STOCHASTIC PROCESSES, DETECTION AND ESTIMATION

6.432 Course Notes

Alan S. Willsky, Gregory W. Wornell, and Jeffrey H. Shapiro

Department of Electrical Engineering and Computer Science
Massachusetts Institute of Technology
Cambridge, MA 02139

Fall 2003

4

Stochastic Processes and Systems

In preceding chapters we have focussed on random variables and random vectors, and their manipulation in problems of detection and estimation. We now want to broaden our development of detection and estimation theory to accomodate sequences and waveforms that are random as well. Sequences and waveforms of this type are referred to as *random* or *stochastic* processes—the two terms are used interchangeably.

We typically refer to random sequences as *discrete-time* stochastic processes, and random waveforms as *continuous-time* stochastic processes. Note that this implies we are restricting our attention to functions of a single variable which we view for convenience as "time." However, it is important to appreciate that in many applications of interest the function is of one of multiple variables. Examples include the intensity values of an image, which are functions of the two spatial variables; a video signal, which is a function of three variables—two spatial and one temporal; and the temperature distribution of the ocean, which is a function of three spatial variables.

We also restrict our attention to scalar-valued functions, though there are again many applications where vector-valued functions (of, say, time) arise rather naturally. Examples include the received acoustic signals at an underwater multi-sensor array, or the imagery corresponding to distinct spectral bands from a satellite performing passive remote-sensing of the earth.

It is also worth remarking that although we don't consider them in our treatment, many of the techniques and concepts we develop do have natural extensions to these multi-dimensional and multi-valued processes.

In this chapter we introduce and develop the concept of stochastic processes in both discrete and continuous time. For example, we'll develop convenient ways of viewing and analyzing such processes based on their probabilistic descriptions, and we'll explore how probabilistic descriptions of such processes are transformed when applied as inputs to various classes of systems.

4.1 STOCHASTIC PROCESSES AND SAMPLE SPACE

There are a several ways to view a stochastic process, each of which affords its own insight. From one perspective, for example, we may view a stochastic process as a collection of random variables indexed in time. For a discrete-time stochastic process, $x[n_0]$ is the random variable associated with the time $n = n_0$. Since time is integer-valued in the discrete-time case, there are a countably infinite number of such random variables. For a continuous-time stochastic process, $x(t_0)$ is the random variable associated with the time $t = t_0$, and since time is real-valued, a continuum of (i.e., uncountably many) random variables is involved.

To develop a more general perspective, recall that a random variable x can be viewed as a function on a probability space $(\Omega, \Pr[\cdot])$, where for any $\omega \in \Omega$, the value $x(\omega)$ is a *realization* of the random variable, i.e., the result of a particular experiment corresponding to the generation of the random variable x. In a similar manner, a stochastic process can be expressed as a function $x(t, \omega)$ of the time index t, and the event index ω jointly. From this perspective, with $t = t_0$ fixed we have our earlier insight—that $x(t_0, \omega)$ corresponds to a function of ω, i.e., a random variable at the corresponding instant in time. However, if we fix $\omega = \omega_0$ and view $x(t, \omega_0)$ as a function of t, we obtain a particular realization of the stochastic process, which we generally refer to as a *sample path* of the process. Hence, we can think of the event space Ω as describing experiments, the outcomes of which yield entire, observed sample paths.

Of course, in a similar manner we can view a discrete-time stochastic process as a function of both n and ω, i.e., $x[n, \omega)$, for ω fixed is a realization of the process, namely a sequence of values for $n = 0, \pm 1, \pm 2, \ldots$, which collectively form a sample path of the process.

In general we will suppress in our notation the explicit dependency on the underlying random mechanism that generates a stochastic process. For example, just as we have denoted random variables as x rather than $x(\omega)$, we will generally write $x(t)$ rather than $x(t, \omega)$. Nevertheless the sample path perspective will frequently prove important to us in much of our development of this topic. For example, this perspective highlights the structure of temporal evolution of the process and the relationships among values of the process corresponding to different times. In many situations, for instance, the continuous-time random phenomena that we encounter evolve smoothly, and as a result the stochastic processes corresponding to these phenomena have continuous sample paths. On the other hand,

we will also encounter important stochastic processes whose sample paths are not continuous, but which are, for example, piecewise-constant. In such cases the sample path perspective is particularly valuable. Finally, the sample path perspective is useful in understanding the effects of filtering and other forms of processing on such signals.

4.2 CHARACTERIZATIONS OF STOCHASTIC PROCESSES

In general, probabilistic characterizations of a stochastic process involve specifying the *joint* probabilistic description of the process at different points in time. A remarkably broad class of stochastic processes are, in fact, completely characterized by the joint probability density functions for arbitrary collections of samples of the process. In particular, *separable* stochastic processes are completely characterized by the collection of N*th-order probability densities* of the form

$$p_{x(t_1),x(t_2),\ldots,x(t_N)}(x_1, x_2, \ldots, x_N) \tag{4.1}$$

for *every* possible choice of N and the time instants $t_1, t_2, \ldots, t_N$. We refer to the densities (4.1) as the finite-dimensional densities of the process. We also note that the obvious counterparts to this characterization apply to discrete-time stochastic processes $x[n]$.

It is relatively easy to see that the finite-dimensional distributions constitute a complete probabilitistic description for discrete-time processes defined over finite time intervals, since the complete probabilistic descriptions of a such a process $x[n]$ where $n = 0, 1, \ldots, M-1$ is obviously given by

$$p_{x[0],x[1],\ldots,x[M-1]}(x_0, x_1, \ldots, x_{M-1})$$

where M is the size of the time interval. However, to show that finite-dimensional distributions can completely characterize stochastic processes defined on infinite time intervals—not only in discrete-time but also in continuous-time—is comparatively much more difficult. In fact, this represents a rather deep result in the theory of stochastic processes; we'll discuss some aspects of this issue later in the chapter.

The various finite-dimensional distributions of a random process are in general strongly interrelated. In particular, dependencies result from consistency constraints, i.e., from the fact that lower-order densities can be obtained from higher-order ones via integration; for example, from a third-order density we can obtain second-order densities as marginals, e.g.,

$$p_{x(t_1),x(t_2)}(x_1, x_2) = \int_{-\infty}^{+\infty} p_{x(t_1),x(t_2),x(t_3)}(x_1, x_2, x_3)\, dx_3. \tag{4.2}$$

Nevertheless, despite the constraints imposed by relationships such as (4.2), the explicit characterization of a stochastic process in terms of its finite-dimensional distributions is in general a rather unwieldy description. Consequently, in practice the development of stochastic processes follows one of two approaches: either

only partial statistical descriptions of the processes are pursued, or the focus is restricted to processes with special structure or properties that substantially simplify their description. We will consider both of these strategies in subsequent sections of this chapter. Let us begin, however, with some preliminary examples.

One of the simplest stochastic processes is *discrete-time white noise*. A discrete-time white noise is defined to be a process whose samples are uncorrelated but otherwise have arbitrary (and possibly time-varying) means and variances. The motivation for this terminology is in fact not obvious at this point, but will become more apparent as we more fully develop the concept of white noise in the chapter.

Example 4.1

A particularly simple discrete-time white noise corresponds to the case in which the samples $x[n]$ are zero-mean and have identical variances, i.e.,

$$E[x[n]] = 0 \tag{4.3a}$$

$$E[x[n]x[m]] = \sigma^2\delta[n-m] \tag{4.3b}$$

where $\delta[n]$ is the discrete-time unit impulse.

It is important to emphasize that the characterization of $x[n]$ given by (4.3) is *not* complete. In fact, it doesn't even specify the complete probability density for $x[n]$ at any point in time. Nevertheless this type of partial characterization, which is referred to as a second-order characterization, is of considerable practical importance, and in fact is a significant focus in this and subsequent chapters.

An example of a particular process satisfying (4.3) with an extremely compact complete characterization corresponds to the case in which the $x[n]$ are independent Gaussian random variables, i.e., $x[n] \sim N(0,\sigma^2)$, and in this case the resulting process is referred to as a Gaussian white noise. A typical sample path of this process is depicted in Fig. 4.1. In this case the finite-dimensional distributions of $x[n]$ can be readily constructed via

$$p_{x[n_1],x[n_2],\ldots,x[n_N]}(x_1,x_2,\ldots,x_N) = \prod_{i=1}^{N} N(x_i;0;\sigma^2), \tag{4.4}$$

which verifies that the characterization is complete. Later within this chapter we will fully develop the concept of a Gaussian random process, and as we will see the complete description of such processes is always conveniently compact.

Example 4.2

A powerful mechanism for constructing or describing random processes is to define them in terms of simpler processes. For example, suppose that $x[n]$ is a discrete-time white noise process satisfying (4.3), and consider a second stochastic process $y[n]$ which is the output of a linear system driven by $x[n]$; specifically,

$$y[n] = \frac{1}{2}y[n-1] + x[n] \qquad n = 1,2,\ldots, \tag{4.5}$$

where $y[0]$ has zero-mean and variance σ_0^2, and is uncorrelated with $x[n]$ for all n.

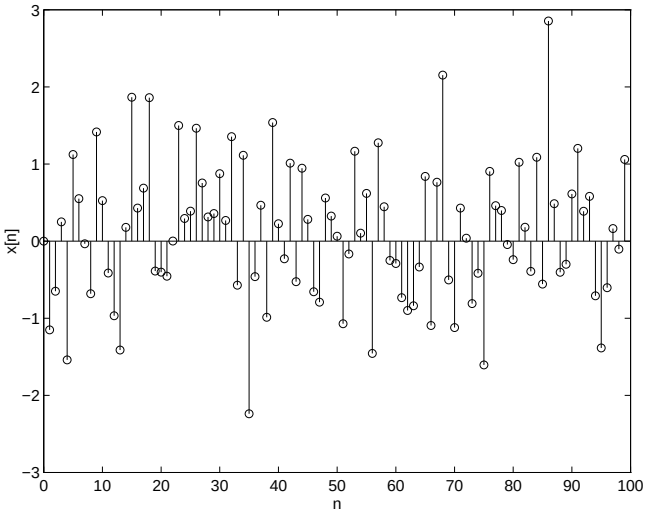

Figure 4.1. A typical sample function of a particular discrete-time Gaussian white noise process with $\mathrm{var}\, x[n] = \sigma^2 = 1$.

Since the system (4.5) is linear, we may always express a vector of arbitrary samples of the output as a linear transformation of an appropriate set of input samples together with the initial condition. In particular, for an arbitrary set of time instants $0 \leq n_1 < n_2 < \cdots < n_N$ we can write

$$\begin{bmatrix} y[n_1] \\ y[n_2] \\ \vdots \\ y[n_N] \end{bmatrix} = \mathbf{A} \begin{bmatrix} y[0] \\ x[1] \\ \vdots \\ x[n_N] \end{bmatrix}, \tag{4.6}$$

where the matrix $\mathbf{A}$ is derived from (4.5). As we will develop in this chapter, for such linear systems the second-order description of $y[n]$ is relatively easy to obtain from the second-order description of the input $x[n]$. Furthermore, as we'll see, when the processes involved are Gaussian, such second-order characterizations provide complete characterizations.

The stochastic processes introduced in the preceding examples have a significant amount of randomness in their evolution over time. In contrast, there are also important classes of stochastic processes with far more constrained behavior, as the following example illustrates.

Example 4.3

Consider the continuous-time sinusoidal signal

$$x(t) = A\cos(2\pi f_0 t + \Theta), \tag{4.7}$$

where one or more of the constants A, f_0, and Θ might be random. For example sinusoidal signals subject to uncertain amplitude modulation might be modeled as in (4.7) with A taken as a random variable A. Similarly frequency modulation effects can be captured through the probabilistic description of f_0, and random phase effects through Θ. In any of these cases, however the nature of the sample paths of the

process—i.e., realizations $x(t)$ of $x(t)$, corresponding to particular values A, f_0, and Θ of A, f_0, and Θ, respectively—are quite regular in form.

For example, consider the particular case in which A and f_0 are known constants but Θ is uniformly distributed between 0 and 2π. In this case using the method of derived distributions we find that the first-order density for $x(t)$ is given by

$$p_{x(t)}(x) = \begin{cases} 1/\left(\pi\sqrt{A^2 - x^2}\right) & |x| < A \\ 0 & \text{otherwise} \end{cases}. \tag{4.8}$$

Next, to obtain the joint probability density for $x(t)$ at two times t_1 and t_2, we first express the density as

$$p_{x(t_1),x(t_2)}(x_1, x_2) = p_{x(t_1)}(x_1)\, p_{x(t_2)|x(t_1)}(x_2|x_1) \tag{4.9}$$

and then recognize that it remains only to obtain the conditional density in (4.9). In this case the constrained nature of the sample path plays a very significant role, since both of the values $x(t_1)$ and $x(t_2)$ must lie on the same sinusoid with known amplitude and frequency and with only the phase being random. Thus given the value of $x(t_1)$, it is straightforward to verify that there are only *two* possible values for $x(t_2)$. Specifically, using some basic trigonometric identities we obtain

$$\begin{aligned} x(t_2) &= A\cos(2\pi f_0 t_2 + \Theta) \\ &= A\cos[(2\pi f_0(t_2 - t_1)) + (2\pi f_0 t_1 + \Theta)] \\ &= A\cos(2\pi f_0 t_1 + \Theta)\cos(2\pi f_0(t_2 - t_1)) \\ &\quad - A\sin(2\pi f_0 t_1 + \Theta)\sin(2\pi f_0(t_2 - t_1)) \\ &= x(t_1)\cos(2\pi f_0(t_2 - t_1)) \pm \sqrt{A^2 - x^2(t_1)}\sin(2\pi f_0(t_2 - t_1)). \end{aligned} \tag{4.10}$$

In turn, from (4.10) we can conclude that the conditional density in (4.9) consists of two impulses at the two possible values for $x(t_2)$.

Finally, it is equally straightforward to verify that the third-order densities are even more highly constrained. In particular, provided the three sample times t_1, t_2, and t_3 are not separated from each other by a multiple of the Nyquist sampling period for this sinusoid, i.e., $1/(2f_0)$, then the values of $x(t_1)$ and $x(t_2)$ *uniquely determine* $x(t_3)$. As a result, the density $p_{x(t_3)|x(t_1),x(t_2)}(x_3|x_1, x_2)$ generally consists of a single impulse.

Example 4.4

There are also important classes of stochastic processes that take on only a discrete set of values. A well-studied example of such a process is the (homogeneous) *Poisson counting process* $N(t)$. As a counting process, the sample paths of a Poisson counting process have the property that they are nondecreasing and take on only nonnegative integer values. Each "count" corresponds to a jump in the value of the process to the next larger integer. A typical sample path of a Poisson counting process is depicted in Fig. 4.2.

The Poisson counting process in particular has the additional property that the number of counts over any interval in time—i.e., the increment to be added to $N(t)$ over that interval—is independent of the behavior of $N(t)$ prior to that interval.

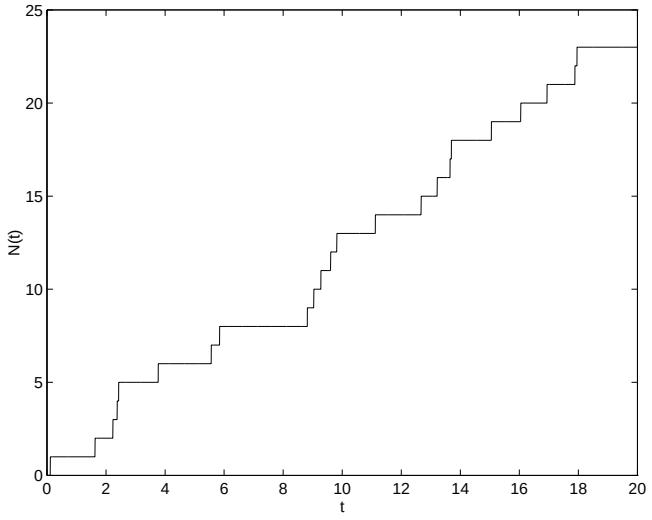

Figure 4.2. A typical sample function of a Poisson counting process.

Formally, the homogeneous Poisson Counting Process is defined as follows. The process starts at 0, i.e., $N(0) = 0$, and for any times t and s such that $t > s \geq 0$, the increment $N(t) - N(s)$ is independent of $N(\tau)$ for all $\tau \leq s$. Furthermore, this increment has the following Poisson distribution

$$\Pr[N(t) - N(s) = k \mid N(\tau), \tau \leq s] = \Pr[N(t) - N(s) = k] \tag{4.11}$$

$$= \frac{[\lambda(t-s)]^k e^{-\lambda(t-s)}}{k!}. \tag{4.12}$$

Several properties of $N(t)$ follow from this definition. For example, since $N(0) = 0$, (4.12) immediately implies that

$$\Pr[N(t) = k] = \frac{(\lambda t)^k e^{-\lambda t}}{k!}. \tag{4.13}$$

The parameter λ is referred to as the *rate* of the Poisson process.[1] To see that this is rather natural, observe that when $t - s = \delta t$ is sufficiently small, then expanding $e^{-\lambda\,\delta t}$ in (4.12) via a Taylor series yields[2]

$$\Pr[N(t+\delta t) - N(t) = k] = \begin{cases} 1 - \lambda\delta t + o(\delta t) & k = 0 \\ \lambda\delta t + o(\delta t) & k = 1 \\ o(\delta t) & k \geq 2 \end{cases} \tag{4.14}$$

[1]There are also classes of Poisson counting processes where the rate parameter is time varying, i.e., $\lambda = \lambda(t)$. These are referred to as inhomogeneous Poisson counting processes, and they will be explored in some detail in Section 4.7.5.

[2]The order notation $o(\cdot)$ denotes terms of higher than first order, so that

$$\lim_{\delta t \to 0} \frac{o(\delta t)}{\delta t} = 0.$$

Thus, for sufficiently small δt, (4.14) implies

$$\Pr\left[N(t+\delta t)-N(t)=k\right] \approx \begin{cases} 1-\lambda\delta t & k=0 \\ \lambda\delta t & k=1 \end{cases} \tag{4.15}$$

Exploiting (4.11), we can readily obtain a complete statistical characterization of the process $N(t)$. In particular, it is straightforward to determine arbitrary finite-dimensional distributions of $N(t)$. For example, if $0=t_0\le t_1\le t_2\le\cdots\le t_N$ and $k_0=0$, then

$$\begin{aligned} &\Pr\left[N(t_1)=k_1, N(t_2)=k_2,\ldots,N(t_N)=k_N\right] \\ &\quad= \Pr\left[N(t_1)=k_1, N(t_2)-N(t_1)=k_2-k_1,\ldots,N(t_N)-N(t_{N-1})=k_N-k_{N-1}\right] \\ &\quad= \prod_{i=1}^{N}\Pr\left[N(t_i)-N(t_{i-1})=k_i-k_{i-1}\right] = e^{-\lambda t_N}\prod_{i=1}^{N}\frac{[\lambda(t_i-t_{i-1})]^{k_i-k_{i-1}}}{(k_i-k_{i-1})!}. \end{aligned} \tag{4.16}$$

However, as we will see, because of the special structure of the sample paths of the counting process, some other equivalent but more powerful statistic characterizations can be developed for these particular processes based on descriptions of *complete* sample paths rather than finite sets of sample times.

Example 4.5

Another important discrete-valued process is the *random telegraph wave* $x(t)$, which is defined directly in terms of the Poisson process. Specifically $x(t)$ takes on only the values of ±1, and each of these two values is equally likely at the initial time $t=0$. Furthermore $x(t)$ switches values whenever an underlying Poisson process $N(t)$ of the type described in Example 4.4 changes value. More precisely,

$$x(t)=x(0)(-1)^{N(t)}=\begin{cases} +x(0) & \text{if } N(t) \text{ is even} \\ -x(0) & \text{if } N(t) \text{ is odd} \end{cases}. \tag{4.17}$$

where $N(t)$ and $x(0)$ are independent. Once again our characterization of this process is complete since we can construct arbitrary finite-dimensional distributions from the specifications given. For example, for $t_2\ge t_1\ge 0$

$$\begin{aligned} &\Pr\left[x(0)=1, x(t_1)=1, x(t_2)=-1\right] \\ &\quad= \Pr\left[x(0)=1\right]\Pr\left[N(t_1) \text{ is even}\right]\Pr\left[N(t_2)-N(t_1) \text{ is odd}\right] \\ &\quad= \frac{1}{2}\cosh(\lambda t_1)\sinh[\lambda(t_2-t_1)]e^{-\lambda t_2}, \end{aligned} \tag{4.18}$$

where we have used

$$\begin{aligned} \Pr\left[N(t)-N(s) \text{ is even}\right] &= \sum_{k:\text{even}}\frac{[\lambda(t-s)]^k e^{-\lambda(t-s)}}{k!} \\ &= e^{-\lambda(t-s)}\cosh[\lambda(t-s)] \end{aligned} \tag{4.19}$$

$$\begin{aligned} \Pr\left[N(t)-N(s) \text{ is odd}\right] &= \sum_{k:\text{odd}}\frac{[\lambda(t-s)]^k e^{-\lambda(t-s)}}{k!} \\ &= e^{-\lambda(t-s)}\sinh[\lambda(t-s)]. \end{aligned} \tag{4.20}$$

A typical sample path of a random telegraph wave is depicted in Fig. 4.3.

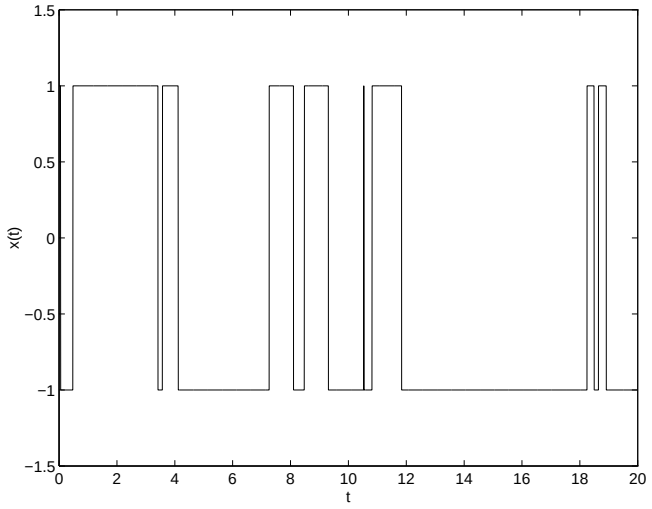

Figure 4.3. A typical sample function of a random telegraph wave.

4.3 SECOND-ORDER CHARACTERIZATIONS

The second-order characterization for a continuous-time stochastic process $x(t)$ consists of two deterministic functions: the *mean function*

$$m_x(t) = E[x(t)] \tag{4.21}$$

and the *(auto)covariance function*

$$K_{xx}(t,s) = \operatorname{cov}(x(t), x(s)). \tag{4.22}$$

An equivalent characterization is in terms of the mean function (4.21) and the *autocorrelation function*

$$R_{xx}(t,s) = E[x(t)x(s)]. \tag{4.23}$$

That the two characterizations are equivalent follows immediately from the readily verified relation

$$R_{xx}(t,s) = K_{xx}(t,s) + m_x(t)m_x(s). \tag{4.24}$$

Similarly the second-order characterization of a discrete-time stochastic process $x[n]$ consists of the mean sequence

$$m_x[n] = E[x[n]] \tag{4.25}$$

and (auto)covariance sequence

$$K_{xx}[n,m] = \operatorname{cov}(x[n], x[m]) \tag{4.26}$$

or autocorrelation sequence

$$R_{xx}[n,m] = E[x[n]x[m]] = K_{xx}[n,m] + m_x[n]m_x[m] \tag{4.27}$$

The second-order description of a stochastic process is analogous to a second-order description of a random vector, i.e., one in terms of its mean vector and

covariance matrix. From this perspective, it is clear that second-order characterizations consitute only partial specifications of the process. Nevertheless, as was the case for random vectors, these partial characterizations are both convenient and useful. For example, we saw that for random vectors, second-order characterizations were sufficient to solve Bayesian least-squares estimation problems when linear constraints were imposed on the estimator. As we will see, second-order characterizations play an analogous role for stochastic processes. In particular, when we develop LLS estimators for stochastic processes in a later chapter, we will see that only second-order characterizations of the processes are required. As a simple example, if we use $\hat{x}_{\rm LLS}(t|s)$ to denote the LLS estimate of $x(t)$ based on $x(s)$, where t and s are two distinct time instants, then using LLS estimation theory we developed in Chapter 3 we obtain

$$\hat{x}_{\rm LLS}(t|s) = m_x(t) + \frac{K_{xx}(t,s)}{K_{xx}(s,s)}[x(s) - m_x(s)]. \tag{4.28}$$

Furthermore, the corresponding error variance, which we denote using $\lambda_{\rm LLS}(t|s)$, takes the form

$$\lambda_{\rm LLS}(t|s) = K_{xx}(t,t) - \frac{K_{xx}^2(t,s)}{K_{xx}(s,s)}. \tag{4.29}$$

Example 4.6

Referring back to Example 4.1, we now see that what we specified was the second-order characterization for what we referred to as a discrete-time white noise process. Rewriting (4.3) using our new notation we have for this process $x[n]$ that

$$m_x[n] = 0 \tag{4.30}$$

$$K_{xx}[n,m] = \sigma^2\delta[n-m]. \tag{4.31}$$

4.3.1 Uncorrelated Increments Processes

One class of stochastic processes whose second-order description has some special structure are referred to as *uncorrelated increments processes*. As we'll develop in this section, it is this class of processes that give rise to the concept of white noise. For the moment, we restrict our attention to continuous-time processes; the discrete-time case will be summarized later.

To begin, we refer to quantities of the form $x(t_1) - x(t_2)$ for any $t_1 > t_2$ as an *increment* of the process $x(t)$. Formally, we define an uncorrelated increments process as follows: a continuous-time stochastic process $x(t)$ is said to have *uncorrelated increments* if for every choice of t, s, and τ such that $t > s \geq \tau$ we have that the increment $x(t) - x(s)$ is uncorrelated with $x(\tau)$, i.e.,

$$\text{cov}\,(x(t) - x(s), x(\tau)) = 0 \qquad \text{for all } t > s \geq \tau. \tag{4.32}$$

Uncorrelated increments processes have a number of special properties. For example, as one immediate consequence of their definition we have that increments defined on non-overlapping time intervals are uncorrelated with one another—a property which motivates their name. Specifically, let t_1, s_1, t_2, s_2 be arbitrary time instants such that $[s_2, t_2) \cap [s_1, t_1) = \emptyset$, i.e., $t_1 > s_1 \geq t_2 > s_2$. Then exploiting both the linearity of the covariance operator and (4.32) we have

$$\begin{aligned}&\operatorname{cov}\left(x(t_1) - x(s_1), x(t_2) - x(s_2)\right)\\ &\quad = \operatorname{cov}\left(x(t_1) - x(s_1), x(t_2)\right) - \operatorname{cov}\left(x(t_1) - x(s_1), x(s_2)\right) = 0\end{aligned}$$

as claimed.

The covariance function for such a process has the particularly simple form

$$K_{xx}(t, s) = \operatorname{var}\left[x(\min(t, s))\right] = K_{xx}(\min(t, s), \min(t, s)). \tag{4.33}$$

To verify (4.33), we begin by considering the case for which $t > s$. Then

$$\begin{aligned}K_{xx}(t, s) &= \operatorname{cov}\left(x(t), x(s)\right) = \operatorname{cov}\left(x(s) + [x(t) - x(s)], x(s)\right)\\ &= \operatorname{cov}\left(x(s), x(s)\right) + \operatorname{cov}\left(x(t) - x(s), x(s)\right)\\ &= \operatorname{cov}\left(x(s), x(s)\right) = \operatorname{var} x(s)\end{aligned} \tag{4.34}$$

where we have again used (4.32) with $\tau = s$. An analogous derivation establishes that if $t < s$, then

$$K_{xx}(t, s) = \operatorname{var} x(t). \tag{4.35}$$

Finally, combining (4.34) and (4.35) we obtain (4.33).

The definition of an uncorrelated increments process in discrete-time closely parallels that for continuous-time. In particular, we have the following definition: a discrete-time stochastic process $x[n]$ is said to have *uncorrelated increments* if for every choice of n, m, and k such that $n > m \geq k$ we have that the increment $x[n] - x[m]$ is uncorrelated with $x[k]$, i.e.,

$$\operatorname{cov}\left(x[n] - x[m], x[k]\right) = 0 \qquad \text{for } n > m \geq k. \tag{4.36}$$

With this definition, the covariance function of a discrete-time uncorrelated increments process $x[n]$ takes the correspondingly simple form

$$K_{xx}[n, m] = \operatorname{var}\left[x[\min(n, m)]\right] = K_{xx}[\min(n, m), \min(n, m)]. \tag{4.37}$$

Eq. (4.37) can be readily derived in a manner analogous to that for the continuous-time case above.

Let's now consider some examples. We begin with the discrete-time case.

Example 4.7

A simple example of an uncorrelated increments process is the *discrete-time random walk*, which is defined via the recursion

$$x[n] = x[n-1] + w[n] \tag{4.38}$$

where $w[n]$ is the discrete-time white noise of Example 4.1, i.e., $w[n]$ satisfies

$$E[w[n]] = 0 \tag{4.39a}$$

$$E[w[n]w[m]] = \sigma^2\delta[n-m], \tag{4.39b}$$

and the initial condition $x[0]$ is uncorrelated with $w[n]$. Let us assume that $x[0]$ has zero mean and variance σ_0^2.[3]

The random walk $x[n]$, which via (4.38) we can view as the output of a particular linear system (an accumulator) driven by the white noise $w[n]$, can be re-expressed in the form

$$x[k] = x[0] + \sum_{i=1}^{k} w[i]. \tag{4.40}$$

Similarly, an increment of the process $x[n] - x[m]$ for $n > m$ can be expressed in the form

$$x[n] - x[m] = \sum_{i=m+1}^{n} w[i]. \tag{4.41}$$

Since the samples $w[n]$ are uncorrelated random variables we can then readily establish that $x[n]$ is indeed an uncorrelated increments process: for any $n > m \geq k$ we have, using (4.41) and (4.40),

$$E[(x[n] - x[m])x[k]] = \sum_{i=m+1}^{n} E[w[i]x[0]] + \sum_{i=m+1}^{n}\sum_{j=1}^{k} E[w[i]w[j]] = 0.$$

In addition, from (4.41) and the fact that the samples $w[n]$ have zero mean and identical variances, we have that the increments of $x[n]$ also have zero mean and have a variance proportional to the length of the increment interval, i.e., again for $n > m$,

$$E[x[n] - x[m]] = 0 \tag{4.42a}$$

$$E\left[(x[n] - x[m])^2\right] = \sigma^2(n-m). \tag{4.42b}$$

In turn, using (4.42) and the fact that $x[n]-x[0]$ is uncorrelated with $x[0]$ we can verify that the variance of the random walk grows linearly with time n; specifically, for $n \geq 0$,

$$\operatorname{var} x[n] = \operatorname{var}[x[0] + (x[n] - x[0])] = \operatorname{var} x[0] + \operatorname{var}[x[n] - x[0]] = \sigma_0^2 + \sigma^2 n. \tag{4.43}$$

Hence, using (4.37) we have that for the discrete-time random walk

$$K_{xx}[n,m] = \sigma_0^2 + \sigma^2 \min(n,m). \tag{4.44}$$

Finally, we comment that when the white noise $w[n]$ consists of independent Gaussian random variables, the corresponding random walk $x[n]$ is referred to as a discrete-time *Wiener* process or discrete-time *Brownian motion*. A typical sample path of this discrete-time Wiener process is depicted in Fig. 4.4. In this case the random walk will have other important properties as will become apparent later in the chapter.

[3]Sometimes it is convenient to take $x[0] = 0$, in which case $\sigma_0^2 = 0$.

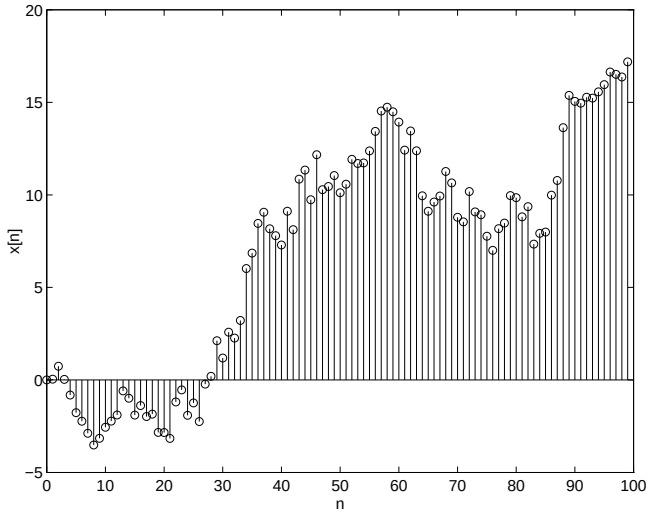

Figure 4.4. A typical sample function of a discrete-time Wiener process with $\sigma^2 = 1$ and $x[0] = 0$.

Before we proceed to a continuous-time example, it is worth pointing out that in fact all uncorrelated increments processes can be described in the form developed in Example 4.7. In particular, it is relatively straightforward to verify that every process $x[n]$ that satisfies (4.36) can be expressed in the form (4.38) where the samples $w[n]$ are all uncorrelated, and uncorrelated with $x[0]$. However, it is important to emphasize that the samples $w[n]$ need not be zero-mean nor have identical variances for (4.38) to hold.

Since we refer to any random process $w[n]$ consisting of uncorrelated random variables as discrete-time white noise, it is clear that in discrete time the concept of white noise is intimately connected to the concept of an uncorrelated increments process. In fact, the first difference of a discrete-time uncorrelated increments process is a discrete-time white noise process, i.e.,

$$w[n] = x[n] - x[n-1]$$

In fact, a similar relationship holds in continuous-time, but the development involves some important subtleties that didn't arise in the discrete-time case. To develop the key ideas, we consider our next example.

Example 4.8

The continuous-time counterpart $x(t)$ of the process considered in Example 4.7 is referred to as a *continuous-time random walk*, and is a process whose increments have zero-mean and have variance proportional to the length of the increment interval, i.e., for $t > s$

$$E\left[(x(t) - x(s))^2\right] = \sigma^2(t-s). \tag{4.45}$$

With $x(0)$ a zero-mean random variable with variance σ_0^2, we then obtain in a manner analogous to the discrete-time case, that the variance of a continuous-time random walk also grows linearly with time; specifically,

$$\operatorname{var} x(t) = \sigma_0^2 + \sigma^2 t \tag{4.46}$$

Using (4.46) in (4.33) we also get immediately that this process has covariance function

$$K_{xx}(t,s) = \sigma_0^2 + \sigma^2 \min(t,s). \tag{4.47}$$

In a manner analogous to the discrete-time case, we can view the evolution of the continuous-time random walk as the accumulation over time of uncorrelated increments. If we attempt to carry this perspective to a limit, letting the increment sizes get increasingly small, then we can interpret what we obtain as trying to represent the process $x(t)$ by the running integral of its derivative, which we denote using $w(t)$, i.e.,

$$x(t) = x(0) + \int_0^t w(\tau)\, d\tau.$$

However, while this will be conceptually useful to us, it is important to realize that for uncorrelated increments processes this limit does not in fact exist in the usual sense. To see this, let us define

$$w_\Delta(t) = \frac{x(t) - x(t-\Delta)}{\Delta} \tag{4.48}$$

and note that if the derivative of $x(t)$ exists, then

$$w(t) = \lim_{\Delta \to 0} w_\Delta(t).$$

Now $w_\Delta(t)$ as defined in (4.48) is zero-mean and, because increments are uncorrelated,

$$E\left[w_\Delta(t) w_\Delta(\tau)\right] = 0, \qquad \text{for } |t-\tau| > \Delta. \tag{4.49}$$

so that as $\Delta \to 0$, arbitrary closely spaced but distinct samples of $w_\Delta(t)$ are uncorrelated. However, from (4.45), we see that $w_\Delta(t)$ has variance

$$E\left[w_\Delta^2(t)\right] = \frac{\sigma^2 \Delta}{\Delta^2} = \frac{\sigma^2}{\Delta}, \tag{4.50}$$

which diverges as $\Delta \to 0$, and thus the derivative of $x(t)$ cannot exist in any usual sense. Nevertheless, uncorrelated increments processes do have a derivative

$$w(t) = \frac{dx(t)}{dt} \tag{4.51}$$

in a generalized sense, and it is this derivative that is continuous-time white noise. Consistent with (4.49) and (4.50) $w(t)$ does indeed have infinite variance and the property that $w(t)$ and $w(\tau)$ are uncorrelated for any distinct values of t and τ. We will develop useful ways of characterizing this process in the next section, but we stress that the concept of white noise will prove extremely useful and frequently represents a useful idealization of physical processes. Indeed, white noise plays as important a role in the analysis of stochastic processes and systems as the unit impulse does for the analysis of deterministic signals and systems.

An important example of a random walk corresponds to the case in which the increments are specifically independent Gaussian random variables. Analogous to the discrete-time case, this particular random walk is referred to as continuous-time Brownian motion or the continuous-time Wiener process. A typical sample path of a continuous-time Wiener process is depicted in Fig. 4.5. Note that the highly irregular nature of the sample path. In fact, Wiener processes have the special property that their sample paths are continuous everywhere but nowhere differentiable!

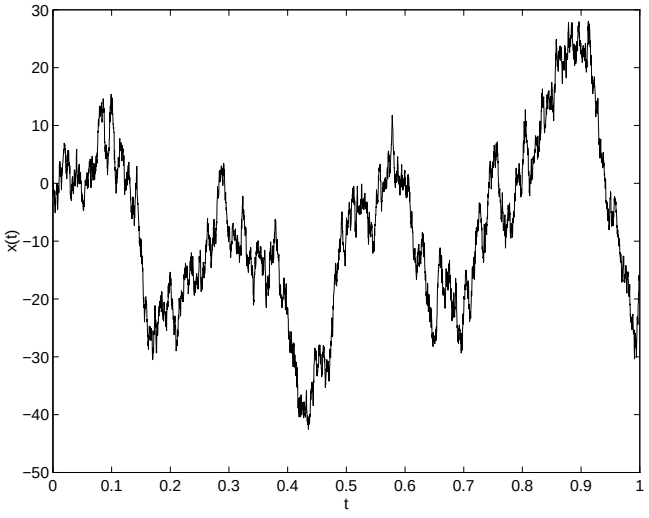

Figure 4.5. A typical sample function of a continuous-time Wiener process with $x[0] = 0$.

Example 4.9

Another example of a discrete-time random walk as described in Example 4.7 corresponds to the case in which $x[0] = 0$ and the underlying white noise $w[n]$ is a Bernoulli process, i.e., the samples $w[n]$ are independent and identically distributed with, for some parameter p,

$$\Pr[w[n] = i] = \begin{cases} p & i = 1 \\ 1 - p & i = 0 \end{cases}. \tag{4.52}$$

In this case, the random walk $x[n]$ is a counting process—each sample path is nondecreasing and takes on only non-negative integer values. Furthermore, since

$$E[w[n]] = p \tag{4.53a}$$

$$\operatorname{var} w[n] = p(1 - p) \tag{4.53b}$$

the random walk $x[n]$ has a variance and mean that grows linearly with time, i.e.,

$$E[x[n]] = np \tag{4.54a}$$

$$\operatorname{var} x[n] = np(1 - p). \tag{4.54b}$$

A typical sample function of this Bernoulli counting process, which is more naturally denoted using $N[n]$ rather than $x[n]$, is depicted in Fig. 4.6.

Example 4.10

The continuous-time counterpart to the Bernoulli counting process $N[n]$ is the Poisson counting process $N(t)$, which we began to explore in Example 4.4. Since we defined $N(t)$ as a process with independent increments, it is also an uncorrelated increments process. Let's now explore some additional properties of this process as an uncorrelated increments process.

First, since $N(t)$ at any fixed time t is Poisson-distributed as given in (4.13), we obtain via straightforward computations

$$m_N(t) = E[N(t)] = \lambda t \tag{4.55}$$

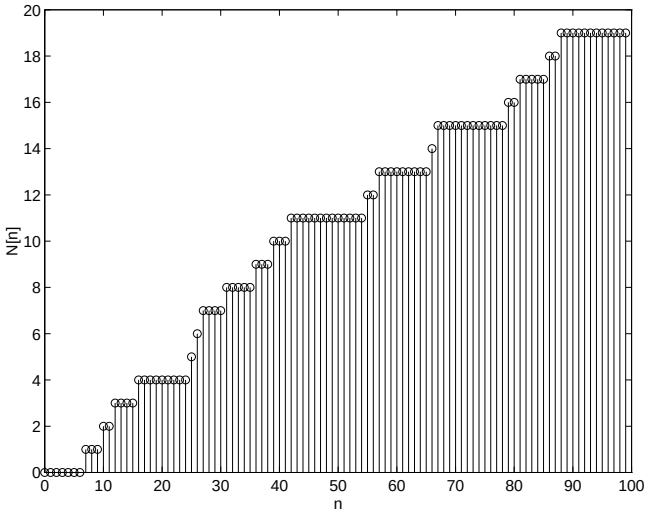

Figure 4.6. A typical sample function of a Bernoulli counting process with $p = 0.2$.

and

$$\operatorname{var} N(t) = \lambda t. \tag{4.56}$$

Then, since $N(t)$ is an uncorrelated increments process, we immediately obtain its covariance function via (4.33) as

$$K_{NN}(t, s) = \lambda \min(t, s). \tag{4.57}$$

We again note that although the continuous-time Wiener process and the Poisson counting process have identical covariance functions (to within a scale factor), their sample paths are strikingly different. This reemphasizes the fact that second order descriptions provide only a partial characterization of a process, and that second-order statistics alone often tell us surprisingly little about the sample paths of a process. For example, as we commented earlier, the sample paths of a continuous-time Wiener process are continuous functions, while the sample paths of the Poisson counting process have many discontinuities.

Finally, note that many issues associated with defining a derivative for uncorrelated increments process are readily apparent in the case of the Poisson process. In particular, since the sample paths of the Poisson counting process have discontinuities, a usual derivative is not defined. However, when we admit generalized functions in our treatment, we see that since the sample paths are piecewise constant with unit height jumps at the locations of the count times, the corresponding generalized derivative is a process consisting of a train of unit-area impulses at the count times. This "random impulse train" is in fact a conceptually very important continuous-time white noise model for many physical problems. These arise, for example, in modeling optical communication systems and electronic noises.

4.3.2 Continuous-Time White Noise

As the discussion and examples in the last section has suggested, continuous-time uncorrelated increments processes have sufficiently irregular sample paths that

they are not differentiable in the usual sense. Nevertheless, these processes invariably have a generalized derivative, and this derivative is referred to as a *white noise process*. In this section we develop the covariance properties of this generalized derivative.

We begin by first noting that if $x'(t)$ denotes the derivative of a random process $x(t)$, then

$$K_{x'x'}(t,s) = E\left[x'(t)x'(s)\right] = E\left[\frac{dx(t)}{dt}\frac{dx(s)}{ds}\right] = \frac{\partial^2}{\partial t \partial s}K_{xx}(t,s). \tag{4.58}$$

where we have interchanged the order of differentiation and expectation. Next, using $\lambda_x(t)$ to denote the variance of $x(t)$ we have that when $x(t)$ has uncorrelated increments,

$$K_{xx}(t,s) = \lambda_x(\min(t,s)) = \begin{cases} \lambda_x(s) & s < t \\ \lambda_x(t) & \text{otherwise} \end{cases}. \tag{4.59}$$

Substituting (4.59) into (4.58) and using $w(t)$ to denote the derivative of $x(t)$ we obtain

$$\begin{aligned} K_{ww}(t,s) &= \frac{\partial}{\partial t}\left[\frac{\partial}{\partial s}\lambda_x(\min(t,s))\right] \\ &= \frac{\partial}{\partial t}\left[\lambda_x'(s)\, u(t-s)\right] \\ &= \lambda_x'(s)\,\delta(t-s) \end{aligned} \tag{4.60}$$

where $\lambda_x'(t) = d\lambda_x(t)/dt$, where $u(t)$ is the unit-step function, i.e.,

$$u(t) = \begin{cases} 1 & t > 0 \\ 0 & \text{otherwise} \end{cases},$$

and where $\delta(t) = du(t)/dt$ is the unit-impulse.

From (4.60) we see that as one would expect from a process for which increments corresponding to arbitrarily short intervals are uncorrelated, arbitrarily closely spaced samples of the derivative process are uncorrelated, i.e., for every t and s such that $t \neq s$ we have that $w(t)$ and $w(s)$ are uncorrelated. That the derivative doesn't exist in the usual sense is readily apparent from the fact that (4.60) implies that $\lambda_w(t)$, the variance of $w(t)$, is unbounded. These results are, of course, all consistent with our observations about derivatives of uncorrelated increments processes in the examples of the last section.

As a final remark, note that when the variance of $x(t)$ is a linear (affine) function of t the covariance function of corresponding white noise, i.e., (4.60) is a function only of $t-s$. As we'll discuss, $w(t)$ in this case is referred to as a "wide-sense stationary" white noise process, and the corresponding $x(t)$ is said to have wide-sense stationary and uncorrelated increments.

4.3.3 Properties of Covariance Functions

As is the case with covariance matrices, covariance functions of both continuous-time and discrete-time stochastic processes are not arbitrary functions.

We begin with the continuous-time case, and define the notion of a positive definite function. These are a special class of symmetric functions: a function of two variables $F(t,s)$ is *symmetric* if for all t and s we have

$$F(t,s) = F(s,t).$$

Specifically, a symmetric function of two variables $F(t,s)$ is a *positive semidefinite function* if for all choices of deterministic functions $a(t)$ we have

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} a(t)\, F(t,s)\, a(s)\, dt\, ds \geq 0. \tag{4.61}$$

In addition, we say that $F(t,s)$ is a *positive definite function* if for all choices of deterministic functions $a(t)$ such that $a(t) \not\equiv 0$ we have

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} a(t)\, F(t,s)\, a(s)\, dt\, ds > 0. \tag{4.62}$$

With these definitions, we then have the following main result: a bivariate function $F(t,s)$ is a valid covariance function if and only if $F(t,s)$ is positive semidefinite.

To verify the "if" part of this statement will require some properties of positive definite functions we develop in the next chapter of these notes. However, to establish the "only if" part of this statement, it suffices to note that if we define a random variable z in terms of a random process $x(t)$ via

$$z = \int_{-\infty}^{+\infty} a(t)\, x(t)\, dt,$$

where $a(t)$ is an arbitrary deterministic function, then

$$0 \leq \operatorname{var} z = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} a(t)\, K_{xx}(t,s)\, a(s)\, dt\, ds.$$

Note that when we choose $a(t)$ to be of the form

$$a(t) = \sum_{i=1}^{N} a_i \delta(t - t_i) \tag{4.63}$$

where N, the constants a_i, and the time instants t_i are all arbitrary, we see that an immediate consequence of the fact that $K_{xx}(t,s)$ is a positive semidefinite function is that all matrices of the form

$$\begin{bmatrix} K_{xx}(t_1,t_1) & K_{xx}(t_1,t_2) & \cdots & K_{xx}(t_1,t_N) \\ K_{xx}(t_2,t_1) & K_{xx}(t_2,t_2) & \cdots & K_{xx}(t_2,t_N) \\ \vdots & \vdots & \ddots & \vdots \\ K_{xx}(t_N,t_1) & K_{xx}(t_N,t_2) & \cdots & K_{xx}(t_N,t_N) \end{bmatrix} \tag{4.64}$$

must be positive semidefinite. We would of course expect this to be a necessary condition—indeed (4.64) is the covariance matrix $\mathbf{\Lambda_z}$ of the random vector composed of the corresponding samples of the process, i.e.,

$$\mathbf{z} = \begin{bmatrix} x(t_1) & x(t_2) & \cdots & x(t_N) \end{bmatrix}^{\mathrm{T}}.$$

Furthermore, an obvious special case of this result is that for all t we have $K_{xx}(t,t) = \operatorname{var} x(t) \geq 0$.

Note too that from the Cauchy-Schwarz inequality we have for any t and s,

$$K_{xx}^2(t,s) \leq K_{xx}(t,t)K_{xx}(s,s) \tag{4.65}$$

with equality if and only if $x(t) = ax(s) + b$ for some constants a and b.

In addition to the second-order statistics of an individual process, it is frequently important to consider the joint second-order characterization of a pair of processes $x(t)$ and $y(t)$. This characterization consists of the individual second-order characterizations of the processes, i.e., their means $m_x(t)$ and $m_y(t)$, covariance functions $K_{xx}(t,s)$ and $K_{yy}(t,s)$, and their *cross-covariance function*, which is defined as

$$K_{xy}(t,s) = \operatorname{cov}\left(x(t), y(s)\right). \tag{4.66}$$

As an alternative to (4.66) we may equivalently use the *cross-correlation function* for $x(t)$ and $y(t)$, which is defined as

$$R_{xy}(t,s) = E\left[x(t)y(s)\right] = K_{xy}(t,s) + m_x(t)m_y(s). \tag{4.67}$$

Like a cross-covariance matrix, a cross-covariance function has no constraints on its form—it is an arbitrary bivariate function. However, there are relationships between different cross-covariance functions. In particular,

$$K_{xy}(t,s) = K_{yx}(s,t) \tag{4.68}$$

The correlation function $R_{xy}(t,s)$ obeys an analogous relationship.

Finally, it is straightforward to verify via the Cauchy-Schwarz inequality that the cross-covariance function, in the continuous-time case, satisfies the bound

$$K_{xy}^2(t,s) \leq K_{xx}(t,t)\, K_{yy}(s,s), \tag{4.69}$$

with equality if and only if $x(t) = ay(s) + b$ for some constants a and b.

We have an analogous set of results in the discrete-time case. In particular, a bivariate sequence $F[n,m]$ is said to be symmetric if for all n and m we have $F[n,m] = F[m,n]$. A symmetric bivariate sequence $F[n,m]$ is a positive semidefinite sequence if for all choices of $a[n]$ we have

$$\sum_{n=-\infty}^{+\infty} \sum_{m=-\infty}^{+\infty} a[n]\, F[n,m]\, a[m] \geq 0.$$

Furthermore, $F[n,m]$ is a positive definite sequence if for all choices of $a[n]$ such that $a[n] \not\equiv 0$ we have

$$\sum_{n=-\infty}^{+\infty} \sum_{m=-\infty}^{+\infty} a[n]\, F[n,m]\, a[m] > 0.$$

With these definitions we have that $F[n,m]$ is a valid covariance function for a discrete-time stochastic process if and only if $F[n,m]$ is a positive semidefinite sequence.

For discrete-time processes, the cross-covariance function takes the corresponding form

$$K_{\mathsf{xy}}[n,m] = \operatorname{cov}\left(\mathsf{x}[n], \mathsf{y}[m]\right), \tag{4.70}$$

as does the cross-correlation function,

$$R_{\mathsf{xy}}[n,m] = E\left[\mathsf{x}[n]\mathsf{y}[m]\right] = K_{\mathsf{xy}}[n,m] + m_{\mathsf{x}}[n] m_{\mathsf{y}}[m], \tag{4.71}$$

and the following relationship exists among cross-covariance functions:

$$K_{\mathsf{xy}}[n,m] = K_{\mathsf{yx}}[m,n]. \tag{4.72}$$

We conclude this section by exploring the properties of a particular example of a covariance function.

Example 4.11

An important class of discrete-time stochastic processes that will arise frequently in our examples has a covariance function of the form

$$K_{\mathsf{xx}}[n,m] = \sigma^2 \alpha^{|n-m|} \tag{4.73}$$

where α is a parameter that satisfies $|\alpha| < 1$. For future convenience, we will refer to (4.73) as a *first-order* covariance function, where motivation for this terminology will become clearer in subsequent sections of the chapter.

As is apparent from (4.73), processes with first-order covariance functions have the property that the magnitude of their correlation between samples of the process decays exponentially with their separation. The magnitude of the parameter α controls the rate of this exponential decay, and to be precise the constant $1/\ln|\alpha|$ is the associated time constant of this decay. When $\alpha \simeq +1$, samples of the process far apart in time remain highly correlated. For $\alpha \simeq 0$, samples of the process comparatively close in time exhibit little correlation, and in fact we see from (4.73) that for $\alpha = 0$ the process specializes to white noise. Note that when $\alpha < 0$, samples of the process separated in time by an odd number of samples are negatively correlated, and for $\alpha \simeq -1$ this negative correlation is strong.

The continuous-time counterpart of the covariance function (4.73) takes the form

$$K_{\mathsf{xx}}(t,s) = \sigma^2 e^{-\lambda|t-s|} \tag{4.74}$$

where λ is a parameter satisfying $\lambda > 0$. By analogy we refer to (4.74) as a *continuous-time* first-order covariance function. In this case, the correlation between samples of

the process also decays exponentially with their separation, and this time $1/\lambda$ is the time constant of the decay. Hence, larger values of λ correspond to more rapid decay. The time constant $1/\lambda$ is a popular measure of the what is called *correlation time* for the process, and in this case corresponds to the temporal separation by which the correlation coefficient has decreased to e^{-1}.

An example of a process we have already encountered that has a covariance function of the form (4.74) is the random telegraph wave of Example 4.5. To show this, we first obtain using the properties of the process that

$$m_x(t) = 1 \cdot \Pr[x(t) = +1] - 1 \cdot \Pr[x(t) = -1] = 0.$$

In turn, when $t > s$, we find that

$$\begin{aligned} K_{xx}(t,s) &= 1 \cdot \Pr[N(t) - N(s) \text{ is even}] - 1 \cdot \Pr[N(t) - N(s) \text{ is odd}] \\ &= e^{-2\lambda(t-s)}. \end{aligned} \tag{4.75}$$

After performing an analogous calculation for $t < s$, we can then conclude that

$$K_{xx}(t,s) = e^{-2\lambda|t-s|}. \tag{4.76}$$

which is precisely of the form (4.74) with $\sigma^2 = 1$ and λ replaced by 2λ.

However, it is important to again emphasize that since (4.73) and (4.74) are only partial characterizations of a process, one can construct a variety of very different processes with covariance functions of this type. For example, as Fig. 4.3 indicates, the random telegraph wave has piecewise-constant sample paths. However, later in the chapter it will be apparent that other processes with first-order covariance functions have, for example, continuous sample paths and Gaussian samples.

4.4 STATIONARITY

The first-order covariance function in Example 4.11 had a rather special property—that the covariance between two samples of the process $x(t)$ and $x(s)$ depended only on the temporal separation of these samples. In other words, only the relative temporal location and not the absolute temporal location of these samples impacted their correlation. In particular, any other pair of samples of the process with the same temporal separation, i.e., $x(t+\Delta)$ and $x(s+\Delta)$ for any Δ, have the same correlation. This kind of "statistical time-invariance," which is termed *stationarity*, is in fact a rather natural property of many physical stochastic processes. In this section, we explore this concept and its implications.

Before we proceed, it is worth noting that many of the processes we have encountered already do not have this statistical time-invariance property. For example, the continuous-time random walk of Example 4.8 has variance that grows linearly with time, i.e., (cf. (4.46))

$$\operatorname{var} x(t) = \sigma_0^2 + \sigma^2 t.$$

For a process of this type, given statistics of even a single sample $x(t_0)$ we can infer information about the value of the time instant t_0 and, in turn, the location of the time origin of the process.

For some physical processes, the notion of an absolute time origin is rather natural—for example, the weakening atmospheric forces and fluctuations affecting a rocket from launch through orbital insertion. However, there are also many physical processes that do not have such absolute time references. For example, the statistics of noise in a communications receiver typically remain the same from day to day.

The strongest notion of stationarity is referred to as *strict-sense* stationarity, and is defined in terms of the complete statistical characterization of the process. Specifically, a stochastic process $x(t)$ is said to be strict-sense stationary (SSS) if all its finite-dimensional distributions are time-invariant, i.e., if for all choices of N and time instants $t_1, t_2, \ldots, t_N$ we have, for any τ,

$$p_{x(t_1),x(t_2),\ldots,x(t_N)}(x_1, x_2, \ldots, x_N) = p_{x(t_1+\tau),x(t_2+\tau),\ldots,x(t_N+\tau)}(x_1, x_2, \ldots, x_N). \tag{4.77}$$

A weaker notion of stationarity is referred to as *Nth-order stationarity*, and is defined in terms of a partial statistical characterization of the process. Specifically, a stochastic process $x(t)$ is said to be Nth-order stationary if all N-dimensional distributions of the process are time-invariant. For example, a process $x(t)$ is first-order stationary if the distribution for a particular sample $x(t_0)$ does not depend on t_0:

$$p_{x(t_0)}(x_0) = p_{x(0)}(x_0).$$

Likewise, a process $x(t)$ is second-order stationary if the joint distribution for a pair of samples $x(t)$ and $x(s)$ depends only on $t - s$:

$$p_{x(s),x(t)}(x_0, x_1) = p_{x(0),x(t-s)}(x_0, x_1).$$

Note that if a process is Nth-order stationary, it must also be kth-order stationary for all $k < N$. To verify this it suffices to integrate out $N - k$ of the variables in (4.77).

A still weaker notion of stationarity—but one that is typically simpler to verify in practice—is the concept of *wide-sense* stationarity. Specifically, we say a process $x(t)$ is wide-sense stationary (WSS) if its second-order characterization is time-invariant, i.e., if its mean does not vary with time

$$m_x(t) = m_x \tag{4.78}$$

and if its covariance function $K_{xx}(t, s)$ depends only on $t - s$, the temporal separation of the corresponding samples. As a result of this property, for WSS processes we will find it convenient to slightly abuse notation and express K_{xx} as a function of only a single variable, viz.,

$$K_{xx}(t, s) \triangleq K_{xx}(t - s). \tag{4.79}$$

Hence for a WSS process $x(t)$, we will adopt the notation

$$m_x = E\left[x(t)\right] \tag{4.80a}$$

$$K_{xx}(\tau) = \operatorname{cov}\left(x(t), x(t-\tau)\right) = \operatorname{cov}\left(x(t+\tau), x(t)\right), \tag{4.80b}$$

which emphasizes the fact that these statistics don't depend on t.

Some important relationships between these notions of stationarity are worth emphasizing. For example, strict-sense stationarity clearly implies second-order stationarity, which in turn implies wide-sense stationarity—since $m_x(t)$ and $K_{xx}(t,s)$ are moments of second-order densities for $x(t)$. However, these implications do not generally hold in reverse. In fact, there are many WSS processes that are not even first-order stationary, i.e., WSS processes for which the density of $x(t)$ varies with time. Indeed, the WSS property only ensures that the densities of $x(t)$ for different values of t all have the same mean and variance. These observations are of course consistent with the fact that second-order characterizations of stochastic processes are only partial characterizations.

Corresponding notions of stationarity and their interrelationships apply in discrete-time. For example, a discrete-time process is Nth-order stationary if for all choices of $n_1, n_2, \ldots, n_N$ we have, for any k,

$$p_{x[n_1],x[n_2],\ldots,x[n_N]}(x_1, x_2, \ldots, x_N) = p_{x[n_1+k],x[n_2+k],\ldots,x[n_N+k]}(x_1, x_2, \ldots, x_N). \tag{4.81}$$

Likewise, a discrete-time process is WSS if we can express its second-order characterization in the form

$$m_x = E\left[x[n]\right] \tag{4.82a}$$

$$K_{xx}[k] = \operatorname{cov}\left(x[n], x[n-k]\right) = \operatorname{cov}\left(x[n+k], x[n]\right) \tag{4.82b}$$

where in particular there is no dependence on n.

Example 4.12

Consider a white noise process $x[n]$ that is a sequence of independent random variables where for n odd, $x[n]$ is uniformly distributed on $[-\sqrt{3}, \sqrt{3}]$, i.e., for n odd,

$$p_{x[n]}(x) = \begin{cases} 1/(2\sqrt{3}) & |x| < \sqrt{3} \\ 0 & \text{otherwise} \end{cases}.$$

For n even, the $x[n]$ are zero-mean, unit-variance Gaussian random variables, i.e.,

$$p_{x[n]}(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

This process is WSS, since

$$m_x[n] = 0$$

$$K_{xx}[n,m] = \delta[n-m]$$

However, it clearly is not even first-order stationary; e.g.,

$$p_{x[0]}(x) \neq p_{x[1]}(x).$$

There are many examples of SSS random processes. A simple example is the discrete-time white noise process consisting of independent, identically distributed Gaussian random variables. A second example is the random sinusoid

(4.7) in Example 4.3, where A and f_0 are known constants, and Θ is a random phase uniformly distributed between 0 and 2π. To see this, it suffices to note that from (4.7) we have

$$x(t - t_0) = A\cos(2\pi f_0(t - t_0) + \Theta) = A\cos(2\pi f_0 t + \Phi) \tag{4.83}$$

where

$$\Phi = (\Theta - 2\pi f_0 t_0) \bmod 2\pi. \tag{4.84}$$

Since Θ is uniformly distributed, so is Φ, and thus we conclude that $x(t)$ in (4.7) is SSS. Similarly the random telegraph wave of Example 4.5 is also SSS.

4.4.1 Properties of Wide-Sense Stationary Processes

In Section 4.3.3, we showed that covariance functions of stochastic processes were not arbitrary functions, but were specifically positive semidefinite functions. Similar constraints therefore apply to the covariance functions of specifically WSS processes. In particular, for a continuous-time WSS process $x(t)$ we have that its covariance function must be symmetric, i.e.,

$$K_{xx}(\tau) = K_{xx}(-\tau), \tag{4.85}$$

and positive semidefinite, i.e., for all choices of $a(t) \not\equiv 0$,

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} a(t)\, K_{xx}(t - s)\, a(s)\, dt\, ds \geq 0. \tag{4.86}$$

Moreover, choosing $a(t)$ to be of the form (4.63), we get as a special case of (4.86), for all N and a_i not all zero,

$$\sum_{i=1}^{N} \sum_{j=1}^{N} a_i K_{xx}(t_i - t_j) a_j \geq 0. \tag{4.87}$$

For $N = 1$ in particular we have the obvious special case $K_{xx}(0) = \operatorname{var} x(t) \geq 0$. Note too, that specializing (4.65) for WSS processes we get the following additional bound on the values of a covariance function:

$$|K_{xx}(\tau)| \leq K_{xx}(0) \tag{4.88}$$

with equality for all τ if and only if $x(t)$ is the same random variable for all t, i.e., the random process $x(t)$ is a random constant. Other processes satisfy (4.88) with equality only for certain values of τ—an example is the periodic process of Example 4.3. For a large class of processes, however, $K_{xx}(\tau)$ decays to zero as $|\tau| \to \infty$, indicating that $x(t)$ and $x(s)$ become increasingly decorrelated as their temporal separation increases. An example of a process with this behavior is one with the first-order covariance function (4.74).

As a final comment on WSS processes, it is worth remarking that the constant mean condition (4.78) is not of primary interest. In particular, since the mean of a

process $x(t)$ represents the known, deterministic component of $x(t)$, at least conceptually it can always be removed from the process to yield a zero-mean process

$$\tilde{x}(t) = x(t) - m_x(t) \tag{4.89}$$

with the same covariance structure. Hence, even if (4.78) is not satisfied by the process, as long as the covariance function satisfies (4.79) then the equivalent zero-mean process (4.89) is WSS. Consequently, the essential quantity of interest in analyzing a WSS process is its covariance function.

Again, we emphasize that all the quantities and properties we developed in this section for continuous-time processes have counterparts for discrete-time processes. In particular, the covariance function of a discrete-time WSS process is symmetric, i.e.,

$$K_{xx}[k] = K_{xx}[-k], \tag{4.90}$$

and positive semidefinite, i.e., for all $a[n] \not\equiv 0$,

$$\sum_{n=-\infty}^{+\infty} \sum_{m=-\infty}^{+\infty} a[n]\, K_{xx}[n-m]\, a[m] \geq 0. \tag{4.91}$$

Likewise, the covariance function for such processes satisfies the bound

$$|K_{xx}[k]| \leq K_{xx}[0]. \tag{4.92}$$

4.4.2 Jointly Stationary Stochastic Processes

A broader concept of stationarity involves the statistical properties of a collection of stochastic processes. For simplicity of exposition, we'll restrict our attention in this section to a pair of stochastic processes. In particular, a pair of processes $x(t)$ and $y(t)$ are said to be *jointly strict-sense stationary* if their *joint* statistics are time-invariant , i.e., if for all N and M, and times $t_1, t_2, \ldots, t_N$ and $s_1, s_2, \ldots, s_M$, we have, for any τ,

$$p_{x(t_1),x(t_2),\ldots,x(t_N),y(s_1),y(s_2),\ldots,y(s_M)}(x_1, x_2, \ldots, x_N, y_1, y_2, \ldots, y_M) =$$
$$p_{x(t_1+\tau),x(t_2+\tau),\ldots,x(t_N+\tau),y(s_1+\tau),y(s_2+\tau),\ldots,y(s_M+\tau)}(x_1, x_2, \ldots, x_N, y_1, y_2, \ldots, y_M).$$

Analogously, there is a weaker notion of joint stationarity that involves only the joint second-order characterization of a pair of processes. In particular, $x(t)$ and $y(t)$ are said to be *jointly wide-sense stationary* if they are individually WSS and if their cross-covariance function can be expressed in the form

$$K_{xy}(t, s) = K_{xy}(t - s) \tag{4.93}$$

where we have again adopted a slight modification in notation.

Specializing (4.68), the cross-covariance functions of jointly WSS processes are related according to

$$K_{xy}(\tau) = \operatorname{cov}(x(t), y(t - \tau)) = K_{yx}(-\tau). \tag{4.94}$$

Similarly, specializing (4.69), we see that $K_{xy}(\tau)$ also satisfies the bound

$$K_{xy}^2(\tau) \leq K_{xx}(0)\, K_{yy}(0). \tag{4.95}$$

It is important to emphasize that there are many processes that are individually WSS but not jointly WSS. For example, suppose that $x(t)$ is an arbitrary WSS process, and define

$$y(t) = x(-t). \tag{4.96}$$

Then it is straightforward to verify that $y(t)$ is WSS as well. However, $x(t)$ and $y(t)$ are not jointly WSS. To see this, it suffices to note that

$$K_{xy}(t,s) = \text{cov}\,(x(t), y(s)) = \text{cov}\,(x(t), x(-s)) = K_{xx}(t+s), \tag{4.97}$$

i.e., $K_{xy}(t,s)$ is not a function of $t-s$ alone. Note too, that it is similarly straightforward to construct processes that are individually SSS but not jointly SSS. In particular, it suffices to let $x(t)$ in the preceding example be SSS.

Once again, analogous definitions and properties apply in the case of discrete-time stochastic processes. For example, a pair of processes $x[n]$ and $y[n]$ are jointly SSS if for all N and M, and times $n_1, n_2, \ldots, n_N$ and $m_1, m_2, \ldots, m_M$, we have, for any k,

$$p_{x[n_1],x[n_2],\ldots,x[n_N],y[m_1],y[m_2],\ldots,y[m_M]}(x_1, x_2, \ldots, x_N, y_1, y_2, \ldots, y_M) =$$
$$p_{x[n_1+k],x[n_2+k],\ldots,x[n_N+k],y[m_1+k],y[m_2+k],\ldots,y[m_M+k]}(x_1, x_2, \ldots, x_N, y_1, y_2, \ldots, y_M).$$

Similarly, $x[n]$ and $y[n]$ are jointly WSS if they are individually WSS and if

$$K_{xy}[n,m] = K_{xy}[n-m], \tag{4.98}$$

in which case we have

$$K_{xy}[k] = \text{cov}\,(x[n], y[n-k]) = K_{yx}[-k] \tag{4.99}$$

and

$$K_{xy}^2[k] \leq K_{xx}[0]\, K_{yy}[0]. \tag{4.100}$$

4.5 STOCHASTIC PROCESSES THROUGH LINEAR SYSTEMS

In many applications, the stochastic processes of interest are the inputs or outputs of various kinds of linear and nonlinear systems—either natural or engineered. When a stochastic process is the input to a system, the output is also a stochastic process. In general, while a complete characterization of the output process can be determined from a complete characterization of the system and input process, this can be very difficult in practice in many cases.

In this section, we begin to explore the effects of specifically linear systems on stochastic processes. As we will now see, for linear systems the second-order characterizations of output processes can always be obtained in a relatively straightforward manner from the corresponding second-order characterizations of input

processes. In our treatment of this topic, we assume familiarity with basic techniques for representing and analyzing linear systems in both continuous-time and discrete-time. A very brief summary of the some key definitions, methods, and results is provided in Appendix 4.A.

We begin by considering a general continuous-time linear system of the form

$$y(t) = \int_{-\infty}^{+\infty} h(t,\tau)\, x(\tau)\, d\tau, \tag{4.101}$$

where $x(t)$ is a stochastic process with mean function $m_x(t)$ and covariance function $K_{xx}(t,s)$. It is straightforward to calculate the second-order statistics of the output $y(t)$ and the cross-covariance between the input and output.

In particular, taking expected values of both sides of (4.101), we first see that

$$m_y(t) = \int_{-\infty}^{+\infty} h(t,\tau) m_x(\tau)\, d\tau \tag{4.102}$$

If we then define the zero-mean processes

$$\tilde{x}(t) = x(t) - m_x(t) \tag{4.103a}$$
$$\tilde{y}(t) = y(t) - m_y(t). \tag{4.103b}$$

we see that $\tilde{x}(t)$ and $\tilde{y}(t)$ are related via

$$\tilde{y}(t) = \int_{-\infty}^{+\infty} h(t,\tau)\, \tilde{x}(\tau)\, d\tau. \tag{4.104}$$

Next, let $z(t)$ be a third and arbitrary stochastic process and define

$$\tilde{z}(t) = z(t) - m_z(t).$$

Then using (4.104) we have that

$$K_{yz}(t,s) = E\left[\tilde{y}(t)\tilde{z}(s)\right] = \int_{-\infty}^{+\infty} h(t,\tau)\, K_{xz}(\tau,s)\, d\tau. \tag{4.105}$$

Similarly, using the cross-covariance property (4.68), we find that

$$K_{zy}(t,s) = K_{yz}(s,t) = \int_{-\infty}^{+\infty} h(s,\tau)\, K_{xz}(\tau,t)\, d\tau = \int_{-\infty}^{+\infty} h(s,\tau)\, K_{zx}(t,\tau)\, d\tau \tag{4.106}$$

As a special case, when $z(t) = x(t)$ we have that (4.105) and (4.106) become, respectively,

$$K_{yx}(t,s) = \int_{-\infty}^{+\infty} h(t,\tau)\, K_{xx}(\tau,s)\, d\tau \tag{4.107}$$

and

$$K_{xy}(t,s) = \int_{-\infty}^{+\infty} h(s,\tau)\, K_{xx}(t,\tau)\, d\tau. \tag{4.108}$$

As a second special case, when $z(t) = y(t)$ we have that (4.105) becomes

$$K_{yy}(t,s) = \int_{-\infty}^{+\infty} h(t,\tau)\, K_{xy}(\tau,s)\, d\tau, \tag{4.109}$$

and when we substitute (4.108) into (4.109) we obtain

$$K_{yy}(t,s) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} h(t,\tau)\, h(s,\sigma)\, K_{xx}(\tau,\sigma)\, d\tau\, d\sigma. \tag{4.110}$$

More generally, suppose that $z(t)$ is the output of a linear system with arbitrary kernel $g(t,\tau)$ and arbitrary input $w(t)$ so

$$z(t) = \int_{-\infty}^{+\infty} g(t,\tau)\, w(\tau)\, d\tau. \tag{4.111}$$

Then the cross-covariance of outputs of the systems (4.101) and (4.111) are related to the cross-covariance of their inputs via

$$K_{yz}(t,s) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} h(t,\tau)\, g(s,\sigma)\, K_{xw}(\tau,\sigma)\, d\tau\, d\sigma \tag{4.112}$$

For discrete-time processes and systems, we have analogous results. In particular, for a linear system with kernel $h[n,k]$ whose response to an input process $x[n]$ is

$$y[n] = \sum_{k=-\infty}^{+\infty} h[n,k]\, x[k], \tag{4.113}$$

we have

$$m_y[n] = \sum_{k=-\infty}^{+\infty} h[n,k]\, m_x[k] \tag{4.114}$$

$$K_{yz}[n,m] = \sum_{k=-\infty}^{+\infty} h[n,k]\, K_{xz}[k,m] \tag{4.115}$$

$$K_{zy}[n,m] = \sum_{k=-\infty}^{+\infty} h[m,k]\, K_{zx}[n,k] \tag{4.116}$$

$$K_{yx}[n,m] = \sum_{k=-\infty}^{+\infty} h[n,k]\, K_{xx}[k,m] \tag{4.117}$$

$$K_{xy}[n,m] = \sum_{k=-\infty}^{+\infty} h[m,k]\, K_{xx}[n,k] \tag{4.118}$$

$$K_{yy}[n,m] = \sum_{j=-\infty}^{+\infty} \sum_{k=-\infty}^{+\infty} h[n,k]\, h[m,j]\, K_{xx}[k,j] \tag{4.119}$$

Similarly, if $z[n]$ is the response of a linear system with kernel $g[n,k]$ to input $w[n]$, i.e.,

$$z[n] = \sum_{k=-\infty}^{+\infty} g[n,k]\, w[k], \tag{4.120}$$

then the cross-covariance of outputs of the systems (4.113) and (4.120) are related to the cross-variance of their inputs via

$$K_{yz}[n,m] = \sum_{j=-\infty}^{+\infty} \sum_{k=-\infty}^{+\infty} h[n,k]\, g[m,j]\, K_{xw}[k,j]. \tag{4.121}$$

4.5.1 Jointly WSS Processes and LTI Systems

The preceding relations in both continuous-time and discrete-time simplify substantially when the linear systems are also time-invariant and the processes are jointly WSS, as we now develop. We begin with the continuous-time case.

Suppose the system in (4.101) is time-invariant, so that the modified notation

$$h(t,\tau) = h(t-\tau)$$

may be used and hence

$$y(t) = \int_{-\infty}^{+\infty} h(t-\tau)\, x(\tau)\, d\tau. \tag{4.122}$$

Furthermore, suppose that the input process is WSS. Then since the input has constant mean, i.e.,

$$m_x(t) = m_x,$$

so does the output, i.e.,

$$m_y(t) = \int_{-\infty}^{+\infty} h(t-\tau)\, m_x\, d\tau = H(0)\, m_x, \tag{4.123}$$

where

$$H(s) = \int_{-\infty}^{+\infty} h(t)\, e^{-st}\, dt \tag{4.124}$$

so that

$$H(0) = H(s)\big|_{s=0} = \int_{-\infty}^{+\infty} h(t)\, dt$$

is the DC gain of the system. Note that for the mean (4.123) to be well-defined our system must be stable in the BIBO sense, i.e.,

$$\int_{-\infty}^{+\infty} |h(t)|\, dt < \infty. \tag{4.125}$$

Otherwise, $s = 0$ would not be in the domain of convergence of $H(s)$ and thus the DC gain would not be defined.

Next, we note that if $x(t)$ and $z(t)$ are jointly WSS, so are $y(t)$ and $z(t)$. In particular, since $K_{xz}(t,s) = K_{xz}(t-s)$, we have

$$\begin{aligned} K_{yz}(t,s) &= \int_{-\infty}^{+\infty} h(t-\tau)\, K_{xz}(\tau - s)\, d\tau \\ &= \int_{-\infty}^{+\infty} h(\sigma)\, K_{xz}(t-s-\sigma)\, d\sigma \end{aligned} \tag{4.126}$$

where the change of variables $\sigma = t - \tau$ was used to obtain the second equality in (4.126). That $y(t)$ and $z(t)$ are jointly WSS follows from the fact that the right-hand side of (4.126) is a function only of $t - s$. Thus we may rewrite (4.126) in the more convenient form

$$K_{yz}(\tau) = \int_{-\infty}^{+\infty} h(t)\, K_{xz}(\tau - t)\, dt = h(\tau) * K_{xz}(\tau) \tag{4.127}$$

where $*$ denotes convolution. Similarly, we obtain that

$$\begin{aligned} K_{zy}(\tau) = K_{yz}(-\tau) &= \int_{-\infty}^{+\infty} h(t)\, K_{xz}(-\tau - t)\, dt \\ &= \int_{-\infty}^{+\infty} h(-t)\, K_{zx}(\tau - t)\, dt \\ &= h(-\tau) * K_{zx}(\tau). \end{aligned} \tag{4.128}$$

Via the preceding results, we can establish that when $y(t)$ and $x(t)$ are related via (4.122) and $x(t)$ is WSS, then $x(t)$ and $y(t)$ are also jointly WSS. To see this, we first note that via the special case of (4.127) with $z(t) = x(t)$, the cross-covariance between $x(t)$ and $y(t)$ satisfies

$$K_{yx}(\tau) = \int_{-\infty}^{+\infty} h(t)\, K_{xx}(\tau - t)\, dt = h(\tau) * K_{xx}(\tau) \tag{4.129}$$

$$K_{xy}(\tau) = \int_{-\infty}^{+\infty} h(t)\, K_{xx}(\tau + t)\, dt = h(-\tau) * K_{xx}(\tau). \tag{4.130}$$

Next, when we let $z(t) = y(t)$ in (4.127) and then use (4.130), we find that

$$\begin{aligned} K_{yy}(\tau) &= \int_{-\infty}^{+\infty} h(t)\, K_{xy}(\tau - t)\, dt \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} h(t)\, h(s)\, K_{xx}(\tau - t + s)\, dt\, ds. \\ &= h(\tau) * h(-\tau) * K_{xx}(\tau) \end{aligned} \tag{4.131}$$

Thus, (4.129) and (4.131) collectively verify that $x(t)$ and $y(t)$ must be jointly WSS.

Again, we emphasize that for the joint second-order characterization of $x(t)$ and $y(t)$ to be well-defined, the system must be stable, i.e., satisfy (4.125). In fact, more generally we also require that an additional notion of stability be satified. To see this, suppose that $x(t)$ is WSS unit-intensity white noise, so that $K_{xx}(\tau) = \delta(\tau)$. Then from (4.131) we see that

$$\operatorname{var} y(t) = K_{yy}(0) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} h(t)\, h(s)\, \delta(s-t)\, dt\, ds = \int_{-\infty}^{+\infty} h^2(t)\, dt. \tag{4.132}$$

Thus, for this variance to be well-defined, the stability condition

$$\int_{-\infty}^{+\infty} h^2(t)\, dt < \infty \tag{4.133}$$

must be satisfied. However, it is worth noting that for systems with, for example, rational system functions, the two notions of stability, i.e., (4.125) and (4.133) are equivalent, and corresponds to the condition that the domain of convergence of $H(s)$ include the $j\omega$-axis.

It is also worth remarking that (4.125) is a necessary and sufficient condition for the output to have well-defined second-order statistics. In particular, if the system is not stable then for at least some (and typically most) WSS inputs, $y(t)$ will not have well-defined second-order statistics. To verify that (4.125) is sufficient to guarantee that the output has well-defined second-order statistics for any WSS input, we simply note, using (4.131) and (4.88), that

$$\begin{aligned} |K_{yy}(\tau)| \le K_{yy}(0) &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} h(t)\, h(s)\, K_{xx}(s-t)\, ds\, dt \\ &\le \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} |h(t)||h(s)||K_{xx}(s-t)|\, ds\, dt \\ &\le K_{xx}(0) \left(\int_{-\infty}^{+\infty} |h(t)|\, dt \right)^2 < \infty. \end{aligned} \tag{4.134}$$

As a final result, suppose $z(t)$ is the output of an LTI system with impulse response $g(t)$ and arbitrary input $w(t)$ so (4.111) specializes to

$$z(t) = \int_{-\infty}^{+\infty} g(t-\tau)\, w(\tau)\, d\tau. \tag{4.135}$$

Then if input processes $w(t)$ and $x(t)$ are jointly WSS, so are the output processes $z(t)$ and $y(t)$, and, specializing (4.112), their cross-covariance functions are related via

$$K_{yz}(\tau) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} h(t)\, g(s)\, K_{xw}(\tau - t + s)\, dt\, ds = h(\tau) * g(-\tau) * K_{xw}(\tau). \tag{4.136}$$

Again, analogous results apply in the discrete-time case. Specifically if $x[n]$ and $z[n]$ are jointly WSS and $x[n]$ is the input to a stable LTI system with impulse

response $h[n]$, then the output

$$y[n] = \sum_{k=-\infty}^{+\infty} h[n-k]\, x[k] \tag{4.137}$$

is jointly WSS with both $x[n]$ and $z[n]$. The mean of the output process is

$$m_y = H(1)\, m_x, \tag{4.138}$$

where

$$H(1) = H(z)|_{z=1} = \sum_{n=-\infty}^{+\infty} h[n]$$

is the DC gain of the discrete-time system. Furthermore,

$$K_{yz}[k] = \sum_{n=-\infty}^{+\infty} h[n]\, K_{xz}[k-n] = h[k] * K_{xz}[k] \tag{4.139}$$

$$K_{zy}[k] = \sum_{n=-\infty}^{+\infty} h[n]\, K_{zx}[k+n] = h[-k] * K_{zx}[k] \tag{4.140}$$

$$K_{yx}[k] = \sum_{n=-\infty}^{+\infty} h[n]\, K_{xx}[k-n] = h[k] * K_{xx}[k] \tag{4.141}$$

$$K_{xy}[k] = \sum_{n=-\infty}^{+\infty} h[n]\, K_{xx}[k+n] = h[-k] * K_{xx}[k] \tag{4.142}$$

$$K_{yy}[k] = \sum_{n=-\infty}^{+\infty} \sum_{m=-\infty}^{+\infty} h[n]\, h[m]\, K_{xx}[k-n+m] = h[k] * h[-k] * K_{xx}[k] \tag{4.143}$$

Suppose, in addition, that $z[n]$ is the output of a discrete-time LTI system with impulse response $g[n]$ and arbitrary input $w[n]$ so (4.120) specializes to

$$z[n] = \sum_{k=-\infty}^{+\infty} g[n-k]\, w[k]. \tag{4.144}$$

Then if input processes $w[n]$ and $x[n]$ are jointly WSS, so are the output processes $z[n]$ and $y[n]$, and, specializing (4.121), their cross-covariance functions are related via

$$K_{yz}[k] = \sum_{n=-\infty}^{+\infty} \sum_{m=-\infty}^{+\infty} h[n]\, g[m]\, K_{xw}[k-n+m] = h[k] * g[-k] * K_{xw}[k]. \tag{4.145}$$

4.6 FREQUENCY-DOMAIN CHARACTERIZATIONS

Several important additional insights into the relationships developed in the previous section are obtained by exploring frequency domain characterizations of

WSS stochastic processes, which we explore in the this section. We begin by developing the notion of the power spectrum for a WSS process.

4.6.1 Power Spectral Densities

We are often interested in how power in a random process is distributed as a function of frequency. A natural measure of this distribution is obtained via the power spectral density (psd) of the process, which is also frequently referred to as simply the power spectrum. Let's first consider the continuous-time case, where we use $S_{xx}(j\omega)$ to denote the power spectral density of a random process $x(t)$ at frequency ω.

From an engineering perspective, a natural way to define the power spectral density of a process is in terms of what a spectrum analyzer would measure, i.e., the power at the output of a narrow bandpass filter centered at the frequency of interest whose input is $x(t)$. In particular, we choose the following definition.[4]

Definition 4.1 (Power Spectral Density) *Let $h_\epsilon(t)$ be the (complex-valued) impulse response of a unit-energy, ideal bandpass filter whose frequency response is*

$$H_\epsilon(j\omega) = \begin{cases} \sqrt{2\pi/\epsilon} & |\omega - \omega_0| < \epsilon/2 \\ 0 & \textit{otherwise} \end{cases},$$

and let $x_\epsilon(t)$ denote the output of this filter when its input is the wide-sense stationary random process $x(t)$. Then the power spectral density of the process $x(t)$ at frequency $\omega = \omega_0$ is[5]

$$S_{xx}(j\omega_0) \triangleq \lim_{\epsilon \to 0} \operatorname{var} x_\epsilon(t).$$

Power spectral densities have a number of important properties. For example, we note immediately that the power spectral density is a real, nonnegative quantity. Other properties will become apparent from two alternative interpretations of the power spectral density that are important consequences of this definition, as we now develop.

---

[4] This definition is in fact useful for a broader class of processes than the wide-sense stationary ones. However, for simplicity we'll restrict our attention to this class.

[5] Since we are now dealing with complex-valued random variables, we need to generalize our definition of covariance. In particular when u and v are complex-valued random variables, their covariance takes the form

$$\operatorname{cov}(u, v) = E\left[(u - m_u)(v - m_v)^*\right]$$

where * denotes complex conjugation. Note that as a consequence we also have that when u is complex, its variance is

$$\operatorname{var} u = E\left[|u - m_u|^2\right].$$

As the first of these interpretations of the power spectral density, we have the following celebrated theorem, variously attributed to both Wiener and Khinchine, which relates the physical notion of power spectrum given by Definition 4.1 to the autocorrelation function of the process.

Theorem 4.1 (Wiener-Khinchine) *Let $S_{xx}(j\omega)$ be the power spectral density of a wide-sense stationary random process $x(t)$ in the sense of Definition 4.1. Then $S_{xx}(j\omega)$ is related to the autocovariance function $K_{xx}(\tau)$ of the process via*

$$S_{xx}(j\omega) = \mathcal{F}\{K_{xx}(\tau)\} = \int_{-\infty}^{+\infty} K_{xx}(\tau)\, e^{-j\omega\tau}\, d\tau, \tag{4.146}$$

i.e., the autocovariance function and power spectral density are Fourier transform pairs.

To prove this result, let us use $T_{xx}(j\omega)$ to denote the Fourier transform of $K_{xx}(\tau)$ and proceed to show that $T_{xx}(j\omega)$ must equal the power spectral density at frequency ω. We first note that

$$K_{x_\epsilon x_\epsilon}(\tau) = h_\epsilon(\tau) * h_\epsilon^*(-\tau) * K_{xx}(\tau). \tag{4.147}$$

Letting $T_{x_\epsilon x_\epsilon}(j\omega)$ denote the Fourier transform of $K_{x_\epsilon x_\epsilon}(\tau)$, we can use the convolution property of Fourier transforms to rewrite (4.147) as

$$T_{x_\epsilon x_\epsilon}(j\omega) = |H_\epsilon(j\omega)|^2\, T_{xx}(j\omega). \tag{4.148}$$

Now

$$\operatorname{var} x_\epsilon(t) = K_{x_\epsilon x_\epsilon}(0) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} T_{x_\epsilon x_\epsilon}(j\omega)\, d\omega,$$

which when we substitute (4.148) yields

$$\operatorname{var} x_\epsilon(t) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} |H_\epsilon(j\omega)|^2\, T_{xx}(j\omega)\, d\omega. \tag{4.149}$$

Taking limits in (4.149) and recognizing that

$$\lim_{\epsilon\to 0} |H_\epsilon(j\omega)|^2 = 2\pi\delta(\omega - \omega_0)$$

we obtain, via the sifting property of impulses,

$$S_{xx}(j\omega_0) = \lim_{\epsilon\to 0} \operatorname{var} x_\epsilon(t) = T_{xx}(j\omega_0).$$

Note that one immediate consequence of Theorem 4.1 comes from the inverse Fourier transform relationship

$$K_{xx}(\tau) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} S_{xx}(j\omega)\, e^{j\omega\tau}\, d\omega. \tag{4.150}$$

Evaluating (4.150) at $\tau = 0$, we have that

$$E\left[\tilde{x}(t)^2\right] = \operatorname{var} x(t) = K_{xx}(0) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} S_{xx}(j\omega)\, d\omega \tag{4.151}$$

where $\tilde{x}(t) = x(t) - m_x(t)$. Thus, viewing $E[\tilde{x}(t)^2]$ as the ensemble-average instantaneous *power* in the truly random (i.e., zero-mean) component of $x(t)$, we have, consistent with Definition 4.1, that this power is a normalized integral over all frequencies of the power per unit frequency, i.e., the power spectral density $S_{xx}(j\omega)$. A useful method for evaluating (4.151) when $S_{xx}(j\omega)$ is rational is described in Appendix 4.B.

A second important interpretation for the power spectral density in Definition 4.1 is provided by the following theorem, which relates the power spectral density to the Fourier transform of windowed versions of the random process.

Theorem 4.2 *Let $X_T(j\omega)$ denote the Fourier transform of a time-limited version of a wide-sense stationary random process $x(t)$; specifically,*

$$X_T(j\omega) = \mathcal{F}\{x(t)\, w_T(t)\},$$

where

$$w_T(t) = \begin{cases} 1 & -T/2 < t < T/2 \\ 0 & \textit{otherwise} \end{cases}.$$

Then the power spectral density $S_{xx}(j\omega)$ is related to the ensemble-average squared-magnitude of $X_T(j\omega)$. In particular,

$$S_{xx}(j\omega) = \lim_{T\to\infty} \frac{1}{T} \operatorname{var} X_T(j\omega). \tag{4.152}$$

To see this result, we note that

$$\tilde{X}_T(j\omega) = X_T(j\omega) - E[X_T(j\omega)] = \int_{-T/2}^{+T/2} \tilde{x}(t)\, e^{-j\omega t}\, dt$$

where

$$\tilde{x}(t) = x(t) - m_x(t),$$

so that

$$|\tilde{X}_T(j\omega)|^2 = \int_{-T/2}^{+T/2} \int_{-T/2}^{+T/2} \tilde{x}(t)\, \tilde{x}(\tau)\, e^{-j\omega(t-\tau)}\, dt\, d\tau. \tag{4.153}$$

Taking expectations in (4.153) we obtain

$$\frac{1}{T} E\left[|\tilde{X}_T(j\omega)|^2\right] = \frac{1}{T} \int_{-T/2}^{+T/2} \int_{-T/2}^{+T/2} K_{xx}(t-\tau)\, e^{-j\omega(t-\tau)}\, dt\, d\tau. \tag{4.154}$$

Next, we use the following change of variables with unit Jacobian

$$\begin{aligned} t &= v + u/2 \\ \tau &= v - u/2 \end{aligned} \tag{4.155}$$

to effect a 45 degree coordinate rotation and scaling. Hence, the region of integration in (4.154) is transformed from a square to a diamond, and we obtain[6]

$$\begin{aligned}\frac{1}{T}E\left[|\tilde{X}_T(j\omega)|^2\right] &= \frac{1}{T}\int_{-T}^{T}\int_{-(T-|u|)/2}^{(T-|u|)/2} K_{xx}(u)\,e^{-j\omega u}\,dv\,du \\ &= \int_{-T}^{T}\left(1-\frac{|u|}{T}\right)K_{xx}(u)\,e^{-j\omega u}\,du \\ &= \int_{-\infty}^{+\infty} g_T(u)K_{xx}(u)\,e^{-j\omega u}\,du, \end{aligned} \tag{4.156}$$

where

$$g_T(t) = \frac{1}{T}w_T(t) * w_T(t) = \begin{cases} 1-|t|/T & |t|\le T \\ 0 & \text{otherwise}\end{cases}. \tag{4.157}$$

Finally, using the convolution theorem we can transform (4.156) from the Fourier transform of the product to the following convolution of Fourier transforms:

$$\frac{2\pi}{T}E\left[|\tilde{X}_T(j\omega)|^2\right] = \int_{-\infty}^{+\infty} S_{xx}(j(\omega-\nu))\,G_T(j\nu)\,d\nu \tag{4.158}$$

where

$$G_T(\omega) = \frac{1}{T}\left|W_T(j\omega)\right|^2 = T\,\mathrm{sinc}^2(T\omega). \tag{4.159}$$

Finally, using that $G_T(\omega) \to 2\pi\delta(\omega)$ as $T\to\infty$, we obtain from (4.158) the desired result, viz.,

$$\frac{1}{T}E\left[|\tilde{X}_T(j\omega)|^2\right] = S_{xx}(j\omega). \tag{4.160}$$

While Theorem 4.2 makes clear that the Fourier transform of a WSS random process doesn't exist in the usual sense, in fact it is possible to talk about the Fourier transform $X(j\omega)$ of a WSS random process $x(t)$ in a *generalized* sense. In particular, the Fourier transform of a WSS random process exists in much the same way that the Fourier transform of a constant 1 doesn't exist in the usual sense, but has a generalized Fourier transform that is the unit impulse $\delta(\omega)$. The random process counterpart is given by the following theorem.

Theorem 4.3 (Spectral Representation) *The Fourier transform of a wide-sense stationary random process $x(t)$, i.e.,*

$$X(j\omega) = \int_{-\infty}^{+\infty} x(t)\,e^{-j\omega t}\,dt \tag{4.161}$$

is a nonstationary white noise process (in frequency) with autocovariance function

$$K_{XX}(j\omega, j\nu) = \mathrm{cov}\left(X(j\omega), X(j\nu)\right) = 2\pi S_{xx}(j\omega)\,\delta(\omega-\nu). \tag{4.162}$$

[6]...leaving the few intermediate steps as an exercise.

Before deriving this result, we stress that $X(j\omega)$ itself is now a random process, but one that is a function of frequency now rather than time.

A derivation of Theorem 4.3 is as follows. Restricting our attention to zero-mean $x(t)$, we first see from (4.161) that $X(j\omega)$ is also a zero-mean random process. Next, let's look at its correlation function. In particular we have

$$\begin{aligned}\operatorname{cov}\left(X(j\omega), X(j\nu)\right) &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} K_{xx}(\tau - t)\, e^{-j(\omega\tau - \nu t)}\, d\tau\, dt \\ &= \int_{-\infty}^{+\infty} e^{j\nu t}\, dt \int_{-\infty}^{+\infty} K_{xx}(\tau - t)\, e^{-j\omega\tau}\, d\tau. \end{aligned} \tag{4.163}$$

Using the change of variables $\alpha = \tau - t$ in (4.163) we then get

$$\begin{aligned}\operatorname{cov}\left(X(j\omega), X(j\nu)\right) &= \int_{-\infty}^{+\infty} e^{j\nu t} e^{-j\omega t}\, dt \int_{-\infty}^{+\infty} K_{xx}(\alpha)\, e^{-j\omega\alpha}\, d\alpha \\ &= \int_{-\infty}^{+\infty} e^{j\nu t} e^{-j\omega t}\, S_{xx}(j\omega)\, dt \\ &= S_{xx}(j\omega)\, \mathcal{F}\left\{e^{j\nu t}\right\} \\ &= 2\pi S_{xx}(j\omega)\, \delta(\omega - \nu), \end{aligned} \tag{4.164}$$

which completes our derivation.

Several comments are worth making. First, we emphasize that from (4.162) we see that $X(j\omega)$ is a nonstationary white noise process. An important consequence of this is that for wide-sense stationary random processes, frequency components of the process corresponding to distinct frequencies are uncorrelated! As we'll see, this result has a major impact on the frequency domain structure and interpretation of many optimal estimators for WSS random processes.

Furthermore, as a white noise, it has infinite variance, i.e.,

$$\operatorname{var} X(j\omega) = \operatorname{cov}\left(X(j\omega), X(j\omega)\right) = \infty,$$

which we would anticipate since we don't expect the integral (4.161) to exist in the usual sense. However, as you recall from our earlier discussion of white noise, although white noise doesn't exist in a formal mathematical sense, it is the generalized derivative of a well-defined uncorrelated increments process. There is an analogous quantity here, which we call the integrated Fourier transform of $x(t)$, which we denote using $Z(j\omega)$. In particular, $Z(j\omega)$, which is given by

$$Z(j\omega) = \frac{1}{2\pi}\int_{-\infty}^{\omega} X(j\upsilon)\, d\upsilon = \int_{-\infty}^{+\infty} x(t)\, \left[\frac{1}{2\pi}\int_{-\infty}^{\omega} e^{-j\upsilon t}\, d\upsilon\right]\, dt,$$

is a well-defined random process. It has, for example, variance

$$\begin{aligned}\operatorname{var} Z(j\omega) &= \frac{1}{(2\pi)^2}\int_{-\infty}^{\omega}\int_{-\infty}^{\omega} E\left[\tilde{X}(j\alpha)\tilde{X}^*(j\beta)\right]\, d\alpha\, d\beta \\ &= \frac{1}{2\pi}\int_{-\infty}^{\omega} S_{xx}(j\alpha)\, d\alpha, \end{aligned} \tag{4.165}$$

where we have used (4.162) to obtain the last equality in (4.165). Note from (4.165) that $Z(\omega)$ is a nonstationary random process whose variance grows monotonically from 0 at $\omega = -\infty$ to $R_{xx}(0)$, the total power in the process, at $\omega = +\infty$.

Again we emphasize that $Z(j\omega)$ is an uncorrelated increments process, as we would expect of a process having white noise as a derivative. In particular, for any $-\infty \le \xi \le \nu \le \omega \le +\infty$ we have that

$$\operatorname{cov}\left([Z(j\omega) - Z(j\nu)], Z(j\xi)\right) = 0,$$

which can be verified from

$$Z(j\omega) - Z(j\nu) = \frac{1}{2\pi}\int_{\nu}^{\omega} X(j\upsilon)\, d\upsilon$$
$$Z(j\xi) = \frac{1}{2\pi}\int_{-\infty}^{\xi} X(j\upsilon)\, d\upsilon.$$

since $X(j\omega)$ is a white process. In turn, since $Z(j\omega)$ is an uncorrelated increments process, this means that

$$\begin{aligned} K_{ZZ}(j\omega, j\nu) &= \operatorname{cov}\left(Z(j\omega), Z(j\nu)\right) \\ &= \operatorname{var}\left[Z(j\min(\omega,\nu))\right] \\ &= \frac{1}{2\pi}\int_{-\infty}^{\min(\omega,\nu)} S_{xx}(j\alpha)\, d\alpha \end{aligned}$$

where the last equality follows from (4.165).

Let's now briefly summarize the corresponding results of this section for discrete-time processes. In discrete-time, we use $S_{xx}(e^{j\omega})$ to denote the (2π-periodic) power spectral density of a random process $x[n]$, where ω is the (discrete-time) frequency variable, which is defined analogously to the continuous-time case, as follows.

Definition 4.2 (Power Spectral Density) *Let $h_\epsilon[n]$ be the (complex-valued) impulse response of a unit-energy, ideal bandpass filter whose frequency response is*

$$H_\epsilon(e^{j\omega}) = \begin{cases} \sqrt{2\pi/\epsilon} & |\omega - \omega_0 + 2\pi k| < \epsilon/2 \text{ for } k = 0, \pm 1, \pm 2, \ldots \\ 0 & \text{otherwise} \end{cases},$$

and let $x_\epsilon[n]$ denote the output of this filter when its input is the wide-sense stationary random process $x[n]$. Then the power spectral density of the process $x[n]$ at frequency $\omega = \omega_0$ is

$$S_{xx}(e^{j\omega_0}) \triangleq \lim_{\epsilon \to 0} \operatorname{var} x_\epsilon[n].$$

With this definition, we have the following interpretations.

Theorem 4.4 (Wiener-Khinchine) *Let $S_{xx}(e^{j\omega})$ be the power spectral density of a wide-sense stationary random process $x[n]$ in the sense of Definition 4.2. Then $S_{xx}(e^{j\omega})$ is related to the autocovariance function $K_{xx}[k]$ of the process via*

$$S_{xx}(e^{j\omega}) = \mathcal{F}\{K_{xx}[k]\} = \sum_{k=-\infty}^{+\infty} K_{xx}[k]\, e^{-j\omega k} \tag{4.166}$$

i.e., the autocovariance function and power spectral density are (discrete-time) Fourier transform pairs.

Again an immediate consequence of Theorem 4.4 comes from the inverse Fourier transform relationship

$$K_{xx}[k] = \frac{1}{2\pi} \int_{-\pi}^{\pi} S_{xx}(e^{j\omega})\, e^{j\omega k}\, d\omega. \tag{4.167}$$

Evaluating (4.167) at $k = 0$, we have that the total power in the process is

$$E\left[\tilde{x}[n]^2\right] = \operatorname{var} x[n] = K_{xx}[0] = \frac{1}{2\pi} \int_{-\pi}^{\pi} S_{xx}(e^{j\omega})\, d\omega \tag{4.168}$$

where $\tilde{x}[n] = x[n] - m_x[n]$.

The second interpretation for the power spectral density again relates the power spectral density to the Fourier transform of windowed versions of the random process.

Theorem 4.5 *Let $X_N(e^{j\omega})$ denote the Fourier transform of a time-limited version of a wide-sense stationary random process $x[n]$; specifically,*

$$X_N(e^{j\omega}) = \mathcal{F}\{x[n]\, w_N[n]\},$$

where

$$w_N[n] = \begin{cases} 1 & -N \le n \le N \\ 0 & \textit{otherwise} \end{cases}.$$

Then the power spectral density $S_{xx}(e^{j\omega})$ is related to the ensemble-average squared-magnitude of $X_N(e^{j\omega})$. In particular,

$$S_{xx}(e^{j\omega}) = \lim_{N\to\infty} \frac{1}{2N+1} \operatorname{var} X_N(e^{j\omega}) \tag{4.169}$$

In turn, the generalized Fourier transform of a discrete-time WSS process is defined as follows.

Theorem 4.6 (Spectral Representation) *The discrete-time Fourier transform of a wide-sense stationary random process $x[n]$, i.e.,*

$$X(e^{j\omega}) = \sum_{n=-\infty}^{+\infty} x[n]\, e^{-j\omega n} \tag{4.170}$$

is a periodic process in frequency that is white over any frequency interval of length 2π*, and has autocovariance function*

$$K_{XX}(e^{j\omega}, e^{j\nu}) = \operatorname{cov}\left(X(e^{j\omega}), X(e^{j\nu})\right) = 2\pi S_{xx}(e^{j\omega}) \sum_{k=-\infty}^{+\infty} \delta(\omega - \nu + 2\pi k). \quad (4.171)$$

4.6.2 Power Spectral Densities in the Complex Plane

In many situations we will often find it more convenient to work not with the Fourier transform (4.146) but with the corresponding bilateral Laplace transform

$$S_{xx}(s) = \int_{-\infty}^{+\infty} K_{xx}(\tau)\, e^{-s\tau}\, d\tau. \quad (4.172)$$

where s is complex. While we will use the terminology power spectral density for $S_{xx}(s)$, it is worth mentioning that in some texts this term is strictly reserved for the Fourier transform of $K_{xx}(\tau)$, i.e., $S_{xx}(j\omega)$. In this case $S_{xx}(s)$ is either avoided or referred to as the "complex power spectral density." We will use the term power spectral density interchangeably for $S_{xx}(s)$ and for $S_{xx}(j\omega)$.

Since the covariance function is symmetric, i.e., $K_{xx}(\tau) = K_{xx}(-\tau)$, one important property of power spectral densities is their (diagonal) symmetry in the s-plane, i.e.,

$$S_{xx}(s) = S_{xx}(-s). \quad (4.173)$$

This implies that the value of the power spectral density at a point $s_0 = \sigma_0 + j\omega_0$ is identical to its value at a point corresponding to reflecting s_0 first across the imaginary axis, and then across the real axis, yielding $-s_0 = -\sigma_0 - j\omega$.

It is also straightforward to verify that the power spectrum is real on the $j\omega$-axis, which of course was an immediate consequence of Definition 4.1. To see this using transform analysis, we first note that since $K_{xx}(\tau)$ is real-valued, we must have that

$$S_{xx}(j\omega) = S_{xx}^*(-j\omega). \quad (4.174)$$

But when (4.173) is specialized to the the $j\omega$-axis we obtain

$$S_{xx}(j\omega) = S_{xx}(-j\omega). \quad (4.175)$$

Thus, combining (4.174) with (4.175) we obtain

$$S_{xx}(j\omega) = S_{xx}^*(j\omega), \quad (4.176)$$

i.e., $S_{xx}(j\omega)$ is real-valued. It is important to note, however, that away from the $j\omega$-axis, the spectrum $S_{xx}(s)$ is in general *not* real valued.

More generally, the symmetry property (4.173) places strong constraints on the pole-zero patterns of power spectral densities that are rational functions of s.

Specifically, if s_0 is a pole (zero) of $S_{xx}(s)$, then so is $-s_0$. Furthermore, since $K_{xx}(t)$ is real, if s_0 is a pole (zero) of $S_{xx}(s)$, then so is its complex conjugate s_0^*. That is,

> The poles and zeros of $S_{xx}(s)$ have *quadrantal* symmetry: if s_0 is a pole (zero) then so are $s_0^*, -s_0$ and $-s_0^*$. (4.177)

Finally, we can obtain a complete characterization of the pole-zero pattern of a rational power spectral density $S_{xx}(s)$. First, since $\operatorname{var} x(t)$ is finite, we see from (4.151) that $S_{xx}(j\omega)$ must be well-behaved and in particular cannot have singularities. Thus $S_{xx}(s)$ cannot have any poles on the $j\omega$-axis. Furthermore, while $S_{xx}(s)$ can have zeros on the $j\omega$-axis, these zeros must occur with even multiplicity. This is a consequence of the fact that $S_{xx}(s)$ is real and nonnegative on the $j\omega$-axis. Specifically, if $S_{xx}(j\omega_0) = 0$ at some frequency ω_0, then a Taylor series expansion of the power spectrum about this point takes the form

$$S_{xx}(j\omega) = a_k(\omega - \omega_0)^k + a_{k+1}(\omega - \omega_0)^{k+1} + \cdots \tag{4.178}$$

where $k \geq 1$ since $S_{xx}(j\omega_0) = 0$. Then, since we can take ω sufficiently close to ω_0 so that the first term in (4.178) dominates, we see that the only way to guarantee that $S_{xx}(j\omega)$ remains nonnegative in this interval is to have k even and $a_k > 0$. In turn, this implies that $S_{xx}(j\omega)$ must be of the form

$$S_{xx}(j\omega) = (\omega - \omega_0)^{2m} F(j\omega) \tag{4.179}$$

where $F(j\omega)$ has neither poles nor zeros at ω_0. Thus, summarizing, we have the following:

> A rational power spectral density $S_{xx}(s)$ has poles and zeros with quadrantal symmetry. Furthermore $S_{xx}(s)$ has no poles on the $j\omega$-axis and its $j\omega$-axis zeros occur with even multiplicity. (4.180)

This characterization has profound implications for the modeling, analysis, and processing of WSS processes, some of which we will explore not only in this chapter but in subsequent chapters as well.

Example 4.13

Consider again the WSS process $x(t)$ introduced in Example 4.11 with first-order covariance function

$$K_{xx}(\tau) = \sigma^2 e^{-\lambda|\tau|} \tag{4.181}$$

Taking the Laplace transform of (4.181) we see that $x(t)$ has a power spectral density of the form

$$S_{xx}(s) = \frac{\sigma^2}{\lambda + s} + \frac{\sigma^2}{\lambda - s} = \frac{2\lambda\sigma^2}{\lambda^2 - s^2}, \tag{4.182}$$

which we will refer to as a "first-order" power spectral density. Note that $S_{xx}(s)$ has the basic properties we have identified for power spectral densities. For example,

$$S_{xx}(j\omega) = \frac{2\lambda\sigma^2}{\lambda^2 + \omega^2} \geq 0 \qquad \text{for all } \omega. \tag{4.183}$$

Note, too, that the spectrum (4.183) for $x(t)$ is effectively bandlimited, with a 3dB cut-off frequency of $\omega = \lambda$ and a spectral density $2\sigma^2/\lambda$ for frequencies sufficiently below the cut-off. Comparing this with (4.181) we see the familiar Fourier transform uncertainty principle in effect: as λ increases, $K_{xx}(\tau)$ becomes increasingly compressed in time while $S_{xx}(j\omega)$ becomes increasingly dilated in frequency. The compression of $K_{xx}(\tau)$ corresponds to a decrease in the correlation time of the process, which leads to more rapid fluctuations in the sample paths of the process and is reflected in the broadening of the spectrum $S_{xx}(j\omega)$.

Example 4.14

Note that as $\lambda \to +\infty$, the characteristics of the process $x(t)$ in Example 4.13 approach those of the ideal white noise process $x(t)$, for which

$$K_{xx}(\tau) = \sigma^2\delta(\tau). \tag{4.184}$$

Taking the Laplace transform of (4.184), we see that the associated power spectral density is constant, i.e.,

$$S_{xx}(s) = \sigma^2. \tag{4.185}$$

It is this characteristic of white noise—that its power spectral density is flat over all frequency—that motivates the designation "white noise," since white light has an analogous decomposition in terms of the electromagnetic frequency spectrum. For future reference, we remark that the quantity σ^2 is alternately referred to as the *intensity* or the *spectral height* of the white noise process. We stress that σ^2 is *not* the total power or, equivalently, variance of the process. Indeed, these quantities are infinite in the case of white noise.

Let's again briefly summarize the discrete-time counterparts of the results in this section. In discrete-time, we more generally work with the bilateral z-transform

$$S_{xx}(z) = \sum_{k=-\infty}^{+\infty} K_{xx}[k]\, z^{-k} \tag{4.186}$$

where z is complex, and again use the terminology power spectral density for both $S_{xx}(z)$ and its evaluation on the unit-circle $S_{xx}(e^{j\omega})$.

Since the covariance function is symmetric, i.e., $K_{xx}[k] = K_{xx}[-k]$, the power spectral density also has symmetry across the unit circle ($|z| = 1$) in the z-plane, i.e.,

$$S_{xx}(z) = S_{xx}(1/z). \tag{4.187}$$

In particular, the value of the power spectral density at a point $z_0 = r_0 e^{j\theta_0}$ is the same as its value at a point corresponding to first reflecting z_0 across the unit-circle to a radius of $1/r_0$ and then across the imaginary axis so that the phase is reversed, yielding $1/z_0 = (1/r_0)e^{-j\theta_0}$.

While the power spectral density is in general complex-valued in the z-plane, from its definition it is real-valued on the unit circle. In the case of rational power

spectra, this implies that if z_0 is a pole or zero of $S_{xx}(z)$, so is its complex-conjugate z_0^*. Together with the symmetry condition (4.187), this implies:

> A rational power spectral density $S_{xx}(z)$ has poles and zeros appearing in symmetric patterns: if z_0 is a pole (zero), then so are z_0^*, $1/z_0$, and $1/z_0^*$. Furthermore, $S_{xx}(z)$ has no poles on the unit circle and its unit circle zeros occur with even multiplicity. (4.188)

Example 4.15

Consider the WSS process $x[n]$ introduced in Example 4.11 with first-order covariance function

$$K_{xx}[k] = \sigma^2 \alpha^{|k|} \tag{4.189}$$

with $|\alpha| < 1$. Taking the z-transform of (4.189) we obtain

$$\begin{aligned} S_{xx}(z) &= \frac{\sigma^2}{(1-\alpha z^{-1})} + \frac{\sigma^2}{1-\alpha z} - \sigma^2 \\ &= \frac{\sigma^2(1-\alpha^2)}{(1-\alpha z^{-1})(1-\alpha z)} \end{aligned} \tag{4.190}$$

which we refer to as a first-order discrete-time spectrum.

Note that $S_{xx}(z)$ has the basic properties we have identified for power spectral densities: $S_{xx}(z) = S_{xx}(z^{-1})$, the poles of $S_{xx}(z)$ are at α and $1/\alpha$, and $S_{xx}(e^{j\omega})$, as given below, is nonnegative. Specifically,

$$S_{xx}(e^{j\omega}) = \frac{\sigma^2(1-\alpha^2)}{1+\alpha^2-2\alpha\cos\omega} \geq 0. \tag{4.191}$$

If $\alpha > 0$, $S_{xx}(e^{j\omega})$ is a low-pass spectrum with peak at $\omega = 0$ and with decreasing bandwidth as $\alpha \to 1$. If $\alpha < 0$, $S_{xx}(e^{j\omega})$ is a high-pass spectrum with peak at $\omega = \pm\pi$ and with decreasing bandwidth as $\alpha \to -1$. Again we have the usual time-frequency tradeoff that is a consequence of the Fourier transform uncertainty principle: smaller values of $|\alpha|$ correspond to processes with shorter correlation times and larger bandwidths.

Example 4.16

Note that when $\alpha \to 0$, the process $x[n]$ in Example 4.15 becomes a discrete-time white noise process $x[n]$ as introduced in Example 4.6, i.e.,

$$K_{xx}[n] = \sigma^2 \delta[n] \tag{4.192}$$

Recall that in contrast to its continuous-time counterpart, discrete-time white noise is a well-defined process with no analytical subtleties. Taking the z-transform of (4.192) we see that

$$S_{xx}(z) = \sigma^2 \tag{4.193}$$

so that $x[n]$ has a flat power spectral density over all frequencies.

Power spectral densities and the interpretations we've developed will prove to be extremely useful to us in analyzing WSS processes and the effects of LTI systems on them. In particular, we'll see that as in the deterministic analysis of LTI systems, frequency-domain techniques based on the power spectral density are particularly well-suited to the analysis of WSS stochastic processes and their second-order statistics through LTI systems.

4.6.3 Cross Power Spectra and LTI Systems

In the following we again first carry out our analysis for continuous-time processes and systems and then briefly describe the corresponding discrete-time results.

If $x(t)$ and $y(t)$ are jointly WSS processes with cross-covariance $K_{xy}(\tau)$, then the *cross spectral density* (or simply *cross-spectrum*) $S_{xy}(s)$ is defined as the Laplace transform

$$S_{xy}(s) = \int_{-\infty}^{+\infty} K_{xy}(\tau)\, e^{-s\tau}\, d\tau, \tag{4.194}$$

and once again we will refer to both $S_{xy}(s)$ and $S_{xy}(j\omega)$ as cross-spectral densities.

Unlike the power spectrum $S_{xx}(s)$, the cross spectrum $S_{xy}(s)$ is, in general, complex-valued both on the $j\omega$-axis and elsewhere in the s-plane. In addition, we note that one important relationship among cross-spectra is

$$S_{yx}(s) = S_{xy}(-s), \tag{4.195}$$

which is a consequence of the symmetry property among cross-covariance functions (cf. (4.94)).

The cross-spectrum has a number of interpretations analogous to those developed in Section 4.6.1. For example, we can readily express the cross-correlation between the generalized Fourier transforms of $x(t)$ and $y(t)$ in terms of $S_{xy}(j\omega)$; specifically, analogous to (4.171) we have

$$K_{XY}(j\omega, j\nu) = \operatorname{cov}\left(X(j\omega), Y(j\nu)\right) = 2\pi S_{xy}(j\omega)\, \delta(\omega - \nu). \tag{4.196}$$

As another interpretation, the cross-spectrum $S_{xy}(\omega)$ provides an important frequency-domain measure of the correlation between a pair of jointly WSS random processes $x(t)$ and $y(t)$. In practice, one usually works with a normalized version of the cross-spectrum in assessing this dependence. In particular, the (complex-valued) *coherence* between $x(t)$ and $y(t)$ is defined as

$$C_{xy}(j\omega) = \frac{S_{xy}(j\omega)}{\sqrt{S_{xx}(j\omega)}\sqrt{S_{yy}(j\omega)}}. \tag{4.197}$$

The magnitude-squared coherence (MSC) between $x(t)$ and $y(t)$ satisfies

$$0 \le |C_{xy}(j\omega)|^2 \le 1. \tag{4.198}$$

In the remainder of this section, we explore the use of the cross spectrum in the analysis of WSS processes filtered by LTI systems. To begin, recall that frequency domain analysis is particularly well-suited to LTI systems because complex exponentials, i.e., functions of the form e^{st} where s is complex, are eigenfunctions of LTI systems. The consequence of this result is the convolution property

$$a(t) * h(t) \stackrel{\mathcal{L}}{\longleftrightarrow} A(s)\, H(s). \tag{4.199}$$

Analogous properties arise in the analysis of WSS processes with LTI systems. Specifically, using the convolution property (4.199) and the time-reversal property

$$b(t) = a(-t) \stackrel{\mathcal{L}}{\longleftrightarrow} B(s) = A(-s) \tag{4.200}$$

we obtain the frequency domain counterparts of (4.127)–(4.131). Specifically, when $x(t)$ is the input to a stable LTI system with system function $H(s)$ and output $y(t)$, and when $x(t)$ and $z(t)$ are jointly WSS, we have

$$S_{yz}(s) = H(s)\, S_{xz}(s) \tag{4.201}$$
$$S_{zy}(s) = H(-s)\, S_{zx}(s) \tag{4.202}$$
$$S_{yx}(s) = H(s)\, S_{xx}(s) \tag{4.203}$$
$$S_{xy}(s) = H(-s)\, S_{xx}(s) \tag{4.204}$$
$$S_{yy}(s) = H(s)\, H(-s) S_{xx}(s). \tag{4.205}$$

Furthermore, as the frequency domain counterpart to (4.136), when $z(t)$ is the response of a stable LTI system with system function $G(s)$ whose input is $w(t)$, and when $w(t)$ and $x(t)$ are jointly WSS, we have that

$$S_{yz}(s) = H(s)\, G(-s)\, S_{xw}(s). \tag{4.206}$$

As a final comment, note that the results (4.201)–(4.206) can also be verified using relations of the form (4.162) and (4.196) that involve generalized Fourier transforms, together with the frequency domain input-output descriptions

$$Y(j\omega) = H(j\omega)\, X(j\omega)$$
$$Z(j\omega) = G(j\omega)\, W(j\omega).$$

Before concluding this section, let us briefly summarize the discrete-time counterparts. First, if $x[n]$ and $y[n]$ are jointly WSS processes with cross-covariance $K_{xy}[k]$, then the cross spectral density $S_{xy}(z)$ is defined as the z-transform

$$S_{xy}(z) = \sum_{k=-\infty}^{+\infty} K_{xy}[k]\, z^{-k} \tag{4.207}$$

and once again we will refer to both $S_{xy}(z)$ and $S_{xy}(e^{j\omega})$ as cross-spectral densities.

The cross spectrum $S_{xy}(z)$ is, in general, complex-valued both on the unit circle and elsewhere in the z-plane. In addition, we have the following important relationship among cross-spectra

$$S_{yx}(z) = S_{xy}(1/z) \tag{4.208}$$

which is a consequence of the symmetry property among cross-covariance functions (cf. (4.99)). Similarly, we may readily express the cross-correlation between the generalized Fourier transforms of $x[n]$ and $y[n]$ in terms of $S_{xy}(e^{j\omega})$:

$$K_{XY}(e^{j\omega}, e^{j\nu}) = \operatorname{cov}\left(X(e^{j\omega}), Y(e^{j\nu})\right) = 2\pi S_{xy}(e^{j\omega}) \sum_{k=-\infty}^{+\infty} \delta(\omega - \nu + 2\pi k). \quad (4.209)$$

We also define the (complex-valued) *coherence* between $x[n]$ and $y[n]$ as

$$C_{xy}(e^{j\omega}) = \frac{S_{xy}(e^{j\omega})}{\sqrt{S_{xx}(e^{j\omega})}\sqrt{S_{yy}(e^{j\omega})}}, \quad (4.210)$$

and note that the magnitude-squared coherence (MSC) between $x[n]$ and $y[n]$ satisfies

$$0 \leq |C_{xy}(e^{j\omega})|^2 \leq 1. \quad (4.211)$$

As was the case for continuous-time systems, the cross spectra for discrete-time processes is particularly useful in the analysis of discrete-time WSS processes filtered by LTI systems. Again, frequency domain analysis is particularly well-suited to discrete-time LTI systems because complex exponentials, i.e., functions of the form z^n where z is complex, are eigenfunctions of such systems. The consequence of this result is the convolution property

$$a[n] * h[n] \overset{\mathcal{Z}}{\longleftrightarrow} A(z)\, H(z). \quad (4.212)$$

Using the convolution property (4.212) and the time-reversal property

$$b[n] = a[-n] \overset{\mathcal{Z}}{\longleftrightarrow} B(z) = A(1/z) \quad (4.213)$$

we obtain the frequency domain counterparts of (4.139)–(4.143). Specifically, when $x[n]$ is the input to a stable LTI system with system function $H(z)$ and output $y[n]$, and when $x[n]$ and $z[n]$ are jointly WSS, we have

$$S_{yz}(z) = H(z)\, S_{xz}(z) \quad (4.214)$$
$$S_{zy}(z) = H(1/z)\, S_{zx}(z) \quad (4.215)$$
$$S_{yx}(z) = H(z)\, S_{xx}(z) \quad (4.216)$$
$$S_{xy}(z) = H(1/z)\, S_{xx}(z) \quad (4.217)$$
$$S_{yy}(z) = H(z)\, H(1/z) S_{xx}(z). \quad (4.218)$$

Furthermore, as the frequency domain counterpart to (4.145), when $z[n]$ is the response of a stable LTI system with system function $G(z)$ whose input is $w[n]$, and when $w[n]$ and $x[n]$ are jointly WSS, we have that

$$S_{yz}(z) = H(z)\, G(1/z)\, S_{xw}(z). \quad (4.219)$$

Similarly, the results (4.214)–(4.219) can also be verified using relations of the form (4.171) and (4.209) that involve generalized Fourier transforms, together with the frequency domain input-output descriptions

$$Y(e^{j\omega}) = H(e^{j\omega})\, X(e^{j\omega})$$
$$Z(e^{j\omega}) = G(e^{j\omega})\, W(e^{j\omega}).$$

4.6.4 Spectral Factorization, Shaping and Whitening Filters

As an application of the preceding results, in this section we explore how the concept of white noise plays an important role in the modeling of WSS processes. To develop this perspective, we begin with the continuous-time case and suppose that $w(t)$ is a white noise process with unit spectral height, so $S_{ww}(s) = 1$. Furthermore, suppose we let $w(t)$ be the input to an LTI system with system function $G(s)$. Then the output of this system $x(t)$ has, via (4.205), a power spectral density of the form

$$S_{xx}(s) = G(s)\,G(-s), \tag{4.220}$$

which, on the $j\omega$-axis, yields

$$S_{xx}(j\omega) = |G(j\omega)|^2. \tag{4.221}$$

Thus the power spectrum (and hence second-order characterization) of $x(t)$ is completely determined by the system function $G(s)$.

By construction, we have shown that for every choice of $G(s)$, (4.220) is a valid power spectrum. It is also possible to show the converse—that every rational power spectrum $S_{xx}(s)$ can be expressed in the form

$$S_{xx}(s) = G(s)\,G(-s) \tag{4.222}$$

for a suitable choice of $G(s)$, which implies that any second-order characterization of this type may be obtained as the output of a system with system function $G(s)$ whose input is unit intensity WSS white noise. The filter $G(s)$ is referred to as a *shaping* or *synthesis* filter for $x(t)$.

To verify (4.222), we exploit (4.180). Specifically, we construct $G(s)$ by assigning half of the poles and zeros to $G(s)$, and their mirror images with respect to the $j\omega$-axis to $G(-s)$. Likewise, the gain $S_{xx}(0) = G^2(0)$ is split evenly between the two terms. We stress that with this procedure, since $S_{xx}(s)$ has no poles on the $j\omega$-axis, neither does $G(s)$. Furthermore since any zero on the $j\omega$-axis occurs with even multiplicity we can put half of these zeros in $G(s)$ and the other half in $G(-s)$.

As a final comment before we proceed to an example, the result (4.222) in fact holds for a broader class of spectral densities than just those with rational spectra. Moreover the following celebrated theorem that provides a concise necessary and sufficient condition for there to exist a specifically *causal* $G(s)$.

Theorem 4.7 (Paley-Wiener) *The power spectral density $S_{xx}(s)$ of a finite variance wide-sense stationary process $x(t)$ can be expressed in the form* (4.222)*, i.e.,*

$$S_{xx}(s) = G(s)\,G(-s),$$

where $G(s)$ is the system function of a causal system, if and only if

$$\int_{-\infty}^{+\infty} \left| \frac{\ln S_{xx}(j\omega)}{1+\omega^2} \right| d\omega < \infty. \tag{4.223}$$

Example 4.17

Consider again the first-order spectrum (4.181). In this case we can write

$$S_{xx}(s) = \left(\frac{\sqrt{2\lambda\sigma^2}}{\lambda + s}\right)\left(\frac{\sqrt{2\lambda\sigma^2}}{\lambda - s}\right). \tag{4.224}$$

Thus we can think of the process $x(t)$ as the output of a *first-order* system driven by unit intensity white noise, where two possible choices for the first-order system correspond to taking as $G(s)$ either the first or second factors on the right-hand side of (4.224). Since we require that $G(s)$ correspond to a stable system, taking $G(s)$ to be the first factor corresponds to a causal first-order system, while taking $G(s)$ to be the second corresponds to an anticausal system. If we want a causal shaping filter, we can then take

$$G(s) = \frac{\sqrt{2\lambda\sigma^2}}{s + \lambda}, \tag{4.225}$$

for which

$$g(t) = \sqrt{2\lambda\sigma^2}e^{-\lambda t}u(t).$$

The process of factoring $S_{xx}(s)$ in the form $G(s)G(-s)$ is referred to as *spectral factorization*. As the preceding example indicates, there is flexibility in choosing which poles and zeros we include in the factor $G(s)$. In fact, even if we restrict our attention to causal systems $G(s)$, there are many possibilities. For example, this condition only constrains the poles we assign to $G(s)$; zeros may be evenly partitioned among $G(s)$ and $G(-s)$ in any manner and in particular $G(s)$ need not be assigned all left half plane zeros. More generally, and more importantly, there are always *infinitely* many possible spectral factorizations. In particular, if $G(s)$ is a valid spectral factor, so is

$$G'(s) = G(s)\,A(s) \tag{4.226}$$

where $A(s)$ is any stable *all-pass* system, i.e., any system for which

$$|A(j\omega)|^2 = 1. \tag{4.227}$$

or equivalently

$$A(s)\,A(-s) = 1. \tag{4.228}$$

To verify that $G'(s)$ is also a valid spectral factor, it suffices to note, using (4.228), that

$$G'(s)\,G'(-s) = G(s)\,A(s)\,A(-s)\,G(-s) = G(s)\,G(-s) = S_{xx}(s).$$

In effect, we see that all-pass terms $A(s)$ provide a mechanism for selecting the phase of the spectral factor. In general, specializing (4.222) to the $j\omega$-axis we see that the spectral factorization relation

$$S_{xx}(j\omega) = |G(j\omega)|^2$$

constrains only the magnitude of the spectral factor.

Example 4.18

Consider a WSS process $x(t)$ with covariance function

$$K_{xx}(\tau) = 2\delta(\tau) + 3e^{-|\tau|}. \tag{4.229}$$

The power spectral density for $x(t)$ then is

$$S_{xx}(s) = 2 + \frac{6}{1-s^2} = \frac{2(4-s^2)}{1-s^2} = \frac{2(2-s)(2+s)}{(1-s)(1+s)} \tag{4.230}$$

If we want to obtain a spectral factorization of the form (4.222) where $G(s)$ is causal, two possible choices are

$$G_1(s) = \frac{\sqrt{2}(2-s)}{(1+s)} \qquad \text{or} \qquad G_2(s) = \frac{\sqrt{2}(2+s)}{(1+s)}. \tag{4.231}$$

However, other possibilities exist. For example, choosing

$$A(s) = \frac{3-s}{3+s}$$

we obtain (cf. (4.226)) that

$$G_3(s) = A(s)\,G_1(s) = \frac{\sqrt{2}(2-s)(3-s)}{(1+s)(3+s)} \tag{4.232}$$

is also a valid spectral factor.

Again we emphasize that if unit intensity white noise is applied as input to *any* of the causal systems in (4.231) or (4.232), the output will have spectrum (4.230).

In addition to its role in modeling the second-order characteristics of WSS processes, spectral factorization is also used in the construction of *whitening filters*. Specifically, a whitening filter for a WSS process $x(t)$ is an LTI system $H(s)$ such that when $x(t)$ is applied as input, the resulting output $w(t)$ is unit-intensity WSS white noise.[7]

A whitening filter for $x(t)$ can be readily constructed from a spectral factorization of the process into the form (4.222). In particular, if we apply $x(t)$ as an input to the LTI system with system function $H(s) = 1/G(s)$, then the resulting output $w(t)$ has spectrum

$$S_{ww}(s) = H(s)H(-s)S_{xx}(s) = \frac{1}{G(s)}\frac{1}{G(-s)}G(s)G(-s) = 1 \tag{4.233}$$

Note that the poles of $H(s) = 1/G(s)$ are precisely the zeros of $G(s)$, and thus if a causal whitening filter is desired, we must choose a spectral factor $G(s)$ having all its *zeros* in the left-half plane. Again, however, even causal whitening filters for

[7]The unit intensity constraint is not necessary but often convenient. More generally, a whitening filter is any LTI system producing a flat but nonzero output spectrum when $x(t)$ is the input.

a process are not uniquely defined, and in particular all-pass terms can always be included in $H(s)$.

Although there are in general infinitely many whitening filters for a given power spectrum, in some problems—particularly several we will encounter in constructing optimal estimators for stochastic processes in a subsequent chapter—the whitening filter of interest corresponds to the spectral factor with all its poles and all its zeros in the left-half plane. This unique[8] whitening filter is causal and stable, and has a causal and stable inverse, and is denoted by $S_{xx}^+(s)$, i.e.,

$$S_{xx}^+(s) \text{ is the spectral factor of } S_{xx}(s) \text{ with all its poles and zeros in the left-half plane.} \tag{4.234}$$

Similarly, $S_{xx}^-(s)$ is the spectral factor with all poles and zeros in the right-half plane, and thus via the symmetry property of $S_{xx}(s)$ we have

$$S_{xx}^-(s) = S_{xx}^+(-s) \tag{4.235}$$

and

$$S_{xx}(s) = S_{xx}^+(s)S_{xx}^-(s) = S_{xx}^+(s)S_{xx}^+(-s). \tag{4.236}$$

Example 4.19

For the spectrum (4.230) in Example 4.18 we have

$$S_{xx}^+(s) = G_2(s) = \frac{\sqrt{2}(2+s)}{1+s}. \tag{4.237}$$

As a final comment, $S_{xx}^+(s)$ is the system function of what is referred to as a minimum-phase system, and for this reason it is typically referred to as the *minimum-phase spectral factor* of $S_{xx}(s)$. Similarly, $S_{xx}^-(s)$ is referred to as the *maximum-phase spectral factor*. One of the attractive features of the minimum-phase whitening filter $H(s) = 1/S_{xx}^+(s)$ is that not only can we construct a white noise process $w(t)$ from $x(t)$ using this causal filter, but we can also subsequently recover $x(t)$ from $w(t)$ by further causal processing with the corresponding inverse $H^{-1}(s) = S_{xx}^+(s)$. Thus, we say that $x(t)$ and $w(t)$ in this case are causally equivalent, in the sense that we can obtain one from the other by causal operations; for any time t the information in $\{x(s), s \le t\}$ is identical to the information in $\{w(s), s \le t\}$.

Once again, we conclude this section with the discrete-time counterparts of our results. In particular, any rational spectrum $S_{xx}(z)$ can be factored into the form

$$S_{xx}(z) = G(z)\,G(1/z) \tag{4.238}$$

for some choice of $G(z)$. Moreover, if we seek causal spectral factors $G(z)$, then the discrete-time version of the Paley-Wiener theorem provides a necessary and sufficient condition for one to exist.

[8] to within a sign

Theorem 4.8 (Paley-Wiener) *The power spectral density $S_{xx}(z)$ of a finite variance wide-sense stationary process $x[n]$ can be expressed in the form*

$$S_{xx}(z) = G(z)\,G(1/z), \tag{4.239}$$

where $G(z)$ is the system function of a causal system, if and only if

$$\int_{-\pi}^{\pi} \left|\ln S_{xx}(e^{j\omega})\right| \, d\omega < \infty. \tag{4.240}$$

Again in the discrete-time case, the Paley-Wiener condition (4.240) is satisfied by all rational power spectra, and we refer to any filter satisfying (4.238) as a shaping or synthesis filter for $S_{xx}(z)$. The inverse of $G(z)$, i.e., $H(z) = 1/G(z)$ is referred to as a whitening filter for $S_{xx}(z)$ since if $x[n]$ is the input to an LTI system with system function $H(z)$, the output $w[n]$ is a white process.

Again, if a spectral factor $G(z)$ exists, it is not unique since (4.238) constrains the magnitude of $G(e^{j\omega})$ but not the phase, since

$$S_{xx}(e^{j\omega}) = |G(e^{j\omega})|^2.$$

In particular,

$$G'(z) = A(z)\,G(z)$$

is also a valid spectral factor whenever $A(z)$ is a stable all-pass system, i.e., whenever

$$A(z)\,A(1/z) = 1.$$

However, in the case of rational spectra, the spectral factor that results from assigning all poles and zeros of $S_{xx}(e^{j\omega})$ that are *inside* the unit circle to $G(z)$ is unique and is given the special notation $S^+_{xx}(z)$ and is referred to as the minimum-phase spectral factor, i.e.,

$$\begin{aligned} &S^+_{xx}(z) \text{ is the spectral factor of } S_{xx}(z) \text{ with all poles and} \\ &\text{zeros inside the unit circle} \end{aligned} \tag{4.241}$$

With $S^-_{xx}(z)$ denoting the corresponding (maximum-phase) spectral factor, which has all the poles and zeros outside the unit circle, we then have

$$S^-_{xx}(z) = S^+_{xx}(1/z) \tag{4.242}$$

and

$$S_{xx}(z) = S^+_{xx}(z)\,S^-_{xx}(z) = S^+_{xx}(z)\,S^+_{xx}(1/z). \tag{4.243}$$

The minimum phase spectral factor $G(z) = S^+_{xx}(e^{j\omega})$ has the special property that it is stable and causal with a stable and causal inverse, which is the whitening filter $H(e^{j\omega}) = 1/S^+_{xx}(e^{j\omega})$. With this choice, if $w[n]$ is the output of the whitening filter $1/G(z)$ when $x[n]$ is input, then $w[n]$ and $x[n]$ are causally equivalent—for any time n, $\{w[n], m \le n\}$ and $\{x[n], m \le n\}$ carry exactly the same information.

We conclude this section with an example.

Example 4.20

Consider again the first-order spectrum (4.190). In this case we can immediately calculate the minimum phase spectral factor

$$S_{xx}^{+}(z) = \frac{\sigma\sqrt{1-\alpha^2}}{(1-\alpha z^{-1})} \tag{4.244}$$

Thus $x[n]$ can be viewed as the output of the first-order system specified by (4.244) driven by unit strength white noise $w[n]$.

4.6.5 Continuous-Time White Noise Revisited

At this point, we can make some final observations about the concept of white noise. We restrict our attention to the case of wide-sense stationary white noise for simplicity.

First, we point out that white noise is a useful idealized model for many phenomena. In particular, a random process can be well-modeled as a white-noise process as long as it has a flat spectrum out to frequencies beyond the effective bandwidth of any system of interest. That is, in any application the physical system of interest "rolls off" at high enough frequencies, i.e., $|H(j\omega)| \approx 0$ for $|w| > 2\pi B$ for some value of the bandwidth B. Then any process $w(t)$ with a spectrum that is flat out to at least this frequency, i.e.,

$$S_{ww}(j\omega) = \sigma^2 \qquad |\omega| < 2\pi B \tag{4.245}$$

acts, in effect, as a white noise process for this system and thus can be replaced by the ideal white noise model. From a time-domain perspective, the process $w(t)$ behaves like white noise because it decorrelates over time intervals that are negligibly small compared to the time constants of the system $H(s)$.

It is also important to realize that even in applications where bandwidths and correlation times are such that the noise processes are not well-modeled as white, we can still use white noise as the basis for our noise model. In particular, we can use spectral factorization concepts to model such noise as the *output* of a linear system driven by white noise process. Thus, more generally, white noise represents an important and extremely useful idealization.

Example 4.21

A natural example illustrating the concepts in this section is *bandlimited white noise*, i.e., a WSS process $w(t)$ with

$$S_{ww}(j\omega) = \begin{cases} \sigma^2 & |\omega| < 2\pi B \\ 0 & \text{otherwise} \end{cases}. \tag{4.246}$$

Thus $w(t)$ acts like an idealized white noise process when applied to systems with frequency responses that are essentially 0 for all $|\omega| > 2\pi B$. The corresponding

covariance function of $w(t)$ is

$$R_{ww}(\tau) = \frac{\sigma^2 \sin 2\pi B\tau}{\pi\tau} = 2B\sigma^2 \operatorname{sinc} 2B\tau, \tag{4.247}$$

which is zero for $t = \pm 1/(2B), \pm 2/(2B), \ldots$. Thus the set of samples, $w(n/(2B))$ separated in time by $1/(2B)$ are uncorrelated. As expected the process decorrelates more quickly as B is increased.

4.7 SPECIAL CLASSES OF STOCHASTIC PROCESSES

At the outset of this chapter, we pointed out that complete characterizations of arbitrary stochastic processes are often unwieldy in many applications. For this reason, the study of stochastic processes is often limited to partial descriptions, such as the second-order characterization we have focussed on thus far, or to classes of stochastic processes having sufficient special structure that their complete characterizations have a particularly compact form. In this section we explore some important classes of stochastic processes in the latter category.

4.7.1 Separable Processes

At the outset of the chapter we expressed a complete characterization of a stochastic process in terms of a specification of all finite-dimensional densities for the process, i.e., the specification of the Nth-order density

$$p_{x(t_1),x(t_2),\ldots,x(t_N)}(x_1, x_2, \ldots, x_N) \tag{4.248}$$

for all values of N and all possible sets of times $t_1, t_2, \ldots, t_N$, and had an analogous definition in discrete-time.

In our development, we pointed out that the characterization in terms of distributions of the form (4.248) was complete for the class of so-called separable processes. In this section, we explore the properties of this class of processes in more detail.

In general a complete characterization of a stochastic process must provide us with the complete probabilistic description of entire sample paths. However, by specifying the finite-dimensional distributions (4.248) we in essence provide a probabilistic description of sample paths only at *discrete* sets of points. As a result, at least in continuous-time, for such finite-dimensional characterizations to be complete there must be in principle relatively strong constraints on the structure of sample paths of the processes. Of course, in practice, physical observations of a point process are typically limited to a discrete set of samples. Consequently, from a practical point of view, it is natural to restrict our attention, as we have done, to stochastic processes for which the finite-dimensional densities are a complete description. Nevertheless, it is worth emphasizing that such assumptions have deep mathematical implications.

We illustrate these ideas with an example.

Example 4.22

Let z be a uniformly distributed random variable taking values $0 \leq z \leq 1$, and consider the following two stochastic processes $x(t)$ and $y(t)$ for $0 \leq t \leq 1$

$$x(t) = 0 \qquad \text{for all } t \tag{4.249}$$

$$y(t) = \begin{cases} 1 & t = z \\ 0 & t \neq z \end{cases} \tag{4.250}$$

so that each sample path of $y(t)$ is zero everywhere except at one random instant, namely $t = z$.

In this case it is relatively straightforward to show that $x(t)$ and $y(t)$ have identical finite-dimensional densities; specifically,

$$\begin{aligned} & p_{x(t_1),x(t_2),\ldots,x(t_N)}(x_1, x_2, \ldots, x_N) \\ &= p_{y(t_1),y(t_2),\ldots,y(t_N)}(x_1, x_2, \ldots, x_N) = \delta(x_1)\delta(x_2)\cdots\delta(x_N) \end{aligned} \tag{4.251}$$

since

$$\begin{aligned} & \Pr\left[y(t_1) = y(t_2) = \cdots = y(t_N) = 0\right] = \\ &= \Pr\left[z \neq t_1, z \neq t_2, \ldots, z \neq t_N\right] = 1. \end{aligned} \tag{4.252}$$

However, if we consider events that involve a *continuum* of values of t, it is clear these processes do not have identical statistics. For example,

$$\Pr\left[x(t) \leq 1/2 \text{ for all } t\right] = 1 \tag{4.253}$$

$$\Pr\left[y(t) \leq 1/2 \text{ for all } t\right] = 0. \tag{4.254}$$

One of the most fundamental results in the mathematical theory of stochastic processes is that a separable stochastic process can be constructed for any set of consistent finite-dimensional densities, where by consistent, we mean that lower-order densities can be obtained consistently from higher-order densities by integration. From this perspective, we see that the process $x(t)$ in Example 4.22 is separable, but the process $y(t)$ is not, since its distinguishing characteristics appear only when we consider a continuum of time points. However, as our example also illustrates, any non-separable process such as $y(t)$ is equivalent in the sense of finite-dimensional densities to a separable process—in this case $x(t)$.

4.7.2 Gaussian Processes

An important subclass of separable stochastic processes is referred to as the class of Gaussian processes. These processes are defined in a manner analogous to the way in which we defined Gaussian random vectors. Specifically, a continuous-time stochastic process $x(t)$ is defined to be a Gaussian process when for all choices of the deterministic function $a(t)$ the scalar

$$z = \int_{-\infty}^{+\infty} a(t)\, x(t)\, dt \tag{4.255}$$

is a Gaussian random variable.[9]

This definition has several implications. For example, consider a function $a(t)$ of the form

$$a(t) = \sum_{i=1}^{N} a_i \delta(t - t_i) \tag{4.256}$$

where the a_i are arbitrary constants, the t_i are an arbitrary set of time instants, and N is any positive integer. Substituting (4.256) with (4.255) we can conclude that for a Gaussian stochastic process $x(t)$, every collection of samples of the form

$$x(t_1), x(t_2), \ldots, x(t_N) \tag{4.257}$$

are a set of jointly Gaussian random variables. Phrased differently, for Gaussian processes, all finite-dimensional distributions are Gaussian. In fact, for regular Gaussian processes, this constitutes an equivalent characterization. Note, however, that our definition also accommodates generalized processes, and in particular allows us to define Gaussian white noise.

The above observations have an important consequence. Specifically, since Gaussian processes are characterized by Gaussian finite-dimensional distributions, this means that a complete specification of a Gaussian process can be constructed from its second-order characterization. For example, the arbitrary collection of samples (4.257) must be distributed according to

$$\begin{bmatrix} x(t_1) \\ x(t_2) \\ \vdots \\ x(t_N) \end{bmatrix} \sim N \left(\begin{bmatrix} m_x(t_1) \\ m_x(t_2) \\ \vdots \\ m_x(t_N) \end{bmatrix}, \begin{bmatrix} K_{xx}(t_1,t_1) & K_{xx}(t_1,t_2) & \cdots & K_{xx}(t_1,t_N) \\ K_{xx}(t_2,t_1) & K_{xx}(t_2,t_2) & \cdots & K_{xx}(t_2,t_N) \\ \vdots & \vdots & \ddots & \vdots \\ K_{xx}(t_N,t_1) & K_{xx}(t_N,t_2) & \cdots & K_{xx}(t_N,t_N) \end{bmatrix} \right) \tag{4.258}$$

Thus knowledge of $m_x(t)$ and $K_{xx}(t,s)$ is sufficient to completely and uniquely characterize a Gaussian process.

In discrete-time, Gaussian processes are defined in an analogous manner. In particular, a discrete-time stochastic process $x[n]$ is defined to be a Gaussian process when for all choices of the deterministic sequence $a[n]$ the scalar

$$z = \sum_{n=-\infty}^{+\infty} a[n]\, x[n] \tag{4.259}$$

is a Gaussian random variable. Using this definition we can derive corresponding results to those we have just developed in the continuous-time case. For example, all finite-dimensional distributions of discrete-time Gaussian processes are Gaussian, and as a result knowledge of $m_x[n]$ and $K_{xx}[n,m]$ is sufficient to completely characterize a discrete-time Gaussian process.

---

[9] More precisely, we must restrict our attention to all choices of $a(t)$ for which $E\left[z^2\right] < \infty$.

Example 4.23

One of the most well studied examples of a Gaussian stochastic process is the Wiener process. This is the Gaussian random walk, i.e., the unique zero-mean Gaussian process whose covariance function is given by (4.47) in Example 4.8. To verify that the white noise $w(t)$ that is the generalized derivative of a Wiener process $x(t)$ is Gaussian we simply note, using integration by parts, that

$$z = \int_{-\infty}^{+\infty} a(t)\, w(t)\, dt = [a(t)x(t)]\Big|_{-\infty}^{+\infty} - \int_{-\infty}^{+\infty} a'(t)\, x(t)\, dt. \tag{4.260}$$

In particular, since $x(t)$ is a Gaussian random process, the right-hand side of (4.260) is a Gaussian random variable for every choice of $a(t)$.

A very special, important and useful property of Gaussian random processes relates to the effects of filtering on these processes. In particular, if a linear system with kernel $h(t,\tau)$ has a Gaussian random process as its input $x(t)$, then the output

$$y(t) = \int_{-\infty}^{+\infty} h(t,\tau)\, x(\tau)\, d\tau \tag{4.261}$$

of this system is also a Gaussian random process. Before we proceed to derive this result, we note that a special case corresponds to linear systems that are also time invariant.

To derive the general result, we simply note using (4.261) that for any choice of $a(t)$,

$$z = \int_{-\infty}^{+\infty} a(t)\, y(t)\, dt = \int_{-\infty}^{+\infty} b(\tau)\, x(\tau)\, d\tau \tag{4.262}$$

where

$$b(\tau) = \int_{-\infty}^{+\infty} a(t)\, h(t,\tau)\, dt. \tag{4.263}$$

Since $x(t)$ is a Gaussian process, the right-hand side of (4.262) is a Gaussian random variable for every choice of $b(\tau)$. Thus $y(t)$ must also be a Gaussian process.

Analogous results also apply to discrete-time Gaussian processes and discrete-time linear systems. Collectively, these results are of course the random process counterparts of results we had for random vectors. In particular, we showed that linear transformations of Gaussian random vectors were also Gaussian random vectors.

An important consequence of the definition of Gaussian processes is that Gaussian processes that are wide-sense stationary are also strict-sense stationary—the two notions of stationarity are equivalent in the Gaussian case. This result follows immediately from the fact that, as we've indicated, a complete characterization for a Gaussian process in terms of its finite-dimensional distributions can be constructed from its second-order characterization. Hence, when this second-order characterization is time-invariant, so are all finite-dimensional distributions.

In particular, an arbitrary collection of samples $x(t_1), x(t_2), \ldots, x(t_N)$ of a Gaussian process has a multivariate Gaussian density of the form (4.258), which when $x(t)$ is WSS becomes

$$\begin{bmatrix} x(t_1) \\ x(t_2) \\ \vdots \\ x(t_N) \end{bmatrix} \sim N\left(\begin{bmatrix} m_x \\ m_x \\ \vdots \\ m_x \end{bmatrix}, \begin{bmatrix} K_{xx}(0) & K_{xx}(\Delta t_2) & \cdots & K_{xx}(\Delta t_N) \\ K_{xx}(\Delta t_2) & K_{xx}(0) & \cdots & K_{xx}(\Delta t_N - \Delta t_2) \\ \vdots & \vdots & \ddots & \vdots \\ K_{xx}(\Delta t_N) & K_{xx}(\Delta t_N - \Delta t_2) & \cdots & K_{xx}(0) \end{bmatrix} \right) \tag{4.264}$$

where $\Delta t_i = t_i - t_1$. Thus, since (4.264) is a function only of $\Delta t_2 = t_2 - t_1, \Delta t_3 = t_3 - t_1, \ldots, \Delta t_N = t_N - t_1$, the process $x(t)$ must be SSS as well. Again, analogous results apply in the case of discrete-time Gaussian processes.

It is important to stress, however, that for non-Gaussian processes, WSS need not imply SSS, as the following example illustrates.

Example 4.24

Consider the non-Gaussian stochastic process $x[n]$ defined for $n \geq 0$ through the following recursion

$$x[n+1] = \frac{1}{2}x[n] + (1 + x[n])\, w[n] \tag{4.265}$$

where $w[n]$ is a zero-mean WSS white Gaussian noise sequence with

$$E\left[w^2[n]\right] = \frac{5}{8} \tag{4.266}$$

and $x[0]$ is a Gaussian random variable with zero mean and variance 5 that is independent of $w[n]$.

We begin by observing that at any instant $n \geq 0$, the sample $x[n]$ is a function of $x[0]$ and $w[0], w[1], \ldots, w[n-1]$, and thus, because $w[n]$ is independent of all of these quantities, we see that $x[n]$ and $w[n]$ are independent.

Taking expectations of both sides of (4.265) we then see that the mean function satisfies

$$m_x[n+1] = \frac{1}{2}m_x[n] + (1 + m_x[n])E\left[w[n]\right] = \frac{1}{2}m_x[n]. \tag{4.267}$$

Since $m_x[0] = 0$, (4.267) implies that $x[n]$ is zero-mean for all n, i.e.,

$$m_x[n] \equiv 0. \tag{4.268}$$

Turn our attention next to the variance of the process

$$\lambda_x[n] = \operatorname{var} x[n] = E\left[x^2[n]\right] \tag{4.269}$$

we have, squaring both sides of (4.265) and taking expectations,

$$\begin{aligned} \lambda_x[n+1] &= E\left[x^2[n+1]\right] = \frac{1}{4}E\left[x^2[n]\right] + E\left[x[n]\,(1 + x[n])\, w[n]\right] \\ &\quad + E\left[(1 + x[n])^2\, w^2[n]\right]. \end{aligned} \tag{4.270}$$

Combining (4.270) with (4.269) and exploiting the independence of $x[n]$ and $w[n]$ together with the fact that both $x[n]$ and $w[n]$ are zero mean, we obtain

$$\begin{aligned}\lambda_x[n+1] &= \frac{1}{4}\lambda_x[n] + 0 + (1+\lambda_x[n])E\left[w^2[n]\right] \\ &= \frac{7}{8}\lambda_x[n] + \frac{5}{8} \end{aligned} \tag{4.271}$$

where we have also used (4.266). Finally, since $\lambda_x[0] = 5$, we can conclude from (4.271) that $\lambda_x[n] = 5$ for all n, i.e.,

$$\lambda_x[n] \equiv 5. \tag{4.272}$$

Let us next compute the covariance function $K_{xx}[n,m]$ for $x[n]$ with $n \geq m$. Solving (4.265) from the initial time m to the later time n, we obtain

$$x[n] = \left(\frac{1}{2}\right)^{n-m} x[m] + \sum_{k=m}^{n-1} \left(\frac{1}{2}\right)^{n-k-1} (1+x[k])\, w[k]. \tag{4.273}$$

Multiplying both sides of (4.273) by $x[m]$ and taking expectations we obtain

$$K_{xx}[n,m] = E\left[x[n]x[m]\right] = \left(\frac{1}{2}\right)^{n-m} E\left[x^2[m]\right] = 5\left(\frac{1}{2}\right)^{n-m} \tag{4.274}$$

where we have used the fact that $x[m]$ is independent of $w[k]$ at all future times, i.e., $k \geq m$. A similar approach yields the corresponding result for $m \geq n$, and as a result we have, for any m and n

$$K_{xx}[n,m] = 5\left(\frac{1}{2}\right)^{|n-m|}. \tag{4.275}$$

Thus $x[n]$ is WSS.

Hence, (4.268), (4.272), and (4.275) we conclude that $x[n]$ is WSS. However, $x[n]$ is in fact not even first-order stationary. In particular, the third moment

$$\xi_x[n] \triangleq E\left[x^3[n]\right]$$

grows without bound. To see this, if we take the third power of both sides of (4.265) and then take expectations, we obtain

$$\begin{aligned}\xi_x[n+1] &= \left(\frac{1}{2}\right)^3 \xi_x[n] + 3\left(\frac{1}{2}\right)^2 E[x^2[n](1+x[n])\, w[n]] \\ &+ 3\left(\frac{1}{2}\right) E\left[x[n]\,(1+x[n])^2\, w^2[n]\right] \\ &+ E\left[(1+x[n])^3\, w^3[n]\right]. \end{aligned} \tag{4.276}$$

Since $x[n]$ and $w[n]$ are independent, the expectations in (4.276) can be simplified. Furthermore, since $w[n]$ is zero-mean and Gaussian, its odd-order moments vanish. Thus the second and fourth terms on the right-hand side of (4.276) vanish, leaving

$$\begin{aligned}\xi_x[n+1] &= \frac{1}{8}\xi_x[n] + \frac{3}{2}E\left[x[n] + 2x^2[n] + x^3[n]\right] E\left[w^2[n]\right] \\ &= \frac{1}{8}\xi_x[n] + \frac{15}{16}\{2\lambda_x[n] + \xi_x[n]\} \\ &= \frac{17}{16}\xi_x[n] + \frac{75}{8} \end{aligned} \tag{4.277}$$

where we have used (4.266) and the fact that $\lambda_x[n] = 5$. Since $x[0]$ is zero-mean Gaussian, $\xi_x[0] = 0$. However, from (4.276) we see that $\xi_x[n]$ grows without bound as $n \to \infty$.

Jointly Gaussian Processes

Let us now consider some important generalizations of the preceding definitions and results. In particular, we now develop the notion of jointly Gaussian random processes. Two continuous-time random processes $x(t)$ and $y(t)$ are defined to be jointly Gaussian if for all choices of $a(t)$ and $b(t)$ the scalar

$$z = \int_{-\infty}^{+\infty} [a(t)\, x(t) + b(t)\, y(t)]\, dt \tag{4.278}$$

is a Gaussian random variable.

Several consequences of this definition are straightforward to verify using the results of this section. For example, if $x(t)$ and $y(t)$ are jointly Gaussian processes, then they are individually Gaussian processes. Moreover, every collection of samples of the form

$$x(t_1), x(t_2), \ldots, x(t_N), y(s_1), y(s_2), \ldots, y(s_M)$$

are a set of jointly Gaussian random variables, i.e., all joint finite-dimensional distributions are Gaussian. In turn, this implies that jointly Gaussian random processes are completely characterized by their joint second-order characterization—specifically their means $m_x(t)$ and $m_y(t)$, their autocovariance functions $K_{xx}(t, s)$ and $K_{yy}(t, s)$, and their cross-covariance function $K_{xy}(t, s)$.

We also point out that the above results may be generalized to multiple random processes. For example, we say a vector random process

$$\mathbf{x}(t) = \begin{bmatrix} x_1(t) & x_2(t) & \cdots & x_N(t) \end{bmatrix}^{\mathrm{T}}$$

is a Gaussian process if its constituent processes $x_1(t), x_2(t), \ldots, x_N(t)$ are jointly Gaussian processes, i.e., if for every choice of vector function

$$\mathbf{a}(t) = \begin{bmatrix} a_1(t) & a_2(t) & \cdots & a_N(t) \end{bmatrix}^{\mathrm{T}}$$

the random variable

$$z = \int_{-\infty}^{+\infty} \mathbf{a}^{\mathrm{T}}(t)\, \mathbf{x}(t)\, dt$$

is Gaussian. Again, analogous definitions and results apply in the case of discrete-time stochastic processes.

We also have an important stationarity result for jointly Gaussian processes in both continuous-time and discrete-time. Specifically, if a set of jointly Gaussian processes is jointly WSS, then it is also jointly SSS. Again, for jointly Gaussian processes, the two notions of stationarity are equivalent. This result can be verified

in a manner similar to the corresponding result in the previous section. Again we stress that the converse is not true—jointly WSS processes that are not jointly Gaussian are in general not SSS. In particular, it is possible to construct a pair of random processes $x(t)$ and $y(t)$ that are jointly WSS and individually Gaussian, but not jointly Gaussian and not jointly SSS.

4.7.3 Independent Increments Processes

An important subset of uncorrelated increments processes are those having *independent* increments. In particular, $x(t)$ is an independent increments process if for any t and s such that $t > s$, the increment $x(t) - x(s)$ is independent of the past, i.e., of $\{x(\tau), \text{for all } \tau \leq s\}$.

Discrete-time examples of independent increments processes are the Bernoulli counting process of Example 4.9 and the discrete-time Wiener process. Continuous-time examples of independent increments processes are the Poisson counting process of Example 4.10 and the continuous-time Wiener process.

Note that except in degenerate cases, independent increment processes are *not* SSS, nor are they even first-order stationary. Indeed, like all uncorrelated increments processes they have a variance that grows linearly with time.

Although independent increment processes are not stationary, these processes can have stationary *increments*, which is a very different property. In particular, $x(t)$ is said to be a process with *stationary independent increments* if it is an independent increment process and if for any $t > s$ the probability density of the increment $x(t) - x(s)$ depends only on the increment length $t - s$, and not on the absolute location of the increment along the time axis. An important property of processes with stationary independent increments is that their generalized derivatives are SSS white noises.

Examples of processes with stationary independent increments are the Poisson process of Example 4.4 (cf. (4.14)), as well as the discrete-time and continuous-time Wiener processes, i.e., the Gaussian processes with second-order characterizations that are the random walks of Examples 4.7 and 4.8, respectively.

Processes with stationary independent increments must have variances that grow linearly with time. In fact this is more generally a property of processes with WSS and uncorrelated increments. To verify this result, we begin by defining

$$q(t) = \text{var}\,[x(s+t) - x(s)] \tag{4.279}$$

as the variance of the increment, so that $q(0) = 0$. Then for any Δ we have

$$\begin{aligned} q(t+\Delta) &= \text{var}\,[x(s+t+\Delta) - x(s)] \\ &= \text{var}\,[[x(s+t+\Delta) - x(s+\Delta)] + [x(s+\Delta) - x(s)]] \\ &= \text{var}\,[x(s+t+\Delta) - x(s+\Delta)] + \text{var}\,[x(s+\Delta) - x(s)] \\ &= q(t) + q(\Delta). \end{aligned} \tag{4.280}$$

In turn, (4.280) implies that

$$\frac{dq(t)}{dt} = \lim_{\Delta \to 0} \frac{q(t+\Delta) - q(t)}{\Delta} = \lim_{\Delta \to 0} \frac{q(\Delta) - q(0)}{\Delta} = \left. \frac{dq(t)}{dt} \right|_{t=0} \triangleq a, \tag{4.281}$$

i.e., $dq(t)/dt$ is a constant, from which we conclude that

$$\operatorname{var} x(t) = \operatorname{var} [[x(t) - x(0)] + x(0)] = q(t) + \operatorname{var} x(0) = at + b. \tag{4.282}$$

As a final comment, by modifying the construction of discrete-time Brownian motion we can readily construct independent increment processes that do not have stationary increments. The resulting processes are referred to as *inhomogeneous* Wiener processes or Brownian motions. In particular, if we let $w[n]$ be a sequence of independent, zero-mean Gaussian random variables with time-varying variances of the form

$$E\left[w^2[n]\right] = \sigma_n^2, \tag{4.283}$$

then the running sum of the $w[n]$'s, i.e.,

$$x[n] = \sum_{k=1}^{n} w[k] \tag{4.284}$$

is an independent increment process. Furthermore, since for $n > m$ we have

$$E\left[(x[n] - x[m])^2\right] = \sum_{k=m+1}^{n} \sigma_k^2 \tag{4.285}$$

which is not a function of $n - m$ alone, we can conclude that the increments are indeed not stationary. Another class of processes with nonstationary independent increments are the inhomogeneous Poisson processes, which we explore in detail in Section 4.7.5.

4.7.4 Markov Processes

Another important class of stochastic processes with special structure is the class of *Markov processes*. A Markov process $x(t)$ has the defining characteristic that the current value of the process captures all of the memory in the process so that there is no additional information in the past that can be used to predict the future. More precisely $x(t)$ is Markov if for any set of ordered times $t_1 < t_2 < \cdots < t_N$,

$$p_{x(t_N)|x(t_1),x(t_2),\ldots,x(t_{N-1})}(x_N|x_1, x_2, \ldots, x_{N-1}) = p_{x(t_N)|x(t_{N-1})}(x_N|x_{N-1}). \tag{4.286}$$

Given the values of $x(t)$ at a number of points in the past, $t_1, t_2, \ldots, t_{N-1}$, we need keep only the most recent value $x(t_{N-1})$, as the earlier values of the process provide no additional information useful in predicting the future values $x(t_N)$.

Markov processes have the attractive feature that their complete characteristizations are especially compact. The key quantity that arises in the probabilistic description of these processes is the *transition density*, i.e., the conditional density

$$p_{x(t)|x(s)}(x|z) \tag{4.287}$$

for $s < t$. The transition density describes how probability densities transition from one time instant to another, i.e.,

$$\begin{aligned} p_{x(t)}(x) &= \int_{-\infty}^{+\infty} p_{x(t),x(s)}(x,z)\,dz \\ &= \int_{-\infty}^{+\infty} p_{x(t)|x(s)}(x|z)\,p_{x(s)}(z)\,dz. \end{aligned} \tag{4.288}$$

To see that the transition density, in effect, fully specifies a Markov process, we simply note that if we order the times $t_1 < t_2 < \cdots < t_N$ in any Nth-order density, then

$$\begin{aligned} &p_{x(t_1),x(t_2),\ldots,x(t_N)}(x_1,x_2,\ldots,x_N) \\ &\quad = p_{x(t_1)}(x_1)p_{x(t_2)|x(t_1)}(x_2|x_1)\cdots p_{x(t_N)|x(t_1),x(t_2),\ldots,x(t_{N-1})}(x_N|x_1,x_2,\ldots,x_{N-1}) \\ &\quad = p_{x(t_1)}(x_1)\,p_{x(t_2)|x(t_1)}(x_2|x_1)\cdots p_{x(t_N)|x(t_{N-1})}(x_N|x_{N-1}), \end{aligned} \tag{4.289}$$

where the second equality follows from the Markov property (4.286). Hence, more precisely we see that a complete specification of a Markov process consists of the specification of the first-order density $p_{x(t)}(x)$ together with the transition density (4.287).

In the case of Markov processes, the consistency condition for finite-dimensional distributions—i.e., that lower-order distributions can be obtained from higher order distributions—is equivalent to the so-called *Chapman-Kolmogorov* condition. For three arbitrary ordered time instants $s < t < u$, the Chapman-Kolmogorov condition takes the form

$$\begin{aligned} p_{x(u)|x(s)}(x|z) &= \int_{-\infty}^{+\infty} p_{x(u),x(t)|x(s)}(x,y|z)\,dy \\ &= \int_{-\infty}^{+\infty} p_{x(u)|x(t)}(x|y)\,p_{x(t)|x(s)}(y|z)\,dy, \end{aligned} \tag{4.290}$$

where we have used the Markov property to obtain the second equality. It is worth mentioning that while we have stated our results for continuous-time Markov processes, analogous definitions and results hold for discrete-time Markov processes.

There are many important examples of Markov processes. In fact, every independent increment process is a Markov process. To see this, let $x(t)$ be an independent increment process and take any set of times $t_1 < t_2 < \cdots < t_N$. Then we can write $x(t_N)$ as

$$x(t_N) = x(t_{N-1}) + [x(t_N) - x(t_{N-1})]. \tag{4.291}$$

In turn, if we are given $x(t_{N-1})$, the remaining uncertainty in $x(t_N)$ is obviously that in the increment $[x(t_N) - x(t_{N-1})]$. However, this increment is *independent* of $x(t_1), x(t_2), \ldots, x(t_{N-1})$. Consequently given $x(t_{N-1})$, there is no additional useful information concerning $x(t_N)$ in the values of $x(t_1), x(t_2), \ldots, x(t_{N-2})$. Hence $x(t)$ is Markov.

While every independent increments process is Markov, there are many Markov processes that do not have independent increments.

Example 4.25

Consider the process $y[n]$ introduced in Example 4.2, which satisfies the first-order difference equation

$$y[n] = \frac{1}{2}y[n-1] + x[n]. \tag{4.292}$$

Suppose that $x[n]$ is a zero-mean white Gaussian noise process, and suppose that the initial condition $y[0]$ is Gaussian and independent of $x[n]$. To see that this process is Markov, note that via the recursion (4.292), $y[n]$ is a function of $y[0]$ and $x[k]$ for $k \leq n$. Furthermore, for any $n > m$ we can write $y[n]$ as a function of $y[m]$ and the subsequent input values $x[k]$ for $m+1 \leq k \leq n$:

$$y[n] = \left(\frac{1}{2}\right)^{n-m} y[m] + \sum_{k=m+1}^{n} \left(\frac{1}{2}\right)^{n-k} x[k]. \tag{4.293}$$

Thus if we are given $y[m]$, the remaining uncertainty in $y[n]$ is due to subsequent values of the input noise, which are independent of the entire past. Hence, there is no additional useful information in values of $y[k]$ for $k < m$, and thus $y[n]$ is Markov. On the other hand, to see that it does not have independent increments, it suffices to note from (4.292) that we have the following expression for one-step increments:

$$y[n] - y[n-1] = -\frac{1}{2}y[n-1] + x[n] \tag{4.294}$$

which is obviously not independent of $y[n-1]$.

Example 4.26

Another example of a Markov process that does not have independent increments is the random telegraph wave introduced in Example 4.5. Recall that $x(t)$ switches between values of ± 1 where the switches occur at points in time when a Poisson process $N(t)$ changes value and where the initial value $x(0)$ is independent of $N(t)$ and takes on the values ± 1 with equal probability. That is,

$$x(t) = x(0)(-1)^{N(t)}. \tag{4.295}$$

If we take any values of s and t, with $s < t$, then from (4.295) we see that

$$x(t) = x(s)(-1)^{[N(t)-N(s)]}. \tag{4.296}$$

If we are then given the value of $x(s)$, we see from (4.296) that the remaining uncertainty in the value of $x(t)$ is determined by $N(t) - N(s)$. However, since $N(t)$ is an independent increments process independent of $x(0)$, we see that $N(t) - N(s)$ is independent of the entire past through time s. Consequently if we are given $x(s)$, there is no additional useful information in the values of $x(\tau)$ for $\tau \leq s$ concerning the values of $x(t)$. Hence $x(t)$ is Markov. On the other hand, $x(t)$ does *not* have independent increments, since the value of $x(t) - x(s)$ is certainly *not* independent of $x(s)$. In particular, if $x(s) = +1$, then (since $x(t)$ can only take on the values of ± 1) we know that $x(t) - x(s) \leq 0$. Similarly if $x(s) = -1$, we know that $x(t) - x(s) \geq 0$.

4.7.5 Inhomogeneous Poisson Counting Processes

In this section we take a more detailed look at one particular class of processes that is both of great practical importance and also provides an excellent example of a process in which it is far more useful to look at sample paths directly. Specifically, we consider a generalization of the process introduced in Example 4.4.

In particular an *inhomogeneous* Poisson Counting Process $\mathsf{N}(t)$ with rate $\lambda(t)$ is defined by the following properties:

- $\mathsf{N}(0) = 0$
- $\mathsf{N}(t)$ is an independent increments process
- For $t > s$, $\mathsf{N}(t) - \mathsf{N}(s)$ is a Poisson random variable with mean $\int_s^t \lambda(\tau)\, d\tau$ so that

$$\Pr\left[\mathsf{N}(t) - \mathsf{N}(s) = k\right] = \frac{[\int_s^t \lambda(\tau)\, d\tau]^k e^{-\int_s^t \lambda(\tau)\, d\tau}}{k!}. \tag{4.297}$$

Note that if $\lambda(t) = \lambda$ is constant, then this model reduces to that of a homogeneous Poisson counting process introduced in Example 4.4. Also, it is important to emphasize that there is actually something that must be checked to make sure that the specification we have just given makes sense. Specifically, we must make sure that the increment distribution in (4.297) is consistent (which corresponds to checking the Chapman-Kolmogorov equation). Specifically, if $s < t < u$, we must have that

$$\Pr\left[\mathsf{N}(u) - \mathsf{N}(s) = k\right] = \sum_{j=0}^{k} \Pr\left[\mathsf{N}(t) - \mathsf{N}(s) = j\right] \Pr\left[\mathsf{N}(u) - \mathsf{N}(t) = k - j\right], \tag{4.298}$$

a calculation we leave to the reader.

Let us make several preliminary calculations for $\mathsf{N}(t)$. Note first that by setting $s = 0$ in (4.297) and using the fact that $\mathsf{N}(0) = 0$, we see that

$$\Pr\left[\mathsf{N}(t) = k\right] = \frac{[\int_0^t \lambda(\tau)\, d\tau]^k e^{-\int_0^t \lambda(\tau)\, d\tau}}{k!}. \tag{4.299}$$

Also for Δ small, from (4.297) we see that

$$\begin{aligned}\Pr\left[\mathsf{N}(t+\Delta) - \mathsf{N}(t) = k\right] &\approx \frac{\lambda(t)^k \Delta^k}{k!} e^{-\lambda(t)\Delta} \\ &= \frac{\lambda(t)^k \Delta^k}{k!}[1 - \lambda(t)\Delta + \frac{\lambda^2(t)\Delta^2}{2} - \cdots],\end{aligned} \tag{4.300}$$

i.e.,

$$\Pr\left[\mathsf{N}(t+\Delta) - \mathsf{N}(t) = k\right] = \begin{cases} 1 - \lambda(t)\Delta + o(\Delta), & k = 0 \\ \lambda(t)\Delta + o(\Delta), & k = 1 \\ o(\Delta), & k > 1 \end{cases} \tag{4.301}$$

where

$$\lim_{\Delta \to 0} \frac{o(\Delta)}{\Delta} = 0. \tag{4.302}$$

If we then define $m_{\mathsf{N}}(t) = E[\mathsf{N}(t)]$, we see that

$$\begin{aligned} m_{\mathsf{N}}(t+\Delta) &= m_{\mathsf{N}}(t) + E[\mathsf{N}(t+\Delta) - \mathsf{N}(t)] \\ &= m_{\mathsf{N}}(t) + 0[1 - \lambda(t)\Delta + o(\Delta)] + 1 \cdot \lambda(t)\Delta + o(\Delta) \end{aligned} \tag{4.303}$$

so that

$$\dot{m}_{\mathsf{N}}(t) = \lim_{\Delta \to 0} \frac{m_{\mathsf{N}}(t+\Delta) - m_{\mathsf{N}}(t)}{\Delta} = \lambda(t). \tag{4.304}$$

Since $m_{\mathsf{N}}(0) = 0$, we see that indeed

$$m_{\mathsf{N}}(t) = \int_0^t \lambda(\tau)\, d\tau. \tag{4.305}$$

Also, since $\mathsf{N}(t)$ is an independent increments process, we know from (4.44) that

$$K_{\mathsf{NN}}(t,\tau) = \operatorname{var}[\mathsf{N}(\min(t,s))]. \tag{4.306}$$

To find $\operatorname{var} \mathsf{N}(t)$ we proceed in an analogous fashion to our calculation of $m_{\mathsf{N}}(t)$. Specifically, we can write $\mathsf{N}(t+\Delta)$ as a sum of two independent random variables

$$\mathsf{N}(t+\Delta) = \mathsf{N}(t) + [\mathsf{N}(t+\Delta) - \mathsf{N}(t)] \tag{4.307}$$

so that

$$\begin{aligned} \operatorname{var} \mathsf{N}(t+\Delta) &= \operatorname{var} \mathsf{N}(t) + \operatorname{var}[\mathsf{N}(t+\Delta) - \mathsf{N}(t)] \\ &= \operatorname{var} \mathsf{N}(t) + E\left[(\mathsf{N}(t+\Delta) - \mathsf{N}(t))^2\right] \\ &\quad - (E[\mathsf{N}(t+\Delta) - \mathsf{N}(t)])^2 \\ &= \operatorname{var} \mathsf{N}(t) + 0 \cdot [1 - \lambda(t)\Delta + o(\Delta)] + 1 \cdot \lambda(t)\Delta + o(\Delta) \\ &\quad - (\lambda(t)\Delta)^2 + o^2(\Delta). \end{aligned} \tag{4.308}$$

Thus

$$\frac{d \operatorname{var} \mathsf{N}(t)}{dt} = \lim_{\Delta \to 0} \frac{\operatorname{var} \mathsf{N}(t+\Delta) - \operatorname{var} \mathsf{N}(t)}{\Delta} = \lambda(t) \tag{4.309}$$

and since $\operatorname{var} \mathsf{N}(0) = 0$, we see that

$$\operatorname{var} \mathsf{N}(t) = \int_0^t \lambda(\tau)\, d\tau. \tag{4.310}$$

While the preceding calculations provide us with the second-order statistics of $\mathsf{N}(t)$, as we have discussed, these statistics provide us with surprisingly little information about the sample paths of $\mathsf{N}(t)$. On the other hand, the sample paths of $\mathsf{N}(t)$ are very simply described in terms of a *finite* number of random variables, namely the number of jumps $\mathsf{N}(t)$ experiences and when they occur. Specifically,

let $W_i, i = 1, 2, \ldots$ denote the ith *occurrence time* or *arrival time*—i.e., W_i is the time that $N(t)$ changes value from $i-1$ to i, so that

$$N(t) = \begin{cases} 0 & 0 \le t < W_1 \\ 1 & W_1 < t < W_2 \\ 2 & W_2 < t < W_3 \\ \vdots & \end{cases}. \tag{4.311}$$

Let us first determine the joint probability density for the first k occurrence times $W_1, W_2, \ldots, W_k$. Specifically, note that from the definition of probability densitites,

$$\begin{aligned} &\Pr\left[w_1 < W_1 < w_1 + \Delta, w_2 < W_2 < w_2 + \Delta, \ldots, w_k < W_k < w_k + \Delta\right] \\ &= \; p_{W_1, W_2, \ldots, W_k}(w_1, w_2, \ldots, w_k)\Delta^k + o(\Delta^k). \end{aligned} \tag{4.312}$$

On the other hand, the left hand side of (4.312) can be rewritten in terms of events involving $N(t)$. Specifically, the left hand side of (4.312) is the product of

$$\begin{aligned} \Pr\left[(N(w_1) = 0\right] &= e^{-\int_0^{w_1} \lambda(\tau)\, d\tau} \\ \Pr\left[N(w_1 + \Delta) - N(w_1) = 1\right] &= \lambda(w_1)\Delta + o(\Delta) \\ \Pr\left[N(w_2) - N(w_1 + \Delta) = 0\right] &= e^{-\int_{w_1+\Delta}^{w_2} \lambda(\tau)\, d\tau} \\ \Pr\left[N(w_2 + \Delta) - N(w_2) = 1\right] &= \lambda(w_2)\Delta + o(\Delta) \\ &\vdots \\ \Pr\left[N(w_k + \Delta) - N(w_k) = 1\right] &= \lambda(w_k)\Delta + o(\Delta). \end{aligned}$$

That is, matching (4.312) and (4.313) we see that

$$p_{W_1, W_2, \ldots, W_k}(w_1, w_2, \ldots, w_k) = \begin{cases} \left[\prod_{i=1}^k \lambda(w_i)\right] e^{-\int_0^{w_k} \lambda(\tau)\, d\tau} & 0 < w_1 < w_2 < \cdots < w_k \\ 0 & \text{otherwise} \end{cases}. \tag{4.313}$$

Note that if $\lambda(t) = \lambda$, a constant, then

$$p_{W_1, W_2, \ldots, W_k}(w_1, w_2, \ldots, w_k) = \begin{cases} \lambda^k e^{-\lambda w_k} & 0 < w_1 < w_2 < \cdots < w_k \\ 0 & \text{otherwise} \end{cases}. \tag{4.314}$$

Note that these densities *are* functions of $w_1, w_2, \ldots, w_k$ since a nonzero value for the density requires that $0 < w_1 < w_2 < \cdots < w_k$. Also, by integrating out $w_1, w_2, \ldots, w_{k-1}$ we can obtain the marginal density for w_k. In particular for the general inhomogeneous process

$$p_{W_k}(w_k) = \begin{cases} \left(\int_0^{w_k} \lambda(\tau)\, d\tau\right)^{k-1} \lambda(w_k) e^{-\int_0^{w_k} \lambda(\tau)\, d\tau}/(k-1)! & w_k \ge 0 \\ 0 & \text{otherwise} \end{cases} \tag{4.315}$$

and for the homogenous case when $\lambda(t) = \lambda$,

$$p_{W_k}(w_k) = \begin{cases} \frac{\lambda^k w_k^{k-1} e^{-\lambda w_k}}{(k-1)!} & w_k \ge 0 \\ 0 & \text{otherwise} \end{cases}. \tag{4.316}$$

Now let us turn to the complete probabilistic description of a sample path of $N(t)$ over the time interval $0 \leq t \leq T$. In particular, as we have stated what we need to specify in order to completely specify this sample path is the number of jumps of $N(t)$, over the interval, i.e., the value of $N(T)$ and the times at which these jumps occur. That is for each $k \geq 0$ we need to specify

$$\begin{aligned} & p_{W_1,W_2,\ldots,W_k,N(T)}(w_1, w_2, \ldots, w_k, k) \\ &= p_{W_1,W_2,\ldots,W_k}(w_1, w_2, \ldots, w_k) \Pr[N(T) = k \mid w_1, w_2, \ldots, w_k] \\ &= p_{W_1,W_2,\ldots,W_k}(w_1, w_2, \ldots, w_k) \Pr[N(T) - N(w_k) = 0] \\ &= \begin{cases} \left[\prod_{i=1}^k \lambda(w_i)\right] e^{-\int_0^T \lambda(\tau)\, d\tau} & 0 < w_1 < w_2 < \cdots < w_k < T \\ 0 & \text{otherwise} \end{cases}. \end{aligned} \tag{4.317}$$

where we have used the fact that $N(t)$ is an independent increment process in the second equality and (4.313) in the third equality. Note also that in the homogeneous case, when $\lambda(t) = \lambda$, then

$$p_{W_1,W_2,\ldots,W_k,N(T)}(w_1, w_2, \ldots, w_k, k) = \begin{cases} \lambda^k e^{-\lambda T} & 0 < w_1 < \cdots < w_k < T \\ 0 & \text{otherwise} \end{cases}. \tag{4.318}$$

Eqs. (4.317) and (4.318) provide us with a complete probabilistic descriptions of sample paths of $N(t)$ in the inhomogeneous and homogeneous cases. In addition, with a bit more work we can obtain additional insight into the nature of Poisson processes. Specifically, a useful way in which to think about generating a sample path of $N(t)$ over the interval $0 \leq t \leq T$ is first to generate $N(T)$, the number of jumps in the interval and then to generate the times of the jumps $W_1, W_2, \ldots$ conditioned on the number of jumps, $N(T)$, that occur. From (4.299) we see that generating $N(T)$ corresponds to generating a Poisson random variable with mean $\int_0^T \lambda(\tau)\, d\tau$. To generate $W_1, W_2, \ldots, W_k$ conditioned on knowledge of $N(T)$ we need the following density:

$$\begin{aligned} & p_{W_1,W_2,\ldots,W_k|N(T)}(w_1, w_2, \ldots, w_k|k) \\ &= \frac{p_{W_1,W_2,\ldots,W_k,N(T)}(w_1, w_2, \ldots, w_k, k)}{\Pr[N(T) = k]} \\ &= \begin{cases} k! \prod_{i=1}^k \lambda(w_i) / \left[\int_0^T \lambda(\tau)\, d\tau\right]^k & 0 \leq w_1 < w_2 \cdots \leq w_k \leq T \\ 0 & \text{otherwise} \end{cases} \end{aligned} \tag{4.319}$$

where we have used (4.299) and (4.317) to obtain the last line of this equation.

An explanation of (4.319) tells us that generating $W_1, W_2, \ldots, W_k$ given $N(T) = k$ is temptingly close to being a simple task. In particular, the value of this density over the range where it is nonzero appears to be simply a product of functions of the individual w_i—i.e., the W_i almost look independent. Of course they are far from independent, since for example, if we know that $W_1 = w_1$, then we also know for certain that $W_2 > w_1$. That is, whether the value of the density of (4.319) is 0

or takes on the nonzero product form in (4.319) depends on whether or not the values $w_1, w_2, \ldots, w_k$ are *ordered* or not. This is made even simpler to see if we look at the homogeneous case, when $\lambda(t) = \lambda$, in which case (4.319) becomes

$$p_{W_1,W_2,\ldots,W_k|N(T)}(w_1, w_2, \ldots, w_k|k) = \begin{cases} k!/T^k & 0 \le w_1 < w_2 < \cdots < w_k \le T \\ 0 & \text{otherwise} \end{cases}. \quad (4.320)$$

That is, the joint probability density function for $W_1, W_2, \ldots, W_k$ given that $N(T) = k$ takes on a *constant* value, namely $(k!/T^k)$, as long as the w_i are ordered and a different constant value, namely 0, otherwise.

The fact that the densities in (4.319) and (4.320) are temptingly close to having very simple forms except for the nuisance of requiring that the W_i be ordered suggests what may at first appear to be a curious step but in fact leads to a dramatic simplification of the description of a Poisson sample path. Specifically, the idea is to *randomly permute* the ordered occurrence times so that order is no longer a constraint.

To be precise, suppose that $N(T) = k$ and consider the set of all possible permutations of the numbers 1 through k. Note that there are exactly $k!$ such permutation. The *unordered occurrence* times $U_1, U_2, \ldots, U_k$ are then obtained by taking the ordered occurrence time $W_1, W_2, \ldots, W_k$ and applying a randomly chosen permutation to them—i.e., we choose one of the $k!$ possible permutations with equal probability $(= 1/k!)$ and apply it to $W_1, W_2, \ldots, W_k$. Thus there is a random permutation $i_1, i_2, \ldots, i_k$ of the numbers $1, 2, \ldots, k$ such that

$$\begin{aligned} W_1 &= U_{i_1} \\ W_2 &= U_{i_2} \\ &\vdots \\ W_k &= U_{i_k}. \end{aligned} \quad (4.321)$$

Let us now look at the density for $U_1, U_2, \ldots, U_k$ given that $N(T) = k$. First of all, note that the U_i are not ordered. Consequently the only constraint on each U_i is the same, namely $0 \le U_i \le T$. Also note that given hypothesized values for $U_1, U_2, \ldots, U_k$, we obviously *can* order them – i.e., we can undo the random permutation as in (4.321) to determine the corresponding values of the *ordered* occurrence times. Given all of this, we can then compute

$$\begin{aligned} & p_{U_1,U_2,\ldots,U_k|N(T)}(u_1, u_2, \ldots, u_k|k) \\ &= p_{W_1,W_2,\ldots,W_k|N(T)}(u_{i_1}, u_{i_2}, \ldots, u_{i_k}|k) \\ &\quad \times \Pr[\text{this particular permutation chosen}] \\ &= \frac{1}{k!} p_{W_1,W_2,\ldots,W_k|N(T)}(u_{i_1}, u_{i_2}, \ldots, u_{i_k}|k) \\ &= \begin{cases} \prod_{i=1}^k \lambda(u_i) / \left(\int_0^T \lambda(\tau)\, d\tau\right)^k & 0 \le u_i \le T \\ 0 & \text{otherwise} \end{cases} \end{aligned} \quad (4.322)$$

where in the last equality we have used (4.319) plus the fact that $u_{i_1}, u_{i_2}, \ldots, u_{i_k}$ have been ordered.

Note that all we have done is to take the probability density for the ordered occurrence times and split it evenly among all of the possible permutations of these times. Thus for example if λ is constant and $N(2) = 2$, $p_{u_1,u_2|N(2)}(.7, 1.3|2)$ and $p_{u_1,u_2|N(2)}(1.3, .7|2)$ have the same value, namely $1/4$, which is $1/2$ of the value of the ordered occurrence time $p_{w_1,w_2|N(2)}(.7, 1.3|2) = 1/2$. Similarly if $N(2) = 3$, $p_{u_1,u_2,u_3|N(2)}(u_1, u_2, u_3|3)$ is $1/6$ times the value of the density for the ordered occurrence times since there are 6 possible permutations among which to split the probability.

The most important implication of the preceding analysis can be seen from (4.322). Specifically, given that $N(t) = k$, then the unordered occurrence times $u_1, u_2, \ldots, u_k$ are independent and identically distributed with common density

$$p_{u|N(T)}(u|k) = \begin{cases} \lambda(u) / \int_0^T \lambda(\tau)\, d\tau & 0 \le u \le T \\ 0 & \text{otherwise} \end{cases}. \tag{4.323}$$

Thus to generate a sample path of a Poisson process $N(t)$ for $0 \le t \le T$,

1. Generate a Poisson random variable $N(T)$ with mean $\int_0^T \lambda(\tau)\, d\tau$.
2. Given that $N(T) = k$, we generate k independent, identically distributed random variables $u_1, u_2, \ldots, u_k$ with common density given by (4.323).
3. Order $u_1, u_2, \ldots, u_k$, yielding the ordered occurrence times so that $N(t)$ is completely defined.

Note that, as is reasonable, (4.323) states that jumps are more likely to occur at times at which the rate $\lambda(t)$ is large. If λ is constant, however, (4.323) reduces to

$$p_{u|N(T)}(u|k) = \begin{cases} 1/T & 0 \le u \le T \\ 0 & \text{otherwise} \end{cases} \tag{4.324}$$

so that the unordered occurrence times are *uniformly distributed* over $0 \le t \le T$.

There is one last concept that we wish to briefly introduce, namely the concept of *interarrival times*. Specifically, we define the ith interarrival time t_i as the time between the $(i-1)$st jump of $N(T)$ and the ith jump. That is

$$\begin{aligned} t_1 &= w_1 \\ t_2 &= w_2 - w_1 \\ &\vdots \\ t_i &= w_i - w_{i-1}. \end{aligned} \tag{4.325}$$

Since $w_i = t_1 + t_2 + \cdots + t_i$, the joint density for the first k interarrival times can readily[10] be determined from that for $w_1, w_2, \ldots, w_k$. In particular, we have (when all $t_i \geq 0$):

$$\begin{aligned} & p_{t_1,t_2,\ldots,t_k}(t_1, t_2, \ldots, t_k) \\ &= p_{w_1,w_2,\ldots,w_k}(t_1, t_1 + t_2, \ldots, t_1 + t_2 + \cdots + t_k) \\ &= \lambda(t_1)\,\lambda(t_1 + t_2) \cdots \lambda(t_1 + t_2 + \cdots + t_k)\, e^{-\int_0^{t_1+t_2+\cdots+t_k} \lambda(\tau)\, d\tau} \end{aligned} \tag{4.326}$$

where we have used (4.313). Note that in general, (4.326) does *not* reduce to a product of functions of the individual t_i so that the t_i are *not* independent. This makes sense, since if $\lambda(t)$ is time varying, the likely values for each t_i will depend on the time, i.e., it will depend on $w_{i-1} = t_1 + t_2 + \cdots + t_{i-1}$. However, if $\lambda(t)$ is constant, then specializing (4.326) to this case, we see that $t_1, t_2, \ldots, t_k$ are independent and identically distributed random variables with common *exponential density*

$$p_{t_i}(t) = \begin{cases} \lambda e^{-\lambda t} & t \geq 0 \\ 0 & \text{otherwise} \end{cases}. \tag{4.327}$$

4.8 MEASUREMENT OF PROCESS CHARACTERISTICS AND ERGODICITY

An important aspect of all of the techniques developed in this course is that we use probabilistic *models* of stochastic phenomena in order to develop optimal processing algorithms, methods for analyzing processes and systems, and measures of performance in various signal processing tasks. This, of course, raises the question of how those models are constructed in the first place. The construction of probabilistic models is, in fact, an enormous field encompassing much of statistics, adaptive signal processing, and system identification. Furthermore, the formulations and methods that have been developed are numerous and varied, reflecting the many different contexts in which probabilistic models are desired. For example, in some cases physical laws and reasoning may provide us with most of the probabilistic description of a random process with the exception of a few parameters (e.g., a first-order Gauss-Markov process with unknown mean, variance, and correlation time or a Poisson process with unknown mean). In such situations we may wish to use techniques such as ML estimation to estimate these parameters. In other contexts we may know comparatively little about a phenomenon and thus cannot use such a constrained model but rather must use more of a "black box" approach to estimating some of the statistics of the process in question.

In this section we introduce perhaps the most basic and well-known approach to estimating process statistics which is based on the simple concept of

[10]Note that what we are doing here is calculating the joint density of k functions of k random variables. In general this requires evaluating the determinant of the Jacobian of the transformation. However, in this case a straightforward calculation shows that this determinant always equals 1.

computing time-averaged characteristics of an observed sample function and using these as estimates of process statistics. That is, we use *time-averaged* quantities to estimate *ensemble-averaged* statistics. The deep mathematical connection between time and ensemble averages is the subject of ergodic theory, and while we will not investigate that theory in any depth or with mathematical rigor, we will be able to give a few important insights into the nature of the subject and to its practical significance.

The key to our analysis is the investigation of a very simple time-averaging process, namely the calculation of the sample mean of a process. To begin, consider the discrete-time case in which we observe a sample path of the stochastic process $x[n], n = 0, \pm 1, \pm 2, \ldots$ and suppose that for any positive, odd integer N we define the N-point centered average

$$\hat{m}_x[n; N] = \frac{1}{N} \sum_{m=n-\frac{(N-1)}{2}}^{n+\frac{(N-1)}{2}} x[m]. \tag{4.328}$$

Intuitively we would expect $\hat{m}_x[n; N]$ to be a "reasonable" estimate of the mean of $x[n]$. However, what do we mean by "reasonable?" Also, how do we choose N? We might expect that choosing N large would have the beneficial effect of smoothing or averaging out the randomness in $x[n]$. But is that always the case and, more precisely, how quickly does the uncertainty in this estimate decrease with increasing N? Such considerations may be important if there are reasons to limit the size of N such as the expense of data collection, the time delay in producing an estimate, or the possibility that the mean of $x[n]$ might be changing with time. Indeed in the latter case, we might expect that there would be a strong desire to keep N as *small* as possible.

To make the preceding discussion more concrete, consider the situation in which $x[n]$ may have time varying mean $m_x[n]$ but it has a stationary covariance function $K_{xx}[n] = \mathrm{cov}\,(x[m], x[n+m])$. That is, we can think of $x[n]$ as being made up of the sum

$$x[n] = m_x[n] + \tilde{x}[n] \tag{4.329}$$

where $\tilde{x}[n]$ is WSS with zero-mean and covariance $K_{xx}[n]$.

With these assumptions we can proceed, as we have in the past, to evaluate the bias and variance in the sample mean estimator (4.328). Specifically,

$$E\left[\hat{m}_x[n; N]\right] = \frac{1}{N} \sum_{m=n-\frac{(N-1)}{2}}^{n+\frac{(N-1)}{2}} m_x[m], \tag{4.330}$$

so that the bias in this estimate is given by

$$b[n] = m_x[n] - E\left[\hat{m}_x[n; N]\right] = m_x[n] - \frac{1}{N} \sum_{m=n-\frac{(N-1)}{2}}^{n+\frac{(N-1)}{2}} m_x[m]. \tag{4.331}$$

From this we can conclude the following:

1. If $x[n]$ itself is WSS so that $m_x[n] = m_x$ is constant, then $E[\hat{m}_x[n;N]] = m_x$, so that $\hat{m}_x[n;N]$ is an unbiased estimator.
2. In the trivial case in which $N = 1$, i.e., when $\hat{m}_x[n;1] = x[n]$ so that no averaging occurs, the estimate is unbiased even if $m_x[n]$ varies with time. However, for $N > 1$, $\hat{m}_x[n;N]$ is in general a biased estimator if the mean of $x[n]$ varies with time.

Since the averaging operation is a low-pass filter we see that fluctuations in $m_x[n]$ that are fast compared to the length N of the averaging window will be averaged out, leading to a bias in the estimator. Thus, if time-varying fluctuations in the mean are expected, there is clear motivation to keep the averaging window small in order to be able to follow these fluctations with little or no bias. On the other hand, as we show next, reducing the *variance* in the running mean generally requires *increasing* N.

Specifically, as we have seen before, the mean-squared error in an estimate is the sum of the estimate variance and the square of the estimate bias:

$$\begin{aligned} E\left[(m_x[n] - \hat{m}_x[n;N])^2\right] &= (m_x[n] - E[\hat{m}_x[n;N]])^2 \\ &\quad + E\left[(\hat{m}_x[n;N] - E[\hat{m}_x[n;N]])^2\right] \\ &= b^2[n] + \operatorname{var}\hat{m}_x[n;N]. \end{aligned} \tag{4.332}$$

Since $\hat{m}_x[n;N]$ is simply the output of an LTI system driven by a random input $x[n]$, the calculation of the variance of the output $\hat{m}_x[n;N]$ is a simple application of the results of Section 4.5 on the propagation of second-order statistics through linear systems. Alternatively, we can recognize that

$$\hat{m}_x[n;N] - E[\hat{m}_x[n;N]] = \frac{1}{N} \sum_{m=n-\frac{(N-1)}{2}}^{n+\frac{(N-1)}{2}} \tilde{x}[m], \tag{4.333}$$

so that

$$\begin{aligned} \operatorname{var}\hat{m}_x[n;N] &= E\left[\left(\frac{1}{N} \sum_{m=n-\frac{(N-1)}{2}}^{n+\frac{(N-1)}{2}} \tilde{x}[m]\right)^2\right] \\ &= \frac{1}{N^2} \sum_m \sum_k K_{xx}[m-k], \end{aligned} \tag{4.334}$$

where the limits of both sums in (4.334) are $n - (N-1)/2$ and $n + (N-1)/2$. Examining this summation, we see that there are N pairs of (m,k) values with $m = k, N-1$ pairs with $m = k+1$ and another $N-1$ with $m = k-1$, and, more

generally, $N - r$ pairs with $m = k + r$ and another $N - r$ values with $m = k - r$. Thus,

$$\begin{aligned}\operatorname{var} \hat{m}_x[n; N] &= \frac{1}{N^2} \sum_{k=1-N}^{N-1} (N - |k|) K_{xx}[k] \\ &= \frac{1}{N} \sum_{k=1-N}^{N-1} \left(1 - \frac{|k|}{N}\right) K_{xx}[k]. \end{aligned} \tag{4.335}$$

Thus

$$E\left[(m_x[n] - \hat{m}_x[n; N])^2\right] = b^2[n] + \frac{1}{N} \sum_{k=1-N}^{N-1} \left(1 - \frac{|k|}{N}\right) K_{xx}[k] \tag{4.336}$$

where $b[n]$ is given in (4.331).

Let us take a look at several specific cases.

Example 4.27

Consider the case in which $x[n]$ is white noise ($K_{xx}[k] = \sigma^2 \delta[k]$) with a constant mean m_x. In this case we know that $b[n] = 0$ and, from (4.336) we see that

$$E\left[(\hat{m}_x[n; N] - m_x)^2\right] = \operatorname{var} \hat{m}_x[n; N] = \frac{\sigma^2}{N}. \tag{4.337}$$

There are several points to note about this example: (1) the estimate is unbiased; (2) if $x[n]$ is Gaussian, then $\hat{m}_x[n; N]$ is precisely the ML estimate of m_x based on the observation of $x[m], m = n - (N-1)/2, \ldots, n + (N-1)/2$; and (3) the mean-squared estimation error in (4.337) decreases with increasing N, and, in fact,

$$\lim_{N \to \infty} E\left[(\hat{m}_x[n; N] - m_x)^2\right] = 0. \tag{4.338}$$

Example 4.28

Consider a variation on the preceding example, where $K_{xx}[n] = \sigma^2 \delta[n]$ but where $m_x[n]$ may be time-varying. In this case, from (4.336) and (4.337) we have that

$$E\left[(m_x[n] - \hat{m}_x[n; N])^2\right] = b^2[n] + \frac{\sigma^2}{N} \tag{4.339}$$

which is the prototypical example of the tradeoff involved in the choice of N. In particular, as we have pointed out, the process of averaging is a smoothing, or low-pass operation, and this is reflected in: (a) a reduction in the effect of "noise", i.e., of the random fluctuations $\tilde{x}[n]$ as reflected in the variance term, σ^2/N ; and (b) the possible presence of a bias $b[n]$ in the estimate due to the inability of the low-pass filter to follow high-frequency fluctations in $m_x[n]$. Stated another way, increasing N reduces the effect of noise but also reduces the temporal *resolution* of our estimate—i.e., $\hat{m}_x[n; N]$ provides an estimate of the *average* of $m_x[n]$ over intervals of length N. Thus by adjusting N we can control the noise/resolution tradeoff inherent in sample mean estimate.

Example 4.29

Suppose now that $\tilde{x}[n]$ is not white but in fact has some temporal correlation. In particular, suppose that

$$K_{xx}[n] = \sigma^2 \alpha^{|n|} \tag{4.340}$$

where $|\alpha| < 1$. In this case, evaluating (4.335) yields

$$\begin{aligned} \operatorname{var} \hat{m}_x[n;N] &= \frac{\sigma^2}{N} \sum_{k=1-N}^{N-1} \left(1 - \frac{|k|}{N}\right) \alpha^{|k|} \\ &= \frac{\sigma^2}{N^2(1-\alpha)} \left[N(1+\alpha) - \frac{2\alpha(1-\alpha^N)}{1-\alpha}\right]. \end{aligned} \tag{4.341}$$

Note that as long as N is large enough, e.g.,

$$N \gg \frac{2\alpha}{1-\alpha^2} \tag{4.342}$$

the first term in brackets in (4.341) dominates, so that

$$\operatorname{var} \hat{m}_x[n;N] \approx \frac{\sigma^2}{N} \left(\frac{1+\alpha}{1-\alpha}\right). \tag{4.343}$$

Contrasting this with (4.337), we see that once again the variance of the estimate decreases to 0 as $N \to \infty$. However, the proportionality factor of $(1+\alpha)/(1-\alpha)$ is indicative of the fact that the temporal correlation in $x[n]$ has an influence on the overall variance. For example, if $\alpha = 1/2$, then the variance in (4.343) equals $3\sigma^2/N$ as compared to σ^2/N for the case of white noise. In the white noise case, since each measurement is uncorrelated with the others, each provides a distinct "degree of freedom" in the estimate $\hat{m}_x[n;N]$, where the variance in this estimate equals the unaveraged variance, σ^2, divided by the number of degrees of freedom. Eq. (4.341) states that the effective number of degrees of freedom—i.e., the effective number of uncorrelated samples—when $K_{xx}[n]$ is as in (4.340) is given by $N(1-\alpha)/(1+\alpha)$.

Note also that if m_x is constant, then (4.343) also represents an approximate (and (4.341) the exact) value of the mean-squared estimation error, which still goes to 0 as $N \to \infty$. Note, however, that unlike the white noise case, $\hat{m}_x[n;N]$ is *not* the ML estimate of m_x in this case if $x[n]$ is Gaussian. Nevertheless, it is a consistent estimator since it is unbiased and has vanishing variance as $N \to \infty$. On the other hand, if $m_x[n]$ is time-varying from (4.336) and (4.343) is given by

$$E\left[(m_x[n] - \hat{m}_x[n;N])^2\right] \approx b^2[n] + \frac{\sigma^2}{N}\left(\frac{1+\alpha}{1-\alpha}\right), \tag{4.344}$$

and we once again have that choosing N controls the tradeoff between noise and resolution.

Suppose that we now focus on the case in which $x[n]$ is WSS so that m_x is constant and $\hat{m}_x[n;N]$ is an unbiased estimator of m_x. In this case

$$\begin{aligned} E\left[(\hat{m}_x[n;N] - m_x)^2\right] &= \operatorname{var} \hat{m}_x[n;N] \\ &= \frac{1}{N} \sum_{k=1-N}^{N-1} \left(1 - \frac{|k|}{N}\right) K_{xx}[k], \end{aligned} \tag{4.345}$$

and an important question to ask is if this mean-squared error goes to 0 as $N \to \infty$, i.e., is $\hat{m}_x[n;N]$ a consistent estimate? For the two cases considered in Examples 4.27 and 4.29 this *is* the case (see(4.337) and (4.343). However, as the following example illustrates, this is *not* always the case.

Example 4.30

Suppose that $x[n]$ is a WSS stochastic process with *constant* sample paths, i.e., $x[n] = x[0]$ for *all* n. In this case

$$K_{xx}[n] = \text{cov}\,(x[n], x[0]) = \text{var}\, x[0] = \sigma^2 \tag{4.346}$$

so that from (4.336), or more easily [11] from (4.334), we see that

$$\text{var}\, \hat{m}_x[n;N] = \sigma^2, \tag{4.347}$$

which is *constant independent of* N. This, of course, makes sense, since if $x[n]$ is constant, averaging accomplishes nothing, i.e., $\hat{m}_x[n;N] = x[n]$ independent of N.

Example 4.30 illustrates the critical assumption involved if we believe that temporal averages can provide us with useful information about statistical averages. Specifically, for temporal averaging to be effective, it must be the case that there is a sufficient level of variability and fluctuation in each sample path so that ensemble behavior can be inferred from temporal averaging. Loosely speaking, this is the general concept of ergodicity. In particular, we can now define the simplest (and weakest) notion of ergodicity: a WSS stochastic process $x[n]$ is *ergodic in the mean* if $\hat{m}_x[n;N]$ is a consistent estimator, i.e., if

$$\lim_{N\to\infty} \frac{1}{N} \sum_{k=1-N}^{N-1} \left(1 - \frac{|k|}{N}\right) K_{xx}(k) = 0. \tag{4.348}$$

Roughly speaking what ergodicity in the mean requires is that $x[n]$ decorrelate fast enough as $n \to \infty$, so that the sum on the right-hand side of (4.348) converges (or at worst diverges slower than $O(N)$).

Note that we can perform exactly analogous analysis for continuous-time stochastic processes. In particular the T-second centered average of a stochastic process $x(t)$ is defined as

$$\hat{m}_x(t;T) = \frac{1}{T} \int_{t-T/2}^{t+T/2} x(s)\, ds, \tag{4.349}$$

and in this case, if $x(s)$ has mean $m_x(s)$

$$E\left[\hat{m}_x(t;T)\right] = \frac{1}{T} \int_{t-T/2}^{t+T/2} m_x(s)\, ds \tag{4.350}$$

[11] There are N^2 terms in the double sum in (4.334), and each term equals σ^2.

so that $\hat{m}_x(t;T)$ is an unbiased estimate if m_x is constant and otherwise has a bias $b(t) = m_x(t) - E[\hat{m}_x(t;T)]$ due to the low-pass nature of the averaging operation in (4.349) which prevents $\hat{m}_x(t;T)$ from following high frequency fluctuations in $m_x(t)$.

If $x(t)$ has stationary covariance function $K_{xx}(t)$, then

$$\begin{aligned}\operatorname{var}\hat{m}_x(t;T) &= \frac{1}{T^2}\int_{t-T/2}^{t+T/2}\int_{t-T/2}^{t+T/2} K_{xx}(\tau-\sigma)\,d\tau\,d\sigma \\ &= \frac{1}{T}\int_{-T}^{T}\left(1-\frac{|\tau|}{T}\right)K_{xx}(\tau)\,d\tau \end{aligned} \tag{4.351}$$

so that the mean-squared error is given by

$$E\left[[m_x(t)-\hat{m}_x(t;T)]^2\right] = b^2(t) + \operatorname{var}\hat{m}_x(t;T) \tag{4.352}$$

which exhibits the same fundamental noise/resolution tradeoff as described previously.

If $x(t)$ is WSS so that $\hat{m}(t;T)$ is an unbiased estimate, we can again ask the question of consistency, or ergodicity in the mean: does $\operatorname{var}\hat{m}_x(t;T) \to 0$ as $T \to \infty$? In analogy to the three examples given previously in discrete time, we have the following in continuous time:

Example 4.31

Suppose first that $\tilde{x}(t) = x(t) - m_x$ is white noise, i.e., $K_{xx}(t) = \sigma^2\delta(t)$. In this case

$$\operatorname{var}\hat{m}_x(t;T) = \frac{\sigma^2}{T} \tag{4.353}$$

which clearly goes to 0 as $T \to \infty$. Next, suppose that $K_{xx}(t) = \sigma^2 e^{-\lambda|t|}$. In this case

$$\operatorname{var}\hat{m}_x(t;T) = \frac{2\sigma^2}{\lambda T}\left[1-\frac{1-e^{-\lambda T}}{\lambda T}\right] \approx \frac{2\sigma^2}{\lambda T} \tag{4.354}$$

where the approximation is valid as long as $\lambda T \gg 1$. Note that once again $\operatorname{var}\hat{m}_x(t;T) \to 0$ as $T \to \infty$. In addition note that $2/\lambda$ is the width of $K_{xx}(t)$ between the points at which it falls off by a factor of e from its peak value. Thus $\lambda T/2 = T/(\lambda/2)$ has the interpretation as the effective number of degrees of freedom (i.e., the effective number of uncorrelated samples) in the measurement of $x(t)$ over an interval of length T.

Finally, once again if $x(t)$ is a process with constant sample paths, so that $K_{xx}(t) = K_{xx}(0) = \sigma^2$, then

$$\operatorname{var}\hat{m}_x(t;T) = \sigma^2 \tag{4.355}$$

so that $x(t)$ is *not* ergodic in the mean.

4.A LINEAR SYSTEMS

In this appendix we briefly summarize some of the basic results and concepts we will use. We begin with the continuous-time case.

4.A.1 Continuous-Time Systems

A general linear system with input $x(t)$ and output $y(t)$ has the form

$$y(t) = \int_{-\infty}^{+\infty} h(t,\tau)\, x(\tau)\, d\tau \tag{4.356}$$

where $h(t,\tau)$ is referred to as the *impulse response* or *kernel* of the system, i.e., if $x(t) = \delta(t - t_0)$ (where $\delta(\tau)$ denotes the unit impulse), then $y(t) = h(t, t_0)$. The system is *causal* if

$$h(t,\tau) = 0 \qquad \text{for } t < \tau \tag{4.357}$$

so that

$$y(t) = \int_{-\infty}^{t} h(t,\tau)\, x(\tau)\, d\tau. \tag{4.358}$$

The system is time-invariant if

$$h(t,\tau) = h(t-\tau, 0) \triangleq h(t-\tau) \tag{4.359}$$

where the last equality is an abuse of notation that we introduce for convenience. If (4.356) is time-invariant, then $y(t)$ is the convolution of $h(t)$ and $x(t)$:

$$y(t) = \int_{-\infty}^{+\infty} h(t-\tau)\, x(\tau)\, d\tau = \int_{-\infty}^{+\infty} h(\tau)\, x(t-\tau)\, d\tau. \tag{4.360}$$

Such a system is referred to as a *linear time-invariant* (LTI) system. Note that an LTI system is causal if and only if $h(t) = 0$ for all $t < 0$.

Let $X(s)$ denote the bilateral Laplace transform of the function $x(t)$:

$$X(s) = \mathcal{L}\{x(t)\} \triangleq \int_{-\infty}^{+\infty} x(t)\, e^{-st}\, dt. \tag{4.361}$$

For a time-invariant system, with $y(t)$, $h(t)$ and $x(t)$ related by (4.360) we have that

$$Y(s) = H(s)X(s). \tag{4.362}$$

The *Fourier transform* of $x(t)$ is simply $X(j\omega)$

$$X(j\omega) = \int_{-\infty}^{+\infty} x(t)\, e^{-j\omega t}\, dt \tag{4.363}$$

so that for an LTI system

$$Y(j\omega) = H(j\omega)\, X(j\omega). \tag{4.364}$$

The inverse Fourier transform is given by

$$x(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} X(j\omega)\, e^{j\omega t}\, d\omega. \tag{4.365}$$

Note that

$$X(0) = \int_{-\infty}^{+\infty} x(t)\, dt. \tag{4.366}$$

$$x(0) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} X(j\omega)\, d\omega. \tag{4.367}$$

Several properties of Fourier and Laplace transforms are as follows:

- If $x(t)$ is real then $X(-j\omega) = X^*(j\omega)$ where $*$ denotes complex conjugate.
- If $x(t)$ is even, i.e., if $x(-t) = x(t)$, then $X(-s) = X(s)$.
- If $x(t)$ is real and even, then so is $X(j\omega)$.
- If $x(t) = e^{j\omega_0 t}$, then $X(j\omega) = 2\pi\delta(\omega - \omega_0)$.
- If $x(t) = A\cos\omega_0 t$, then

  $$y(t) = A|H(j\omega_0)|\cos(\omega_0 t + \measuredangle H(j\omega_0))$$

  where the notation $|\cdot|$ and $\measuredangle\cdot$ denote magnitude and phase, respectively, of a complex quantity.
- If $x(t)$ has Laplace transform $X(s)$, then $sX(s)$ is the Laplace transform of

  $$\frac{dx(t)}{dt}.$$

- If $x(t) = \delta(t - t_0)$, then $X(s) = e^{-st_0}$.
- If $x(t) = \delta^{(n)}(t)$—where $\delta^{(n)}(t)$ is the nth derivative of $\delta(t)$—then $X(s) = s^n$.

Consider the signal

$$x_1(t) = e^{at}u(t) \tag{4.368}$$

where $u(t)$ is the unit step, i.e.,

$$u(t) = \begin{cases} 1 & t \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{4.369}$$

(here a could be complex). Its Laplace transform is given by

$$X_1(s) = \frac{1}{s - a}. \tag{4.370}$$

Note that $X_1(s)$ has a *pole* at $s = a$ (i.e., its denominator is zero at $s = a$). If we substitute (4.368) into (4.361), we find that the transform (4.361) converges and equals (4.370) as long as $\mathrm{Re}(s) > \mathrm{Re}(a)$. That is, the complete specifications of the Laplace transform of (4.368) requires the additional specification of the *domain of convergence* (DOC) for the transform.

To emphasize this point, consider a second signal

$$x_2(t) = -e^{at}\, u(-t). \tag{4.371}$$

The Laplace transform of $x_2(t)$ has the same form as (4.370), i.e.,

$$X_2(s) = \frac{1}{s-a} \tag{4.372}$$

although in this case the DOC is $\mathrm{Re}(s) < \mathrm{Re}(a)$. In general, then, the Laplace transform *together with* the specification of the DOC are required to uniquely characterize a signal. However, in the context of interest to us here, we will always be dealing with signals and functions that decay (or at least do not diverge) as $t \to \pm\infty$. Because of this fact, we will not explicitly specify the DOC, as there will be a unique choice of the DOC consistent with decaying functions. In particular consider the system function

$$X(s) = \frac{1}{s-a} \tag{4.373}$$

which has a pole at $s = \mathrm{Re}(a)$ and which has two possible DOC's: $\mathrm{Re}(s) > \mathrm{Re}(a)$ (corresponding to $x_1(t)$ in (4.368)) or $\mathrm{Re}(s) < \mathrm{Re}(a)$ (corresponding to $x_2(t)$ in (4.371)). If $\mathrm{Re}(a) < 0$, so that the pole is in the left-half plane, $x_1(t)$ decays as $t \to +\infty$, but $x_2(t)$ diverges as $t \to -\infty$. Similarly if $\mathrm{Re}(a) > 0$, diverges as $t \to +\infty$ but $x_2(t)$ decays as $t \to -\infty$. Thus, consistent with our convention here of focusing exclusively on signals that do not diverge, we have the inverse transform pair

$$X(s) = \frac{1}{s-a} \overset{\mathcal{L}}{\longleftrightarrow} x(t) = \begin{cases} e^{at}u(t) & \text{if } \mathrm{Re}(a) < 0 \\ -e^{at}u(-t) & \text{if } \mathrm{Re}(a) > 0 \end{cases}. \tag{4.374}$$

More generally for rational system functions with several poles, our convention will be that poles in the left-half plane correspond to signals, such as $x_1(t)$ in (4.368), that decay as $t \to +\infty$, while poles in the right-half plane correspond to signals like $x_2(t)$ in (4.371) that decay as $t \to -\infty$.

Continuing, let

$$x_3(t) = te^{at}\,u(t). \tag{4.375}$$

Then

$$X_3(s) = \frac{1}{(s-a)^2}. \tag{4.376}$$

More generally, if

$$x_4(t) = \frac{t^{n-1}}{(n-1)!}e^{at}u(t), \tag{4.377}$$

then

$$X_4(s) = \frac{1}{(s-a)^n}. \tag{4.378}$$

Also, if

$$x_5(t) = -\frac{t^{n-1}}{(n-1)!}e^{at}u(-t), \tag{4.379}$$

then $X_5(s)$ is exactly of the same form as in (4.378), the difference being that the DOC for $X_4(s)$ is $\mathrm{Re}(s) > \mathrm{Re}(a)$ while that for $X_5(s)$ is $\mathrm{Re}(s) < \mathrm{Re}(a)$. Once again

we will always be focusing on nondiverging signals so that by convention the transform

$$X(s) = \frac{1}{(s-a)^n} \tag{4.380}$$

will correspond to $x_4(t)$ in (4.377) if $\mathrm{Re}(a) < 0$ and to $x_5(t)$ in (4.379) if $\mathrm{Re}(a) > 0$.

These examples provide us with the basis of inverting rational transforms, i.e., transforms of the form

$$X(s) = \frac{c_m s^m + c_{m-1}s^{m-1} + \cdots + c_1 s + c_0}{s^n + d_{n-1}s^{n-1} + \cdots + d_1 s + d_0}. \tag{4.381}$$

If $m \geq n$, we can reduce $X(s)$ to the form

$$X(s) = \beta_{m-n}s^{m-n} + \beta_{m-n-1}s^{m-n-1} + \cdots + \beta_1 s + \beta_0 + X_1(s) \tag{4.382}$$

where $X_1(s)$ is a *proper* rational fraction

$$X_1(s) = \frac{\alpha_{n-1}s^{n-1} + \cdots + \alpha_1 s + \alpha_0}{s^n + d_{n-1}s^{n-1} + \cdots + d_1 s + d_0}. \tag{4.383}$$

Thus

$$x(t) = \beta_{m-n}\delta^{(m-n)}(t) + \cdots + \beta_1\delta'(t) + \beta_0\delta(t) + x_1(t) \tag{4.384}$$

where $x_1(t)$ is the inverse transform $X_1(s)$. To find this, suppose that the denominator is factored as

$$s^n + d_{n-1}s^{n-1} + \cdots + d_1 s + d_0 = (s-\lambda_1)^{k_1}(s-\lambda_2)^{k_2}\cdots(s-\lambda_r)^{k_r}. \tag{4.385}$$

Then $X_1(s)$ can be rewritten as

$$X_1(s) = \frac{A_{11}}{(s-\lambda_1)} + \frac{A_{12}}{(s-\lambda_1)^2} + \cdots + \frac{A_{1k_1}}{(s-\lambda_1)^{k_1}} + \frac{A_{21}}{(s-\lambda_2)} + \cdots + \frac{A_{rk_r}}{(s-\lambda_r)^{kr}} \tag{4.386}$$

where the A_{ij} can be obtained, for example by equating the two expressions in (4.383), (4.386), clearing denominators, and matching coefficients of each power of s. An alternative, closed-form solution is

$$A_{ij} = \frac{1}{(k_i - j)!}\left\{\frac{d^{(k_i-j)}}{ds^{k_i-j}}\left[(s-\lambda_i)^{k_i}X_1(s)\right]\right\}\Bigg|_{s=\lambda_i}. \tag{4.387}$$

The inverse transform of (4.386) can then be obtained term-by-term by inspection using our convention. For example, if

$$X_1(s) = \frac{2}{s+1} + \frac{1}{s-1} - \frac{3}{(s-1)^2} \tag{4.388}$$

then

$$x_1(t) = 2e^{-t}u(t) + \left(3te^t - e^t\right)u(-t). \tag{4.389}$$

Next consider a causal LTI system whose input $x(t)$ and output $y(t)$ are related by the following linear constant coefficient differential equation:

$$\frac{d^n y(t)}{dt^n} + a_{n-1}\frac{d^{n-1}y(t)}{dt^{n-1}} + \cdots + a_0 y(t) = b_m\frac{d^m x(t)}{dt^m} + b_{m-1}\frac{d^{m-1}x(t)}{dt^{m-1}} + \cdots b_0 x(t). \tag{4.390}$$

Applying the differentiation property repeatedly and solving we find that the system function for this system is given by

$$H(s) = \frac{Y(s)}{X(s)} = \frac{b_m s^m + b_{m-1} s^{m-1} + \cdots + b_0}{s^n + a_{n-1} s^{n-1} + \cdots + a_0}. \tag{4.391}$$

Note that $H(s)$ is rational so that $h(t)$ consists of terms as in (4.368), (4.371), (4.377), etc., where the a_i's determine the poles of $H(s)$, i.e., the zeros of the denominator polynomial $s^n + a_{n-1}s^{n-1} + \cdots + a_0$.

Consider an LTI system described in (4.360). We can call such a system *bounded-input/bounded-output* (BIBO) *stable* if whenever $x(t)$ is bounded (i.e., $|x(t)| \leq K < \infty$ for some K and for all t), then so is $y(t)$ (i.e., $|y(t)| \leq M < \infty$ for some M and for all t). The system (4.360) is BIBO stable if and only if

$$\int_{-\infty}^{+\infty} |h(t)|\, dt < \infty. \tag{4.392}$$

Note that this condition requires that $h(t)$ decay as $t \to \pm\infty$, which is consistent with our convention in inverting Laplace transforms, i.e., in identifying left-half plane poles with *causal* functions that decay as $t \to +\infty$ and right-half plane poles with *anticausal* functions that decay as $t \to -\infty$. As our use of this convention suggests we will be focusing attention on stable systems.

4.A.2 Discrete-Time Systems

We now turn our attention to discrete-time linear systems. We use the time index n which now takes on integer values only. A general linear system with input $x[n]$ and output $y[n]$ has the form

$$y[n] = \sum_{m=-\infty}^{+\infty} h[n,m]\, x[m] \tag{4.393}$$

where $h[n,m]$ is referred to as the *impulse response* or *weighting function* of the system, i.e., if $x[n] = \delta[n - n_0]$, then $y[n] = h[n, n_0]$ where $\delta[n]$ is the discrete-time unit impulse[12]

$$\delta[n] = \begin{cases} 1 & n = 0 \\ 0 & \text{otherwise} \end{cases}. \tag{4.394}$$

The system (4.393) is *causal* if

$$h[n,m] = 0 \qquad \text{for } n < m \tag{4.395}$$

[12]Sometimes the Kronecker delta

$$\delta_{ij} = \begin{cases} 1 & i = j \\ 0 & \text{otherwise} \end{cases}$$

is used as an alternative to the discrete-time impulse. Hence, $\delta_{ij} = \delta[i-j]$ and $\delta[n] = \delta_{n0}$.

so that

$$y[n] = \sum_{m=-\infty}^{n} h[n,m]\, x[m]. \tag{4.396}$$

The system is time-invariant if

$$h[n,m] = h[n-m,0] \triangleq h[n-m]. \tag{4.397}$$

If (4.393) is time-invariant, then $y[n]$ is the convolution of $h[n]$ and $x[n]$:

$$y[n] = \sum_{m=-\infty}^{+\infty} h[n-m]\, x[m] = \sum_{m=-\infty}^{+\infty} h[m]\, x[n-m]. \tag{4.398}$$

Such a system is referred to as a *linear time-invariant* (LTI) system. Note that an LTI system is causal if and only if $h[n] = 0$ for all $n < 0$.

Let $X(z)$ denote the (bilateral) *z-transform* of the function $x[n]$:

$$X(z) = \mathcal{Z}\{x[n]\} = \sum_{n=-\infty}^{+\infty} x[n] z^{-n}. \tag{4.399}$$

For a time-invariant system, with $y[n]$, $h[n]$ and $x[n]$ related by (4.398), we have that

$$Y(z) = H(z)X(z). \tag{4.400}$$

The *Fourier transform* of $x[n]$ is simply $X(e^{j\omega})$:

$$X(e^{j\omega}) = \sum_{n=-\infty}^{+\infty} x[n] e^{-j\omega n}, \tag{4.401}$$

so that for an LTI system

$$Y(e^{j\omega}) = H(e^{j\omega})\, X(e^{j\omega}). \tag{4.402}$$

The inverse Fourier transform is given by

$$x[n] = \frac{1}{2\pi} \int_{-\pi}^{\pi} X(e^{j\omega})\, e^{j\omega n}\, d\omega. \tag{4.403}$$

Note that

$$X(1) = \sum_{n=-\infty}^{+\infty} x[n] \tag{4.404}$$

$$x[0] = \frac{1}{2\pi} \int_{-\pi}^{\pi} X(e^{j\omega})\, d\omega. \tag{4.405}$$

Note also that since $X(e^{j\omega})$ is periodic in ω, with period 2π, the integrals in (4.403) and (4.405) could actually be computed over *any* interval of length 2π.

Several properties of Fourier and z-transforms are as follows:

- If $x[n]$ is real, then $X(e^{-j\omega}) = X^*(e^{j\omega})$.
- If $x[n]$ is even, then $X(z^{-1}) = X(z)$ and consequently $X(e^{-j\omega}) = X(e^{j\omega})$.
- If $x[n]$ is real and even, so is $X(e^{j\omega})$, viewed as a function of ω.
- If $x[n] = e^{j\omega_0 n}$, then $X(j\omega) = 2\pi \sum_{k=-\infty}^{+\infty} \delta(\omega - \omega_0 - 2\pi k)$.
- If $x[n] = A\cos\omega_0 n$, then

$$y[n] = A|H(e^{j\omega_0})|\cos(\omega_0 n + \measuredangle H(e^{j\omega_0}))$$

.

- If $x[n]$ has z-transform $X(z)$, then $zX(z)$ is the z-transform of $x[n+1]$.
- If $x[n] = \delta[n - n_0]$, then $X(z) = z^{-n_0}$.

Consider the signal

$$x_1[n] = \alpha^n u[n] \tag{4.406}$$

where α could be complex and where $u[n]$ is the discrete-time unit step

$$u[n] = \begin{cases} 1 & n \geq 0 \\ 0 & \text{otherwise} \end{cases}. \tag{4.407}$$

The z-transform of (4.406) is given by

$$X_1(z) = \frac{1}{1 - \alpha z^{-1}} = \frac{z}{z - \alpha}. \tag{4.408}$$

Note that $X(z)$ has a pole at $z = \alpha$. Furthermore, substituting (4.406) into (4.399) we see that for the z-transform to converge, we must have $|z| > |\alpha|$, so that the full specification of the z-transform of $x_1[n]$ is the formula (4.408) together with the DOC $|z| > |\alpha|$. Note also that the signal

$$x_2[n] = -\alpha^n u[-n-1] \tag{4.409}$$

also has the z-transform given by the formula (4.408) although in this case with the DOC $|z| < |\alpha|$. As in continuous time our convention in discrete time will be to associate z-transforms with signals that do not diverge. Thus poles inside the unit circle $|z| < 1$ will correspond to signals that decay as $n \to +\infty$, while poles outside the unit circle correspond to signals that decay as $n \to -\infty$. For example,

$$X(z) = \frac{1}{1 - \alpha z^{-1}} \overset{z}{\longleftrightarrow} x[n] = \begin{cases} \alpha^n u[n] & \text{if } |\alpha| < 1 \\ -\alpha^n u[-n-1] & \text{if } |\alpha| > 1 \end{cases}. \tag{4.410}$$

Furthermore, using the shifting property

$$X(z) = \frac{1}{z - \alpha} \overset{z}{\longleftrightarrow} x[n] = \begin{cases} \alpha^{n-1} u[n-1] & \text{if } |\alpha| < 1 \\ -\alpha^{n-1} u[-n] & \text{if } |\alpha| > 1 \end{cases}.$$

Continuing, let

$$x_3[n] = n\alpha^n u[n]. \tag{4.411}$$

Then

$$X_3(z) = \frac{\alpha z^{-1}}{(1-\alpha z^{-1})^2} = \frac{\alpha z}{(z-\alpha)^2} \tag{4.412}$$

with DOC $|z| > |\alpha|$. Furthermore, the signal

$$x_4[n] = -n\alpha^n u[-n-1] \tag{4.413}$$

has the same z-transform as in (4.412) although, in this case DOC is $|z| < |\alpha|$. Applying our convention we then have

$$X(z) = \frac{\alpha z^{-1}}{(1-\alpha z^{-1})^2} \overset{z}{\longleftrightarrow} x[n] = \begin{cases} n\alpha^n u[n] & \text{if } |\alpha| < 1 \\ -n\alpha^n u[-n-1] & \text{if } |\alpha| > 1 \end{cases}. \tag{4.414}$$

More generally, the z-transforms of $n^{k-1}\alpha^n u[n]$ and $-n^{k-1}\alpha^n u[-n-1]$ have denominators equal to $(z-\alpha)^k$ (although the algebra in deriving the exact expression is somewhat tedious).

The preceding analysis provides us with the basis for inverting rational z-transforms. Specifically, suppose

$$X(z) = \frac{c_m z^m + c_{m-1} z^{m-1} + \cdots + c_1 z + c_0}{z^k + d_{k-1} z^{k-1} + \cdots + d_1 z + d_0}. \tag{4.415}$$

In this case we can proceed exactly as in continuous-time, writing

$$X(z) = \beta_{m-k} z^{m-k} + \cdots + \beta_1 z + \beta_0 + X_1(z)$$

where

$$X_1(z) = \frac{\alpha_{k-1} z^{k-1} + \cdots + \alpha_1 z + \alpha_0}{z^k + d_{k-1} z^{k-1} + \cdots + d_1 z + d_0} \tag{4.416}$$

Thus

$$x[n] = \beta_{m-k}\delta[n+m-k] + \cdots + \beta_1\delta[n+1] + \beta_0\delta[n] + x_1[n]$$

where $x_1[n]$ is the inverse transform of $X_1(z)$. To find this, suppose that the denominator of (4.415) is factored as

$$z^k + d_{k-1} z^{k-1} + \cdots + d_1 z + d_0 = (z-\lambda_1)^{k_1}(z-\lambda_2)^{k_2}\cdots(z-\lambda_r)^{k_r} \tag{4.417}$$

Thus $X_1(z)$ can be written as

$$X_1(z) = \frac{A_{11}}{(z-\lambda_1)} + \cdots + \frac{A_{1k_1}}{(z-\lambda_1)^{k_1}} + \cdots + \frac{A_{rk_r}}{(z-\lambda_r)^{k_r}} \tag{4.418}$$

where the A_{ij} can be obtained by equating (4.416) and (4.418) and clearing denominators or by the same formula as in the Laplace transform case

$$A_{ij} = \frac{1}{(k_i-j)!}\left\{\frac{d^{k_i-j}}{dz^{(k_i-j)}}\left[(z-\lambda_i)^{k_i} X(z)\right]\right\}\Bigg|_{z=\lambda_i}. \tag{4.419}$$

The inverse transform of (4.417) can then be obtain term-by-term using our convention. For example, if

$$X(z) = \frac{2z^{-1}}{1-(1/2)z^{-1}} - \frac{(1/2)z^{-1}}{(1-(1/2)z^{-1})^2} - \frac{6z^{-1}}{(1-2z^{-1})^2},$$

then

$$x[n] = 2\left(\frac{1}{2}\right)^{n-1} u[n-1] - n\left(\frac{1}{2}\right)^{n} u[n] + 3(2)^n u[-n-1].$$

Next, consider a causal LTI system whose input $x[n]$ and output $y[n]$ are related by the following linear constant coefficient difference equation

$$y[n] + a_1 y[n-1] + \cdots + a_k y[n-k] = b_0 x[n] + b_1 x[n-1] + \cdots + b_m x[n-m]. \quad (4.420)$$

Applying the time shift property repeatedly and solving, we find that the system function for this system is given by

$$H(z) = \frac{Y(z)}{X(z)} = \frac{b_0 + b_1 z^{-1} + \cdots + b_m z^{-m}}{1 + a_1 z^{-1} + \cdots + a_k z^{-k}}. \quad (4.421)$$

Note that $H(z)$ is rational so that $h[n]$ consists of terms of the form $\alpha^n u[n], n\alpha^n u[-n]$, etc., where the values of α are the poles of $H(z)$.

Consider the general LTI system described by (4.398). Such a system is BIBO stable if and only if

$$\sum_{n=-\infty}^{+\infty} |h[n]| < \infty. \quad (4.422)$$

Once again this condition, which requires that $h[n]$ decay, is consistent with our convention in inverting z-transforms, i.e., in identifying poles inside the unit circle with functions that decay as $n \to +\infty$ and poles outside the unit circle with functions that decay as $n \to -\infty$.

4.B POWER CALCULATIONS FOR RATIONAL SPECTRA

In this appendix, we develop a useful algebraic procedure for calculating the variance of a WSS $x(t)$ from its power spectral density $S_{xx}(s)$ when $S_{xx}(s)$ is rational.

Since $\operatorname{var} x(t) = K_{xx}(0)$, it suffices to calculate the inverse Laplace transform of $S_{xx}(s)$ at $t = 0$. Before proceeding, however, we note that for $x(t)$ to have finite variance we require the degree of the numerator of $S_{xx}(s)$ be strictly smaller than the degree of its denominator—otherwise $x(t)$ would have an infinite variance white noise component. Assuming that $S_{xx}(s)$ is strictly proper, we can perform a partial fraction expansion which takes on a particularly simple form. Specifically, if $-p_1, -p_2, \ldots, -p_m$ denote the distinct left-hand plane poles of $S_{xx}(s)$ with multiplicities $k_1, k_2, \ldots, k_n$, then using $S_{xx}(-s) = S_{xx}(s)$, we have that

$$S_{xx}(s) = \sum_{i=1}^{m} \sum_{j=1}^{k_i} \left[\frac{A_{ij}}{(p_i + s)^j} + \frac{A_{ij}}{(p_i - s)^j} \right] \quad (4.423)$$

Then, since $K_{xx}(-t) = K_{xx}(t)$, we need only look at the positive time part of $S_{xx}(s)$, i.e., the terms in (4.423) with poles in the left-half plane. That is the Laplace transform of the positive time portion of $K_{xx}(t)$ is given by

$$\sum_{i=1}^{m} \sum_{j=1}^{k_i} \frac{A_{ij}}{(p_i + s)^j}. \tag{4.424}$$

Since

$$\frac{a}{(s+p)^j} \overset{\mathcal{L}}{\longleftrightarrow} \frac{at^{j-1}}{(j-1)!} e^{-pt} \tag{4.425}$$

we then see that

$$\operatorname{var} x(t) = K_{xx}(0) = \sum_{i=1}^{m} A_{i1}. \tag{4.426}$$

Example 4.32

Consider the process $x(t)$ resulting from applying a white noise process with spectral height 2 to the system with system function

$$H(s) = \frac{(4-s)}{(s+1)(s+2)}. \tag{4.427}$$

Thus the power spectral density of $x(t)$ is

$$\begin{aligned} S_{xx}(s) = 2H(s)H(-s) &= \frac{2(4-s)(4+s)}{(1+s)(2+s)(1-s)(2-s)} \\ &= \frac{5}{s+1} - \frac{2}{s+2} + \frac{5}{1-s} - \frac{2}{2-s}. \end{aligned} \tag{4.428}$$

Thus

$$\operatorname{var} x(t) = 5 - 2 = 3. \tag{4.429}$$