CHAPTER 9

Random Processes

INTRODUCTION

Much of your background in signals and systems is assumed to have focused on the effect of LTI systems on deterministic signals, developing tools for analyzing this class of signals and systems, and using what you learned in order to understand applications in communication (e.g., AM and FM modulation), control (e.g., stability of feedback systems), and signal processing (e.g., filtering). It is important to develop a comparable understanding and associated tools for treating the effect of LTI systems on signals modeled as the outcome of probabilistic experiments, i.e., a class of signals referred to as random signals (alternatively referred to as random processes or stochastic processes). Such signals play a central role in signal and system design and analysis, and throughout the remainder of this text. In this chapter we define random processes via the associated ensemble of signals, and begin to explore their properties. In successive chapters we use random processes as models for random or uncertain signals that arise in communication, control and signal processing applications.

9.1 DEFINITION AND EXAMPLES OF A RANDOM PROCESS

In Section 7.3 we defined a random variable X as a function that maps each outcome of a probabilistic experiment to a real number. In a similar manner, a real-valued CT or DT random process, $X(t)$ or $X[n]$ respectively, is a function that maps each outcome of a probabilistic experiment to a real CT or DT signal respectively, termed the realization of the random process in that experiment. For any fixed time instant $t = t_0$ or $n = n_0$, the quantities $X(t_0)$ and $X[n_0]$ are just random variables. The collection of signals that can be produced by the random process is referred to as the ensemble of signals in the random process.

EXAMPLE 9.1 Random Oscillators

As an example of a random process, imagine a warehouse containing N harmonic oscillators, each producing a sinusoidal waveform of some specific amplitude, frequency, and phase, all of which may be different for the different oscillators. The probabilistic experiment that results in the ensemble of signals consists of selecting an oscillator according to some probability mass function (PMF) that assigns a probability to each of the numbers from 1 to N, so that the ith oscillator is picked

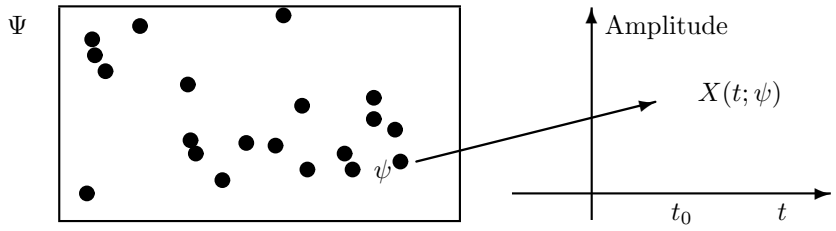

FIGURE 9.1 A random process.

with probability p_i. Associated with each outcome of this experiment is a specific sinusoidal waveform.

In Example 9.1, before an oscillator is chosen, there is uncertainty about what the amplitude, frequency and phase of the outcome of the experiment will be. Consequently, for this example, we might express the random process as

$$X(t) = A\sin(\omega t + \phi)$$

where the amplitude A, frequency ω and phase ϕ are all random variables. The value $X(t_1)$ at some specific time t_1 is also a random variable. In the context of this experiment, knowing the PMF associated with each of the numbers 1 to N involved in choosing an oscillator, as well as the specific amplitude, frequency and phase of each oscillator, we could determine the probability distributions of any of the underlying random variables A, ω, ϕ or $X(t_1)$ mentioned above.

Throughout this and later chapters, we will be considering many other examples of random processes. What is important at this point, however, is to develop a good mental picture of what a random process is. A random process is not just one signal but rather an ensemble of signals, as illustrated schematically in Figure 9.2 below, for which the outcome of the probabilistic experiment could be any of the four waveforms indicated. Each waveform is deterministic, but the process is probabilistic or random because it is not known *a priori* which waveform will be generated by the probabilistic experiment. Consequently, prior to obtaining the outcome of the probabilistic experiment, many aspects of the signal are unpredictable, since there is uncertainty associated with which signal will be produced. After the experiment, or *a posteriori*, the outcome is totally determined.

If we focus on the values that a random process $X(t)$ can take at a particular instant of time, say t_1 — i.e., if we look down the entire ensemble at a fixed time — what we have is a random variable, namely $X(t_1)$. If we focus on the ensemble of values taken at an arbitrary collection of ℓ fixed time instants $t_1 < t_2 < \cdots < t_\ell$ for some arbitrary integer ℓ, we are dealing with a set of ℓ jointly distributed random variables $X(t_1)$, $X(t_2)$, $\cdots$, $X(t_\ell)$, all determined together by the outcome of the underlying probabilistic experiment. From this point of view, a random process

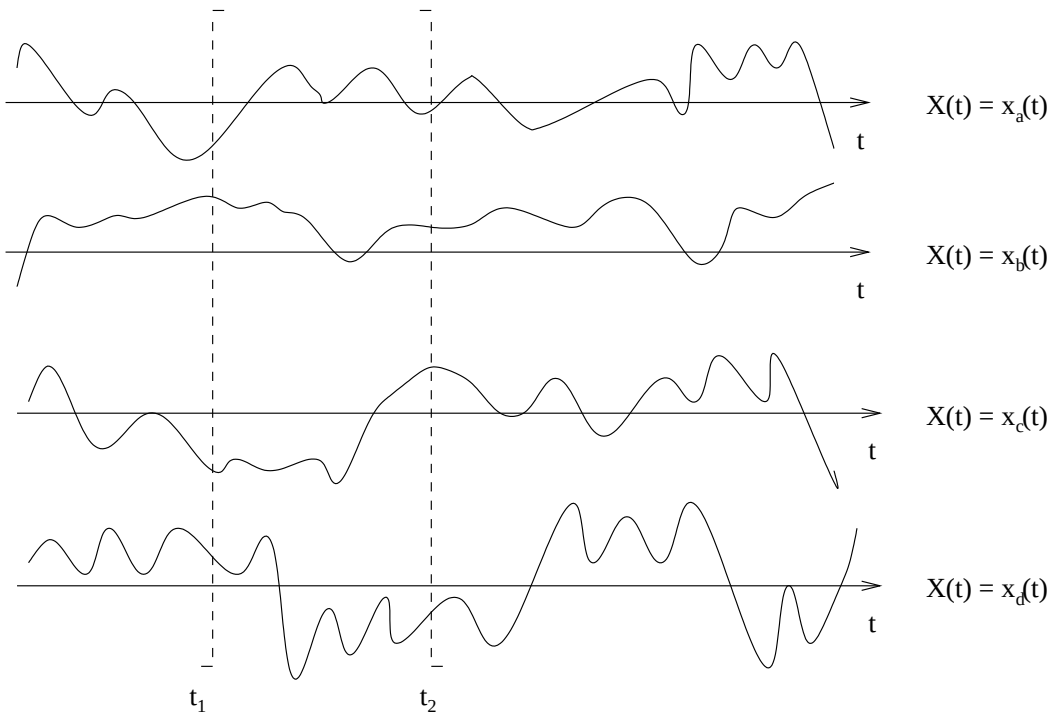

FIGURE 9.2 Realizations of the random process $X(t)$

can be thought of as a family of jointly distributed random variables indexed by t (or n in the DT case). A full probabilistic characterization of this collection of random variables would require the joint PDFs of multiple samples of the signal, taken at arbitrary times:

$$f_{X(t_1),X(t_2),\cdots,X(t_\ell)}(x_1, x_2, \cdots, x_\ell)$$

for all ℓ and all $t_1, t_2, \cdots, t_\ell$.

An important set of questions that arises as we work with random processes in later chapters of this book is whether, by observing just part of the outcome of a random process, we can determine the complete outcome. The answer will depend on the details of the random process, but in general the answer is no. For some random processes, having observed the outcome in a given time interval might provide sufficient information to know exactly which ensemble member was determined. In other cases it would not be sufficient. We will be exploring some of these aspects in more detail later, but we conclude this section with two additional examples that

further emphasize these points.

EXAMPLE 9.2 Ensemble of batteries

Consider a collection of N batteries, each providing one voltage out of a given finite set of voltage values. The histogram of voltages (i.e., the number of batteries with a given voltage) is given in Figure 9.3. The probabilistic experiment is to choose

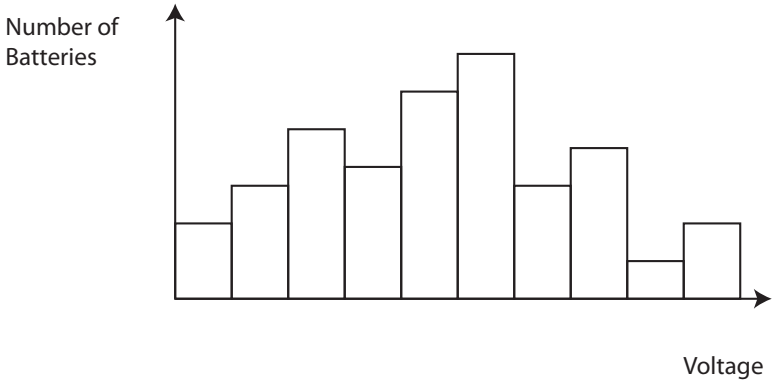

FIGURE 9.3 Histogram of battery distribution for Example 9.2.

one of the batteries, with the probability of picking any specific one being $\frac{1}{N}$, i.e., they are all equally likely to be picked. A little reflection should convince you that if we multiply the histogram in Figure 9.3 by $\frac{1}{N}$, this normalized histogram will represent (or approximate) the PMF for the battery voltage at the outcome of the experiment. Since the battery voltage is a constant signal, this corresponds to a random process, and in fact is similar to the oscillator example discussed earlier, but with $\omega = 0$ and $\phi = 0$, so that only the amplitude is random.

For this example observation of $X(t)$ at any one time is sufficient information to determine the outcome for all time.

EXAMPLE 9.3 Ensemble of coin tossers

Consider N people, each independently having written down a long random string of ones and zeros, with each entry chosen independently of any other entry in their string (similar to a sequence of independent coin tosses). The random process now comprises this ensemble of strings. A realization of the process is obtained by randomly selecting a person (and therefore one of the N strings of ones and zeros), following which the specific ensemble member of the random process is totally determined. The random process described in this example is often referred to as

the Bernoulli process because of the way in which the string of ones and zeros is generated (by independent coin flips).

Now suppose that person shows you only the tenth entry in the string. Can you determine (or predict) the eleventh entry from just that information? Because of the manner in which the string was generated, the answer clearly is no. Similarly if the entire past history up to the tenth entry was revealed to you, could you determine the remaining sequence beyond the tenth? For this example, the answer is again clearly no.

While the entire sequence has been determined by the nature of the experiment, partial observation of a given ensemble member is in general not sufficient to fully specify that member.

Rather than looking at the nth entry of a single ensemble member, we can consider the random variable corresponding to the values from the entire ensemble at the nth entry. Looking down the ensemble at $n = 10$, for example, we would would see ones and zeros with equal probability.

In the above discussion we indicated and emphasized that a random process can be thought of as a family of jointly distributed random variables indexed by t or n. Obviously it would in general be extremely difficult or impossible to represent a random process this way. Fortunately, the most widely used random process models have special structure that permits computation of such a statistical specification. Also, particularly when we are processing our signals with linear systems, we often design the processing or analyze the results by considering only the first and second moments of the process, namely the following functions:

$$\textit{Mean:} \quad \mu_X(t_i) = E[X(t_i)], \tag{9.1}$$

$$\textit{Auto-correlation:} \quad R_{XX}(t_i, t_j) = E[X(t_i)X(t_j)], \text{ and} \tag{9.2}$$

$$\textit{Auto-covariance:} \quad C_{XX}(t_i, t_j) = E[(X(t_i) - \mu_X(t_i))(X(t_j) - \mu_X(t_j))]$$
$$= R_{XX}(t_i, t_j) - \mu_X(t_i)\mu_X(t_j). \tag{9.3}$$

The word "auto" (which is sometime written without the hyphen, and sometimes dropped altogether to simplify the terminology) here refers to the fact that both samples in the correlation function or the covariance function come from the same process; we shall shortly encounter an extension of this idea, where the samples are taken from two different processes.

One case in which the first and second moments actually suffice to completely specify the process is in the case of what is called a Gaussian process, defined as a process whose samples are always jointly Gaussian (the generalization of the bivariate Gaussian to many variables).

We can also consider multiple random processes, e.g., two processes, $X(t)$ and $Y(t)$. For a full stochastic characterization of this, we need the PDFs of all possible combinations of samples from $X(t), Y(t)$. We say that $X(t)$ and $Y(t)$ are independent if every set of samples from $X(t)$ is independent of every set of samples from $Y(t)$,

so that the joint PDF factors as follows:

$$\begin{aligned} & f_{X(t_1),\cdots,X(t_k),Y(t'_1),\cdots,Y(t'_\ell)}(x_1,\cdots,x_k,y_1,\cdots,y_\ell) \\ &= f_{X(t_1),\cdots,X(t_k)}(x_1,\cdots,x_k).f_{Y(t'_1),\cdots,Y(t'_\ell)}(y_1,\cdots,y_\ell)\ . \end{aligned} \tag{9.4}$$

If only first and second moments are of interest, then in addition to the individual first and second moments of $X(t)$ and $Y(t)$ respectively, we need to consider the cross-moment functions:

$$\textit{Cross-correlation: } R_{XY}(t_i,t_j) = E[X(t_i)Y(t_j)], \text{ and} \tag{9.5}$$

$$\begin{aligned} \textit{Cross-covariance: } C_{XY}(t_i,t_j) &= E[(X(t_i)-\mu_X(t_i))(Y(t_j)-\mu_Y(t_j))] \\ &= R_{XY}(t_i,t_j) - \mu_X(t_i)\mu_Y(t_j). \end{aligned} \tag{9.6}$$

If $C_{XY}(t_1,t_2) = 0$ for all t_1, t_2, we say that the processes $X(t)$ and $Y(t)$ are uncorrelated. Note again that the term "uncorrelated" in its common usage means that the processes have zero covariance rather than zero correlation.

Note that everything we have said above can be carried over to the case of DT random processes, except that now the sampling instants are restricted to be discrete time instants. In accordance with our convention of using square brackets $[\cdot]$ around the time argument for DT signals, we will write $\mu_X[n]$ for the mean of a random process $X[\cdot]$ at time n; similarly, we will write $R_{XX}[n_i,n_j]$ for the correlation function involving samples at times n_i and n_j; and so on.

9.2 STRICT-SENSE STATIONARITY

In general, we would expect that the joint PDFs associated with the random variables obtained by sampling a random process at an arbitrary number k of arbitrary times will be time-dependent, i.e., the joint PDF

$$f_{X(t_1),\cdots,X(t_k)}(x_1,\cdots,x_k)$$

will depend on the specific values of $t_1,\cdots,t_k$. If all the joint PDFs stay the same under arbitrary time *shifts*, i.e., if

$$f_{X(t_1),\cdots,X(t_k)}(x_1,\cdots,x_k) = f_{X(t_1+\tau),\cdots,X(t_k+\tau)}(x_1,\cdots,x_k) \tag{9.7}$$

for arbitrary τ, then the random process is said to be strict-sense stationary (SSS). Said another way, for a strict-sense stationary process, the statistics depend only on the relative times at which the samples are taken, not on the absolute times.

EXAMPLE 9.4 Representing an i.i.d. process

Consider a DT random process whose values $X[n]$ may be regarded as independently chosen at each time n from a fixed PDF $f_X(x)$, so the values are independent and identically distributed, thereby yielding what is called an i.i.d. process. Such processes are widely used in modeling and simulation. For instance, if a particular

DT communication channel corrupts a transmitted signal with added noise that takes independent values at each time instant, but with characteristics that seem unchanging over the time window of interest, then the noise may be well modeled as an i.i.d. process. It is also easy to generate an i.i.d. process in a simulation environment, provided one can arrange a random-number generator to produce samples from a specified PDF (and there are several good ways to do this). Processes with more complicated dependence across time samples can then be obtained by filtering or other operations on the i.i.d. process, as we shall see in the next chapter.

For such an i.i.d. process, we can write the joint PDF quite simply:

$$f_{X[n_1],X[n_2],\cdots,X[n_\ell]}(x_1,x_2,\cdots,x_\ell) = f_X(x_1)f_X(x_2)\cdots f_X(x_\ell) \tag{9.8}$$

for any choice of ℓ and $n_1,\cdots,n_\ell$. The process is clearly SSS.

9.3 WIDE-SENSE STATIONARITY

Of particular use to us is a less restricted type of stationarity. Specifically, if the mean value $\mu_X(t_i)$ is independent of time and the autocorrelation $R_{XX}(t_i,t_j)$ or equivalently the autocovariance $C_{XX}(t_i,t_j)$ is dependent only on the time difference $(t_i - t_j)$, then the process is said to be wide-sense stationary (WSS). Clearly a process that is SSS is also WSS. For a WSS random process $X(t)$, therefore, we have

$$\mu_X(t) = \mu_X \tag{9.9}$$

$$\begin{aligned} R_{XX}(t_1,t_2) &= R_{XX}(t_1+\alpha,t_2+\alpha) \text{ for every } \alpha \\ &= R_{XX}(t_1-t_2,0)\ . \end{aligned} \tag{9.10}$$

(Note that for a Gaussian process (i.e., a process whose samples are always jointly Gaussian) WSS implies SSS, because jointly Gaussian variables are entirely determined by the their joint first and second moments.)

Two random processes $X(t)$ and $Y(t)$ are jointly WSS if their first and second moments (including the cross-covariance) are stationary. In this case we use the notation $R_{XY}(\tau)$ to denote $E[X(t+\tau)Y(t)]$.

EXAMPLE 9.5 Random Oscillators Revisited

Consider again the harmonic oscillators as introduced in Example 9.1, i.e.

$$X(t;A,\Theta) = A\cos(\omega_0 t + \Theta)$$

where A and Θ are independent random variables, and now ω_0 is fixed at some known value.

If Θ is actually fixed at the constant value θ_0, then every outcome is of the form $x(t) = A\cos(\omega_0 t + \theta_0)$, and it is straightforward to see that this process is not WSS

(and hence not SSS). For instance, if A has a nonzero mean value, $\mu_A \neq 0$, then the expected value of the process, namely $\mu_A \cos(\omega_0 t + \theta_0)$, is time varying. To argue that the process is not WSS even when $\mu_A = 0$, we can examine the autocorrelation function. Note that $x(t)$ is fixed at the value 0 for all values of t such that $\omega_0 t + \theta_0$ is an odd multiple of $\pi/2$, and takes the values $\pm A$ half-way between such points; the correlation between such samples taken π/ω_0 apart in time can correspondingly be 0 (in the former case) or $-E[A^2]$ (in the latter). The process is thus not WSS.

On the other hand, if Θ is distributed uniformly in $[-\pi, \pi]$, then

$$\mu_X(t) = \mu_A \int_{-\pi}^{\pi} \frac{1}{2\pi} \cos(\omega_0 t + \theta) d\theta = 0 \,, \tag{9.11}$$

$$\begin{aligned} C_{XX}(t_1, t_2) &= R_{XX}(t_1, t_2) \\ &= E[A^2] E[\cos(\omega_0 t_1 + \Theta) \cos(\omega_0 t_2 + \Theta)] \\ &= \frac{E[A^2]}{2} \cos(\omega_0 (t_2 - t_1)) \,, \end{aligned} \tag{9.12}$$

so the process is WSS. It can also be shown to be SSS, though this is not totally straightforward to show formally.

To simplify notation for a WSS process, we write the correlation function as $R_{XX}(t_1 - t_2)$; the argument $t_1 - t_2$ is referred to as the lag at which the correlation is computed. For the most part, the random processes that we treat will be WSS processes. When considering just first and second moments and not entire PDFs or CDFs, it will be less important to distinguish between the random process $X(t)$ and a specific realization $x(t)$ of it — so we shall go one step further in simplifying notation, by using lower case letters to denote the random process itself. We shall thus talk of the random process $x(t)$, and — in the case of a WSS process — denote its mean by μ_x and its correlation function $E\{x(t+\tau)x(t)\}$ by $R_{xx}(\tau)$. Correspondingly, for DT we'll refer to the random process $x[n]$ and (in the WSS case) denote its mean by μ_x and its correlation function $E\{x[n+m]x[n]\}$ by $R_{xx}[m]$.

9.3.1 Some Properties of WSS Correlation and Covariance Functions

It is easily shown that for real-valued WSS processes $x(t)$ and $y(t)$ the correlation and covariance functions have the following symmetry properties:

$$R_{xx}(\tau) = R_{xx}(-\tau) \,, \qquad C_{xx}(\tau) = C_{xx}(-\tau) \tag{9.13}$$

$$R_{xy}(\tau) = R_{yx}(-\tau) \,, \qquad C_{xy}(\tau) = C_{yx}(-\tau) \,. \tag{9.14}$$

We see from (9.13) that the autocorrelation and autocovariance have even symmetry. Similar properties hold for DT WSS processes.

Another important property of correlation and covariance functions follows from noting that the correlation coefficient of two random variables has magnitude not

exceeding 1. Applying this fact to the samples $x(t)$ and $x(t+\tau)$ of the random process $x(\cdot)$ directly leads to the conclusion that

$$-C_{xx}(0) \leq C_{xx}(\tau) \leq C_{xx}(0) . \tag{9.15}$$

In other words, the autocovariance function never exceeds in magnitude its value at the origin. Adding μ_x^2 to each term above, we find the following inequality holds for correlation functions:

$$-R_{xx}(0) + 2\mu_x^2 \leq R_{xx}(\tau) \leq R_{xx}(0) . \tag{9.16}$$

In Chapter 10 we will demonstrate that correlation and covariance functions are characterized by the property that their Fourier transforms are real and nonnegative at all frequencies, because these transforms describe the frequency distribution of the expected power in the random process. The above symmetry constraints and bounds will then follow as natural consequences, but they are worth highlighting here already.

9.4 SUMMARY OF DEFINITIONS AND NOTATION

In this section we summarize some of the definitions and notation we have previously introduced. As in Section 9.3, we shall use lower case letters to denote random processes, since we will only be dealing with expectations and not densities. Thus, with $x(t)$ and $y(t)$ denoting (real) random processes, we summarize the following definitions:

$$\textit{mean}: \quad \mu_x(t) \triangleq E\{x(t)\} \tag{9.17}$$

$$\textit{autocorrelation}: \quad R_{xx}(t_1,t_2) \triangleq E\{x(t_1)x(t_2)\} \tag{9.18}$$

$$\textit{cross}-\textit{correlation}: \quad R_{xy}(t_1,t_2) \triangleq E\{x(t_1)y(t_2)\} \tag{9.19}$$

$$\begin{aligned} \textit{autocovariance}: \quad & C_{xx}(t_1,t_2) \triangleq E\{[x(t_1)-\mu_x(t_1)][x(t_2)-\mu_x(t_2)]\} \\ & = R_{xx}(t_1,t_2) - \mu_x(t_1)\mu_x(t_2) \end{aligned} \tag{9.20}$$

$$\begin{aligned} \textit{cross}-\textit{covariance}: \quad & C_{xy}(t_1,t_2) \triangleq E\{[x(t_1)-\mu_x(t_1)][y(t_2)-\mu_y(t_2)]\} \\ & = R_{xy}(t_1,t_2) - \mu_x(t_1)\mu_y(t_2) \end{aligned} \tag{9.21}$$

strict-sense stationary (SSS):	all joint statistics for $x(t_1), x(t_2), \ldots, x(t_\ell)$ for all $\ell > 0$ and all choices of sampling instants $t_1, \cdots, t_\ell$ depend only on the *relative* locations of sampling instants.
wide-sense stationary (WSS):	$\mu_x(t)$ is constant at some value μ_x, and $R_{xx}(t_1, t_2)$ is a function of $(t_1 - t_2)$ only, denoted in this case simply by $R_{xx}(t_1 - t_2)$; hence $C_{xx}(t_1, t_2)$ is a function of $(t_1 - t_2)$ only, and written as $C_{xx}(t_1 - t_2)$.
jointly wide-sense stationary:	$x(t)$ and $y(t)$ are individually WSS and $R_{xy}(t_1, t_2)$ is a function of $(t_1 - t_2)$ only, denoted simply by $R_{xy}(t_1 - t_2)$; hence $C_{xy}(t_1, t_2)$ is a function of $(t_1 - t_2)$ only, and written as $C_{xy}(t_1 - t_2)$.

For WSS processes we have, in continuous-time and with simpler notation,

$$R_{xx}(\tau) = E\{x(t+\tau)x(t)\} = E\{x(t)x(t-\tau)\} \tag{9.22}$$

$$R_{xy}(\tau) = E\{x(t+\tau)y(t)\} = E\{x(t)y(t-\tau)\}, \tag{9.23}$$

and in discrete-time,

$$R_{xx}[m] = E\{x[n+m]x[n]\} = E\{x[n]x[n-m]\} \tag{9.24}$$

$$R_{xy}[m] = E\{x[n+m]y[n]\} = E\{x[n]y[n-m]\}. \tag{9.25}$$

We use corresponding (centered) definitions and notation for *covariances*:

$$C_{xx}(\tau), C_{xy}(\tau), C_{xx}[m], \text{ and } C_{xy}[m] \ .$$

It is worth noting that an alternative convention used elsewhere is to define $R_{xy}(\tau)$ as $R_{xy}(\tau) \triangleq E\{x(t)y(t+\tau)\}$. In our notation, this expectation would be denoted by $R_{xy}(-\tau)$. It's important to be careful to take account of what notational convention is being followed when you read this material elsewhere, and you should also be clear about what notational convention we are using in this text.

9.5 FURTHER EXAMPLES

EXAMPLE 9.6 Bernoulli process

The Bernoulli process, a specific example of which was discussed previously in Example 9.3, is an example of an i.i.d. DT process with

$$\mathrm{P}(x[n] = 1) = p \tag{9.26}$$

$$\mathrm{P}(x[n] = -1) = (1 - p) \tag{9.27}$$

and with the value at each time instant n independent of the values at all other

time instants. A simple calculation results in

$$E\{x[n]\} = 2p - 1 = \mu_x \tag{9.28}$$

$$E\{x[n+m]x[n]\} = \begin{cases} 1 & m = 0 \\ (2p-1)^2 & m \neq 0 \end{cases} \tag{9.29}$$

$$C_{xx}[m] = E\{(x[n+m] - \mu_x)(x[n] - \mu_x)\} \tag{9.30}$$

$$= \{1 - (2p-1)^2\}\delta[m] = 4p(1-p)\delta[m]\ . \tag{9.31}$$

EXAMPLE 9.7 Random telegraph wave

A useful example of a CT random process that we'll make occasional reference to is the random telegraph wave. A representative sample function of a random telegraph wave process is shown in Figure 9.4. The random telegraph wave can be defined through the following two properties:

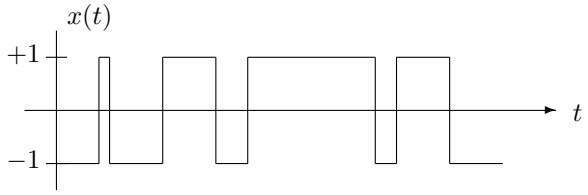

FIGURE 9.4 One realization of a random telegraph wave.

1. $X(0) = \pm 1$ with probability 0.5.
2. $X(t)$ changes polarity at Poisson times, i.e., the probability of k sign changes in a time interval of length T is

$$\mathrm{P}(k \text{ sign changes in an interval of length } T) = \frac{(\lambda T)^k e^{-\lambda T}}{k!}\ . \tag{9.32}$$

Property 2 implies that the probability of a non-negative, even number of sign changes in an interval of length T is

$$\mathrm{P}(\text{a non-negative even \# of sign changes}) = \sum_{\substack{k=0 \\ k \text{ even}}}^{\infty} \frac{(\lambda T)^k e^{-\lambda T}}{k!} = e^{-\lambda T} \sum_{k=0}^{\infty} \frac{1 + (-1)^k}{2} \frac{(\lambda T)^k}{k!} \tag{9.33}$$

Using the identity

$$e^{\lambda T} = \sum_{k=0}^{\infty} \frac{(\lambda T)^k}{k!}$$

equation (9.33) becomes

$$\begin{aligned}\mathrm{P}(\text{a non-negative even \# of sign changes}) &= e^{-\lambda T}\,\frac{(e^{\lambda T}+e^{-\lambda T})}{2}\\ &= \frac{1}{2}(1+e^{-2\lambda T})\,. \end{aligned} \tag{9.34}$$

Similarly, the probability of an odd number of sign changes in an interval of length T is $\frac{1}{2}(1-e^{-2\lambda T})$. It follows that

$$\begin{aligned}\mathrm{P}(X(t)=1) &= \mathrm{P}(X(t)=1|X(0)=1)\mathrm{P}(X(0)=1)\\ &\quad+\mathrm{P}(X(t)=1|X(0)=-1)\mathrm{P}(X(0)=-1)\\ &= \frac{1}{2}\mathrm{P}(\text{even \# of sign changes in } [0,t])\\ &\quad+\frac{1}{2}\mathrm{P}(\text{odd \# of sign changes in } [0,t])\\ &= \frac{1}{2}\left\{\frac{1}{2}(1+e^{-2\lambda t})\right\}+\frac{1}{2}\left\{\frac{1}{2}(1-e^{-2\lambda t})\right\}=\frac{1}{2}\,. \end{aligned} \tag{9.35}$$

Note that because of Property I, the expression in the last line of Eqn. (9.35) is not needed, since the line before that already allows us to conclude that the answer is $\frac{1}{2}$: since the number of sign changes in any interval must be either even or odd, their probabilities add up to 1, so $\mathrm{P}\left(X(t)=1\right)=\frac{1}{2}$. However, if Property 1 is relaxed to allow $\mathrm{P}(X(0)=1)=p_0\neq\frac{1}{2}$, then the above computation must be carried through to the last line, and yields the result

$$\mathrm{P}(X(t)=1)=p_0\left\{\frac{1}{2}(1+e^{-2\lambda t})\right\}+(1-p_0)\left\{\frac{1}{2}(1-e^{-2\lambda t})\right\}=\frac{1}{2}\left\{1+(2p_0-1)e^{-2\lambda t}\right\}\,. \tag{9.36}$$

Returning to the case where Property 1 holds, so $\mathrm{P}(X(t)=1)$, we get

$$\mu_X(t)=0,\ \text{and} \tag{9.37}$$

$$\begin{aligned}R_{XX}(t_1,t_2) &= E[X(t_1)X(t_2)]\\ &= 1\times\mathrm{P}\left(X(t_1)=X(t_2)\right)+(-1)\times\mathrm{P}\left(X(t_1)\neq X(t_2)\right)\\ &= e^{-2\lambda|t_2-t_1|}\,. \end{aligned} \tag{9.38}$$

In other words, the process is exponentially correlated and WSS.

9.6 ERGODICITY

The concept of ergodicity is sophisticated and subtle, but the essential idea is described here. We typically observe the outcome of a random process (e.g., we record a noise waveform) and want to characterize the statistics of the random process by measurements on one ensemble member. For instance, we could consider the time-average of the waveform to represent the mean value of the process (assuming this

mean is constant for all time). We could also construct histograms that represent the fraction of time (rather than the probability-weighted fraction of the ensemble) that the waveform lies in different amplitude bins, and this could be taken to reflect the probability density across the ensemble of the value obtained at a particular sampling time. If the random process is such that the behavior of almost every particular realization over time is representative of the behavior down the ensemble, then the process is called ergodic.

A simple example of a process that is not ergodic is Example 9.2, an ensemble of batteries. Clearly, for this example, the behavior of any realization is not representative of the behavior down the ensemble.

Narrower notions of ergodicity may be defined. For example, if the time average

$$\langle x \rangle = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} x(t)\, dt \tag{9.39}$$

almost always (i.e. for almost every realization or outcome) equals the ensemble average μ_X, then the process is termed *ergodic in the mean*. It can be shown, for instance, that a WSS process with finite variance at each instant and with a covariance function that approaches 0 for large lags is ergodic in the mean. Note that a (nonstationary) process with time-varying mean cannot be ergodic in the mean.

In our discussion of random processes, we will primarily be concerned with first- and second-order moments of random processes. While it is extremely difficult to determine in general whether a random process is ergodic, there are criteria (specified in terms of the moments of the process) that will establish ergodicity in the mean and in the autocorrelation. Frequently, however, such ergodicity is simply assumed for convenience, in the absence of evidence that the assumption is not reasonable. Under this assumption, the mean and autocorrelation can be obtained from time-averaging on a single ensemble member, through the following equalities:

$$E\{x(t)\} = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} x(t) dt \tag{9.40}$$

and

$$E\{x(t)x(t+\tau)\} = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} x(t)x(t+\tau) dt \tag{9.41}$$

A random process for which (9.40) and (9.41) are true is referred as second-order ergodic.

9.7 LINEAR ESTIMATION OF RANDOM PROCESSES

A common class of problems in a variety of aspects of communication, control and signal processing involves the estimation of one random process from observations

of another, or estimating (predicting) future values from the observation of past values. For example, it is common in communication systems that the signal at the receiver is a corrupted (e.g., noisy) version of the transmitted signal, and we would like to estimate the transmitted signal from the received signal. Other examples lie in predicting weather and financial data from past observations. We will be treating this general topic in much more detail in later chapters, but a first look at it here can be beneficial in understanding random processes.

We shall first consider a simple example of linear prediction of a random process, then a more elaborate example of linear FIR filtering of a noise-corrupted process to estimate the underlying random signal. We conclude the section with some further discussion of the basic problem of linear estimation of one random variable from measurements of another.

9.7.1 Linear Prediction

As a simple illustration of linear prediction, consider a discrete-time process $x[n]$. Knowing the value at time n_0 we may wish to predict what the value will be m samples into the future, i.e. at time $n_0 + m$. We limit the prediction strategy to a linear one, i.e., with $\hat{x}[n_0 + m]$ denoting the predicted value, we restrict $\hat{x}[n_0 + m]$ to be of the form

$$\hat{x}[n_0 + m] = ax[n_0] + b \tag{9.42}$$

and choose the prediction parameters a and b to minimize the expected value of the square of the error, i.e., choose a and b to minimize

$$\epsilon = E\{(x[n_0 + m] - \hat{x}[n_0 + m])^2\} \tag{9.43}$$

or

$$\epsilon = E\{(x[n_0 + m] - ax[n_0] - b)^2\}. \tag{9.44}$$

To minimize ϵ we set to zero its partial derivative with respect to each of the two parameters and solve for the parameter values. The resulting equations are

$$E\{(x[n_0 + m] - ax[n_0] - b)x[n_0]\} = E\{(x[n_0 + m] - \widehat{x}[n_0 + m])x[n_0]\} = 0 \tag{9.45a}$$

$$E\{x[n_0 + m] - ax[n_0] - b\} = E\{x[n_0 + m] - \widehat{x}[n_0 + m]\} = 0\,. \tag{9.45b}$$

Equation (9.45a) states that the error $x[n_0 + m] - \widehat{x}[n_0 + m]$ associated with the optimal estimate is orthogonal to the available data $x[n_0]$. Equation (9.45b) states that the estimate is unbiased.

Carrying out the multiplications and expectations in the preceding equations results in the following equations, which can be solved for the desired constants.

$$R_{xx}[n_0 + m, n_0] - aR_{xx}[n_0, n_0] - b\mu_x[n_0] = 0 \tag{9.46a}$$

$$\mu_x[n_0 + m] - a\mu_x[n_0] - b = 0. \tag{9.46b}$$

If we assume that the process is WSS so that $R_{xx}[n_0+m, n_0] = R_{xx}[m]$, $R_{xx}[n_0, n_0] = R_{xx}[0]$, and also assume that it is zero mean, ($\mu_x = 0$), then equations (9.46) reduce to

$$a = R_{xx}[m]/R_{xx}[0] \tag{9.47}$$

$$b = 0 \tag{9.48}$$

so that

$$\widehat{x}[n_0 + m] = \frac{R_{xx}[m]}{R_{xx}[0]} x[n_0]. \tag{9.49}$$

If the process is not zero mean, then it is easy to see that

$$\widehat{x}[n_0 + m] = \mu_x + \frac{C_{xx}[m]}{C_{xx}[0]} (x[n_0] - \mu_x) \ . \tag{9.50}$$

An extension of this problem would consider how to do prediction when measurements of several past values are available. Rather than pursue this case, we illustrate next what to do with several measurements in a slightly different setting.

9.7.2 Linear FIR Filtering

As another example, which we will treat in more generality in chapter 11 on Wiener filtering, consider a discrete-time signal $s[n]$ that has been corrupted by additive noise $d[n]$. For example, $s[n]$ might be a signal transmitted over a channel and $d[n]$ the noise introduced by the channel. The received signal $r[n]$ is then

$$r[n] = s[n] + d[n]. \tag{9.51}$$

Assume that both $s[n]$ and $d[n]$ are zero-mean random processes and are uncorrelated. At the receiver we would like to process $r[n]$ with a causal FIR (finite impulse response) filter to estimate the transmitted signal $s[n]$.

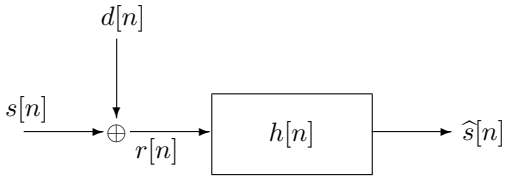

FIGURE 9.5 Estimating the noise corrupted signal.

If $h[n]$ is a causal FIR filter of length L, then

$$\widehat{s}[n] = \sum_{k=0}^{L-1} h[k] r[n-k]. \tag{9.52}$$

We would like to determine the filter coefficients $h[k]$ to minimize the mean square error between $\widehat{s}[n]$ and $s[n]$, i.e., minimize ϵ given by

$$\begin{aligned}\epsilon &= E(s[n]-\widehat{s}[n])^2 \\ &= E(s[n]-\sum_{k=0}^{L-1} h[k]r[n-k])^2. \end{aligned} \tag{9.53}$$

To determine h, we set $\frac{\partial \epsilon}{\partial h[m]} = 0$ for each of the L values of m. Taking this derivative, we get

$$\begin{aligned}\frac{\partial \epsilon}{\partial h[m]} &= -E\{2(s[n]-\sum_{k} h[k]r[n-k])r[n-m]\} \\ &= -E\{2(s[n]-\widehat{s}[n])r[n-m]\} \\ &= 0 \qquad\qquad m=0,1,\cdots,L-1 \end{aligned} \tag{9.54}$$

which is the orthogonality condition we should be expecting: the error $(s[n]-\widehat{s}[n])$ associated with the optimal estimate is orthogonal to the available data, $r[n-m]$.

Carrying out the multiplications in the above equations and taking expectations results in

$$\sum_{k=0}^{L-1} h[k]R_{rr}[m-k] = R_{sr}[m]\ , \quad m=0,1,\cdots,L-1 \tag{9.55}$$

Eqns. (9.55) constitute L equations that can be solved for the L parameters $h[k]$. With $r[n] = s[n] + d[n]$, it is straightforward to show that $R_{sr}[m] = R_{ss}[m] + R_{sd}[m]$ and since we assumed that $s[n]$ and $d[n]$ are uncorrelated, then $R_{sd}[m] = 0$. Similarly, $R_{rr}[m] = R_{ss}[m] + R_{dd}[m]$.

These results are also easily modified for the case where the processes no longer have zero mean.

9.8 THE EFFECT OF LTI SYSTEMS ON WSS PROCESSES

Your prior background in signals and systems, and in the earlier chapters of these notes, has characterized how LTI systems affect the input for deterministic signals.

We will see in later chapters how the correlation properties of a random process, and the effects of LTI systems on these properties, play an important role in understanding and designing systems for such tasks as filtering, signal detection, signal estimation and system identification. We focus in this section on understanding in the time domain how LTI systems shape the correlation properties of a random process. In Chapter 10 we develop a parallel picture in the frequency domain, after establishing that the frequency distribution of the expected power in a random signal is described by the Fourier transform of the autocorrelation function.

Consider an LTI system whose input is a sample function of a WSS random process $x(t)$, i.e., a signal chosen by a probabilistic experiment from the ensemble that constitutes the random process $x(t)$; more simply, we say that the input is the random

process $x(t)$. The WSS input is characterized by its mean and its autocovariance or (equivalently) autocorrelation function.

Among other considerations, we are interested in knowing when the output process $y(t)$ — i.e., the ensemble of signals obtained as responses to the signals in the input ensemble — will itself be WSS, and want to determine its mean and autocovariance or autocorrelation functions, as well as its cross-correlation with the input process. For an LTI system whose impulse response is $h(t)$, the output $y(t)$ is given by the convolution

$$y(t) = \int_{-\infty}^{+\infty} h(v)x(t-v)dv = \int_{-\infty}^{+\infty} x(v)h(t-v)dv \tag{9.56}$$

for any specific input $x(t)$ for which the convolution is well-defined. The convolution is well-defined if, for instance, the input $x(t)$ is bounded and the system is bounded-input bounded-output (BIBO) stable, i.e. $h(t)$ is absolutely integrable. Figure 9.6 indicates what the two components of the integrand in the convolution integral may look like.

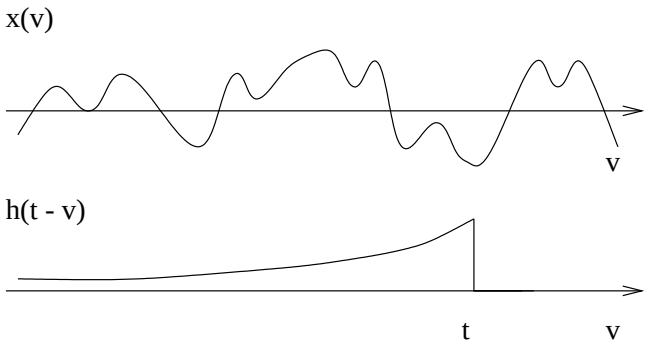

FIGURE 9.6 Illustration of the two terms in the integrand of Eqn. (9.56)

Rather than requiring that every sample function of our input process be bounded, it will suffice for our convolution computations below to assume that $E[x^2(t)] = R_{xx}(0)$ is finite. With this assumption, and also assuming that the system is BIBO stable, we ensure that $y(t)$ is a well-defined random process, and that the formal manipulations we carry out below — for instance, interchanging expectation and convolution — can all be justified more rigorously by methods that are beyond our scope here. In fact, the results we obtain can also be applied, when properly interpreted, to cases where the input process does not have a bounded second moment, e.g., when $x(t)$ is so-called CT white noise, for which $R_{xx}(\tau) = \delta(\tau)$. The results can also be applied to a system that is not BIBO stable, as long as it has a well-defined frequency response $H(j\omega)$, as in the case of an ideal lowpass filter, for example.

We can use the convolution relationship (9.56) to deduce the first- and second-order properties of $y(t)$. What we shall establish is that $y(t)$ is itself WSS, and that

$x(t)$ and $y(t)$ are in fact jointly WSS. We will also develop relationships for the autocorrelation of the output and the cross-correlation between input and output.

First, consider the mean value of the output. Taking the expected value of both sides of (9.56), we find

$$\begin{aligned} E[y(t)] &= E\left\{\int_{-\infty}^{+\infty} h(v)x(t-v)\,dv\right\} \\ &= \int_{-\infty}^{+\infty} h(v)E[x(t-v)]\,dv \\ &= \int_{-\infty}^{+\infty} h(v)\mu_x\,dv \\ &= \mu_x \int_{-\infty}^{+\infty} h(v)\,dv \\ &= H(j0)\,\mu_x = \mu_y\ . \end{aligned} \tag{9.57}$$

In other words, the mean of the output process is *constant*, and equals the mean of the input scaled by the the DC gain of the system. This is also what the response of the system would be if its input were held constant at the value μ_x.

The preceding result and the linearity of the system also allow us to conclude that applying the *zero-mean* WSS process $x(t)-\mu_x$ to the input of the stable LTI system would result in the *zero-mean* process $y(t)-\mu_y$ at the output. This fact will be useful below in converting results that are derived for correlation functions into results that hold for covariance functions.

Next consider the cross-correlation between output and input:

$$\begin{aligned} E\{y(t+\tau)x(t)\} &= E\left\{\left[\int_{-\infty}^{+\infty} h(v)x(t+\tau-v)dv\right]x(t)\right\} \\ &= \int_{-\infty}^{+\infty} h(v)E\{x(t+\tau-v)x(t)\}dv\ . \end{aligned} \tag{9.58}$$

Since $x(t)$ is WSS, $E\{x(t+\tau-v)x(t)\} = R_{xx}(\tau-v)$, so

$$\begin{aligned} E\{y(t+\tau)x(t)\} &= \int_{-\infty}^{+\infty} h(v)R_{xx}(\tau-v)dv \\ &= h(\tau) * R_{xx}(\tau) \\ &= R_{yx}(\tau)\ . \end{aligned} \tag{9.59}$$

Note that the cross-correlation depends only on the lag τ between the sampling instants of the output and input processes, not on both τ and the absolute time location t. Also, this cross-correlation between the output and input is deterministically related to the autocorrelation of the input, and can be viewed as the signal that would result if the system input were the autocorrelation function, as indicated in Figure 9.7.

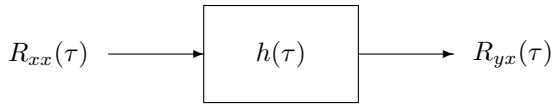

FIGURE 9.7 Representation of Eqn. (9.59)

We can also conclude that

$$R_{xy}(\tau) = R_{yx}(-\tau) = R_{xx}(-\tau) * h(-\tau) = R_{xx}(\tau) * h(-\tau) \ , \tag{9.60}$$

where the second equality follows from Eqn. (9.59) and the fact that time-reversing the two functions in a convolution results in time-reversal of the result, while the last equality follows from the symmetry Eqn. (9.13) of the autocorrelation function.

The above relations can also be expressed in terms of covariance functions, rather than in terms of correlation functions. For this, simply consider the case where the input to the system is the zero-mean WSS process $x(t) - \mu_x$, with corresponding zero-mean output $y(t) - \mu_y$. Since the correlation function for $x(t) - \mu_x$ is the same as the covariance function for $x(t)$, i.e., since

$$R_{x-\mu_x, x-\mu_x}(\tau) = C_{xx}(\tau) \ , \tag{9.61}$$

the results above hold unchanged when every correlation function is replaced by the corresponding covariance function. We therefore have, for instance, that

$$C_{yx}(\tau) = h(\tau) * C_{xx}(\tau) \tag{9.62}$$

Next we consider the autocorrelation of the output $y(t)$:

$$E\{y(t+\tau)y(t)\} = E\left\{ \left[\int_{-\infty}^{+\infty} h(v)x(t+\tau-v)dv \right] y(t) \right\}$$

$$= \int_{-\infty}^{+\infty} h(v) \underbrace{E\{x(t+\tau-v)y(t)\}}_{R_{xy}(\tau-v)} dv$$

$$\begin{aligned} &= \int_{-\infty}^{+\infty} h(v) R_{xy}(\tau - v) dv \\ &= h(\tau) * R_{xy}(\tau) \\ &= R_{yy}(\tau) \ . \end{aligned} \tag{9.63}$$

Note that the autocorrelation of the output depends only on τ, and not on both τ and t. Putting this together with the earlier results, we conclude that $x(t)$ and $y(t)$ are jointly WSS, as claimed.

The corresponding result for covariances is

$$C_{yy}(\tau) = h(\tau) * C_{xy}(\tau) \; . \tag{9.64}$$

Combining (9.63) with (9.60), we find that

$$R_{yy}(\tau) = R_{xx}(\tau) * \underbrace{h(\tau) * h(-\tau)}_{h(\tau)*h(-\tau) \stackrel{\triangle}{=} \overline{R}_{hh}(\tau)} = R_{xx}(\tau) * \overline{R}_{hh}(\tau) \; . \tag{9.65}$$

The function $\overline{R}_{hh}(\tau)$ is typically referred to as the **deterministic autocorrelation function** of $h(t)$, and is given by

$$\overline{R}_{hh}(\tau) = h(\tau) * h(-\tau) = \int_{-\infty}^{+\infty} h(t+\tau)h(t)dt \; . \tag{9.66}$$

For the covariance function version of (9.65), we have

$$C_{yy}(\tau) = C_{xx}(\tau) * \underbrace{h(\tau) * h(-\tau)}_{h(\tau)*h(-\tau) \stackrel{\triangle}{=} \overline{R}_{hh}(\tau)} = C_{xx}(\tau) * \overline{R}_{hh}(\tau) \; . \tag{9.67}$$

Note that the deterministic correlation function of $h(t)$ is still what we use, even when relating the covariances of the input and output. Only the means of the input and output processes get adjusted in arriving at the present result; the impulse response is untouched.

The correlation relations in Eqns. (9.59), (9.60), (9.63) and (9.65), as well as their covariance counterparts, are very powerful, and we will make considerable use of them. Of equal importance are their statements in the Fourier and Laplace transform domains. Denoting the Fourier and Laplace transforms of the correlation function $R_{xx}(\tau)$ by $S_{xx}(j\omega)$ and $S_{xx}(s)$ respectively, and similarly for the other correlation functions of interest, we have:

$$\begin{aligned} S_{yx}(j\omega) &= S_{xx}(j\omega)H(j\omega), & S_{yy}(j\omega) &= S_{xx}(j\omega)|H(j\omega)|^2, \\ S_{yx}(s) &= S_{xx}(s)H(s), & S_{yy}(s) &= S_{xx}(s)H(s)H(-s) \; . \end{aligned} \tag{9.68}$$

We can denote the Fourier and Laplace transforms of the covariance function $C_{xx}(\tau)$ by $D_{xx}(j\omega)$ and $D_{xx}(s)$ respectively, and similarly for the other covariance functions of interest, and then write the same sorts of relationships as above.

Exactly parallel results hold in the DT case. Consider a stable discrete-time LTI system whose impulse response is $h[n]$ and whose input is the WSS random process $x[n]$. Then, as in the continuous-time case, we can conclude that the output process $y[n]$ is *jointly WSS* with the input process $x[n]$, and

$$\mu_y = \mu_x \sum_{-\infty}^{\infty} h[n] \tag{9.69}$$

$$R_{yx}[m] = h[m] * R_{xx}[m] \tag{9.70}$$

$$R_{yy}[m] = R_{xx}[m] * \overline{R}_{hh}[m] \; , \tag{9.71}$$

where $\overline{R}_{hh}[m]$ is the deterministic autocorrelation function of $h[m]$, defined as

$$\overline{R}_{hh}[m] = \sum_{n=-\infty}^{+\infty} h[n+m]h[n] \; . \tag{9.72}$$

The corresponding Fourier and $\mathcal{Z}$-transform statements of these relationships are:

$$\begin{array}{lll} \mu_y = H(e^{j0})\mu_x \; , & S_{yx}(e^{j\Omega}) = S_{xx}(e^{j\Omega})H(e^{j\Omega}) \; , & S_{yy}(e^{j\Omega}) = S_{xx}(e^{j\Omega})|H(e^{j\Omega})|^2, \\ \mu_y = H(1)\mu_x \; , & S_{yx}(z) = S_{xx}(z)H(z) \; , & S_{yy}(z) = S_{xx}(z)H(z)H(1/z). \end{array} \tag{9.73}$$

All of these expressions can also be rewritten for covariances and their transforms.

The basic relationships that we have developed so far in this chapter are extremely powerful. In Chapter 10 we will use these relationships to show that the Fourier transform of the autocorrelation function describes how the expected power of a WSS process is distributed in frequency. For this reason, the Fourier transform of the autocorrelation function is termed the **power spectral density (PSD)** of the process.

The relationships developed in this chapter are also very important in using random processes to measure or identify the impulse response of an LTI system. For example, from (9.70), if the input $x[n]$ to a DT LTI system is a WSS random process with autocorrelation function $R_{xx}[m] = \delta[m]$, then by measuring the cross-correlation between the input and output we obtain a measurement of the system impulse response. It is easy to construct an input process with autocorrelation function $\delta[m]$, for example an i.i.d. process that is equally likely to take the values $+1$ and -1 at each time instant.

As another example, suppose the input $x(t)$ to a CT LTI system is a random telegraph wave, with changes in sign at times that correspond to the arrivals in a Poisson process with rate λ, i.e.,

$$\mathrm{P}(k \text{ switches in an interval of length } T) = \frac{(\lambda T)^k e^{-\lambda T}}{k!} \; . \tag{9.74}$$

Then, assuming $x(0)$ takes the values ± 1 with equal probabilities, we can determine that the process $x(t)$ has zero mean and correlation function $R_{xx}(\tau) = e^{-2\lambda|\tau|}$, so it is WSS (for $t \geq 0$). If we determine the cross-correlation $R_{yx}(\tau)$ with the output $y(t)$ and then use the relation

$$R_{yx}(\tau) = R_{xx}(\tau) * h(\tau) \; , \tag{9.75}$$

we can obtain the system impulse response $h(\tau)$. For example, if $S_{yx}(s), S_{xx}(s)$ and $H(s)$ denote the associated Laplace transforms, then

$$H(s) = \frac{S_{yx}(s)}{S_{xx}(s)} \; . \tag{9.76}$$

Note that $S_{xx}(s)$ is a rather well-behaved function of the complex variable s in this case, whereas any particular sample function of the process $x(t)$ would not have such a well-behaved transform. The same comment applies to $S_{yx}(s)$.

As a third example, suppose that we know the autocorrelation function $R_{xx}[m]$ of the input $x[n]$ to a DT LTI system, but do not have access to $x[n]$ and therefore cannot determine the cross-correlation $R_{yx}[m]$ with the output $y[n]$, but *can* determine the output autocorrelation $R_{yy}[m]$. For example, if

$$R_{xx}[m] = \delta[m] \tag{9.77}$$

and we determine $R_{yy}[m]$ to be $R_{yy}[m] = \left(\frac{1}{2}\right)^{|m|}$, then

$$R_{yy}[m] = \left(\frac{1}{2}\right)^{|m|} = \overline{R}_{hh}[m] = h[m] * h[-m]. \tag{9.78}$$

Equivalently, $H(z)H(z^{-1})$ can be obtained from the $\mathcal{Z}$-transform $S_{yy}(z)$ of $R_{yy}[m]$. Additional assumptions or constraints, for instance on the stability and causality of the system and its inverse, may allow one to recover $H(z)$ from knowledge of $H(z)H(z^{-1})$.

6.011 Introduction to Communication, Control, and Signal Processing
Spring 2010