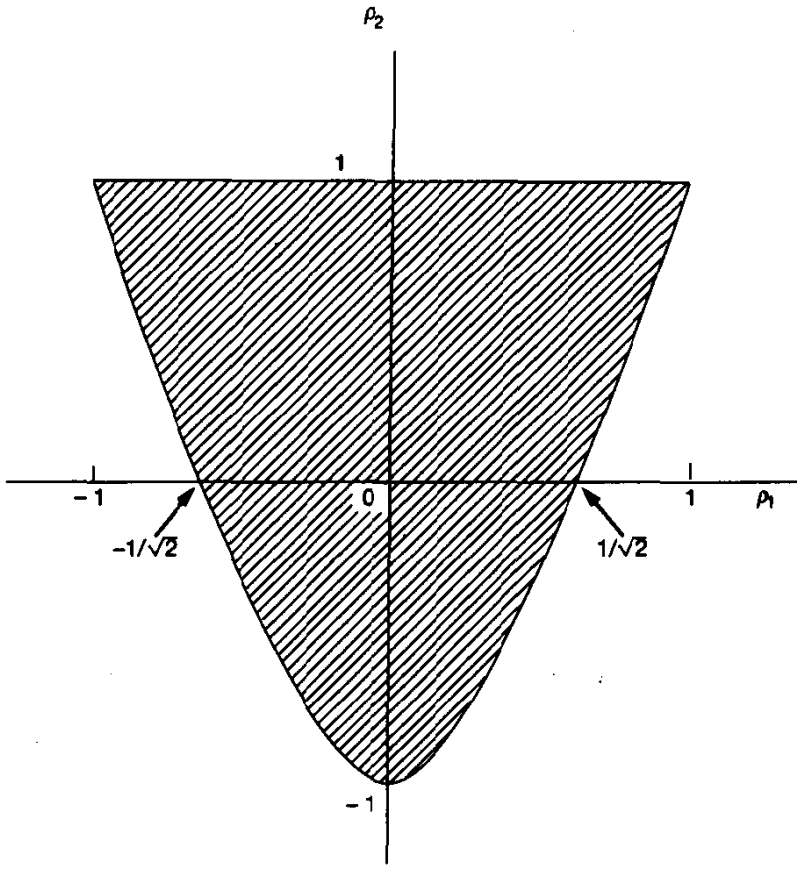

Figure 2.10 Permissible region for parameters of second-order AR process in terms of the normalized correlation coefficients ρ_1 and ρ_2.

Next, substituting Eq. (2.84) in (2.87), and solving for $\sigma_u^2 = r(0)$, we get

$$\sigma_u^2 = \left(\frac{1+a_2}{1-a_2}\right)\frac{\sigma_v^2}{[(1+a_2)^2-a_1^2]} \tag{2.88}$$

For the three sets of AR parameters considered previously, we thus find that the variance of the white noise $v(n)$ has the values given in Table 2.1, assuming that $\sigma_u^2 = 1$.

TABLE 2.1 AR PARAMETERS AND NOISE VARIANCE

a_1	a_2	σ_v^2
−0.10	−0.8	0.27
0.1	−0.8	0.27
−0.975	0.95	0.0731

2.10 SELECTING THE MODEL ORDER

The representation of a stochastic process by a linear model may be used for synthesis or analysis. In *synthesis*, we generate a desired time series by assigning a prescribed set of values to the parameters of the model and feeding it with white noise of zero mean and prescribed variance. In *analysis*, on the other hand, we *estimate* the parameters of the model by processing a given time series of finite length. Insofar as the estimation is statistical, we need an appropriate measure of the fit between the model and the observed data. This implies that unless we have some prior information, the estimation procedure should include a criterion for selecting the *model order* (i.e., the number of independently adjusted parameters in the model). In the case of an AR process defined by Eq. (2.42), the model order equals M. In the case of an MA process defined by Eq. (2.52), the model order equals K. In the case of an ARMA process defined by Eq. (2.53), the model order equals (M, K). Various criteria for model-order selection are described in the literature (Priestley, 1981; Kay, 1988). In this section we describe two important criteria for selecting the order of the model, one of which was pioneered by Akaike (1973, 1974) and the other by Rissanen (1978) and Schwartz (1978); both criteria result from the use of information-theoretic arguments, but in entirely different ways.

An Information-Theoretic Criterion

Let $u_i = u(i)$, $i = 1, 2, \ldots, N$, denote the data obtained by N independent observations of a stationary discrete-time stochastic process, and $g(u_i)$ denote the probability density function of u_i. Let $f_U(u_i|\hat{\boldsymbol{\theta}}_m)$ denote the conditional probability density function of u_i, given $\hat{\boldsymbol{\theta}}_m$, where $\hat{\boldsymbol{\theta}}_m$ is the *estimated* vector of parameters that model the process. Let m be the model order, so that we may write

$$\hat{\boldsymbol{\theta}}_m = \begin{bmatrix} \hat{\theta}_{1m} \\ \hat{\theta}_{2m} \\ \cdot \\ \cdot \\ \cdot \\ \hat{\theta}_{mm} \end{bmatrix} \tag{2.89}$$

We thus have several models that compete with each other to represent the process of interest. An *information-theoretic criterion* (AIC) proposed by Akaike selects the model for which the quantity

$$\text{AIC}(m) = -2L(\hat{\boldsymbol{\theta}}_m) + 2m \tag{2.90}$$

is a minimum. The function $L(\hat{\boldsymbol{\theta}}_m)$ is defined by

$$L(\hat{\boldsymbol{\theta}}_m) = \max \sum_{i=1}^{N} \ln f_U(u_i|\hat{\boldsymbol{\theta}}_m) \tag{2.91}$$

where ln denotes the natural logarithm. The criterion of Eq. (2.91) is derived by minimizing the *Kullback–Leibler mean information*,[3] which is used to provide a measure of the separation or distance between the "unknown" true probability density function $g(u)$ and the conditional probability density function $f_U(u_i|\hat{\boldsymbol{\theta}}_m)$ given by the model in light of the observed data.

The function $L(\hat{\boldsymbol{\theta}}_m)$, constituting the first term on the right-hand side of Eq. (2.90), except for a scalar, is recognized as a *logarithm* of the *maximum-likelihood estimates* of the parameters in the model; for a discussion of the method of maximum likelihood, see Appendix D. The second term, $2m$, represents a *model complexity penalty* that makes AIC(m) an estimate of the Kullback–Leibler mean information.

The first term of Eq. (2.90) tends to decrease rapidly with model order m. On the other hand, the second term increases linearly with m. The result is that if we plot AIC(m) versus model order m, the graph will, in general, show a definite minimum value, and the *optimum order* of the model is determined by that value of m at which AIC(m) attains its minimum value. The minimum value of AIC is called MAIC(minimum AIC).

Minimum Description Length Criterion

Rissanen (1978, 1989) has used an entirely different approach to solve the statistical model identification problem. Specifically, he starts with the notion that a model may be viewed as a device for describing the regular features of a set of observed data, with the objective being that of searching for a model that best captures the regular features or constraints that give the data their special structure. Recognizing that the presence of constraints reduces uncertainty about the data, the objective may equally be that of *encoding* the data in the shortest or least redundant manner; the term "encoding" used here refers to an exact description of the observed data. Accordingly, the number of binary digits needed to encode both the observed data, when advantage is taken of the constraints offered by a model, and the model itself may be used as a criterion for *measuring the amount of the same constraints* and therefore the goodness of the model.

We may thus formally state Rissanen's *minimum description length (MDL) criterion*[4] as follows: Given a data set of interest and a family of competing statistical

[3]In Akaike (1973, 1974, 1977) and Ulrych and Ooe (1983), the criterion of Eq. (2.90) is derived from the principle of minimizing the expectation $E[I(g;f(\cdot|\hat{\boldsymbol{\theta}}_m)]$, where

$$I(g;f(\cdot|\hat{\boldsymbol{\theta}}_m)) = \int_{-\infty}^{\infty} g(u) \ln g(u)\, du - \int_{-\infty}^{\infty} g(u) \ln f_U(u|\hat{\boldsymbol{\theta}}_m)\, du$$

We refer to $I(g;f(\cdot|\hat{\boldsymbol{\theta}}))$ as the *Kullback–Leibler mean information* for discrimination between $g(u)$ and $f_U(u|\hat{\boldsymbol{\theta}}_m)$ (Kullback and Leibler, 1951). The idea is to minimize the information added to the time series by modeling it as an AR, MA, or ARMA process of finite order, since any information added is virtually false information in a real-world situation. Since $g(u)$ is fixed and unknown, the problem reduces to one of maximizing the second term that makes up $I(g;f(\cdot|\hat{\boldsymbol{\theta}}_m))$.

[4]The idea of *minimum description length* of individual recursively definable objects may be traced to Kolmogorov (1968).

models, the best model is the particular one that provides the shortest description length for the data. In mathematical terms, it is defined by[5] (Rissanen, 1978, 1989; Wax, 1995)

$$\text{MDL}(m) = -L(\hat{\boldsymbol{\theta}}_m) + \frac{1}{2} m \ln N \tag{2.92}$$

where m is the number of independently adjusted parameters in the model, and N is the sample size (i.e., the number of observations). As with Akaike's information–theoretic criterion, $L(\hat{\boldsymbol{\theta}}_m)$ is the logarithm of the maximum likelihood estimates of the model parameters. In comparing Eqs. (2.90) and (2.92), we see that the principal difference between the AIC and MDL criterion lies in the structure-dependent term.

According to Rissanen (1989), the MDL criterion offers the following attributes:

- The model permits the shortest encoding of the observed data and captures all the *learnable* properties of the observed data in the best possible manner.
- The MDL criterion is a *consistent* model-order estimator in the sense that it converges to the true model order as the sample size increases.
- The model is optimal in the context of linear regression problems as well as ARMA models.

Perhaps the most significant point to note is the fact that in all of the applications involving the MDL criterion, there has been no anomalous result or a model with undesirable properties reported in the literature.

2.11 COMPLEX GAUSSIAN PROCESSES

Gaussian stochastic processes, or simply *Gaussian processes*, are frequently encountered in both theoretical and applied analysis. In this section we present a summary of some important properties of Gaussian processes that are *complex valued*.[6]

Let $u(n)$ denote a complex Gaussian process consisting of N samples. For the first- and second-order statistics of this process, we assume the following:

1. A *mean* of zero as shown by

$$\mu = E[u(n)] = 0 \qquad \text{for } n = 1, 2, \ldots, N \tag{2.93}$$

[5]Schwartz (1989) has derived a similar result, using a Bayesian approach. In particular, he considers the asymptotic behavior of Bayes estimators under a special class of *priors*. These priors put positive probability on the subspaces that correspond to the competing models. The decision is made by selecting the model that yields the maximum *a posteriori probability*.

It turns out that, in the large sample limit, the two approaches taken by Schwartz and Rissanen yield essentially the same result. However, Rissanen's approach is much more general, whereas Schwartz's approach is restricted to the case that the observations are independent and come from an exponential distribution.

[6]For a detailed treatment of complex Gaussian processes, see the book by Miller (1974). Properties of complex Gaussian processes are also discussed in Kelly et al. (1960), Reed (1962), and McGee (1971).

2. An *autocorrelation function* denoted by

$$r(k) = E[u(n)u^*(n-k)], \quad k = 0, 1, \ldots, N-1 \tag{2.94}$$

The set of autocorrelation functions $\{r(k), k = 0, 1, \ldots, N-1\}$ defines the correlation matrix $\mathbf{R}$ of the Gaussian process $u(n)$.

The shorthand notation $\mathcal{N}(\mathbf{0},\mathbf{R})$ is commonly used to refer to a Gaussian process with a mean vector of zero and correlation matrix $\mathbf{R}$.

Equations (2.93) and (2.94) imply wide-sense stationarity of the process. Knowledge of the mean μ and the autocorrelation function $r(k)$ for varying values of lag k is indeed sufficient for the complete characterization of the complex Gaussian process $u(n)$. In particular, it may be shown that the *joint probability density function* of N samples of the process so described is as follows (Kelly et al., 1960):

$$\mathbf{f}_{\mathbf{U}}(\mathbf{u}) = \frac{1}{(2\pi)^N \det(\Lambda)} \exp\left(-\frac{1}{2}\mathbf{u}^H \Lambda^{-1} \mathbf{u}\right) \tag{2.95}$$

where $\mathbf{u}$ is the N-by-1 data vector; that is,

$$\mathbf{u} = [u(1), u(2), \ldots, u(N)]^T \tag{2.96}$$

and Λ is the N-by-N Hermitian-symmetric *moment matrix* of the process, defined in terms of the correlation matrix $\mathbf{R} = \{r(k)\}$ as

$$\begin{aligned} \Lambda &= \tfrac{1}{2} E[\mathbf{u}\mathbf{u}^H] \\ &= \tfrac{1}{2}\mathbf{R} \end{aligned} \tag{2.97}$$

Note that the joint probability density function $f_{\mathbf{U}}(\mathbf{u})$ is $2N$-dimensional, where the factor 2 accounts for the fact that each of the N samples of the process has a real and an imaginary part. Note also that the probability density function of a single sample $u(n)$ of the process, which is a special case of Eq. (2.95), is given by

$$f_U(u) = \frac{1}{\pi\sigma^2} \exp\left(-\frac{|u|^2}{\sigma^2}\right) \tag{2.98}$$

where $|u|$ is the magnitude of the sample $u(n)$ and σ^2 is its variance.

Based on the representation described herein, we may now summarize some important properties of a *zero-mean complex Gaussian process u(n) that is wide-sense stationary* as follows:

1. The process $u(n)$ is *stationary in the strict sense.*
2. The process $u(n)$ is *circularly complex* in the sense that any two different samples $u(n)$ and $u(k)$ of the process satisfy the condition

$$E[u(n)u(k)] = 0 \qquad \text{for } n \neq k \tag{2.99}$$

It is for this reason that the process $u(n)$ is often referred to as a *circularly complex Gaussian process.*

3. Suppose that $u_n = u(n)$, for $n = 1, 2, \ldots, N$, are samples picked from a zero-mean, complex Gaussian process $u(n)$. We may thus state Property 3 in two parts (Reed, 1962):
   (a) If $k \neq l$, then

$$E[u_{s_1}^* u_{s_2}^* \cdots u_{s_k}^* u_{t_1} u_{t_2} \cdots u_{t_l}] = 0 \tag{2.100}$$

   where s_i and t_j are integers selected from the available set $\{1, 2, \ldots, N\}$.
   (b) If $k = l$, then

$$E[u_{s_1}^* u_{s_2}^* \cdots u_{s_l}^* u_{t_1} u_{t_2} \cdots u_{t_l}] = E[u_{s_{\pi(1)}}^* u_{t_1}] E[u_{s_{\pi(2)}}^* u_{t_2}] \cdots E[u_{s_{\pi(l)}}^* u_{t_l}] \tag{2.101}$$

   where π is a permutation of the set of integers $\{1, 2, \ldots, l\}$, and $\pi(j)$ is the jth element of that permutation. For the set of integers $\{1, 2, \ldots, l\}$ we have a total of $l!$ possible permutations. This means that the right-hand side of Eq. (2.101) consists of the product of $l!$ expectation product terms. Equation (2.101) is called the *Gaussian moment factoring theorem.*

Example 2

Consider the case of $N = 4$, for which the complex Gaussian process $u(n)$ consists of the four samples u_1, u_2, u_3, and u_4. Hence, the use of the Gaussian moment factoring theorem given in Eq. (2.101) yields the following useful identity:

$$E[u_1^* u_2^* u_3 u_4] = E[u_1^* u_3] E[u_2^* u_4] + E[u_2^* u_3] E[u_1^* u_4] \tag{2.102}$$

For other useful identities derived from the Gaussian moment factoring theorem, see Problem 11.

2.12 SUMMARY AND DISCUSSION

In this chapter we studied the partial characterization of a stationary discrete-time stochastic process. Such a characterization is uniquely described in terms of two statistical parameters:

1. The mean, which is a constant
2. The autocorrelation function, which depends only on the time difference between any two samples of the process

The mean of the process may naturally be zero, or it can always be subtracted from the process to yield a new process of zero mean. For this reason, in much of the discussion in subsequent chapters of this book, the mean of the process is assumed to be zero. Thus, given an M-by-1 observation vector $\mathbf{u}(n)$ known to belong to a complex, stationary, discrete-time stochastic process of zero mean, we may partially describe it by defining an M-by-M correlation matrix $\mathbf{R}$ as the statistical expectation of the outer product of $\mathbf{u}(n)$ with

itself. The matrix $\mathbf{R}$ is Hermitian, Toeplitz, and almost always positive definite; the latter property means that $\mathbf{R}$ is almost always nonsingular, and therefore the inverse matrix $\mathbf{R}^{-1}$ exists.

Another topic discussed in the chapter is the notion of a stochastic model, the need for which arises when we are given a set of experimental data known to be of a statistical nature, and the requirement is to analyze the data. In this context, we may mention two general requirements for a suitable model:

1. An *adequate number of adjustable parameters* for the model to capture the essential information content of the input data
2. *Mathematical tractability* of the model

The first requirement, in effect, means that the complexity of the model should closely match the complexity of the underlying physical mechanism responsible for generating the input data; in so doing, problems associated with underfitting or overfitting the input data are avoided. The second requirement is usually satisfied by the choice of a linear model.

Within the family of linear stochastic models, the autoregressive (AR) model is often preferred over the moving average (MA) model and the autoregressive-moving average (ARMA) model for an important reason: unlike an MA or ARMA model, computation of the AR coefficients is governed by a system of linear equations, namely, the Yule–Walker equations. Moreover, except for a predictable component, we may approximate a stationary discrete-time stochastic process by an AR model of sufficiently high order, subject to certain restrictions. To select a suitable value for the model order, we may use an information-theoretic criterion (AIC) according to Akaike or the minimum-description length (MDL) criterion according to Rissanen. A useful feature of the MDL criterion is that it is a consistent model-order estimator.

PROBLEMS

1. The sequences $y(n)$ and $u(n)$ are related by the difference equation

$$y(n) = u(n + a) - u(n - a)$$

where a is a constant. Evaluate the autocorrelation function of $y(n)$ in terms of that of $u(n)$.

2. Consider a correlation matrix $\mathbf{R}$ for which the inverse matrix $\mathbf{R}^{-1}$ exists. Show that $\mathbf{R}^{-1}$ is Hermitian.

3. (a) Equation (2.26) relates the $(M + 1)$-by-$(M + 1)$ correlation matrix $\mathbf{R}_{M+1}$, pertaining to the observation vector $\mathbf{u}_{M+1}(n)$ taken from a stationary stochastic process, to the M-by-M correlation matrix $\mathbf{R}_M$ of the observation vector $\mathbf{u}_M(n)$ taken from the same process. Evaluate the inverse of the correlation matrix $\mathbf{R}_{M+1}$ in terms of the inverse of the correlation matrix $\mathbf{R}_M$.

   (b) Repeat your evaluation using Eq. (2.27).

4. A first-order autoregressive (AR) process $u(n)$, which is real-valued, satisfies the real-valued difference equation

$$u(n) + a_1 u(n-1) = v(n)$$

where a_1 is a constant, and $v(n)$ is a white-noise process of variance σ_v^2.

(a) Show that if $v(n)$ has a nonzero mean, the AR process $u(n)$ is nonstationary.

(b) For the case when $v(n)$ has zero mean, and the constant a_1 satisfies the condition $|a_1| < 1$, show that the variance of $u(n)$ equals

$$\text{var}[u(n)] = \frac{\sigma_v^2}{1 - a_1^2}$$

(c) For the conditions specified in part (b), find the autocorrelation function of the AR process $u(n)$. Sketch this autocorrelation function for the two cases $0 < a_1 < 1$ and $-1 < a_1 < 0$.

5. Consider an autoregressive process $u(n)$ of order 2, described by the difference equation

$$u(n) = u(n-1) - 0.5u(n-2) + v(n)$$

where $v(n)$ is a white-noise process of zero mean and variance 0.5.

(a) Write the Yule–Walker equations for the process.

(b) Solve these two equations for the autocorrelation function values $r(1)$ and $r(2)$.

(c) Find the variance of $u(n)$.

6. Consider a wide-sense stationary process that is modeled as an AR process $u(n)$ of order M. The set of parameters made up of the average power P_0 and the AR coefficients $a_1, a_2, \ldots, a_M$ bear a one-to-one correspondence with the autocorrelation sequence $r(0), r(1), r(2), \ldots, r(M)$, as shown by

$$\{r(0), r(1), r(2), \ldots, r(M)\} \rightleftharpoons \{P_0, a_1, a_2, \ldots, a_M\}$$

Justify the validity of this statement.

7. Evaluate the transfer functions of the following two stochastic models:

(a) The MA model of Fig. 2.3

(b) The ARMA model of Fig. 2.4.

(c) Specify the conditions for which the transfer function of the ARMA model of Fig. 2.4 reduces (1) to that of an AR model, and (2) to that of an MA model.

8. Consider an MA process $x(n)$ of order 2 described by the difference equation

$$x(n) = v(n) + 0.75v(n-1) + 0.25v(n-2)$$

where $v(n)$ is a zero-mean white-noise process of unit variance. The requirement is to approximate this process by an AR process $u(n)$ of order M. Do this approximation for the following orders:

(a) $M = 2$

(b) $M = 5$

(c) $M = 10$

Comment on your results. How big would the order M of the AR process $u(n)$ have to be for it to be equivalent to the MA process $x(n)$ exactly?

9. A time series $u(n)$ obtained from a wide-sense stationary stochastic proces of zero mean and correlation matrix $\mathbf{R}$ is applied to an FIR filter of impulse response w_n. This impulse response defines the coefficient vector $\mathbf{w}$.

(a) Show that the average power of the filter output is equal to $\mathbf{w}^H\mathbf{R}\mathbf{w}$.

(b) How is the result in part (a) modified if the stochastic process at the filter input is a white noise of variance σ^2?

10. A general linear complex-valued process $u(n)$ is described by

$$u(n) = \sum_{k=0}^{\infty} b_k^* v(n-k)$$

where $v(n)$ is a white noise process, and b_k is a complex coefficient. Justify the following statements:

(a) If the process $v(n)$ is Gaussian, then the original process $u(n)$ is also Gaussian.

(b) Conversely, a Gaussian process $u(n)$ implies that the process $v(n)$ is Gaussian.

11. Consider a complex Gaussian process $u(n)$. Let $u(n) = u_n$. Using the Gaussian moment factoring theorem, demonstrate the following identities:

(a) $E[(u_1^* u_2)^k] = k!\,(E[u_1^* u_2])^k$

(b) $E[|u|^{2k}] = k!\,(E[|u|^2])^k$