

Optical Propagation, Detection, and Communication

Jeffrey H. Shapiro
Massachusetts Institute of Technology

©1988,2000

Chapter 4

Random Processes

In this chapter, we make the leap from N joint random variables—a random vector—to an infinite collection of joint random variables—a random waveform. Random process¹ theory is the branch of mathematics that deals with such entities. This theory is useful for modeling real-world situations which possess the following characteristics.

- The three attributes, listed in Chapter 3, for useful application of probabilistic models are present.
- The experimental outcomes are waveforms.

The shot noise and thermal noise currents discussed in our photodetector phenomenology are, of course, the principal candidates for random process modeling in this book. Random process theory is *not* an area with which the reader is assumed to have significant prior familiarity. Yet, even though this field is rich in new concepts, we shall hew to the straight and narrow, limiting our development to the material that is fundamental to succeeding chapters—first and second moment theory, and Gaussian random processes. We begin with some basic definitions.

4.1 Basic Concepts

Consider a real-world experiment, suitable for probabilistic analysis, whose outcomes are waveforms. Let $\mathcal{P} = \{\Omega, \text{Pr}(\cdot)\}$ be a probability-space model for this experiment, and let $\{x(t, \omega) : \omega \in \Omega\}$ be an assignment of *deterministic* waveforms—functions of t —to the sample points $\{\omega\}$, as sketched in Fig. 4.1. This probabilistic construct creates a *random process*, $x(t, \cdot)$, on the probabil-

¹The term *stochastic* process is also used.

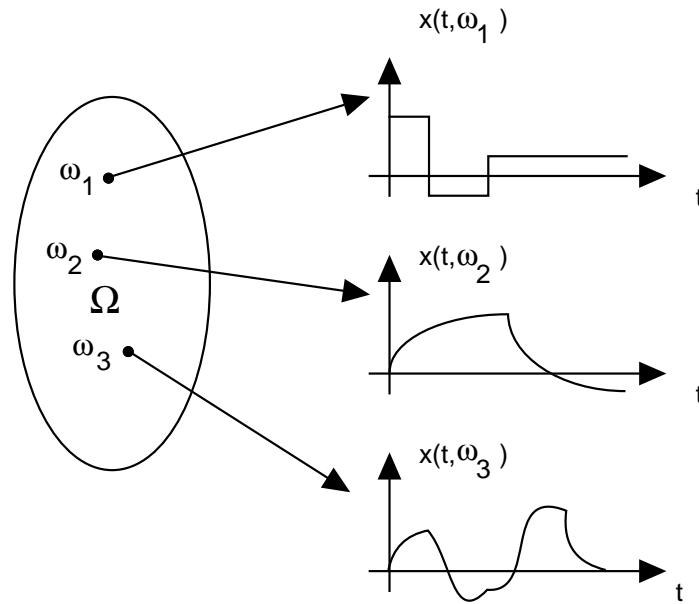


Figure 4.1: Assignment of waveforms to sample points in a probability space

ity space \mathcal{P} , i.e., because of the uncertainty as to which ω will occur when the experiment modeled by \mathcal{P} is performed, there is uncertainty as to which waveform will be produced.

We will soon abandon the full probability-space notation for random processes, just as we quickly did in Chapter 3 for the corresponding case of random variables. Before doing so, however, let us hammer home the preceding definition of a random process by examining some limiting cases of $x(t, \omega)$.

random process With t and ω both regarded as variables, i.e., $-\infty < t < \infty$ and $\omega \in \Omega$, then $x(t, \omega)$ refers to the random process.

sample function With t variable and $\omega = \omega_1$ fixed, then $x(t, \omega_1)$ is a deterministic function of t —the *sample function* of the random process, $x(t, \omega)$, associated with the sample point ω_1 .

sample variable With $t = t_1$ fixed and ω variable, then $x(t_1, \omega)$ is a deterministic mapping from the sample space, Ω , to the real line, \mathcal{R}^1 . It is thus a random variable—the *sample variable* of the random process, $x(t, \omega)$, associated with the time² instant t_1 .

²Strictly speaking, a random process is a collection of joint random variables indexed by an index parameter. Throughout this chapter, we shall use t to denote the index parameter,

sample value With $t = t_1$ and $\omega = \omega_1$ both fixed, then $x(t_1, \omega_1)$ is a number.

This number has two interpretations: it is the time sample at t_1 of the sample function $x(t, \omega_1)$; and it is also the sample value at ω_1 of the random variable $x(t_1, \omega)$.

For the most part, we shall no longer carry along the sample space notation. We shall use $x(t)$ to denote a generic random process, and $x(t_1)$ to refer to the random variable obtained by sampling this process at $t = t_1$. However, when we are sketching typical sample functions of our random-process examples, we shall label such plots $x(t, \omega_1)$ vs. t , etc., to emphasize that they represent the deterministic waveforms associated with specific sample points in some underlying Ω .

If one time sample of a random process, $x(t_1)$, is a random variable, then two such time samples, $x(t_1)$ and $x(t_2)$, must be two joint random variables, and N time samples, $\{x(t_n) : 1 \leq n \leq N\}$, must be N joint random variables, i.e., a random vector

$$\mathbf{x} \equiv \begin{bmatrix} x(t_1) \\ x(t_2) \\ \vdots \\ x(t_N) \end{bmatrix}. \quad (4.1)$$

A *complete statistical characterization* of a random process $x(t)$ is defined to be the information sufficient to deduce the probability density for any random vector, \mathbf{x} , obtained via sampling, as in Eq. 4.1. This must be true for all choices of the sampling times, $\{t_n : 1 \leq n \leq N\}$, and for all dimensionalities, $1 \leq N < \infty$. It is not necessary that this characterization comprise an explicit catalog of densities, $\{p_{\mathbf{x}}(\mathbf{X})\}$, for all choices and dimensionalities of the sample-time vector

$$\mathbf{t} \equiv \begin{bmatrix} t_1 \\ t_2 \\ \vdots \\ t_N \end{bmatrix}. \quad (4.2)$$

Instead, the characterization may be given implicitly, as the following two examples demonstrate.

single-frequency wave Let θ be a random variable that is uniformly distributed on the interval $0 \leq \theta \leq 2\pi$, and let P and f_0 be positive

and call it *time*. Later, we will have occasion to deal with random processes with multidimensional index parameters, e.g., a 2-D spatial vector in the entrance pupil of an optical system.

constants. The single-frequency wave, $x(t)$, is then

$$x(t) \equiv \sqrt{2P} \cos(2\pi f_0 t + \theta). \quad (4.3)$$

Gaussian random process A random process, $x(t)$, is a *Gaussian* random process if, for all \mathbf{t} and N , the random vector, \mathbf{x} , obtained by sampling this process is Gaussian. The statistics of a Gaussian random process are completely characterized³ by knowledge of its *mean function*

$$m_x(t) \equiv E[x(t)], \quad \text{for } -\infty < t < \infty, \quad (4.4)$$

and its *covariance function*

$$K_{xx}(t, s) \equiv E[\Delta x(t)\Delta x(s)], \quad \text{for } -\infty < t, s < \infty, \quad (4.5)$$

where $\Delta x(t) \equiv x(t) - m_x(t)$.

We have sketched a typical sample function of the single-frequency wave in Fig. 4.2. It is a pure tone of amplitude $\sqrt{2P}$, frequency f_0 , and phase $\theta(\omega_1)$. This certainly does not look like a random process—it is not *noisy*. Yet, Eq. 4.3 *does* generate a random process, according to our definition. Let $\mathcal{P} = \{\Omega, \Pr(\cdot)\}$ be the probability space that underlies the random variable θ . Then, Eq. 4.3 implies the deterministic sample-point-to-sample-function mapping

$$x(t, \omega) = \sqrt{2P} \cos[2\pi f_0 t + \theta(\omega)], \quad \text{for } \omega \in \Omega, \quad (4.6)$$

which, with the addition of the probability measure $\Pr(\cdot)$, makes $x(t)$ a random process. Physically, there is only *one* random variable in this random process—the phase of the wave.⁴ Thus, this random process is rather trivial, although it may be used to model the output of an ideal oscillator whose amplitude and frequency are known, but whose phase, with respect to an observer's clock, is completely random.

The Gaussian random process example is much more in keeping with our intuition about noise. For example, in Fig. 4.3 we have sketched a typical

³All time-sample vectors from a Gaussian random process are Gaussian. To find their probability densities we need only supply their mean vectors and their covariance matrices. These can be found from the mean function and covariance function—the continuous-time analogs of the mean vector and covariance matrix—as will be seen below.

⁴As a result, it is a straightforward—but tedious—task to go from the definition of the single-frequency wave to an explicit collection of sample-vector densities. The calculations for $N = 1$ and $N = 2$ will be performed in the home problems for this chapter.

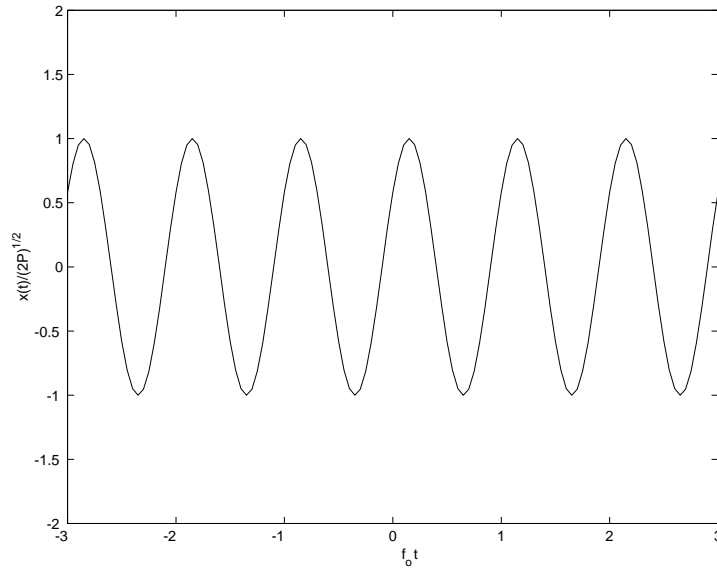


Figure 4.2: Typical sample function of the single-frequency wave

sample function for the Gaussian random process, $x(t)$, whose mean function is

$$m_x(t) = 0, \quad \text{for } -\infty < t < \infty, \quad (4.7)$$

and whose covariance function is

$$K_{xx}(t, s) = P \exp(-\lambda|t - s|), \quad \text{for } -\infty < t, s < \infty, \quad (4.8)$$

where P and λ are positive constants.

Some justification for Fig. 4.3 can be provided from our Chapter 3 knowledge of Gaussian random vectors. For the Gaussian random process whose mean function and covariance function are given by Eqs. 4.7 and 4.8, the probability density for a single time sample, $x(t_1)$, will be Gaussian, with $E[x(t_1)] = m_x(t_1) = 0$, and $\text{var}[x(t_1)] = K_{xx}(t_1, t_1) = P$. Thus, as seen in Fig. 4.3, this time sample will typically fall within a few \sqrt{P} of 0, even though there is some probability that values approaching $\pm\infty$ will occur.

To justify the *dynamics* of the Fig. 4.3 sample function, we need—at the least—to consider the jointly Gaussian probability density for two time samples, viz. $x(t_1)$ and $x(t_2)$. Equivalently, we can suppose that $x(t_1) = X_1$ has occurred, and examine the conditional statistics for $x(t_2)$. We know that this conditional density will be Gaussian, because $x(t_1)$ and $x(t_2)$ are jointly Gaussian; the conditional mean and conditional variance are as follows [cf.

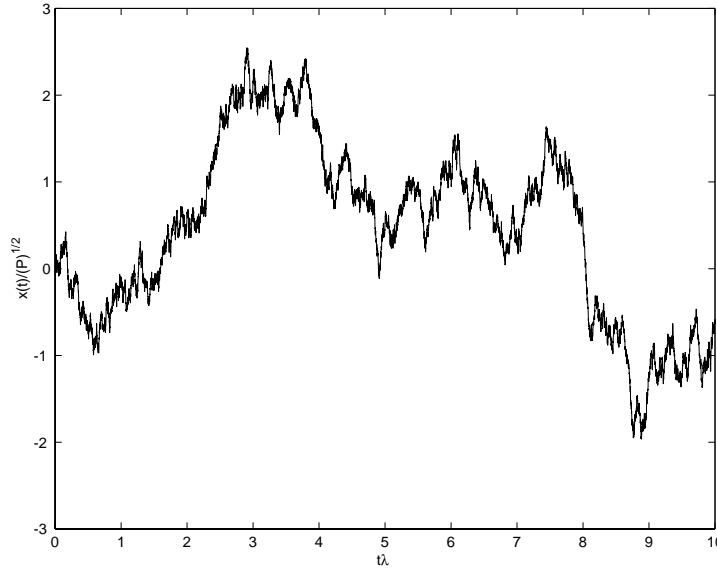


Figure 4.3: Typical sample function for a Gaussian random process with mean function Eq. 4.7 and covariance function Eq. 4.8

Eqs. 3.87 and 3.88]:

$$\begin{aligned} E[x(t_2) | x(t_1) = X_1] &= m_x(t_2) + \frac{K_{xx}(t_2, t_1)}{K_{xx}(t_1, t_1)}[X_1 - m_x(t_1)] \\ &= \exp(-\lambda|t_2 - t_1|)X_1, \end{aligned} \quad (4.9)$$

and

$$\begin{aligned} \text{var}[x(t_2) | x(t_1) = X_1] &= K_{xx}(t_2, t_2) - \frac{K_{xx}(t_2, t_1)^2}{K_{xx}(t_1, t_1)} \\ &= P[1 - \exp(-2\lambda|t_2 - t_1|)]. \end{aligned} \quad (4.10)$$

Equations 4.9 and 4.10 support the waveform behavior shown in Fig. 4.3. Recall that exponents must be dimensionless. Thus if t is time, in units of seconds, then $1/\lambda$ must have these units too. For $|t_2 - t_1| \ll 1/\lambda$, we see that the conditional mean of $x(t_2)$, given $x(t_1) = X_1$ has occurred, is very close to X_1 . Moreover, under this condition, the conditional variance of $x(t_2)$ is much less than its a priori variance. Physically, this means that the process cannot have changed much over the time interval from t_1 to t_2 . Conversely, when $|t_2 - t_1| \gg 1/\lambda$ prevails, we find that the conditional mean and the conditional variance of $x(t_2)$, given $x(t_1) = X_1$ has occurred, are very nearly equal to the unconditional values, i.e., $x(t_2)$ and $x(t_1)$ are approximately statistically

independent for such long time separations. In summary, $1/\lambda$ is a *correlation* time for this process in that $x(t)$ can't change much on time scales appreciably shorter than $1/\lambda$.

More will be said about Gaussian random processes in Section 4.4. We will first turn our attention to an extended treatment of mean functions and covariance functions for arbitrary random processes.

4.2 Second-Order Characterization

Complete statistical characterization of a random process is like knowing the probability density of a single random variable—all meaningful probabilities and expectations can be calculated, but this full information may not really be needed. A substantial, but incomplete, description of a single random variable is knowledge of its mean value and its variance. For a random vector, the corresponding first and second moment partial characterization consists of the mean vector and the covariance matrix. In this section, we shall develop the second-order characterization of a random process, i.e., we shall consider the information provided by the mean function and the covariance function of a random process. As was the case for random variables, this information is incomplete—there are a wide variety of random processes, with wildly different sample functions, that share a common second-order characterization.⁵ Nevertheless, second-order characterizations are extraordinarily useful, because they are relatively simple, and they suffice for linear-filtering/signal-to-noise-ratio calculations.

Suppose $x(t)$ is a random process—not necessarily Gaussian—with mean function $m_x(t)$ and covariance function $K_{xx}(t, s)$, as defined by Eqs. 4.4 and 4.5, respectively. By direct notational translation of material from Chapter 3 we have the following results.

mean function The mean function, $m_x(t)$, of a random process, $x(t)$, is a deterministic function of time whose value at an arbitrary specific time, $t = t_1$, is the mean value of the random variable $x(t_1)$.

covariance function The covariance function, $K_{xx}(t, s)$, of a random process, $x(t)$, is a deterministic function of two time variables; its value at an arbitrary pair of times, $t = t_1$ and $s = s_1$, is the covariance between the random variables $x(t_1)$ and $x(s_1)$.

⁵One trio of such processes will be developed in the home problems for this chapter.

Thus, $m_x(t)$ is the *deterministic* part of the random process, i.e., $\Delta x(t) \equiv x(t) - m_x(t)$ is a zero-mean random process—the *noise* part of $x(t)$ —which satisfies $x(t) = m_x(t) + \Delta x(t)$ by construction. We also know that $\text{var}[x(t)] = E[\Delta x(t)^2] = K_{xx}(t, t)$ measures the mean-square noise strength in the random process as a function of t . Finally, for $t \neq s$, we have that

$$\rho_{xx}(t, s) \equiv \frac{K_{xx}(t, s)}{\sqrt{K_{xx}(t, t)K_{xx}(s, s)}} \quad (4.11)$$

is the correlation coefficient between samples of the random process taken at times t and s . If $\rho_{xx}(t_1, s_1) = 0$, then the time samples $x(t_1)$ and $x(s_1)$ are uncorrelated, and perhaps statistically independent. If $|\rho_{xx}(t_1, s_1)| = 1$, then these time samples are completely correlated, and $x(t_1)$ can be found, with certainty, from $x(s_1)$ [cf. Eq. 3.69].

All of the above properties and interpretations ascribed to the second-order— $m_x(t)$ and $K_{xx}(t, s)$ —characterization of a random process rely on their probabilistic origins. Random processes, however, combine both probabilistic and waveform notions. Thus, our main thrust in this chapter will be to develop some of the latter properties, in the particular context of linear filtering. Before doing so, we must briefly address some mathematical constraints on covariance functions.

Any real-valued deterministic function of a single parameter, $f(\cdot)$, can in principle be the mean function of a random process. Indeed, if we start with the Gaussian random process $x(t)$ whose mean function and covariance function are given by Eqs. 4.7 and 4.8, respectively, and define⁶

$$y(t) = f(t) + x(t), \quad (4.12)$$

then it is a simple matter—using the linearity of expectation—to show that

$$m_y(t) = f(t), \quad (4.13)$$

and

$$K_{yy}(t, s) = K_{xx}(t, s) = P \exp(-\lambda|t - s|). \quad (4.14)$$

We thus obtain a random process with the desired mean function.

An arbitrary real-valued deterministic function of two parameters, $g(\cdot, \cdot)$, may *not* be a possible covariance function for a random process, because of

⁶Equation 4.12 is a transformation of the original random process $x(t)$ into a new random process $y(t)$; in sample-function terms it says that $y(t, \omega) = f(t) + x(t, \omega)$, for $\omega \in \Omega$. Because these random processes are defined on the same probability space, they are *joint* random processes.

the implicit probabilistic constraints on covariance functions that are listed below.

$$K_{xx}(t, s) = \text{cov}[x(t), x(s)] = K_{xx}(s, t), \quad \text{for all } t, s, \quad (4.15)$$

$$K_{xx}(t, t) = \text{var}[x(t)] \geq 0, \quad \text{for all } t, \quad (4.16)$$

$$|K_{xx}(t, s)| \leq \sqrt{K_{xx}(t, t)K_{xx}(s, s)}, \quad \text{for all } t, s. \quad (4.17)$$

Equations 4.15 and 4.16 are self-evident; Eq. 4.17 is a reprise of correlation coefficients never exceeding one in magnitude.

The preceding covariance function constraints comprise *necessary* conditions that a real-valued deterministic $g(\cdot, \cdot)$ must satisfy for it to be a possible $K_{xx}(t, s)$; they are not *sufficient* conditions. Let $x(t)$ be a random process with covariance function $K_{xx}(t, s)$, let $\{t_1, t_2, \dots, t_N\}$ be an arbitrary collection of sampling times, and let $\{a_1, a_2, \dots, a_N\}$ be an arbitrary collection of real constants, and define a random variable z according to

$$z \equiv \sum_{n=1}^N a_n x(t_n). \quad (4.18)$$

Because $\text{var}(z) \geq 0$ must prevail, we have that⁷

$$\text{var}(z) = \sum_{n=1}^N \sum_{m=1}^N a_n a_m K_{xx}(t_n, t_m) \geq 0, \quad \text{for all } \{t_n\}, \{a_n\}, \text{ and } N. \quad (4.19)$$

Equations 4.16 and 4.17 can be shown to be the $N = 1$ and $N = 2$ special cases, respectively, of Eq. 4.19. Functions which obey Eq. 4.19 are said to be *non-negative definite*. More importantly, any real-valued deterministic function of t and s which is a symmetric function of its arguments and is non-negative definite can be the covariance function of a random process—Eqs. 4.15 and 4.19 are the necessary and sufficient conditions for a valid covariance function.⁸

It can be difficult to check whether or not a given function of two variables is non-negative definite. We shall see, in the next section, that there is an important special case—that of wide-sense stationary processes—for which this verification is relatively simple. Thus, we postpone presentation of some key covariance function examples until we have studied the wide-sense stationary case.

⁷The derivation of this formula parallels that of Eq. 3.107.

⁸The preceding arguments can be cast backwards into the random vector arena—any real-valued deterministic N vector can be the mean vector of a random N vector, any real-valued deterministic $N \times N$ matrix that is symmetric and non-negative definite can be the covariance matrix of a random N vector.

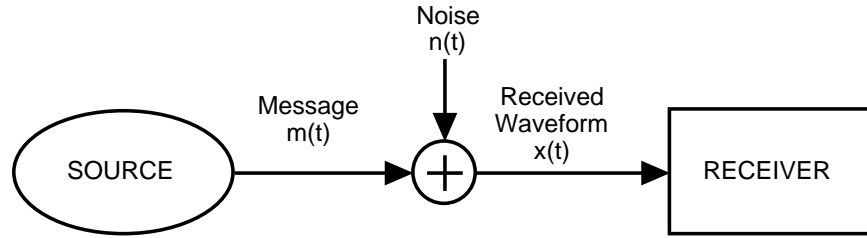


Figure 4.4: A simple communication system

4.3 Linear Filtering of Random Processes

To motivate our treatment of linear filtering of a random process, $x(t)$, let us pursue the signal and noise interpretations assigned to $m_x(t)$ and $\Delta x(t)$ in the context of the highly simplified communication system shown in Fig. 4.4. Here, the source output is a deterministic message waveform, $m(t)$, which is conveyed to the user through an additive-noise channel—the received waveform is $x(t) = m(t) + n(t)$, where $n(t)$ is a zero-mean random process with covariance function $K_{nn}(t, s)$. We have the following second-order characterization for $x(t)$ [cf. Eq. 4.12]

$$m_x(t) = m(t), \quad (4.20)$$

and

$$K_{xx}(t, s) = K_{nn}(t, s). \quad (4.21)$$

Given the structure of Fig. 4.4 and the discussion following Eq. 3.35, it is reasonable to define an instantaneous signal-to-noise ratio, $\text{SNR}(t)$, for this problem by means of

$$\text{SNR}(t) \equiv \frac{m_x(t)^2}{K_{xx}(t, t)} = \frac{m(t)^2}{K_{nn}(t, t)}. \quad (4.22)$$

If $\text{SNR}(t) \gg 1$ prevails for all t , then the Chebyshev inequality guarantees that $x(t)$ will, with high probability, be very close to $m(t)$ at any time.

The above description deals only with the probabilistic structure of the problem. Now, let us add some waveform characteristics. Suppose that the message $m(t)$ is a baseband speech waveform, whose frequency content is from 30 to 3000 Hz, and suppose that the noise $n(t)$ is a broadband thermal noise, with significant frequency content from 10 Hz to 100 MHz. If we find that $\text{SNR}(t) \ll 1$ prevails, the message will be buried in noise. Yet, it is intuitively clear that the received waveform should be narrowband filtered, to pass the signal components and reject the out-of-band noise components of $x(t)$. After

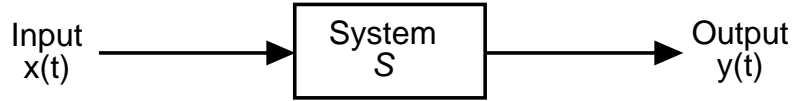


Figure 4.5: Deterministic continuous-time system

such filtering, the signal-to-noise ratio may then obey $\text{SNR}(t) \gg 1$. The random process machinery for analyzing this problem will be developed below, after a brief review of deterministic linear systems.

Deterministic Linear Systems

The reader is expected to have had a basic course in deterministic continuous-time linear systems. Our review of this material will prove immediately useful for linear filtering of random processes. Its translation into functions of 2-D spatial vectors will comprise Chapter 5's linear system—Fourier optics—approach to free-space propagation of quasimonochromatic, paraxial, scalar fields.

Figure 4.3 shows a deterministic continuous-time system, \mathcal{S} , for which application of a deterministic input waveform $x(t)$ produces a deterministic output waveform $y(t)$. Two important properties which \mathcal{S} may possess are as follows.

linearity \mathcal{S} is a linear system if it obeys the *superposition principle*, i.e., if $x_1(t)$ and $x_2(t)$ are arbitrary input waveforms which, when applied to \mathcal{S} , yield corresponding output waveforms $y_1(t)$ and $y_2(t)$, respectively, and if a_1 and a_2 are constants, then applying $a_1x_1(t) + a_2x_2(t)$ to \mathcal{S} results in $a_1y_1(t) + a_2y_2(t)$ as the output.

time invariance \mathcal{S} is a time-invariant system if shifting an input waveform along the time axis—delaying it or retarding it—yields a corresponding output that shifts in an identical manner, i.e.,

$$x_1(t) \xrightarrow{\mathcal{S}} y_1(t) \quad (4.23)$$

implies that

$$x_1(t - T) \xrightarrow{\mathcal{S}} y_1(t - T), \quad (4.24)$$

for arbitrary input waveforms, $x_1(t)$, and all values of the time shift, T .

Linearity and time invariance are *not* tightly coupled properties—a system may be linear or nonlinear, time-invariant or time-varying, in any combination.

However, the confluence of these properties results in an especially important class of systems—the linear time-invariant (LTI) systems. To understand just how special LTI systems are, we need to recall two waveform building-block procedures.

sifting integral Any reasonably-smooth⁹ deterministic waveform, $x(t)$, can be represented as a superposition of impulses via the sifting integral

$$x(t) = \int_{-\infty}^{\infty} x(\tau)\delta(t - \tau) d\tau. \quad (4.25)$$

inverse Fourier transform integral Any reasonably-smooth deterministic waveform, $x(t)$, can be represented as a superposition of complex sinusoids via an inverse Fourier transform integral

$$x(t) = \int_{-\infty}^{\infty} X(f)e^{j2\pi ft} df, \quad (4.26)$$

where

$$X(f) = \int_{-\infty}^{\infty} x(t)e^{-j2\pi ft} dt. \quad (4.27)$$

Suppose \mathcal{S} is an LTI system, with input/output relation $y(t) = \mathcal{S}[x(t)]$. By use of the sifting-integral representation, Eq. 4.25, and the superposition principle, the input/output relation for \mathcal{S} can be reduced to a *superposition integral*

$$y(t) = \int_{-\infty}^{\infty} x(\tau)h(t, \tau) d\tau, \quad (4.28)$$

where $h(t, \tau)$ is the response of \mathcal{S} at time t to a unit-area impulse at time τ . Because of time invariance, $h(t, \tau)$ must equal $h(t - \tau) \equiv h(t - \tau, 0)$, reducing Eq. 4.28 to a *convolution integral*

$$y(t) = \int_{-\infty}^{\infty} x(\tau)h(t - \tau) d\tau. \quad (4.29)$$

Physically, $h(t)$ is the response of the LTI system \mathcal{S} to a unit-area impulsive input at time $t = 0$. Naturally, $h(t)$ is called the *impulse response* of the system.

By Fourier transformation of Eq. 4.29 we can obtain the following alternative LTI-system input/output relation

$$Y(f) = X(f)H(f), \quad (4.30)$$

⁹Reasonably-smooth is a code phrase indicating that we will not be concerned with the rigorous limits of validity of these procedures.

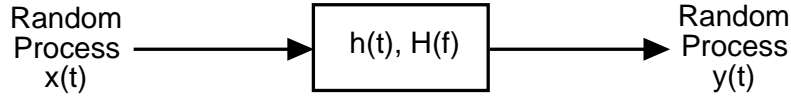


Figure 4.6: LTI system with a random process input

where $Y(f)$ and $H(f)$ are obtained from $y(t)$ and $h(t)$ via equations similar to Eq. 4.27. The fact that Fourier transformation changes convolution into multiplication is an important calculational technique to be cognizant of. Physically, it is more important to understand Eq. 4.30 from the inverse-Fourier-transform approach to signal representation. Specifically, Eqs. 4.26 and 4.30 imply that sinusoids are *eigenfunctions* of LTI systems, i.e., if $A \cos(2\pi ft + \phi)$ is the input to an LTI system, the corresponding output will be $|H(f)|A \cos(2\pi ft + \phi + \arg[H(f)])$. In words, the response of an LTI system to a sinusoid of frequency f is also a sinusoid of frequency f ; the system merely changes the amplitude of the sinusoid by $|H(f)|$ and shifts its phase by $\arg[H(f)]$. $H(f)$ is called the *frequency response* of the system.

Mean and Covariance Propagation

Now let us examine what happens when we apply a *random process*, $x(t)$, at the input to a deterministic LTI system whose impulse response is $h(t)$ and whose frequency response is $H(f)$, as shown in Fig. 4.6. Equation 4.29, which presumes a deterministic input, can be employed on a sample-function basis to show that the *output* of the system will be a random process, $y(t)$, whose sample functions are related to those of the input by convolution with the impulse response, viz.¹⁰

$$y(t, \omega) = \int_{-\infty}^{\infty} x(\tau, \omega) h(t - \tau) d\tau. \quad (4.31)$$

Thus, $y(t)$ and $x(t)$ are joint random processes defined on a common probability space. Moreover, given the second-order characterization of the process $x(t)$ and the *linearity* of the system, we will be able to find the second-order characterization of process $y(t)$ using techniques that we established in our work with random vectors.

Suppose $x(t)$ has mean function $m_x(t)$ and covariance function $K_{xx}(t, s)$. What are the resulting mean function and covariance function of the output

¹⁰It follows from this result that we can use Eq. 4.29 for a random process input. We shall eschew use of the random process version of Eq. 4.30, and postpone introduction of frequency-domain descriptions until we specialize to wide-sense stationary processes.

process $y(t)$? The calculations are continuous-time versions of the Chapter 3 component-notation manipulations for N -D linear transformations. We have that

$$\begin{aligned} m_y(t) &= E \left[\int_{-\infty}^{\infty} x(\tau) h(t - \tau) d\tau \right] \\ &= \int_{-\infty}^{\infty} E[x(\tau) h(t - \tau)] d\tau \end{aligned} \quad (4.32)$$

$$= \int_{-\infty}^{\infty} E[x(\tau)] h(t - \tau) d\tau \quad (4.33)$$

$$= \int_{-\infty}^{\infty} m_x(\tau) h(t - \tau) d\tau = m_x(t) * h(t), \quad (4.34)$$

where $*$ denotes convolution. Equation 4.32 is obtained using “the average of the sum is the sum of the averages”, as in Chapter 3, only here the summation is an integral. Equation 4.33 follows from “the average of a constant times a random quantity equals the constant times the average of the random quantity”, as in Chapter 3, only here the constant and random quantity are indexed by a continuous—rather than a discrete—parameter. It is Eq. 4.34, however, that is of greatest importance; it shows that the mean output of an LTI system driven by a random process is the mean input passed through the system.¹¹

Because the input to and the output from the LTI filter in Fig. 4.6 equal their mean functions plus their noise parts, Eq. 4.34 also tells us that the noise in the output of a random-process driven LTI system is the noise part of the input passed through the system. We then find that the covariance function of the output obeys

$$\begin{aligned} K_{yy}(t, s) &= E \left[\int_{-\infty}^{\infty} \Delta x(\tau) h(t - \tau) d\tau \int_{-\infty}^{\infty} \Delta x(\tau') h(s - \tau') d\tau' \right] \\ &= E \left[\int_{-\infty}^{\infty} d\tau \int_{-\infty}^{\infty} d\tau' \Delta x(\tau) \Delta x(\tau') h(t - \tau) h(s - \tau') \right]. \\ &= \int_{-\infty}^{\infty} d\tau \int_{-\infty}^{\infty} d\tau' E[\Delta x(\tau) \Delta x(\tau') h(t - \tau) h(s - \tau')] \\ &= \int_{-\infty}^{\infty} d\tau \int_{-\infty}^{\infty} d\tau' K_{xx}(\tau, \tau') h(t - \tau) h(s - \tau'). \end{aligned} \quad (4.35)$$

This derivation is the continuous-time analog of the componentwise linear-transformation covariance calculations performed in Chapter 3. Of particular

¹¹This property was seen in Chapter 3. As stated in words, it depends only on the linearity, not on the time invariance, of the system. Time invariance permits the use of a convolution integral, instead of a superposition integral, for the input/output relation.

note, for future similar manipulations, is the employment of different dummy variables of integration so that the convolution integrals for $\Delta y(t)$ and $\Delta y(s)$ could be combined into a double integral amenable to the interchange of expectation and integration.

Equation 4.35 is essentially a double-convolution of the input covariance with the system's impulse response. Fortunately, we will seldom have to explicitly carry out such integrations. Before we turn to frequency-domain considerations, let us augment our second-order characterization by introducing the *cross-covariance function* of the processes $x(t)$ and $y(t)$, namely

$$K_{xy}(t, s) \equiv E[\Delta x(t)\Delta y(s)]. \quad (4.36)$$

The cross-covariance function, $K_{xy}(t, s)$, is a deterministic function of two time values; at $t = t_1$ and $s = s_1$, this function equals the covariance between the random variables $x(t_1)$ and $y(s_1)$.¹² When the process $y(t)$ is obtained from the process $x(t)$ as shown in Fig. 4.6, we have that

$$K_{xy}(t, s) = \int_{-\infty}^{\infty} K_{xx}(t, \tau)h(s - \tau) d\tau. \quad (4.37)$$

The cross-covariance function provides a simple, but imperfect, measure of the degree of statistical dependence between two random processes. We will comment further on this issue in Section 4.5.

Wide-Sense Stationarity

Let us find the mean function and the covariance function of the single-frequency wave, $x(t)$ from Eq. 4.3. These are easily shown to be

$$\begin{aligned} m_x(t) &= E[\sqrt{2P} \cos(2\pi f_0 t + \theta)] \\ &= \int_0^{2\pi} \sqrt{2P} \cos(2\pi f_0 t + \theta) \frac{1}{2\pi} d\theta = 0, \end{aligned} \quad (4.38)$$

and

$$\begin{aligned} K_{xx}(t, s) &= E[2P \cos(2\pi f_0 t + \theta) \cos(2\pi f_0 s + \theta)] \\ &= E\{P \cos[2\pi f_0(t - s)]\} + E\{P \cos[2\pi f_0(t + s) + 2\theta]\} \\ &= P \cos[2\pi f_0(t - s)], \end{aligned} \quad (4.39)$$

¹²The term cross-covariance function is used because this function quantifies the covariances between time samples from *two* random processes. The term auto-covariance is sometimes used for functions like K_{xx} and K_{yy} , because they each specify covariances between time samples from a *single* random process.

respectively. Equations 4.38 and 4.39 have a variety of interesting properties. That the mean function, Eq. 4.38, should be zero at all times is intuitively clear from Eq. 4.3—the sample functions of $x(t)$ comprise all possible phase shifts of an amplitude $2P$, frequency f_0 sinusoid, and these occur with equal probability, because of the uniform distribution of θ . That the correlation coefficient, $\rho_{xx}(t, s)$, associated with the covariance function, Eq. 4.39, should give $\rho_{xx}(t, t + n/f_0) = 1$ for n an integer follows directly from Fig. 4.3—all the sample functions of $x(t)$ are sinusoids of period $1/f_0$.

The very specific characteristics of the single-frequency wave's mean and covariance function are *not* the point we are driving at. Rather it is the fact that the mean function is a *constant*, viz.

$$m_x(t) = m_x(0), \quad \text{for all } t, \quad (4.40)$$

and that its covariance function depends only on time *differences*, namely

$$K_{xx}(t, s) = K_{xx}(t - s, 0), \quad \text{for all } t, s, \quad (4.41)$$

that matters. Equations 4.40 and 4.41 say that the second-order characterization of this random process is time invariant—the mean and variance of any single time sample of the process $x(t)$ are independent of the time at which that sample is taken, and the covariance between two different time samples of the process $x(t)$ depends only on their time separation. We call random processes which obey Eqs. 4.40 and 4.41 *wide-sense stationary* random processes.¹³

The single-frequency wave is wide-sense stationary (WSS)—a sinusoid of known amplitude and frequency but completely random phase certainly has no preferred time origin. The Gaussian random process whose typical sample function was sketched in Fig. 4.3 is also WSS—here the WSS conditions were given at the outset. Not all random processes are wide-sense stationary, however. For example, consider the random-frequency wave, $x(t)$, defined by

$$x(t) \equiv \sqrt{2P} \sin(2\pi ft), \quad (4.42)$$

where P is a positive constant, and f is a random variable which is uniformly distributed on the interval $f_0 \leq f \leq 2f_0$, for f_0 a positive constant. Two typical sample functions for this process have been sketched in Fig. 4.7. Clearly, this process is *not* wide-sense stationary—a preferred time origin is apparent at $t = 0$.

¹³A *strict-sense stationary* random process is one whose complete statistical characterization is time invariant [cf. Section 4.4].

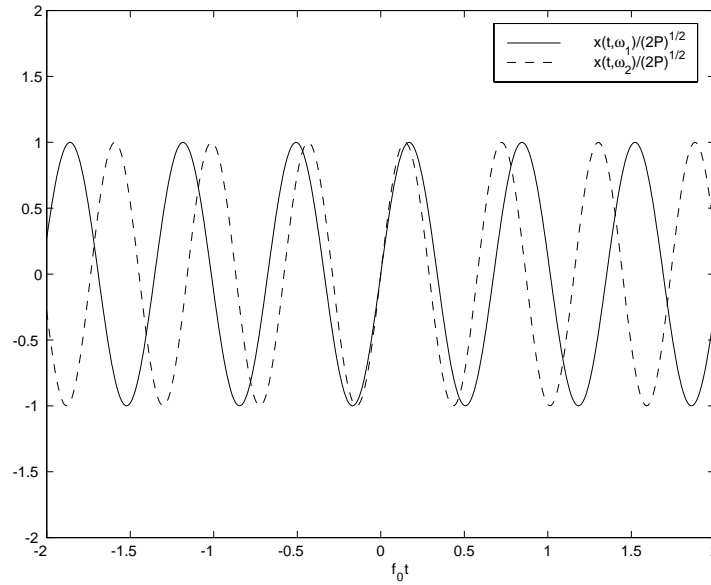


Figure 4.7: Typical sample functions of a random-frequency wave

We make no claim for the physical importance of the preceding random-frequency wave. Neither do we assert that *all* physically interesting random processes must be wide-sense stationary. It seems reasonable, however, to expect that the thermal-noise current of a resistor in thermal equilibrium at temperature T K should be a WSS random process. Likewise, the shot-noise current produced by constant-power illumination of a photodetector should also be wide-sense stationary. Thus the class of WSS processes will be of some interest in the optical communication analyses that follow. Our present task is to examine the implications of Eqs. 4.40 and 4.41 with respect to LTI filtering of a wide-sense stationary random process.

Suppose that the input, $x(t)$, in Fig. 4.6 is a wide-sense stationary process, whose mean function is

$$m_x = E[x(t)], \quad (4.43)$$

and whose covariance function is

$$K_{xx}(\tau) = E[\Delta x(t + \tau)\Delta x(t)], \quad (4.44)$$

where we have exploited Eq. 4.40 in suppressing the time argument of the mean function, and Eq. 4.41 in writing a covariance function that depends only on time difference, τ .¹⁴ We have, from Eqs. 4.34 and 4.35, that the mean

¹⁴An unfortunate recurring problem of technical writing—particularly in multidisciplinary

and covariance function of the output process, $y(t)$, are

$$m_y(t) = \int_{-\infty}^{\infty} m_x h(t - \alpha) d\alpha = m_x H(0) = m_y(0), \quad (4.45)$$

for all t ,

and

$$\begin{aligned} K_{yy}(t, s) &= \int_{-\infty}^{\infty} d\alpha \int_{-\infty}^{\infty} d\beta K_{xx}(\alpha - \beta) h(t - \alpha) h(s - \beta) \\ &= \int_{-\infty}^{\infty} d\alpha \int_{-\infty}^{\infty} d\beta K_{xx}(t - s - \alpha + \beta) h(\alpha) h(\beta), \quad (4.46) \\ &= K_{yy}(t - s, 0), \quad \text{for all } t, s, \end{aligned}$$

where Eq. 4.46 has been obtained via the change of variables $\alpha \rightarrow t - \alpha$, $\beta \rightarrow s - \beta$.

We see that $y(t)$ is a wide-sense stationary random process. This is to be expected. The input process has no preferred time origin in its second-order characterization because it is WSS; the second-order characterization of the output process can be obtained from that of the input because the filter is linear; and the filter imposes no preferred time origin into the propagation of the second-order characterization because the filter is time invariant. In the notation for WSS processes, the above results become

$$m_y = m_x H(0), \quad (4.47)$$

$$K_{yy}(\tau) = \int_{-\infty}^{\infty} d\alpha \int_{-\infty}^{\infty} d\beta K_{xx}(\tau - \alpha + \beta) h(\alpha) h(\beta). \quad (4.48)$$

Furthermore, the qualitative argument given in support of $y(t)$'s being wide-sense stationary extends to the cross-covariance function for which Eq. 4.37 can be reduced to

$$K_{xy}(\tau) \equiv E[\Delta x(t + \tau) \Delta y(t)] = \int_{-\infty}^{\infty} K_{xx}(\tau + \beta) h(\beta) d\beta. \quad (4.49)$$

Two wide-sense stationary random processes whose cross-covariance function depends only on time differences are said to be *jointly* wide-sense stationary (JWSS) processes—such is the case for the $x(t)$ and $y(t)$ processes here.

endeavors like optical communication—is that there is not enough good notation to go around. In deterministic LTI system theory, τ is the preeminent choice for the convolution-integral's dummy variable. In random process theory, τ is the preferred time-difference argument for a covariance function. In what follows we shall replace τ and τ' in Eqs. 4.34, 4.35, and 4.37 with α and β , respectively.

Deeper appreciation for the WSS case can be developed by examining Eqs. 4.47, 4.48, and 4.49 from the frequency domain. The mean function of a WSS random process is a constant—a sinusoid of zero frequency—so that Eq. 4.47 is merely the sinusoidal eigenfunction version of “the mean output of the LTI system is the mean input passed through the system”. The frequency-domain versions of Eqs. 4.48 and 4.49 require that we introduce the Fourier transforms—the *spectral densities*—of the input and output covariance functions, namely

$$S_{xx}(f) \equiv \int_{-\infty}^{\infty} K_{xx}(\tau) e^{-j2\pi f\tau} d\tau, \quad (4.50)$$

$$S_{yy}(f) \equiv \int_{-\infty}^{\infty} K_{yy}(\tau) e^{-j2\pi f\tau} d\tau, \quad (4.51)$$

as well as their *cross-spectral density*, i.e., the Fourier transform of the input/output cross-covariance function

$$S_{xy}(f) \equiv \int_{-\infty}^{\infty} K_{xy}(\tau) e^{-j2\pi f\tau} d\tau. \quad (4.52)$$

We then can reduce the convolution-like integrations involved in the time-domain results Eqs. 4.48 and 4.49 to the following multiplicative relations

$$S_{yy}(f) = S_{xx}(f) |H(f)|^2, \quad (4.53)$$

and

$$S_{xy}(f) = S_{xx}(f) H(f)^*. \quad (4.54)$$

Aside from the calculational advantages of multiplication as opposed to integration, Eq. 4.53 has important mathematical properties and a vital physical interpretation. The 1:1 nature of Fourier transformation tells us that covariance functions can be recovered from their associated spectra by an inverse Fourier integral, e.g.,

$$K_{xx}(\tau) = \int_{-\infty}^{\infty} S_{xx}(f) e^{j2\pi f\tau} df. \quad (4.55)$$

Combining this result with the WSS forms of Eqs. 4.15 and 4.16 yields

$$\left. \begin{array}{l} K_{xx}(-\tau) = K_{xx}(\tau) \\ K_{xx} \text{ real-valued} \end{array} \right\} \longleftrightarrow \left\{ \begin{array}{l} S_{xx}(-f) = S_{xx}(f) \\ S_{xx} \text{ real-valued} \end{array} \right., \quad (4.56)$$

and

$$0 \leq \text{var}[x(t)] = K_{xx}(0) = \int_{-\infty}^{\infty} S_{xx}(f) df. \quad (4.57)$$

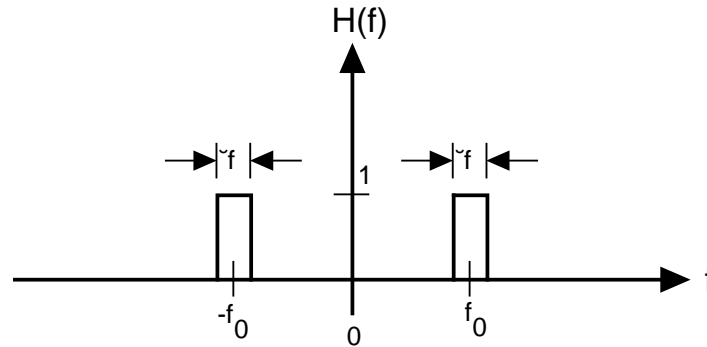


Figure 4.8: Ideal passband filter

As per our discussion following Eqs. 4.15–4.17, the constraints just exhibited for a WSS covariance function and its associated spectrum are necessary but not sufficient conditions for a function of a single variable to be a valid K_{xx} or S_{xx} . Nevertheless, Eq. 4.57 suggests an interpretation of $S_{xx}(f)$ whose validation will lead us to the necessary and sufficient conditions for the WSS case.

We know that $\text{var}[x(t)]$ is the instantaneous mean-square noise strength in the random process $x(t)$. For $x(t)$ wide-sense stationary, this variance can be found—according to Eq. 4.57—by integrating the spectral density $S_{xx}(f)$ over all frequencies. This frequency-domain calculation is consistent with the following property.

spectral-density interpretation For $x(t)$ a WSS random process with spectral density $S_{xx}(f)$, and $f_0 \geq 0$ an arbitrary frequency,¹⁵ $S_{xx}(f_0)$ is the mean-square noise strength per unit bilateral bandwidth in $x(t)$'s frequency f_0 component.

The above property, which we will prove immediately below, certainly justifies referring to $S_{xx}(f)$ as the spectral density of the $x(t)$ process. Its proof is a simple juxtaposition of the physical interpretation and the mathematical analysis of $\text{var}[y(t)]$ for the Fig. 4.6 arrangement when $H(f)$ is the ideal passband filter shown in Fig. 4.8. This ideal filter passes, without distortion, the frequency components of $x(t)$ that lie within a $2\Delta f$ bilateral bandwidth vicinity of frequency f_0 , and completely suppresses all other frequencies.¹⁶ Thus,

¹⁵Strictly speaking, f_0 should be a point of continuity of $S_{xx}(f)$ for this property to hold.

¹⁶Because we are dealing with real-valued time functions and exponential Fourier transforms, $H(-f) = H(f)^*$ must prevail. We shall only refer to *positive* frequencies in discussing the spectral content of the filter's output, but we must employ its *bilateral*—positive and negative frequency—bandwidth in calculating $\text{var}[y(t)]$.

for Δf sufficiently small, we have that $\text{var}[y(t)]/2\Delta f$ is the mean-square noise strength per unit bilateral bandwidth in the frequency f_0 component of the process $x(t)$. On the other hand, a direct mathematical calculation, based on Eqs. 4.57, 4.56, and 4.53, shows that

$$\begin{aligned} \frac{\text{var}[y(t)]}{2\Delta f} &= \frac{K_{yy}(0)}{2\Delta f} = \frac{1}{2\Delta f} \int_{-\infty}^{\infty} S_{yy}(f) df \\ &= \frac{1}{\Delta f} \int_0^{\infty} S_{yy}(f) df \\ &= \frac{1}{\Delta f} \int_0^{\infty} S_{xx}(f) |H(f)|^2 df \\ &= \frac{1}{\Delta f} \int_{f_0-\Delta f/2}^{f_0+\Delta f/2} S_{xx}(f) df \\ &\approx S_{xx}(f_0), \quad \text{for } \Delta f \text{ sufficiently small,} \end{aligned} \quad (4.58)$$

which proves the desired result.

The spectral-density interpretation is the major accomplishment of our second-order characterization work.¹⁷ This interpretation lends itself to treating the simple-minded communication example we used to introduce our development of LTI filtering of random processes. It is also of great value in understanding the temporal content of a random process, albeit within the limitations set by the incompleteness of second-order characterization. In the next subsection we shall present and discuss several $K_{xx} \leftrightarrow S_{xx}$ examples.

Spectral-Density Examples

As a prelude to the examples, we note the following corollary to our spectral-density interpretation: the spectral density of a WSS random process is non-negative,

$$S_{xx}(f) \geq 0, \quad \text{for all } f. \quad (4.59)$$

Moreover, it can be shown that the inverse Fourier transform of a real-valued, even, non-negative function of frequency is a real-valued, even, non-negative definite function of time. Thus, Eqs. 4.56 and 4.59 are necessary and sufficient conditions for an arbitrary deterministic function of frequency to be a valid spectral density for a wide-sense stationary random process. This makes the task of selecting valid $K_{xx} \leftrightarrow S_{xx}$ examples fairly simple—retrieve from our

¹⁷The term *power-spectral density* is often used, with some imprecision. If $x(t)$ has physical units “widgets”, then $S_{xx}(f)$ has units “widgets²/Hz”. Only when widgets² are watts is $S_{xx}(f)$ really a power spectrum. Indeed, the most common spectrum we shall deal with in our photodetection work is that of electrical current; its units are A²/Hz.

storehouse of deterministic linear system theory all Fourier transform pairs for which the frequency function is real-valued, even, and non-negative. The following examples will suffice to illustrate some key points.

single-frequency spectrum The single-frequency wave's covariance function, Eq. 4.39, written in WSS notation,

$$K_{xx}(\tau) = P \cos(2\pi f_0 \tau), \quad (4.60)$$

is associated with the following spectrum,

$$S_{xx}(f) = \frac{P}{2} \delta(f - f_0) + \frac{P}{2} \delta(f + f_0). \quad (4.61)$$

Lorentzian spectrum The exponential covariance function, Eq. 4.8, written in WSS notation,

$$K_{xx}(\tau) = P \exp(-\lambda|\tau|), \quad (4.62)$$

is associated with the Lorentzian spectrum,

$$S_{xx}(f) = \frac{2P\lambda}{(2\pi f)^2 + \lambda^2}. \quad (4.63)$$

bandlimited spectrum The ideal bandlimited spectrum,

$$S_{xx}(f) = \begin{cases} P/2W, & \text{for } |f| \leq W, \\ 0, & \text{otherwise,} \end{cases} \quad (4.64)$$

is associated with the $\sin(x)/x$ covariance function,

$$K_{xx}(\tau) = P \frac{\sin(2\pi W \tau)}{2\pi W \tau}. \quad (4.65)$$

Gaussian spectrum The Gaussian covariance function,

$$K_{xx}(\tau) = P \exp\left(-\frac{\tau^2}{t_c^2}\right), \quad (4.66)$$

is associated with the Gaussian spectrum

$$S_{xx}(f) = P \sqrt{\pi t_c^2} \exp[-(\pi f t_c)^2]. \quad (4.67)$$

white noise The white-noise spectrum,¹⁸

$$S_{xx}(f) = q, \quad \text{for all } f, \quad (4.68)$$

is associated with the impulsive covariance function,

$$K_{xx}(\tau) = q\delta(\tau). \quad (4.69)$$

We have plotted these $K_{xx} \leftrightarrow S_{xx}$ examples in Fig. 4.9. The single-frequency wave's spectrum is fully consistent with our understanding of its sample functions—all the mean-square noise strength in this process is concentrated at $f = f_0$. The Lorentzian, bandlimited, and Gaussian examples all can be assigned reasonably-defined correlation times and bandwidths, as shown in the figure. These evidence the Fourier-transform uncertainty principle, i.e., to make a covariance decay more rapidly we must broaden its associated spectrum proportionally. In physical terms, for two time-samples of a WSS process taken at time-separation τ s to be weakly correlated, the process must contain significant spectral content at or beyond $1/2\pi\tau$ Hz. This is consistent with our earlier discussion of the Gaussian-process sample function shown in Fig. 4.3.

The white-noise spectrum deserves some additional discussion. Its name derives from its having equal mean-square noise density at *all* frequencies. This infinite bandwidth gives it both infinite variance and zero correlation time—both characteristics at odds with physical reality. Yet, white-noise models abound in communication theory generally, and will figure prominently in our study of optical communications. There need be no conflict between realistic modeling and the use of white-noise spectra. If a wide-sense stationary input process in the Fig. 4.6 arrangement has a true spectral density that is very nearly flat over the passband of the filter, no loss in *output-spectrum* accuracy results from replacing the true input spectrum with a white-noise spectrum of the appropriate level. We must remember, when using a white-noise model, that meaningless answers—infinite variance, zero correlation-time—will ensue if no bandlimiting filter is inserted between the source of the noise and our observation point. Any measurement apparatus has some intrinsic bandwidth limitation, so this caution is *not* unduly restrictive.

4.4 Gaussian Random Processes

A random process, $x(t)$, is a collection of random variables indexed by the time parameter, t . A *Gaussian* random process is a collection of *jointly Gaus-*

¹⁸There is not enough good notation to go around; q is *not* the electron charge in this expression.

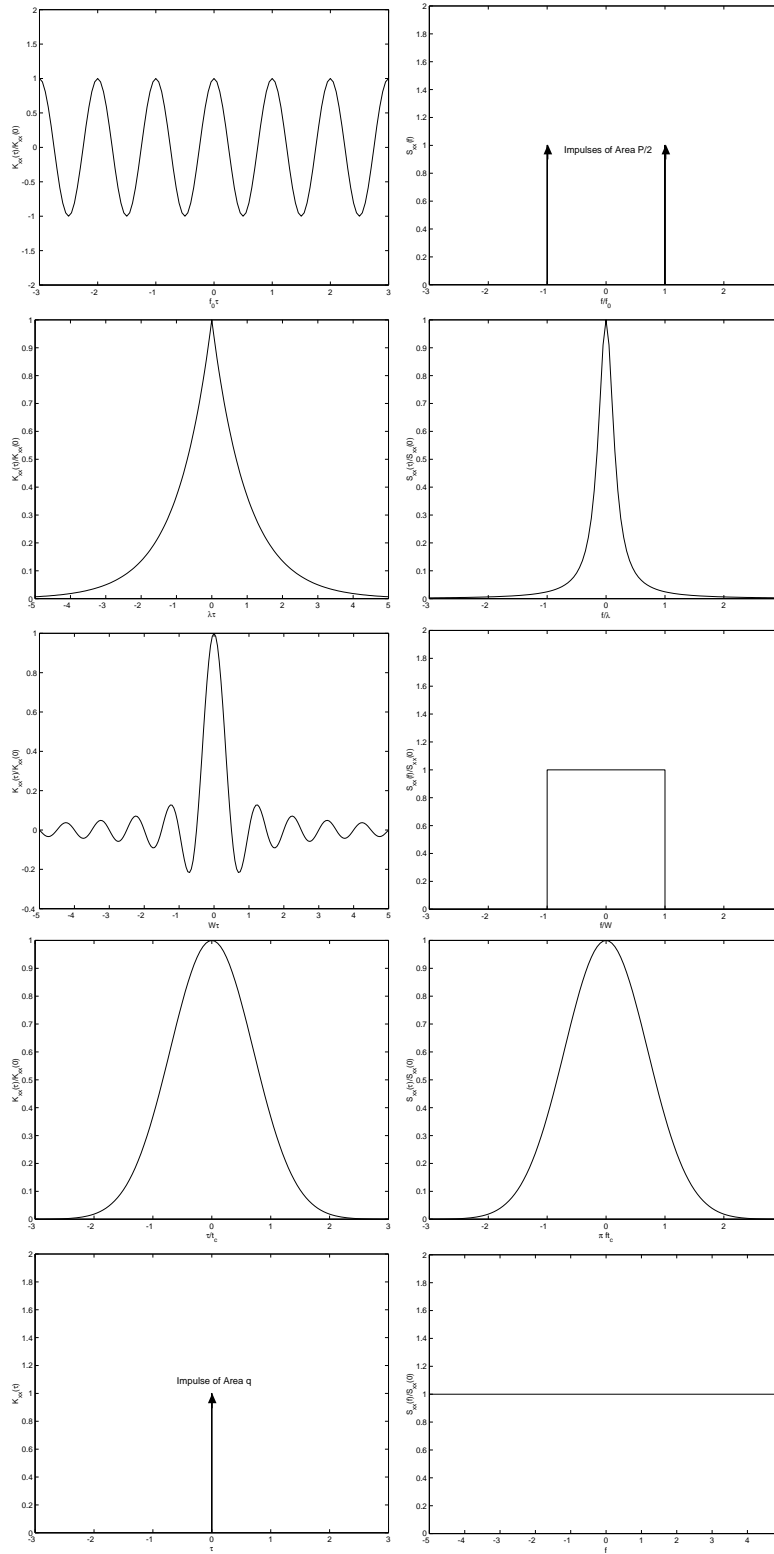


Figure 4.9: Covariance/spectral-density examples, top to bottom: (a) single-frequency wave, (b) Lorentzian spectrum, (c) bandlimited spectrum, (d) Gaussian spectrum, (e) white noise

sian random variables indexed by time. We introduced the Gaussian random process (GRP) early on in this chapter, to provide a quick random-process example whose sample functions looked appropriately *noisy*. Now, having extensively developed the second-order characterization for an arbitrary random process, we return to the Gaussian case. This return will be worthwhile for several reasons: Gaussian random processes are good models for random waveforms whose microscopic composition consists of a large number of more-or-less small, more-or-less independent contributions; and second-order characterization provides complete statistics for a Gaussian random process.

Let $x(t)$ be a GRP with mean function $m_x(t)$ and covariance function $K_{xx}(t, s)$; wide-sense stationarity will not be assumed yet. By definition, the random vector \mathbf{x} , from Eq. 4.1, obtained by sampling this process at the times specified by \mathbf{t} , from Eq. 4.2, is Gaussian distributed. To complete explicit evaluation of the probability density, $p_{\mathbf{x}}(\mathbf{X})$, for this random vector we need only determine its mean vector and covariance matrix; they are given by

$$\mathbf{m}_{\mathbf{x}} = \begin{bmatrix} m_x(t_1) \\ m_x(t_2) \\ \vdots \\ m_x(t_N) \end{bmatrix}, \quad (4.70)$$

and

$$\mathbf{\Lambda}_{\mathbf{x}} = \begin{bmatrix} K_{xx}(t_1, t_1) & K_{xx}(t_1, t_2) & \cdots & K_{xx}(t_1, t_N) \\ K_{xx}(t_2, t_1) & K_{xx}(t_2, t_2) & \cdots & K_{xx}(t_2, t_N) \\ \vdots & \vdots & \ddots & \vdots \\ K_{xx}(t_N, t_1) & K_{xx}(t_N, t_2) & \cdots & K_{xx}(t_N, t_N) \end{bmatrix}. \quad (4.71)$$

Thus, whereas knowledge of mean and covariance functions provides only a *partial* characterization of a general random process, it provides a *complete* characterization of a Gaussian random process. Furthermore, if a GRP $x(t)$ is wide-sense stationary, then it is also strict-sense stationary, i.e., for arbitrary \mathbf{t} , the random vector \mathbf{x} has the *same* probability density function as the random vector \mathbf{x}' defined as follows

$$\mathbf{x}' \equiv \begin{bmatrix} x(0) \\ x(t_2 - t_1) \\ x(t_3 - t_2) \\ \vdots \\ x(t_N - t_1) \end{bmatrix}. \quad (4.72)$$

Arbitrary random processes which are wide-sense stationary need not be strict-sense stationary.

When we introduced jointly Gaussian random variables in Chapter 3, we focused on their closure under linear transformations. The same is true for Gaussian random processes: if the input process in Fig. 4.6 is Gaussian, then the output process is also Gaussian. We shall omit the proof of this property—it merely involves combining the convolution integral input/output relation for the filter with the linear-closure definition of jointly Gaussian random variables to prove that $y(t)$ is a collection of jointly Gaussian random variables indexed by t . Of greater importance is the fact that the mean and covariance propagation results from our second-order characterization now yield complete statistics for the output of an LTI filter that is driven by Gaussian noise.

4.5 Joint Random Processes

The principal focus of the material thus far in this chapter has been on a *single* random process. Nevertheless, we have noted that the random-process input and output in Fig. 4.6 comprise a pair of *joint* random processes on some underlying probability space. We even went so far as to compute their cross-covariance function. Clearly, there will be cases, in our optical communication analyses, when we will use measurements of one random process to infer characteristics of another. Thus, it is germane to briefly examine the complete characterization for joint random processes, and discuss what it means for two random processes to be statistically independent. Likewise, with respect to partial statistics, we ought to understand the joint second-order characterization for two random processes, and what it means for them to be uncorrelated. These tasks will be addressed in this final section. Although the extension to N joint random processes is straightforward, we will restrict our remarks to the 2-D case.

Let $x(t)$ and $y(t)$ be joint random processes. Their complete statistical characterization is the information sufficient to deduce the probability density,

$p_{\mathbf{z}}(\mathbf{Z})$, for the random vector

$$\mathbf{z} \equiv \begin{bmatrix} \mathbf{x} \\ - - - \\ \mathbf{y} \end{bmatrix} = \begin{bmatrix} \begin{bmatrix} x(t_1) \\ x(t_2) \\ \vdots \\ x(t_N) \end{bmatrix} \\ - - - \\ \begin{bmatrix} y(t'_1) \\ y(t'_2) \\ \vdots \\ y(t'_M) \end{bmatrix} \end{bmatrix}, \quad (4.73)$$

for arbitrary $\{t_n, t'_m\}$, N , and M . In words, complete joint-characterization amounts to having the joint statistics for any set of time samples from the two random processes. The joint second-order characterization of the processes $x(t)$ and $y(t)$ consists of their mean functions, covariance functions, and their cross-covariance function. These can always be found from a complete joint characterization; the converse is not generally true.

We can now deal with the final properties of interest.

statistically-independent processes Two random processes are statistically independent if and only if *all* the time samples of one process are statistically independent of *all* the time samples of the other process.

uncorrelated random processes Two random processes are uncorrelated if and only if *all* the time samples of one process are uncorrelated with *all* the time samples of the other process.

For $x(t)$ and $y(t)$ statistically independent random processes, the probability density for a random vector \mathbf{z} , obtained via Eq. 4.73 with arbitrary sample times, factors according to

$$p_{\mathbf{z}}(\mathbf{Z}) = p_{\mathbf{x}}(\mathbf{X})p_{\mathbf{y}}(\mathbf{Y}), \quad \text{for all } \mathbf{Z} = \begin{bmatrix} \mathbf{X} \\ - - - \\ \mathbf{Y} \end{bmatrix}; \quad (4.74)$$

for $x(t)$ and $y(t)$ uncorrelated random processes, we have that

$$K_{xy}(t, s) = 0, \quad \text{for all } t, s. \quad (4.75)$$

Statistically independent random processes are always uncorrelated, but uncorrelated random processes *may* be statistically dependent. In the context

of our photodetector phenomenology, the *physical* independence of the various noise currents—the light current, the dark current, and the thermal current—will lead us to model them as statistically independent random processes.

Although there is a great deal more to be said about random processes, we now have sufficient foundation for our immediate needs. However, before developing the random-process machinery that goes with the generic-photodetector model of Chapter 2, we shall use Chapter 5 to establish a similar analytic beach head in the area of Fourier optics. While not desperately necessary for direct-detection statistical modeling, Fourier optics will be critical to understanding heterodyne detection, and it will serve as the starting point for our coverage of unguided propagation channels.

MIT OpenCourseWare
<https://ocw.mit.edu>

6.453 Quantum Optical Communication
Fall 2016

For information about citing these materials or our Terms of Use, visit: <https://ocw.mit.edu/terms>.