10. Analysis and Processing of Random Signals

In this chapter we introduce methods for analyzing and processing random signals. We cover the following topics:

- Section 10.1 introduces the notion of power spectral density, which allows us to view random processes in the frequency domain.
- Section 10.2 discusses the response of linear systems to random process inputs and introduces methods for filtering random processes.
- Section 10.3 considers two important applications of signal processing: sampling and modulation.
- Sections 10.4 and 10.5 discuss the design of optimum linear systems and introduce the Wiener and Kalman filters.
- Section 10.6 addresses the problem of estimating the power spectral density of a random process.
- Finally, Section 10.7 introduces methods for implementing and simulating the processing of random signals.

10.1 Power Spectral Density

The Fourier series and the Fourier transform allow us to view deterministic time functions as the weighted sum or integral of sinusoidal functions. A time function that varies slowly has the weighting concentrated at the low-frequency sinusoidal components. A time function that varies rapidly has the weighting concentrated at higher-frequency components. Thus the rate at which a deterministic time function varies is related to the weighting function of the Fourier series or transform. This weighting function is called the “spectrum” of the time function.

The notion of a time function as being composed of sinusoidal components is also very useful for random processes. However, since a sample function of a random process can be viewed as being selected from an ensemble of allowable time functions, the weighting function or “spectrum” for a random process must refer in some way to the average rate of change of the ensemble of allowable time functions. Equation (9.66) shows that, for wide-sense stationary processes, the autocorrelation function \( R_X(\tau) \) is an appropriate measure for the average rate of change of a random process. Indeed if a random process changes slowly with time, then it remains correlated with itself for a long period of time, and \( R_X(\tau) \) decreases slowly as a function of \( \tau \). On the other hand, a rapidly varying random process quickly becomes uncorrelated with itself, and \( R_X(\tau) \) decreases rapidly with \( \tau \).

We now present the Einstein-Wiener-Khinchin theorem, which states that the power spectral density of a wide-sense stationary random process is given by the Fourier transform of the autocorrelation function.
10.1.1 Continuous-Time Random Processes

Let \( X(t) \) be a continuous-time WSS random process with mean \( \mu_X \) and autocorrelation function \( R_X(\tau) \). Suppose we take the Fourier transform of a sample of \( X(t) \) in the interval \( 0 < t < T \) as follows

\[
\tilde{x}(f) = \int_0^T X(t') \cdot \exp(-j \cdot 2\pi \cdot f \cdot t') \cdot dt'
\]

We then approximate the power density as a function of frequency by the function:

\[
\tilde{p}_T(f) = \frac{1}{T} \left| \tilde{x}(f) \right|^2 = \frac{1}{T} \cdot \tilde{x}(f) \cdot \tilde{x}(f)^* 
\]

\[
= \frac{1}{T} \left\{ \int_0^T X(t') \cdot \exp(-j \cdot 2\pi \cdot f \cdot t'') \cdot dt' \right\} \cdot \left\{ \int_0^T X(t') \cdot \exp(-j \cdot 2\pi \cdot f \cdot t'') \cdot dt' \right\}^*
\]

where \( * \) denotes the complex conjugate. \( X(t) \) is a random process, so \( \tilde{p}_T(f) \) is also a random process but over a different index set. \( \tilde{p}_T(f) \) is called the **periodogram estimate** and we are interested in the power spectral density of \( X(t) \) which is defined by:

\[
S_X(f) = \lim_{T \to \infty} E[\tilde{p}_T(f)] = \lim_{T \to \infty} \frac{1}{T} \cdot E[|\tilde{x}(f)|^2]
\]

We show at the end of this section that the power spectral density of \( X(t) \) is given by the Fourier transform of the autocorrelation function:

\[
\lim_{T \to \infty} \frac{1}{T} \cdot E[|\tilde{x}(f)|^2] = \lim_{T \to \infty} \frac{1}{T} \cdot \int_0^T \int_0^T \int_0^T \int_0^T E[X(t_1) \cdot X(t_2) \cdot \exp(-j \cdot 2\pi \cdot f \cdot t_1) \cdot \exp(j \cdot 2\pi \cdot f \cdot t_2) \cdot dt_1 \cdot dt_2 
\]

\[
\lim_{T \to \infty} \frac{1}{T} \cdot E[|\tilde{x}(f)|^2] = \lim_{T \to \infty} \frac{1}{T} \cdot \int_0^T \int_0^T R_X(t_1 - t_2) \cdot \exp(-j \cdot 2\pi \cdot f \cdot (t_1 - t_2)) \cdot dt_1 \cdot dt_2
\]

Dr. Bazuin’s “sloppy math derivation follows”

\[
\lim_{T \to \infty} \frac{1}{T} \cdot E[|\tilde{x}(f)|^2] = \lim_{T \to \infty} \frac{1}{T} \cdot \int_0^{T + T_2} R_X(\tau) \cdot \exp(-j \cdot 2\pi \cdot f \cdot \tau) \cdot d\tau \cdot dt_2
\]

\[
\lim_{T \to \infty} \frac{1}{T} \cdot E[|\tilde{x}(f)|^2] = \lim_{T \to \infty} \frac{1}{T} \cdot \int_0^T \Re[R_X(\tau)] \cdot dt_2 = \Re[R_X] \cdot dt_2
\]

\[
S_X(f) = \Re[R_X(\tau)] = \int_{-\infty}^{\infty} R_X(\tau) \cdot \exp(-j \cdot 2\pi \cdot f \cdot \tau) \cdot d\tau
\]

A table of Fourier transforms and its properties is given in Appendix B.
For real-valued random processes, the autocorrelation function is an even function of $\tau$:

$$R_X(\tau) = R_X(-\tau)^*$$

Substitution into Eq. (10.4) implies that for real signals

$$S_X(f) = \int_{-\infty}^{\infty} R_X(\tau) \cdot \cos(2\pi f \cdot \tau) \cdot d\tau$$

since the integral of the product of an even function and an odd function is zero. Equation (10.6) implies that $S_X(f)$ is real-valued and an even function of frequency, $f$. From Eq. (10.2) we have that is nonnegative:

$$S_X(f) \geq 0 \quad \text{for all } f$$

The autocorrelation function can be recovered from the power spectral density by applying the inverse Fourier transform formula to Eq. (10.4):

$$R_X(\tau) = \mathcal{F}^{-1}[S_X(f)] = \int_{-\infty}^{\infty} S_X(f) \cdot \exp(j \cdot 2\pi f \cdot \tau) \cdot df$$

Equation (10.8) is identical to Eq. (4.80), which relates the pdf to its corresponding characteristic function. The last section in this chapter discusses how the FFT can be used to perform numerical calculations for $S_X(f)$ and $R_X(\tau)$.

In electrical engineering it is customary to refer to the second moment of $X(t)$ as the average power of $X(t)$. Equation (10.8) together with Eq. (9.64) gives

$$E[X(t)^2] = R_X(0) = \int_{-\infty}^{\infty} S_X(f) \cdot df$$

Equation (10.9) states that the average power of $X(t)$ is obtained by integrating $S_X(f)$ over all frequencies. This is consistent with the fact that $S_X(f)$ is the “density of power” of $X(t)$ at the frequency $f$.

Since the autocorrelation and autocovariance functions are related by $R_X(\tau) = C_X(\tau) + m_X^2$ the power spectral density is also given by

$$S_X(f) = \mathcal{F}[C_X(\tau) + m_X^2]$$

$$= \mathcal{F}[C_X(\tau)] + \delta(f) \cdot m_X^2$$

where we have used the fact that the Fourier transform of a constant is a delta function. We say the $m_X$ is the “dc” component of $X(t)$.
The notion of power spectral density can be generalized to two jointly wide-sense stationary processes. The cross-power spectral density is defined by

\[ S_{X,Y}(f) = \mathbb{E}[R_{X,Y}(\tau)] \]

where \( R_{X,Y}(\tau) \) is the cross-correlation between \( X(t) \) and \( Y(t) \):

\[ R_{X,Y}(\tau) = \mathbb{E}[X(t + \tau) \cdot Y(t)] \]

In general, \( S_{X,Y}(f) \) is a complex function of frequency, \( f \), even if \( X(t) \) and \( Y(t) \) are both real-valued.

---

**Example 10.1 Random Telegraph Signal**

Find the power spectral density of the random telegraph signal.

In Example 9.24, the autocorrelation function of the random telegraph process was found to be

\[ R_X(\tau) = e^{-2\alpha|\tau|}, \]

where \( \alpha \) is the average transition rate of the signal. Therefore, the power spectral density of the process is

\[ S_X(f) = \int_{-\infty}^{0} e^{2\alpha\tau}e^{-j2\pi f\tau}d\tau + \int_{0}^{\infty} e^{-2\alpha\tau}e^{-j2\pi f\tau}d\tau \]

\[ = \frac{1}{2\alpha - j2\pi f} + \frac{1}{2\alpha + j2\pi f} \]

\[ = \frac{4\alpha}{4\alpha^2 + 4\pi^2 f^2}. \quad (10.13) \]

Figure 10.1 shows the power spectral density for \( \alpha = 1 \) and \( \alpha = 2 \) transitions per second. The process changes two times more quickly when \( \alpha = 2 \); it can be seen from the figure that the power spectral density for \( \alpha = 2 \) has greater high-frequency content.
Example 10.3 White Noise

The power spectral density of a WSS white noise process whose frequency components are limited to the range is shown in Fig. 10.2(a). The process is said to be “white” in analogy to white light, which contains all frequencies in equal amounts. (implying W goes to infinity)

The average power in this process is obtained from Eq. (10.9):

\[ E[X(t)^2] = R_X(0) = \int_{-W}^{W} \frac{N_0}{2} \cdot df = N_0 \cdot W \]

The autocorrelation for this process is obtained from Eq. (10.8):

\[ R_X(\tau) = \mathcal{F}^{-1}[S_X(f)] = \int_{-W}^{W} \frac{N_0}{2} \cdot \exp(j \cdot 2\pi \cdot f \cdot \tau) \cdot df \]
\[ R_X(\tau) = \frac{N_0}{2} \cdot \int_{-\infty}^{\infty} \exp(j \cdot 2\pi \cdot f \cdot \tau) \cdot df \]

\[ R_X(\tau) = \frac{N_0}{2} \cdot \exp(j \cdot 2\pi \cdot f \cdot \tau) \bigg|_{-\infty}^{\infty} \]

\[ R_X(\tau) = \frac{N_0}{2} \cdot \exp(j \cdot 2\pi \cdot W \cdot \tau) - \exp(-j \cdot 2\pi \cdot W \cdot \tau) \]

\[ R_X(\tau) = \frac{N_0}{2} \cdot 2 \cdot j \cdot \sin(2\pi \cdot W \cdot \tau) = N_0 \cdot W \cdot \frac{\sin(2\pi \cdot W \cdot \tau)}{2\pi \cdot W \cdot \tau} \]

\[ R_X(\tau) = N_0 \cdot W \cdot \text{sinc}(2\pi \cdot W \cdot \tau) \]

\[ R_X(\tau) \] is shown in Fig. 10.2(b). Note that \( X(t) \) and \( X(t+\tau) \) are uncorrelated at \( \tau = \pm k/2W, k = 1,2,\ldots \).

The term white noise usually refers to a random process \( W(t) \) whose power spectral density is \( N_0/2 \) for all frequencies:

\[ S_w(f) = \frac{N_0}{2} \quad \text{for all } f \]

Equation (10.15) with \( W \to \infty \) shows that such a process must have infinite average power. By taking the limit \( W \to \infty \) in Eq. (10.16), we find that the autocorrelation of such a process approaches

\[ \lim_{W \to \infty} R_W(\tau) = \frac{N_0}{2} \cdot \delta(\tau) \]

If \( W(t) \) is a Gaussian random process, we then see that \( W(t) \) is the white Gaussian noise process introduced in Example 9.43 with \( \alpha = N_0/2 \).

Example 10.4  Sum of Two Processes

Find the power spectral density of $Z(t) = X(t) + Y(t)$, where $X(t)$ and $Y(t)$ are jointly WSS processes.

The autocorrelation of $Z(t)$ is

$$R_Z(\tau) = E[Z(t + \tau)Z(t)] = E[(X(t + \tau) + Y(t + \tau))(X(t) + Y(t))]$$
$$= R_X(\tau) + R_{XY}(\tau) + R_{YX}(\tau) + R_Y(\tau).$$

The power spectral density is then

$$S_Z(f) = \mathcal{F}\{R_X(\tau) + R_{XY}(\tau) + R_{YX}(\tau) + R_Y(\tau)\}$$
$$= S_X(f) + S_{XY}(f) + S_{YX}(f) + S_Y(f). \quad (10.19)$$

Example 10.5

Let $Y(t) = X(t - d)$, where $d$ is a constant delay and where $X(t)$ is WSS. Find $R_{XY}(\tau)$, $S_{XY}(f)$, $R_Y(\tau)$, and $S_Y(f)$.

The definitions of $R_{XY}(\tau)$, $S_{XY}(f)$, and $R_Y(\tau)$ give

$$R_{XY}(\tau) = E[Y(t + \tau)X(t)] = E[X(t + \tau - d)X(t)] = R_X(\tau - d). \quad (10.20)$$

The time-shifting property of the Fourier transform gives

$$S_{XY}(f) = \mathcal{F}\{R_X(\tau - d)\} = S_X(f)e^{-j2\pi fd}$$
$$= S_X(f)\cos(2\pi fd) - jS_X(f)\sin(2\pi fd). \quad (10.21)$$

Finally,

$$R_Y(\tau) = E[Y(t + \tau)Y(t)] = E[X(t + \tau - d)X(t - d)] = R_X(\tau). \quad (10.22)$$

Equation (10.22) implies that

$$S_Y(f) = \mathcal{F}\{R_Y(\tau)\} = \mathcal{F}\{R_X(\tau)\} = S_X(f). \quad (10.23)$$

Note from Eq. (10.21) that the cross-power spectral density is complex. Note from Eq. (10.23) that $S_X(f) = S_Y(f)$ despite the fact that $X(t) \neq Y(t)$. Thus, $S_X(f) = S_Y(f)$ does not imply that $X(t) = Y(t)$. 


7 of 61
10.1.2 Discrete-Time Random Processes

Let \( X_n \) be a discrete-time WSS random process with mean \( m_X \) and autocorrelation function \( R_X(k) \). The power spectral density of \( X_n \) is defined as the Fourier transform of the autocorrelation sequence

\[
S_X(f) = \mathcal{F}[R_X(k)] = \sum_{k=-\infty}^{\infty} R_X(k) \cdot \exp(-j \cdot 2\pi \cdot f \cdot k)
\]

Note that we need only consider frequencies in the range \(-1/2 < f \leq 1/2\), since \( S_X(f) \) is periodic in \( f \) with period 1. As in the case of continuous random processes, \( S_X(f) \) can be shown to be a real-valued, nonnegative, even function of \( f \).

The inverse Fourier transform formula applied to Eq. (10.23) implies that

\[
R_X(k) = \mathcal{F}^{-1}[S_X(f)] = \int_{-1/2}^{1/2} S_X(f) \cdot \exp(j \cdot 2\pi \cdot f \cdot k) \cdot df
\]

Equations (10.24) and (10.25) are similar to the discrete Fourier transform. In the last section we show how to use the FFT to calculate \( S_X(f) \) and \( R_X(k) \).

The cross-power spectral density \( S_{X,Y}(f) \) of two jointly WSS discrete-time processes \( X_n \) and \( Y_n \) is defined by

\[
S_{X,Y}(f) = \mathcal{F}[R_{X,Y}(k)]
\]

where \( R_{X,Y}(k) \) is the cross-correlation between \( X_n \) and \( Y_n \):

\[
R_{X,Y}(k) = E[X(n+k) \cdot Y(n)]
\]

---

**Example 10.6 White Noise**

Let the process \( X_n \) be a sequence of uncorrelated random variables with zero mean and variance \( \sigma_X^2 \). Find \( S_X(f) \).

The autocorrelation of this process is

\[
R_X(k) = \begin{cases} \sigma_X^2 & k = 0 \\ 0 & k \neq 0. \end{cases}
\]

The power spectral density of the process is found by substituting \( R_X(k) \) into Eq. (10.24):

\[
S_X(f) = \sigma_X^2, \quad -\frac{1}{2} < f < \frac{1}{2}.
\]

Thus the process \( X_n \) contains all possible frequencies in equal measure.
Example 10.7  Moving Average Process

Let the process $Y_n$ be defined by

$$Y_n = X_n + \alpha X_{n-1}, \quad (10.29)$$

where $X_n$ is the white noise process of Example 10.6. Find $S_Y(f)$.

It is easily shown that the mean and autocorrelation of $Y_n$ are given by

$$E[Y_n] = 0,$$

and

$$E[Y_n Y_{n+k}] = \begin{cases} (1 + \alpha^2) \sigma_X^2 & k = 0 \\ \alpha \sigma_X^2 & k = \pm 1 \\ 0 & \text{otherwise.} \end{cases} \quad (10.30)$$

The power spectral density is then

$$S_Y(f) = (1 + \alpha^2) \sigma_X^2 + \alpha \sigma_X^2 \{e^{i2\pi f} + e^{-i2\pi f}\}$$

$$= \sigma_X^2 \{(1 + \alpha^2) + 2\alpha \cos 2\pi f\}. \quad (10.31)$$

$S_Y(f)$ is shown in Fig. 10.3 for $\alpha = 1$.

![Figure 10.3](image)

**FIGURE 10.3**

Power spectral density of moving average process discussed in Example 10.7.
Example 10.8  Signal Plus Noise

Let the observation \( Z_n \) be given by

\[
Z_n = X_n + Y_n,
\]

where \( X_n \) is the signal we wish to observe, \( Y_n \) is a white noise process with power \( \sigma_Y^2 \), and \( X_n \) and \( Y_n \) are independent random processes. Suppose further that \( X_n = A \) for all \( n \), where \( A \) is a random variable with zero mean and variance \( \sigma_A^2 \). Thus \( Z_n \) represents a sequence of noisy measurements of the random variable \( A \). Find the power spectral density of \( Z_n \).

The mean and autocorrelation of \( Z_n \) are

\[
E[Z_n] = E[A] + E[Y_n] = 0
\]

and

\[
E[Z_n Z_{n+k}] = E[(X_n + Y_n)(X_{n+k} + Y_{n+k})]
\]

\[
= E[X_n X_{n+k}] + E[X_n]E[Y_{n+k}]
\]

\[
+ E[X_{n+k}]E[Y_n] + E[Y_n Y_{n+k}]
\]

\[
\]

Thus \( Z_n \) is also a WSS process.

The power spectral density of \( Z_n \) is then

\[
S_Z(f) = E[A^2] \delta(f) + S_Y(f),
\]

where we have used the fact that the Fourier transform of a constant is a delta function.
10.1.3 Power Spectral Density as a Time Average

In the above discussion, we simply stated that the power spectral density is given as the Fourier transform of the autocorrelation without supplying a proof. We now show how the power spectral density arises naturally when we take Fourier transforms of realizations of random processes.

Let \( X_0, X_1, \ldots, X_{k-1} \) be \( k \) observations from the discrete-time, WSS process \( X_n \). Let \( \tilde{x}(f) \) denote the discrete Fourier transform of this sequence:

\[
\tilde{x}(f) = \sum_{m=0}^{k-1} X_m \cdot \exp(-j \cdot 2\pi \cdot f \cdot m)
\]

Note \( \tilde{x}(f) \) that is a complex-valued random variable. The magnitude squared of \( \tilde{x}(f) \) is a measure of the “energy” at the frequency \( f \). If we divide this energy by the total “time” \( k \), we obtain an estimate for the “power” at the frequency \( f \):

\[
\tilde{p}_k(f) = \frac{1}{k} \cdot |\tilde{x}_k(f)|^2 = \frac{1}{k} \cdot \tilde{x}_k(f) \cdot \tilde{x}_k(f)^*
\]

\( \tilde{p}_k(f) \) is called the periodogram estimate for the power spectral density.

Consider the expected value of the periodogram estimate:

\[
E[\tilde{p}_k(f)] = \frac{1}{k} \cdot E\left[ \tilde{x}_k(f) \cdot \tilde{x}_k(f)^* \right]
\]

\[
= \frac{1}{k} \cdot E\left[ \sum_{m=0}^{k-1} X_m \cdot \exp(-j \cdot 2\pi \cdot f \cdot m) \cdot \sum_{i=0}^{k-1} X_i \cdot \exp(j \cdot 2\pi \cdot f \cdot i) \right]
\]

\[
= \frac{1}{k} \cdot \sum_{m=0}^{k-1} \sum_{i=0}^{k-1} E[X_i \cdot X_m] \cdot \exp(-j \cdot 2\pi \cdot f \cdot (m-i))
\]

\[
= \frac{1}{k} \cdot \sum_{m=0}^{k-1} \sum_{i=0}^{k-1} R_X(m-i) \cdot \exp(-j \cdot 2\pi \cdot f \cdot (m-i))
\]

Figure 10.4 shows the range of the double summation in Eq. (10.34).
Note that all the terms along the diagonal \( m' = m - i \) are equal, that \( m' \) ranges from \( -(k - 1) \) to \( (k - 1) \), and that there \( k - |m'| \) are terms along the diagonal \( m' = m - i \). Thus Eq. (10.34) becomes

\[
E[\tilde{p}_k(f)] = \frac{1}{k} \cdot \sum_{m'=-k+1}^{k-1} (k - |m'|) \cdot R_x(m') \cdot \exp(-j \cdot 2\pi \cdot f \cdot (m'))
\]

\[
= \sum_{m'=-k+1}^{k-1} \left( 1 - \frac{|m'|}{k} \right) \cdot R_x(m') \cdot \exp(-j \cdot 2\pi \cdot f \cdot (m'))
\]

Comparison of Eq. (10.35) with Eq. (10.24) shows that the mean of the periodogram estimate is not equal to \( S_X(f) \) for two reasons. First, Eq. (10.34) does not have the term in brackets in Eq. (10.25). Second, the limits of the summation in Eq. (10.35) are not \( \pm \infty \). We say that is a “biased” estimator for \( S_X(f) \). However, as \( k \to \infty \), we see that the term in brackets approaches one, and that the limits of the summation approach \( \pm \infty \). Thus

\[
E[\tilde{p}_k(f)] \to S_X(f) \quad \text{as } k \to \infty
\]

that is, the mean of the periodogram estimate does indeed approach \( S_X(f) \). Note that Eq. (10.36) shows that \( S_X(f) \) is nonnegative for all \( f \), since \( \tilde{p}_k(f) \) is nonnegative for all \( f \).

In order to be useful, the variance of the periodogram estimate should also approach zero. The answer to this question involves looking more closely at the problem of power spectral density estimation. We defer this topic to Section 10.6.

All of the above results hold for a continuous-time WSS random process \( X(t) \) after appropriate changes are made from summations to integrals. The periodogram estimate for \( S_X(f) \), for an observation in the interval \( 0 < t < T \), was defined in Eq. 10.2. The same derivation that led to Eq. (10.35) can be used to show that the mean of the periodogram estimate is given by

\[
E[\tilde{p}_T(f)] = \int_{-\infty}^{\infty} \left( 1 - \frac{\tau}{T} \right) \cdot R_x(\tau) \cdot \exp(-j \cdot 2\pi \cdot f \cdot \tau) \cdot d\tau
\]

It then follows that

\[
E[\tilde{p}_T(f)] \to S_X(f) \quad \text{as } T \to \infty
\]
10.2 Response of Linear Systems to Random Signals

Many applications involve the processing of random signals (i.e., random processes) in order to achieve certain ends. For example, in prediction, we are interested in predicting future values of a signal in terms of past values. In filtering and smoothing, we are interested in recovering signals that have been corrupted by noise. In modulation, we are interested in converting low-frequency information signals into high-frequency transmission signals that propagate more readily through various transmission media.

Signal processing involves converting a signal from one form into another. Thus a signal processing method is simply a transformation or mapping from one time function into another function. If the input to the transformation is a random process, then the output will also be a random process. In the next two sections, we are interested in determining the statistical properties of the output process when the input is a wide-sense stationary random process.

10.2.1 Continuous-Time Systems

Consider a system in which an input signal \( x(t) \) is mapped into the output signal \( y(t) \) by the transformation

\[
y(t) = T[x(t)]
\]

The system is linear if superposition holds, that is,

\[
T[\alpha \cdot x_1(t) + \beta \cdot x_2(t)] = T[\alpha \cdot x_1(t)] + T[\beta \cdot x_2(t)]
\]

where \( x_1(t) \) and \( x_2(t) \) are arbitrary input signals, and \( \alpha \) and \( \beta \) are arbitrary constants. Let \( y(t) \) be the response to input \( x(t) \), then the system is said to be time-invariant if the response to \( x(t+\tau) \) is \( y(t+\tau) \). The impulse response \( h(t) \) of a linear, time-invariant system is defined by

\[
h(t) = T[\delta(t)]
\]

where \( \delta(t) \) is a unit delta function input applied at \( t=0 \). The response of the system to an arbitrary input \( x(t) \) is then

\[
y(t) = h(t) * x(t) = \int_{-\infty}^{\infty} h(\tau) \cdot x(t - \tau) \cdot d\tau = \int_{-\infty}^{\infty} x(\tau) \cdot h(t - \tau) \cdot d\tau
\]

![Figure 10.5](image)

A linear system with a random input signal.
Therefore a linear, time-invariant system is completely specified by its impulse response. The impulse response \( h(t) \) can also be specified by giving its Fourier transform, the transfer function of the system:

\[
H(f) = \mathcal{F}[h(t)] = \int_{-\infty}^{\infty} h(\tau) \cdot \exp(-j \cdot 2\pi \cdot f \cdot \tau) \cdot d\tau
\]

A system is said to be causal if the response at time \( t \) depends only on past values of the input, that is, if \( h(t) = 0 \) for \( t < 0 \).

If the input to a linear, time-invariant system is a random process \( X(t) \) as shown in Fig. 10.5, then the output of the system is the random process given by

\[
Y(t) = h(t) \ast X(t) = \int_{-\infty}^{\infty} h(\tau) \cdot X(t-\tau) \cdot d\tau = \int_{-\infty}^{\infty} X(\tau) \cdot h(t-\tau) \cdot d\tau
\]

We assume that the integrals exist in the mean square sense as discussed in Section 9.7. We now show that if \( X(t) \) is a wide-sense stationary process, then \( Y(t) \) is also wide-sense stationary.

The mean of \( Y(t) \) is given by

\[
E[Y(t)] = E\left[ \int_{-\infty}^{\infty} h(\tau) \cdot X(t-\tau) \cdot d\tau \right] = \int_{-\infty}^{\infty} h(\tau) \cdot E[X(t-\tau)] \cdot d\tau
\]

Now \( m_X = E[X(t-\tau)] \) since \( X(t) \) is wide-sense stationary, so

\[
E[Y(t)] = \int_{-\infty}^{\infty} h(\tau) \cdot m_X \cdot d\tau = m_X \cdot H(0)
\]

where \( H( f ) \) is the transfer function of the system. Thus the mean of the output \( Y(t) \) is the constant

\[
E[Y(t)] = m_Y = m_X \cdot H(0)
\]

The autocorrelation of \( Y(t) \) is given by

\[
E[Y(t) \cdot Y(t+\tau)] = R_Y(\tau) = E\left[ \int_{-\infty}^{\infty} h(s) \cdot X(t-s) \cdot ds \cdot \int_{-\infty}^{\infty} h(r) \cdot X(t+\tau-r) \cdot dr \right]
\]

\[
= E\left[ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(r) \cdot h(s) \cdot X(t-s) \cdot X(t+\tau-r) \cdot ds \cdot dr \right]
\]

\[
= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(r) \cdot h(s) \cdot E[X(t-s) \cdot X(t+\tau-r)] \cdot ds \cdot dr
\]

\[
= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(r) \cdot h(s) \cdot R_X(\tau+s-r) \cdot ds \cdot dr
\]
where we have used the fact that $X(t)$ is wide-sense stationary. The expression on the right-hand side of Eq. (10.42) depends only on $t$. Thus the autocorrelation of $Y(t)$ depends only on $t$ and since the $E[Y(t)]$ is a constant, we conclude that $Y(t)$ is a wide-sense stationary process.

We are now ready to compute the power spectral density of the output of a linear, time-invariant system. Taking the transform of $R_X(\tau)$ as given in Eq. (10.42), we obtain

$$S_Y(f) = \mathfrak{F}[R_Y] = \int_{-\infty}^{\infty} R_Y(\tau) \cdot \exp(-j \cdot 2\pi \cdot f \cdot \tau) \cdot d\tau$$

Change variables, letting $u = \tau + s - r$

$$S_Y(f) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(r) \cdot h(s) \cdot R_X(u) \cdot ds \cdot dr \cdot \exp(-j \cdot 2\pi \cdot f \cdot (u-s+r)) \cdot du$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(r) \cdot e^{-j \cdot 2\pi \cdot f \cdot r} \cdot h(s) \cdot e^{-j \cdot 2\pi \cdot f \cdot s} \cdot R_X(u) \cdot e^{-j \cdot 2\pi \cdot f \cdot u} \cdot ds \cdot dr \cdot du$$

$$= \int_{-\infty}^{\infty} h(r) \cdot e^{-j \cdot 2\pi \cdot f \cdot r} \cdot dr \cdot \int_{-\infty}^{\infty} h(s) \cdot e^{j \cdot 2\pi \cdot f \cdot s} \cdot ds \cdot \int_{-\infty}^{\infty} R_X(u) \cdot e^{-j \cdot 2\pi \cdot f \cdot u} \cdot du$$

$$S_Y(f) = H(f) \cdot H(f)^* \cdot S_X(f)$$

$$S_Y(f) = |H(f)|^2 \cdot S_X(f)$$

where we have used the definition of the transfer function. Equation (10.43) relates the input and output power spectral densities to the system transfer function. Note that can also be found by computing Eq. (10.43) and then taking the inverse Fourier transform.

Hint: you must know the above equation for at least one of the PhD Quals questions …

Equations (10.41) through (10.43) only enable us to determine the mean and autocorrelation function of the output process $Y(t)$. In general this is not enough to determine probabilities of events involving $Y(t)$. However, if the input process is a Gaussian WSS random process, then as discussed in Section 9.7 the output process will also be a Gaussian WSS random process. Thus the mean and autocorrelation function provided by Eqs. (10.41) through (10.43) are enough to determine all joint pdf’s involving the Gaussian random process $Y(t)$.  

Cross-correlation

The cross-correlation between the input and output processes is also of interest:

\[ E[Y(t + \tau) \cdot X(t)] = R_{y,x}(\tau) = E\left[ \int_{-\infty}^{\infty} h(s) \cdot X(t + \tau - s) \cdot ds \cdot X(t) \right] \]

\[ = \int_{-\infty}^{\infty} h(s) \cdot \left[ E[X(t + \tau - s) \cdot X(t)] \right] ds \]

\[ = \int_{-\infty}^{\infty} h(s) \cdot R_X(\tau - s) ds \]

\[ = h(\tau) \ast R_X(\tau) \]

By taking the Fourier transform, we obtain the cross-power spectral density:

\[ S_{Y,X}(f) = \mathcal{F}[R_{y,x}(\tau)] = H(f) \cdot S_X(f) \]

Since \( R_{y,x}(\tau) = R_{x,y}(-\tau) \), we have that

\[ S_{X,Y}(f) = \mathcal{F}[R_{x,y}(\tau)] = S_{y,x}(f)^* \]

\[ S_{X,Y}(f) = H(f)^* \cdot S_X(f) \]
Arbitrary Random Noise Generator:

Example 10.9 provides us with a method for generating WSS processes with arbitrary power spectral density \( S_y(f) \). We simply need to filter white noise through a filter with transfer function \( H(f) = \sqrt{S_y(f)} \). In general this filter will be noncausal.

Generating a causal filter

We can usually, but not always, obtain a causal filter with transfer function \( H(f) \) such that \( S_y(f) = H(f) \cdot H(f)^* \). For example, if \( S_y(f) \) is a rational function, that is, if it consists of the ratio of two polynomials, then it is easy to factor \( S_y(f) \) into the above form, as shown in the next example. Furthermore any power spectral density can be approximated by a rational function. Thus filtered white noise can be used to synthesize WSS random processes with arbitrary power spectral densities, and hence arbitrary autocorrelation functions.
Example 10.10 Ornstein-Uhlenbeck Process

Find the impulse response of a causal filter that can be used to generate a Gaussian random process with output power spectral density and autocorrelation function

\[ S_Y(f) = \frac{\sigma^2}{\alpha^2 + 4\pi^2 f^2} \quad \text{and} \quad R_Y(\tau) = \frac{\sigma^2}{2\alpha} e^{-\frac{\alpha|\tau|}{2\alpha}} \]

This power spectral density factors as follows:

\[ S_Y(f) = \frac{1}{(\alpha - j2\pi f)} \frac{1}{(\alpha + j2\pi f)} \sigma^2. \]

If we let the filter transfer function be \( H(f) = 1/(\alpha + j2\pi f) \), then the impulse response is \( h(t) = e^{-\alpha t} \) for \( t \geq 0 \), which is the response of a causal system. Thus if we filter white Gaussian noise with power spectral density \( \sigma^2 \) using the above filter, we obtain a process with the desired power spectral density.

In Example 9.46, we found the autocorrelation function of the transient response of this filter for a white Gaussian noise input (see Eq. (9.97a)). As was already indicated, when dealing with power spectral densities we assume that the processes are in steady state. Thus as \( t \to \infty \) Eq. (9.97a) approaches Eq. (9.97b).

Signal Frequency Domain Separation

Example 10.11  Ideal Filters

Let $Z(t) = X(t) + Y(t)$, where $X(t)$ and $Y(t)$ are independent random processes with power spectral densities shown in Fig. 10.6(a). Find the output if $Z(t)$ is input into an ideal lowpass filter with transfer function shown in Fig. 10.6(b). Find the output if $Z(t)$ is input into an ideal bandpass filter with transfer function shown in Fig. 10.6(c).

The power spectral density of the output $W(t)$ of the lowpass filter is

$$S_W(f) = |H_{LP}(f)|^2S_X(f) + |H_{LP}(f)|^2S_Y(f) = S_X(f),$$

since $H_{LP}(f) = 1$ for the frequencies where $S_X(f)$ is nonzero, and $H_{LP}(f) = 0$ where $S_Y(f)$ is nonzero. Thus $W(t)$ has the same power spectral density as $X(t)$. As indicated in Example 10.5, this does not imply that $W(t) = X(t)$.

To show that $W(t) = X(t)$, in the mean square sense, consider $D(t) = W(t) - X(t)$. It is easily shown that

$$R_D(\tau) = R_W(\tau) - R_{WX}(\tau) - R_{XW}(\tau) + R_X(\tau).$$

The corresponding power spectral density is

$$S_D(f) = S_W(f) - S_{WX}(f) - S_{XW}(f) + S_X(f)$$

$$= |H_{LP}(f)|^2S_X(f) - H_{LP}(f)S_Y(f) - H_{LP}^*(f)S_Y(f) + S_X(f)$$

$$= 0.$$

Therefore $R_D(\tau) = 0$ for all $\tau$, and $W(t) = X(t)$ in the mean square sense since

$$E[(W(t) - X(t))^2] = E[D^2(t)] = R_D(0) = 0.$$

Thus we have shown that the lowpass filter removes $Y(t)$ and passes $X(t)$. Similarly, the bandpass filter removes $X(t)$ and passes $Y(t)$. 

10.2.2 Discrete-Time Systems

The results obtained above for continuous-time signals also hold for discrete-time signals after appropriate changes are made from integrals to summations.

The response of the system to an arbitrary input random process $X_n$ is then given by

$$Y_n = h_n * X_n = \sum_{j=-\infty}^{\infty} h_j X_{n-j} = \sum_{j=-\infty}^{\infty} h_{n-j} X_j.$$  \hfill (10.48)

Thus discrete-time, linear, time-invariant systems are determined by the unit-sample response $h_n$. The transfer function of such a system is defined by

$$H(f) = \sum_{i=-\infty}^{\infty} h_i e^{-j2\pi fi}.$$  \hfill (10.49)

The derivation from the previous section can be used to show that if $X_n$ is a wide-sense stationary process, then $Y_n$ is also wide-sense stationary. The mean of $Y_n$ is given by

$$m_Y = m_X \sum_{i=-\infty}^{\infty} h_i = m_X H(0).$$  \hfill (10.50)

The autocorrelation of $Y_n$ is given by

$$R_Y(k) = \sum_{j=-\infty}^{\infty} \sum_{i=-\infty}^{\infty} h_j h_i R_X(k + j - i).$$  \hfill (10.51)

By taking the Fourier transform of $R_Y(k)$ it is readily shown that the power spectral density of $Y_n$ is

$$S_Y(f) = |H(f)|^2 S_X(f).$$  \hfill (10.52)

This is the same equation that was found for continuous-time systems.
Example 10.14 First-Order Autoregressive Process

A first-order autoregressive (AR) process with zero mean is defined by

\[ Y_n = \alpha \cdot Y_{n-1} + X_n \]

where \( X_n \) is a zero-mean white noise input random process with average power \( \sigma_X^2 \). Note that \( Y_n \) can be viewed as the output of the system in Fig. 10.7(a) for an iid input \( X_n \). Find the power spectral density and autocorrelation of \( Y_n \).

The unit-sample response can be determined from Eq. (10.54):

\[ h_n = \begin{cases} 0 & n < 0 \\ 1 & n = 0 \\ \alpha^n & n > 0 \end{cases} \]

Note that we require \(|\alpha| < 1\) for the system to be stable. Therefore the transfer function is

\[ H(f) = \sum_{n=0}^{\infty} \alpha^n \cdot \exp(-j \cdot 2\pi \cdot f \cdot n) = \frac{1}{1 - \alpha \cdot \exp(-j \cdot 2\pi \cdot f)} \]

Equation (10.52) then gives

\[ S_Y(f) = |H(f)|^2 \cdot S_X(f) \]

\[ S_Y(f) = \frac{1}{\left|1 - \alpha \cdot \exp(-j \cdot 2\pi \cdot f)\right|^2} \cdot \sigma_X^2 \]

\[ = \left(\frac{1}{1 - \alpha \cdot \exp(-j \cdot 2\pi \cdot f)}\right) \cdot \left(\frac{1}{1 - \alpha \cdot \exp(-j \cdot 2\pi \cdot f)}\right) \cdot \sigma_X^2 \]

\[ = \frac{\sigma_X^2}{1 - 2 \cdot \alpha \cdot \cos(2\pi \cdot f) + \alpha^2} \]
Equation (10.51) gives

\[ R_y(k) = \sum_{j=0}^{\infty} \sum_{i=0}^{\infty} h_i \cdot h_j \cdot R_X(k + j - i) \]

\[ R_y(k) = \sum_{j=0}^{\infty} \alpha^j \cdot \alpha^j \cdot \delta(k + j - i) \cdot \sigma^2_x \]

\[ R_y(k) = \sigma^2_x \cdot \sum_{j=0}^{\infty} \alpha^{k+j} \cdot \alpha^j = \alpha^k \cdot \sigma^2_x \cdot \sum_{j=0}^{\infty} \alpha^{2j} \]

\[ R_y(k) = \frac{\alpha^k \cdot \sigma^2_x}{1 - \alpha^2} \]

which can also be found as an inverse transform of the power spectral density.

**What about an arbitrary digital filter?**

![Diagram of digital filter](image)

**FIGURE 10.7**

(a) Generation of AR process; (b) Generation of ARMA process.

---

**Example 10.15 ARMA Random Process**

An autoregressive moving average (ARMA) process is defined by

\[ Y_n = -\sum_{i=1}^{q} \alpha_i Y_{n-i} + \sum_{j=1}^{p} \beta_j W_{n-j}, \quad (10.55) \]

where \( W_n \) is a WSS, white noise input process. \( Y_n \) can be viewed as the output of the recursive system in Fig. 10.7(b) to the input \( X_n \). It can be shown that the transfer function of the linear system...
defined by the above equation is

\[ H(f) = \frac{\sum_{i=0}^{p} \beta_i e^{-2\pi f i}}{1 + \sum_{i=1}^{q} \alpha_i e^{-2\pi f i}}. \]

The power spectral density of the ARMA process is

\[ S_Y(f) = |H(f)|^2 \sigma_w^2. \]

ARMA models are used extensively in random time series analysis and in signal processing. The general autoregressive process is the special case of the ARMA process with \( \beta_1 = \beta_2 = \cdots = \beta_p = 0. \) The general moving average process is the special case of the ARMA process with \( \alpha_1 = \alpha_2 = \cdots = \alpha_q = 0. \) Octave has a function \texttt{filter}(b, a, x) which takes a set of coefficients \( b = (\beta_1, \beta_2, \ldots, \beta_{p+1}) \) and \( a = (\alpha_1, \alpha_2, \ldots, \alpha_q) \) as coefficient for a filter as in Eq. (10.55) and produces the output corresponding to the input sequence \( x. \) The choice of \( a \) and \( b \) can lead to a broad range of discrete-time filters.

For example, if we let \( a = (1/N, 1/N, \ldots, 1/N) \) we obtain a moving average filter:

\[ Y_n = (W_n + W_{n-1} + \cdots + W_{n-N+1})/N. \]

Figure 10.8 shows a zero-mean, unit-variance Gaussian iid sequence \( W_n \) and the outputs from an \( N = 3 \) and an \( N = 10 \) moving average filter. It can be seen that the \( N = 3 \) filter moderates the extreme variations but generally tracks the fluctuations in \( X_n. \) The \( N = 10 \) filter on the other hand severely limits the variations and only tracks slower longer-lasting trends.

---

**FIGURE 10.8**
Moving average process showing iid Gaussian sequence and corresponding \( N = 3, N = 10 \) moving average processes.
Figures 10.9(a) and (b) show the result of passing an iid Gaussian sequence $X_n$ through first-order autoregressive filters as in Eq. (10.54). The AR sequence with $\alpha = 0.1$ has low correlation between adjacent samples and so the sequence remains similar to the underlying iid random process. The AR sequence with $\alpha = 0.75$ has higher correlation between adjacent samples which tends to cause longer lasting trends as evident in Fig. 10.9(b).
10.3 Bandlimited Random Processes

In this section we consider two important applications that involve random processes with power spectral densities that are nonzero over a finite range of frequencies. The first application involves the sampling theorem, which states that bandlimited random processes can be represented in terms of a sequence of their time samples. This theorem forms the basis for modern digital signal processing systems. The second application involves the modulation of sinusoidal signals by random information signals. Modulation is a key element of all modern communication systems.

10.3.1 Sampling of Bandlimited Random Processes

The two key steps in making signals amenable to digital signal processing are: (1). Convert the continuous-time signals into discrete-time signals by sampling the amplitudes; (2) Representing the samples using a fixed number of bits. In this section we introduce the sampling theorem for wide-sense stationary bandlimited random processes, which addresses the conversion of signals into discrete-time sequences.

Let \( x(t) \) be a deterministic, finite-energy time signal that has Fourier transform \( \tilde{X}(f) = \mathcal{F}[x(t)] \) that is nonzero only in the frequency range \( |f| \leq W \). Suppose we sample \( x(t) \) every \( T \) seconds to obtain the sequence of sample values: \( \{x(-T), x(0), x(T), \ldots\} \). The sampling theorem for deterministic signals states that \( x(t) \) can be recovered exactly from the sequence of samples if \( T \leq 1/2W \) or equivalently \( 1/T \geq 2W \) that is, the sampling rate is at least twice the bandwidth of the signal. The minimum sampling rate \( 1/2W \) is called the Nyquist sampling rate. The sampling theorem provides the following interpolation formula for recovering \( x(t) \) from the samples:

\[
x(t) = \sum_{n=-\infty}^{\infty} x(nT) \cdot p(t - NT) \quad \text{where} \quad p(t) = \frac{\sin(\pi t/T)}{\pi t/T}
\]

Eq. (10.56) provides us with the interesting interpretation depicted in Fig. 10.10(a). The process of sampling \( x(t) \) can be viewed as the multiplication of \( x(t) \) by a train of delta functions spaced \( T \) seconds apart. The sampled function is then represented by:

\[
x_s(t) = \sum_{n=-\infty}^{\infty} x(nT) \cdot \delta(t - NT)
\]
Eq. (10.56) can be viewed as the response of a linear system with impulse response \( p(t) \) to the signal \( x_s(t) \). It is easy to show that the \( p(t) \) in Eq. (10.56) corresponds to the ideal lowpass filter in Fig. 10.6:

\[
P(f) = \mathcal{F}[p(t)] = \begin{cases} 
1 & -W \leq f \leq W \\
0 & |f| > W 
\end{cases}
\]

The proof of the sampling theorem involves the following steps. We show that

\[
\mathcal{F}\left[ \sum_{n=-\infty}^{\infty} x(nT) \cdot p(t - NT) \right] = \frac{1}{T} \cdot P(f) \cdot \sum_{k=-\infty}^{\infty} \tilde{X}(f - k/T)
\]

which consists of the sum of translated versions of \( \tilde{X}(f) = \mathcal{F}[x(t)] \) as shown in Fig. 10.10(b). We then observe that as long as \( 1/T \geq 2W \), then \( P(f) \) in the above expressions selects the \( k=0 \) term in the summation, which corresponds to \( X(f) \). See Problem 10.45 for details.

---

Applying sampling theorem to random variables

Let \( X(t) \) be a WSS process with autocorrelation function \( R_X(\tau) \) and power spectral density \( S_X(f) \). Suppose that \( S_X(f) \) is bandlimited, that is,

\[
S_X(f) = 0 \quad |f| > W
\]

We now show that the sampling theorem can be extended to include r.v. \( X(t) \). Let

\[
\hat{X}(t) = \sum_{n=-\infty}^{\infty} X(nT) \cdot p(t - NT) \quad \text{where} \quad p(t) = \frac{\sin(\pi T)}{\pi T}
\]

then \( \hat{X}(t) = X(t) \) in the mean square sense. Recall that equality in the mean square sense does not imply equality for all sample functions, so this version of the sampling theorem is weaker than the version in Eq. (10.56) for finite energy signals.

To show Eq. (10.59) we first note that since \( S_X(f) \) is bandlimited, we can apply the sampling theorem for deterministic signals to \( R_X(\tau) \):

\[
R_X(\tau) = \sum_{n=-\infty}^{\infty} R_X(nT) \cdot p(t - NT)
\]

Next we consider the mean square error associated with Eq. (10.59):

\[
E\left[\left(X(t) - \hat{X}(t)\right)^2\right] = E\left[\left(X(t) - \hat{X}(t)\right) \cdot X(t)\right] - E\left[\left(X(t) - \hat{X}(t)\right) \cdot \hat{X}(t)\right]
\]

\[
= E\left[\left(X(t) \cdot X(t) - \hat{X}(t) \cdot X(t)\right)\right] - E\left[\left(X(t) \cdot X(t) - \hat{X}(t) \cdot \hat{X}(t)\right)\right]
\]

It is easy to show that Eq. (10.60) implies that each of the terms in brackets is equal to zero. (See Problem 10.48.) We then conclude that \( \hat{X}(t) = X(t) \) in the mean square sense.

---

**Example 10.16 Sampling a WSS Random Process**

Let \( X(t) \) be a WSS process with autocorrelation function \( R_X(\tau) \). Find the mean and covariance functions of the discrete-time sampled process \( X_n = X(nT) \) for \( n = 0, \pm 1, \pm 2, \ldots \).

Since \( X(t) \) is WSS, the mean and covariance functions are:

\[
m_X(n) = E[X(nT)] = m
\]

\[
E[X_n X_m] = E[X(nT)X(mT)] = R_X(nT - mT) = R_X((n - m)T).
\]

This shows \( X_n \) is a WSS discrete-time process.
The sampling theorem provides an important bridge between continuous-time and discrete-time signal processing. It gives us a means for implementing the real as well as the simulated processing of random signals. First, we must sample the random process above its Nyquist sampling rate. We can then perform whatever digital processing is necessary. We can finally recover the continuous-time signal by interpolation.
10.3.2 Amplitude Modulation by Random Signals

Many of the transmission media used in communication systems can be modeled as linear systems and their behavior can be specified by a transfer function $H(f)$, which passes certain frequencies and rejects others. Quite often the information signal $A(t)$ (i.e., a speech or music signal) is not at the frequencies that propagate well. The purpose of a modulator is to map the information signal $A(t)$ into a transmission signal $X(t)$ that is in a frequency range that propagates well over the desired medium. At the receiver, we need to perform an inverse mapping to recover $A(t)$ from $X(t)$. In this section, we discuss two of the amplitude modulation methods.

Let $A(t)$ be a WSS random process that represents an information signal. In general $A(t)$ will be “lowpass” in character, that is, its power spectral density will be concentrated at low frequencies, as shown in Fig. 10.11(a). An amplitude modulation (AM) system produces a transmission signal by multiplying $A(t)$ by a “carrier” signal $\cos(2\pi f_c \cdot t + \Theta)$

$$X(t) = A(t) \cdot \cos(2\pi f_c \cdot t + \Theta)$$

where we assume $\Theta$ is a random variable that is uniformly distributed in the interval $(0,2\pi)$ and $\Theta$ and $A(t)$ are independent.

The autocorrelation of $X(t)$ is

$$E[X(t + \tau) \cdot X(t)] = E[(A(t + \tau) \cdot \cos(2\pi f_c \cdot (t + \tau) + \Theta)) \cdot (A(t) \cdot \cos(2\pi f_c \cdot t + \Theta))]$$

$$E[X(t + \tau) \cdot X(t)] = E[(A(t + \tau) \cdot A(t)) \cdot (\cos(2\pi f_c \cdot (t + \tau) + \Theta)) \cdot (\cos(2\pi f_c \cdot t + \Theta))]$$

$$E[X(t + \tau) \cdot X(t)] = E[A(t + \tau) \cdot A(t)] \cdot E[\cos(2\pi f_c \cdot (t + \tau) + \Theta) \cdot \cos(2\pi f_c \cdot t + \Theta)]$$

$$E[X(t + \tau) \cdot X(t)] = R_A(\tau) \cdot E\left[\frac{1}{2} \cdot \cos(2\pi f_c \cdot t) + \frac{1}{2} \cdot \cos(2\pi f_c \cdot (2t + \tau) + 2\Theta)\right]$$

$$E[X(t + \tau) \cdot X(t)] = \frac{1}{2} \cdot R_A(\tau) \cdot \cos(2\pi f_c \cdot \tau)$$

and as the right hand side is strictly a function of $\tau$,

$$R_X(\tau) = \frac{1}{2} \cdot R_A(\tau) \cdot \cos(2\pi f_c \cdot \tau)$$

where we used the fact that $E[\cos(2\pi f_c \cdot (2t + \tau) + 2\Theta)] = 0$ (see Example 9.10). Thus $X(t)$ is also a wide-sense stationary random process.
The power spectral density of \( X(t) \) is

\[
S_X(f) = \mathcal{F} \left[ R_X(\tau) \right] = \mathcal{F} \left[ \frac{1}{2} \cdot R_A(\tau) \cdot \cos(2\pi f_c \cdot \tau) \right] \\
= \frac{1}{2} \cdot S_A(f) \ast \left( \frac{1}{2} \cdot \delta(f + f_c) + \frac{1}{2} \cdot \delta(f - f_c) \right) \\
= \frac{1}{4} \cdot S_A(f + f_c) + \frac{1}{4} \cdot S_A(f - f_c)
\]

where we used the table of Fourier transforms in Appendix B. Figure 10.11(b) shows \( S_X(f) \). It can be seen that the power spectral density of the information signal has been shifted to the regions around \( \pm f_c \). \( X(t) \) is an example of a bandpass signal. Bandpass signals are characterized as having their power spectral density concentrated about some frequency much greater than zero.
The transmission signal is demodulated by multiplying it by the carrier signal and lowpass filtering, as shown in Fig. 10.12.

Let

\[ Y(t) = [2 \cdot X(t) \cdot \cos(2\pi \cdot f_c \cdot t + \Theta)] \ast h_{LP}(t) \]

Proceeding as above, we find that

\[
S_Y(f) = |H_{LP}(f)|^2 \cdot \left[ S_X(f + f_c) + S_X(f - f_c) \right] \\
S_Y(f) = |H_{LP}(f)|^2 \cdot \left[ \frac{1}{4} \cdot (S_A(f + 2f_c) + S_A(f)) + \frac{1}{4} \cdot (S_A(f) + S_A(f - 2f_c)) \right] \\
= \frac{1}{4} \cdot S_A(f) + \frac{1}{4} \cdot S_A(f) = \frac{1}{2} \cdot S_A(f)
\]

The ideal lowpass filter passes \( S_A(f) \) and blocks \( S_A(f \pm 2f_c) \) which is centered about \( \pm f_c \), so the output of the lowpass filter has power spectral density

\[ S_Y(f) = \frac{1}{2} \cdot S_A(f) \]

In fact, from Example 10.11 we know the output is the original information signal, \( A(t) \).
**Quadrature amplitude modulation (QAM)**

The modulation method in Eq. (10.56) can only produce bandpass signals for which \( S_X(f) \) is locally symmetric about \( f_c \), \( S_X(f_c + \delta f) = S_X(f_c - \delta f) \) for \( |\delta f| < W \), as in Fig. 10.11(b). The method cannot yield real-valued transmission signals whose power spectral density lack this symmetry, such as shown in Fig. 10.13(a). The following quadrature amplitude modulation (QAM) method can be used to produce such signals:

\[
X(t) = A(t) \cdot \cos(2\pi f_c t + \Theta) + B(t) \cdot \sin(2\pi f_c t + \Theta)
\]

where \( A(t) \) and \( B(t) \) are real-valued, jointly wide-sense stationary random processes, and we require that

\[
R_A(\tau) = R_B(\tau) \quad \text{and} \quad R_{BA}(\tau) = -R_{AB}(\tau)
\]

Note that Eq. (10.67a) implies that \( S_A(f) = S_B(f) \), a real-valued, even function of \( f \), as shown in Fig. 10.13(b). Note also that Eq. (10.67b) implies that \( S_{BA}(f) \) is a purely imaginary, odd function of \( f \), as also shown in Fig. 10.13(c) (see Problem 10.57).

Proceeding as before, we can show that \( X(t) \) is a wide-sense stationary random process with autocorrelation function

\[
R_A(\tau) = \frac{1}{2} R_A(\tau) E[\cos(2\pi f_c (t + \tau) + 2\Theta)] + \frac{1}{2} R_{AB}(\tau) E[-\sin(2\pi f_c (t + \tau) + 2\Theta)] + \frac{1}{2} R_{BA}(\tau) E[\sin(2\pi f_c (t + \tau) + 2\Theta)] + \frac{1}{2} R_B(\tau) E[\cos(2\pi f_c (t + \tau) + 2\Theta)]
\]

\[
R_X(\tau) = \frac{1}{2} (R_A(\tau) + R_B(\tau)) \cdot \cos(2\pi f_c \cdot \tau) + \frac{1}{2} (R_{R, A}(\tau) - R_{A, B}(\tau)) \cdot \sin(2\pi f_c \cdot \tau)
\]

\[
R_X(\tau) = R_A(\tau) \cdot \cos(2\pi f_c \cdot \tau) + R_{R, A}(\tau) \cdot \sin(2\pi f_c \cdot \tau)
\]

and power spectral density

\[
S_X(f) = \frac{1}{2} S_A(f - f_c) + \frac{1}{2} S_A(f + f_c) + \frac{1}{2j} S_{R, A}(f - f_c) - \frac{1}{2j} S_{R, A}(f + f_c)
\]

The resulting power spectral density is as shown in Fig. 10.13(a). Thus QAM can be used to generate real-valued bandpass signals with arbitrary power spectral density.

Bandpass random signals, such as those in Fig. 10.13(a), arise in communication systems when wide-sense stationary white noise is filtered by bandpass filters. Let \(N(t)\) be such a process with power spectral density \(S_N(f)\). It can be shown that \(N(t)\) can be represented by

\[
N(t) = N_c(t) \cdot \cos(2\pi f_c t + \Theta) - N_s(t) \cdot \sin(2\pi f_c t + \Theta)
\]

where \(N_c(t)\) and \(N_s(t)\) are jointly wide-sense stationary processes with

\[
S_{N_c}(f) = S_{N_s}(f) = \{S_N(f - f_c) + S_N(f + f_c)\}_L
\]

and

\[
S_{N_c, N_s}(f) = j\{S_N(f - f_c) - S_N(f + f_c)\}_L
\]

where the subscript \(L\) denotes the lowpass portion of the expression in brackets. In words, every real-valued bandpass process can be treated as if it had been generated by a QAM modulator.
Example 10.18  Demodulation of Noisy Signal

The received signal in an AM system is
\[ Y(t) = A(t) \cos(2\pi f_c t + \Theta) + N(t), \]
where \( N(t) \) is a bandlimited white noise process with spectral density
\[ S_N(f) = \begin{cases} 
\frac{N_0}{2} & |f \pm f_c| < W \\
0 & \text{elsewhere}.
\end{cases} \]

Find the signal-to-noise ratio of the recovered signal.
Equation (10.70) allows us to represent the received signal by
\[ Y(t) = \{A(t) + N_c(t)\} \cos(2\pi f_c t + \Theta) - N_i(t) \sin(2\pi f_c t + \Theta). \]

The demodulator in Fig. 10.12 is used to recover \( A(t) \). After multiplication by \( 2 \cos(2\pi f_c t + \Theta) \), we have
\[ 2Y(t) \cos(2\pi f_c t + \Theta) = \{A(t) + N_c(t)\}2 \cos^2(2\pi f_c t + \Theta) \]
\[ - N_i(t)2 \cos(2\pi f_c t + \Theta) \sin(2\pi f_c t + \Theta) \]
\[ = \{A(t) + N_c(t)\} (1 + \cos(4\pi f_c t + 2\Theta)) \]
\[ - N_i(t) \sin(4\pi f_c t + 2\Theta). \]

After lowpass filtering, the recovered signal is
\[ A(t) + N_c(t). \]

The power in the signal and noise components, respectively, are
\[ \sigma_A^2 = \int_{-W}^{W} S_A(f) \, df \]
\[ \sigma_{N_c}^2 = \int_{-W}^{W} S_{N_c}(f) \, df = \int_{-W}^{W} \left( \frac{N_0}{2} + \frac{N_0}{2} \right) \, df = 2WN_0. \]

The output signal-to-noise ratio is then
\[ \text{SNR} = \frac{\sigma_A^2}{2WN_0}. \]
10.4 Optimum Linear Systems

Many problems can be posed in the following way. We observe a discrete-time, zero-mean process $X_\alpha$ over a certain time interval $I = \{t-a, \ldots, t+b\}$ and we are required to use the $a+b+1$ resulting observations $\{X_{t-a}, \ldots, X_t, \ldots, X_{t+b}\}$ to obtain an estimate $Y_t$ for some other (presumably related) zero-mean process $Z_t$. The estimate $Y_t$ is required to be linear, as shown in Fig. 10.14:

$$Y_t = \sum_{\beta=t-a}^{t+b} h_{t-\beta} \cdot X_\beta = \sum_{\beta=t-b}^{t-a} h_{\beta} \cdot X_{t-\beta}$$

The figure of merit for the estimator is the mean square error

$$E[e_t^2] = E[(Z_t - Y_t)^2]$$

and we seek to find the optimum filter, which is characterized by the impulse response $h_\beta$ that minimizes the mean square error.

![Figure 10.14](image)

A linear system for producing an estimate $Y_t$.

Examples 10.19 and 10.20 show that different choices of $Z_t$ and $X_\alpha$ and of observation interval correspond to different estimation problems.
Example 10.19 Filtering and Smoothing Problems

Let the observations be the sum of a “desired signal” \( Z_a \) plus unwanted “noise” \( N_a \):

\[
X_a = Z_a + N_a \quad \alpha \in I.
\]

We are interested in estimating the desired signal at time \( t \). The relation between \( t \) and the observation interval \( I \) gives rise to a variety of estimation problems.

If \( I = (\infty, t) \), that is, \( a = \infty \) and \( b = 0 \), then we have a filtering problem where we estimate \( Z_t \) in terms of noisy observations of the past and present. If \( I = (t - a, t) \), then we have a filtering problem in which we estimate \( Z_t \) in terms of the \( a + 1 \) most recent noisy observations.

If \( I = (\infty, \infty) \), that is, \( a = b = \infty \), then we have a smoothing problem where we are attempting to recover the signal from its entire noisy version. There are applications where this makes sense, for example, if the entire realization \( X_a \) has been recorded and the estimate \( Z_t \) is obtained by “playing back” \( X_a \).

Example 10.20 Prediction

Suppose we want to predict \( Z_t \) in terms of its recent past: \( \{Z_{t-a}, \ldots, Z_{t-1}\} \). The general estimation problem becomes this prediction problem if we let the observation \( X_a \) be the past \( a \) values of the signal \( Z_a \), that is,

\[
X_a = Z_a \quad t - a \leq \alpha \leq t - 1.
\]

The estimate \( Y_t \) is then a linear prediction of \( Z_t \) in terms of its most recent values.

10.4.1 The Orthogonality Condition

It is easy to show that the optimum filter must satisfy the orthogonality condition (see Eq. 6.56), which states that the error must be orthogonal to all the observations \( X_\alpha \) that is,

\[
0 = E[e_t \cdot X_\alpha] \quad \text{for all } \alpha \in I
\]

\[
0 = E[(Z_t - Y_t) \cdot X_\alpha]
\]

or equivalently,

\[
E[Z_t \cdot X_\alpha] = E[Y_t \cdot X_\alpha] \quad \text{for all } \alpha \in I
\]

If we substitute Eq. (10.73) into Eq. (10.76) we find

\[
E[Z_t \cdot X_\alpha] = E\left[\sum_{\beta=-b}^{a} h_\beta \cdot X_{t-\beta} \cdot X_\alpha\right] \quad \text{for all } \alpha \in I
\]

\[
E[Z_t \cdot X_\alpha] = \sum_{\beta=-b}^{a} h_\beta \cdot E[X_{t-\beta} \cdot X_\alpha] \quad \text{for all } \alpha \in I
\]
\[ E[Z_t \cdot X_{\alpha}] = \sum_{\beta = b}^{a} h_{\beta} \cdot R_X(t - \alpha - \beta) \quad \text{for all } \alpha \in I \]

Equation (10.77) shows that \( E[Z_t \cdot X_{\alpha}] \) depends only on \( t - \alpha \) and thus \( X_{\alpha} \) and \( Z_t \) are jointly wide-sense stationary processes. Therefore, we can rewrite Eq. (10.77) as follows:

\[ R_{Z,X}(t - \alpha) = \sum_{\beta = b}^{a} h_{\beta} \cdot R_X(t - \alpha - \beta) \quad \text{for } t - a \leq \alpha \leq t + b \]

Finally, letting \( m = t - \alpha \) we obtain the following key equation:

\[ R_{Z,X}(m) = \sum_{\beta = b}^{a} h_{\beta} \cdot R_X(m - \beta) \quad \text{for } -b \leq m \leq a \]

The optimum linear filter must satisfy the set of \( a + b + 1 \) linear equations given by this equation. Note that it is identical to Eq. (6.60) for estimating a random variable by a linear combination of several random variables. The wide-sense stationarity of the processes reduces this estimation problem to the one considered in Section 6.5.

Note: to compute the filter the auto-correlation \( R_X(k) \) and cross-correlation \( R_{Z,X}(k) \) must be known. In addition, the filter is non-causal unless \( b=0 \). But for digital signal processing this is not a great problem … it is significant when considering continuous time.

Continuous Equivalent

In the above derivation we deliberately used the notation \( Z_t \) instead of \( Z_n \) to suggest that the same development holds for continuous-time estimation. In particular, suppose we seek a linear estimate \( Y(t) \) for the continuous-time random process \( Z(t) \) in terms of observations of the continuous-time random process \( X(\alpha) \) in the time interval \( t - a \leq \alpha \leq t + b \):

\[ Y(t) = \int_{t-a}^{t+b} h(t - \beta) \cdot X(\beta) \cdot d\beta = \int_{-b}^{a} h(\beta) \cdot X(t - \beta) \cdot d\beta \]

It can then be shown that the filter \( h(\beta) \) that minimizes the mean square error is specified by

\[ R_{Z,X}(\tau) = \int_{-b}^{a} h(\beta) \cdot R_X(\tau - \beta) \cdot d\beta \quad \text{for } -b \leq \tau \leq a \]

Thus in the time-continuous case we obtain an integral equation instead of a set of linear equations. The analytic solution of this integral equation can be quite difficult, but the equation can be solved numerically by approximating the integral by a summation.
Mean Square Error Computation

\[ E[e_t^2] = E[(Z_t - Y_t)^2] \]

We now determine the mean square error of the optimum filter. First we note that for the optimum filter, the error and the estimate are orthogonal since

\[ E[e_t \cdot Y_t] = E[e_t \cdot \sum_{\beta=a}^{t+b} h_{t-\beta} \cdot X_\beta] \]

\[ E[e_t \cdot Y_t] = \sum_{\beta=a}^{t+b} h_{t-\beta} \cdot E[e_t \cdot X_\beta] = 0 \]

where the terms inside the last summation are 0 because of orthogonality defined in Eq. (10.75). Since the mean square error is then

\[ E[e_t^2] = E[e_t \cdot (Z_t - Y_t)] = E[e_t \cdot Z_t] - E[e_t \cdot Y_t] = E[e_t \cdot Z_t] \]

since \( e_t \) and \( Y_t \) are orthogonal. Substituting for \( e_t \) again yields

\[ E[e_t^2] = E[(Z_t - Y_t) \cdot Z_t] = E[Z_t \cdot Z_t] - E[Y_t \cdot Z_t] \]

\[ E[e_t^2] = R_Z(0) - E[Z_t \cdot \sum_{\beta=-b}^{a} h_\beta \cdot X_{t-\beta}] \]

\[ E[e_t^2] = R_Z(0) - \sum_{\beta=-b}^{a} h_\beta \cdot E[Z_t \cdot X_{t-\beta}] \]

\[ E[e_t^2] = R_Z(0) - \sum_{\beta=-b}^{a} h_\beta \cdot R_{Z,X}(\beta) \]

Similarly, it can be shown that the mean square error of the optimum filter in the continuous-time case is

\[ E[e_t^2] = R_Z(0) - \int_{-b}^{a} h(\beta) \cdot R_{Z,X}(\beta) \cdot d\beta \]

The following theorems summarize the above results.
Discrete-Time

**Theorem**

Let \( X_t \) and \( Z_t \) be discrete-time, zero-mean, jointly wide-sense stationary processes, and let \( Y_t \) be an estimate for \( Z_t \) of the form

\[
Y_t = \sum_{\beta=-b}^{t+b} h_{t-\beta} X_\beta = \sum_{\beta=-b}^{a} h_{\beta} X_{t-\beta}.
\]

The filter that minimizes \( E[(Z_t - Y_t)^2] \) satisfies the equation

\[
R_{Z,X}(m) = \sum_{\beta=-b}^{a} h_{\beta} R_X(m - \beta) \quad -b \leq m \leq a
\]

and has mean square error given by

\[
E[(Z_t - Y_t)^2] = R_Z(0) - \sum_{\beta=-b}^{a} h_{\beta} R_{Z,X}(\beta).
\]

Continuous-Time

**Theorem**

Let \( X(t) \) and \( Z(t) \) be continuous-time, zero-mean, jointly wide-sense stationary processes, and let \( Y(t) \) be an estimate for \( Z(t) \) of the form

\[
Y(t) = \int_{t-a}^{t+b} h(t - \beta) X(\beta) \, d\beta = \int_{-b}^{a} h(\beta) X(t - \beta) \, d\beta.
\]

The filter \( h(\beta) \) that minimizes \( E[(Z(t) - Y(t))^2] \) satisfies the equation

\[
R_{Z,X}(\tau) = \int_{-b}^{a} h(\beta) R_X(\tau - \beta) \, d\beta \quad -b \leq \tau \leq a
\]

and has mean square error given by

\[
E[(Z(t) - Y(t))^2] = R_Z(0) - \int_{-b}^{a} h(\beta) R_{Z,X}(\beta) \, d\beta.
\]
Example 10.21 Filtering of Signal Plus Noise

Suppose we are interested in estimating the signal $Z_n$ from the $p+1$ most recent noisy observations:

$$X_\alpha = Z_\alpha + N_\alpha \quad \alpha \in I = \{n-p, \ldots, n-1, n\}$$

Find the set of linear equations for the optimum filter if $Z_\alpha$ and $N_\alpha$ are independent random processes.

For this choice of observation interval, Eq. (10.78) becomes

$$R_{Z,X}(m) = \sum_{\beta=0}^p h_\beta \cdot R_X(m-\beta) \quad m \in \{0, 1, \ldots, p\}$$

The cross-correlation terms in Eq. (10.82) are given by

$$R_{Z,X}(m) = E[Z_n \cdot X_{n-m}] = E[Z_n \cdot (Z_{n-m} + N_{n-m})] = R_Z(m)$$

The autocorrelation terms are given by

$$R_X(m-\beta) = E[X_{m-\beta} \cdot X_{n-m}] = E[(Z_{m-\beta} + N_{m-\beta}) \cdot (Z_{n-m} + N_{n-m})]$$

$$= R_Z(m-\beta) + R_{Z,N}(m-\beta) + R_{N,Z}(m-\beta) + R_N(m-\beta)$$

since $Z_\alpha$ and $N_\alpha$ are independent random processes. Thus Eq. (10.82) for the optimum filter becomes

$$R_Z(m) = \sum_{\beta=0}^p h_\beta \cdot (R_Z(m-\beta) + R_N(m-\beta)) R_X(m-\beta) \quad m \in \{0, 1, \ldots, p\}$$

This set of linear equations in unknowns is solved by matrix inversion.

\[
\begin{bmatrix}
R_Z(0) \\
R_Z(1) \\
\vdots \\
R_Z(p)
\end{bmatrix} =
\begin{bmatrix}
R_Z(0) + R_N(0) & R_Z(-1) + R_N(-1) & \cdots & R_Z(-p) + R_N(-p) \\
R_Z(1) + R_N(1) & R_Z(0) + R_N(0) & \cdots & R_Z(-p+1) + R_N(-p+1) \\
\vdots & \vdots & \ddots & \vdots \\
R_Z(p) + R_N(p) & R_Z(p-1) + R_N(p-1) & \cdots & R_Z(0) + R_N(0)
\end{bmatrix}
\begin{bmatrix}
h_0 \\
h_1 \\
\vdots \\
h_p
\end{bmatrix}
\]

\[
\begin{bmatrix}
h_0 \\
h_1 \\
\vdots \\
h_p
\end{bmatrix} =
\begin{bmatrix}
R_Z(0) + R_N(0) & R_Z(-1) + R_N(-1) & \cdots & R_Z(-p) + R_N(-p) \\
R_Z(1) + R_N(1) & R_Z(0) + R_N(0) & \cdots & R_Z(-p+1) + R_N(-p+1) \\
\vdots & \vdots & \ddots & \vdots \\
R_Z(p) + R_N(p) & R_Z(p-1) + R_N(p-1) & \cdots & R_Z(0) + R_N(0)
\end{bmatrix}^{-1}
\begin{bmatrix}
R_Z(0) \\
R_Z(1) \\
\vdots \\
R_Z(p)
\end{bmatrix}
\]

Note that if $N_a$ is an iid white noise process,

$$R_N(m) = \sigma_N^2 \cdot \delta(m)$$

and the solution is simplified when also using $R_z(m) = R_z(-m)$

$$\begin{bmatrix} h_0 \\ h_1 \\ \vdots \\ h_p \end{bmatrix} = \begin{bmatrix} R_Z(0) + \sigma_N^2 & R_Z(1) & \cdots & R_Z(p) \\ R_Z(1) & R_Z(0) + \sigma_N^2 & \cdots & R_Z(p-1) \\ \vdots & \vdots & \ddots & \vdots \\ R_Z(p) & R_Z(p-1) & \cdots & R_Z(0) + \sigma_N^2 \end{bmatrix}^{-1} \begin{bmatrix} R_Z(0) \\ R_Z(1) \\ \vdots \\ R_Z(p) \end{bmatrix}$$

---

**Example 10.22 Filtering of AR Signal Plus Noise**

Find the set of equations for the optimum filter in Example 10.21 if $Z_a$ is a first-order autoregressive process with average power $\sigma_Z^2$ and parameter $r, |r| < 1$, and $N_a$ is a white noise process with average power $\sigma_N^2$.

The autocorrelation for a first-order autoregressive process is given by

$$R_Z(m) = \sigma_Z^2 r^{|m|} \quad m = 0, \pm 1, \pm 2, \ldots.$$ 

(See Problem 10.42.) The autocorrelation for the white noise process is

$$R_N(m) = \sigma_N^2 \delta(m).$$

Substituting $R_Z(m)$ and $R_N(m)$ into Eq. (10.83) yields the following set of linear equations:

$$\sigma_Z^2 r^{|m|} = \sum_{\beta=0}^{p} h_{\beta} (\sigma_Z^2 r^{|m-\beta|} + \sigma_N^2 \delta(m-\beta)) \quad m \in \{0, \ldots, p\}.$$ 

(10.84)

If we divide both sides of Eq. (10.84) by $\sigma_Z^2$ and let $\Gamma = \sigma_N^2/\sigma_Z^2$, we obtain the following matrix equation:

$$\begin{bmatrix} 1 + \Gamma r & r^2 & \cdots & r^p \\ r & 1 + \Gamma r & \cdots & r^{p-1} \\ r^2 & r & \cdots & r^{p-2} \\ \vdots & \vdots & \ddots & \vdots \\ r^p & r^{p-1} & \cdots & 1 + \Gamma \end{bmatrix} \begin{bmatrix} h_0 \\ h_1 \\ \vdots \\ h_p \end{bmatrix} = \begin{bmatrix} 1 \\ r \\ \vdots \\ r^p \end{bmatrix}.$$ 

(10.85)

Note that when the noise power is zero, i.e., $\Gamma = 0$, then the solution is $h_0 = 1, h_j = 0, j = 1, \ldots, p$, that is, no filtering is required to obtain $Z_n$.

Table 10.1 gives the values of the optimal predictor coefficients and the mean square error as the order of the estimator is increased for the first-order autoregressive process with $\sigma_Z^2 = 4, r = 0.9$, and noise variance $\sigma_N^2 = 4$. It can be seen that the predictor places heavier weight on more recent samples, which is consistent with the higher correlation of such samples with the current sample. For smaller values of $r$, the correlation for distant samples drops off more quickly and the coefficients place even lower weighting on them. The mean square error can also be seen to decrease with increasing order $p + 1$ of the estimator. Increasing the first few orders provides significant improvements, but a point of diminishing returns is reached around $p + 1 = 3$. 

10.4.2 Prediction

The linear prediction problem arises in many signal processing applications. In Example 6.31 in Chapter 6, we already discussed the linear prediction of speech signals. In general, we wish to predict \(z_n\) in terms of \(z_{n-1}, z_{n-2}, \ldots, z_{n-p}\)

\[
Y_n = \sum_{\beta=1}^{p} h_{\beta} \cdot z_{n-\beta}
\]

For this problem, \(X_{\alpha} = Z_{\alpha}\), so Eq. (10.79) becomes

\[
R_Z(m) = \sum_{\beta=1}^{a} h_{\beta} \cdot R_Z(m-\beta) \quad m \in \{1, \ldots, p\}
\]

In matrix form this equation becomes

\[
\begin{bmatrix}
R_Z(1) \\
R_Z(2) \\
\vdots \\
R_Z(p)
\end{bmatrix} =
\begin{bmatrix}
R_Z(0) & R_Z(1) & \cdots & R_Z(p) \\
R_Z(1) & R_Z(0) & \cdots & R_Z(p-1) \\
\vdots & \vdots & \ddots & \vdots \\
R_Z(p) & R_Z(p-1) & \cdots & R_Z(0)
\end{bmatrix}
\begin{bmatrix}
h_0 \\
h_1 \\
\vdots \\
h_p
\end{bmatrix}
\]

Equations (10.86a) and (10.86b) are called the Yule-Walker equations.

Equation (10.80) for the mean square error becomes

\[
E[e_t^2] = R_Z(0) - \sum_{\beta=1}^{p} h_{\beta} \cdot R_Z(\beta)
\]

By inverting the matrix we can solve for the vector of filter coefficients \(h\).
Example 10.23  Prediction for Long-Range and Short-Range Dependent Processes

Let $X_1(t)$ be a discrete-time first-order autoregressive process with $\sigma_X^2 = 1$ and $r = 0.7411$, and let $X_2(t)$ be a discrete-time long-range dependent process with autocovariance given by Eq. (9.109), $\sigma_X^2 = 1$, and $H = 0.9$. Both processes have $C_X(1) = 0.7411$, but the autocovariance of $X_1(t)$ decreases exponentially while that of $X_2(t)$ has long-range dependence. Compare the performance of the optimal linear predictor for these processes for short-term as well as long-term predictions.

Table 10.2 below compares the mean square errors and the coefficients of the two processes in the case of short-term prediction. The predictor for $X_1(t)$ attains all of the benefit of prediction with a $p = 1$ system. The optimum predictors for higher-order systems set the other coefficients to zero, and the mean square error remains at 0.4577. The predictor for $X_2(t)$ achieves most of the possible performance with a $p = 1$ system, but small reductions in mean square error do accrue by adding more coefficients. This is due to the persistent correlation among the values in $X_2(t)$.

Table 10.3 shows the dramatic impact of long-range dependence on prediction performance. We modified Eq. (10.86) to provide the optimum linear predictor for $X_1$ based on two observations $X_{t-10}$ and $X_{t-20}$ that are in the relatively remote past. $X_1(t)$ and its previous values are almost uncorrelated, so the best predictor has a mean square error of almost 1, which is the variance of $X_1(t)$. On the other hand, $X_2(t)$ retains significant correlation with its previous values and so the mean square error provides a significant reduction from the unit variance. Note that the second-order predictor places significant weight on the observation 20 samples in the past.

<table>
<thead>
<tr>
<th>TABLE 10.2(a)</th>
<th>Short-term prediction: autoregressive, $r = 0.7411$, $\sigma_X^2 = 1$, $C_X(1) = 0.7411$.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p$</td>
<td>MSE</td>
</tr>
<tr>
<td>1</td>
<td>0.45077</td>
</tr>
<tr>
<td>2</td>
<td>0.45077</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TABLE 10.2(b)</th>
<th>Short-term prediction: long-range dependent process, Hurst = 0.9, $\sigma_X^2 = 1$, $C_X(1) = 0.7411$.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p$</td>
<td>MSE</td>
</tr>
<tr>
<td>1</td>
<td>0.45077</td>
</tr>
<tr>
<td>2</td>
<td>0.43625</td>
</tr>
<tr>
<td>3</td>
<td>0.42712</td>
</tr>
<tr>
<td>4</td>
<td>0.42253</td>
</tr>
<tr>
<td>5</td>
<td>0.41964</td>
</tr>
</tbody>
</table>
### TABLE 10.3(a) Long-term prediction: autoregressive, $r = 0.7411$, $\sigma_X = 1$, $C_X(1) = 0.7411$.

<table>
<thead>
<tr>
<th>$p$</th>
<th>MSE</th>
<th>Coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.99750</td>
<td>0.04977</td>
</tr>
<tr>
<td>2</td>
<td>0.99750</td>
<td>0.04977</td>
</tr>
</tbody>
</table>

### TABLE 10.3(b) Long-term prediction: long-range dependent process, Hurst = 0.9, $\sigma_X = 1$, $C_X(1) = 0.7411$.

<table>
<thead>
<tr>
<th>$p$</th>
<th>MSE</th>
<th>Coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0.79354</td>
<td>0.45438</td>
</tr>
</tbody>
</table>
| 10;20 | 0.74850 | 0.34614     | 0.23822
10.4.3 Estimation Using the Entire Realization of the Observed Process

Suppose that $Z_t$ is to be estimated by a linear function $Y_t$ of the entire realization of $X_t$, that is, $a = b = \infty$ and Eq. (10.73) becomes

$$Y_t = \sum_{\beta=-\infty}^{\infty} h_{\beta} \cdot X_{t-\beta}$$

In the case of continuous-time random processes, we have

$$Y(t) = \int_{-\infty}^{\infty} h(\beta) \cdot X(t - \beta) \cdot d\beta$$

The optimum filters must satisfy Eqs. (10.78) and (10.79), which in this case become

$$R_{Z,X}(m) = \sum_{\beta=-\infty}^{\infty} h_{\beta} \cdot R_X(m - \beta) \quad \text{for all } m$$

$$R_{Z,X}(\tau) = \int_{-\infty}^{\infty} h(\beta) \cdot R_X(\tau - \beta) \cdot d\beta \quad \text{for all } \tau$$

The Fourier transform of the first equation and the Fourier transform of the second equation both yield the same expression:

$$S_{Z,X}(f) = H(f) \cdot S_X(f)$$

which is readily solved for the transfer function of the optimum filter:

$$H(f) = \frac{S_{Z,X}(f)}{S_X(f)}$$

The impulse response of the optimum filter is then obtained by taking the appropriate inverse transform. In general the filter obtained from Eq. (10.89) will be noncausal, that is, its impulse response is nonzero for $t<0$. We already indicated that there are applications where this makes sense, namely, in situations where the entire realization $X_\alpha$ is recorded and the estimate $Z_t$ is obtained in “nonreal time” by “playing back” $X_\alpha$. 
10.4.4 Estimation Using Causal Filters

Now, suppose that $Z_i$ is to be estimated using only the past and present of $X_\alpha$, that is, $I = (-\infty, t)$. Equations (10.78) and (10.79) become

$$R_{Z,X}(m) = \sum_{\beta=0}^{\infty} h_\beta \cdot R_X(m-\beta) \quad \text{for all } m$$

$$R_{Z,X}(\tau) = \int_{-\infty}^{\infty} h(\beta) \cdot R_X(\tau-\beta) \cdot d\beta \quad \text{for all } \tau$$

Equations (10.93a) and (10.93b) are called the Wiener-Hopf equations and, though similar in appearance to Eqs. (10.88a) and (10.88b), are considerably more difficult to solve.

Special case

First, let us consider the special case where the observation process is white, that is, for the discrete-time case $R_X(m) = \delta_m$. Equation (10.93a) is then

$$R_{Z,X}(m) = \sum_{\beta=0}^{\infty} h_\beta \cdot \delta(m - \beta) \quad m \geq 0$$

Thus in this special case, the optimum causal filter has coefficients given by

$$h_m = \begin{cases} 
0 & m < 0 \\
R_{Z,X}(m) & m \geq 0
\end{cases}$$

The corresponding transfer function is

$$H(f) = \sum_{\beta=0}^{\infty} R_{Z,X}(m) \cdot \exp(-j \cdot 2\pi \cdot f \cdot m)$$

Note Eq. (10.95) is not $S_{Z,X}(f)$ since the limits of the Fourier transform in Eq. (10.95) do not extend from $-\infty$ to $\infty$. However, $H(f)$ can be obtained from $S_{Z,X}(f)$ by finding $h_m = \mathcal{F}^{-1}[S_{Z,X}(f)]$, keeping the causal part (i.e., $h_m$ for $m \geq 0$) and setting the noncausal part to 0.

General case

We now show how the solution of the above special case can be used to solve the general case. It can be shown that under very general conditions, the power spectral density of a random process can be factored into the form

$$S_X(f) = |G(f)|^2 = G(f) \cdot G(f)^*$$

where $G(f)$ and $1/G(f)$ are causal filters. This suggests that we can find the optimum filter in two steps, as shown in Fig. 10.15.

Let

$$W(f) = \frac{1}{G(f)}$$
First, we pass the observation process through a “whitening” filter with transfer function to produce a white noise process since
\[ S_X(f) = |W(f)|^2 \cdot S_X(f) = \frac{|G(f)|^2}{|G(f)|^2} = 1 \quad \text{for all } f \]

Second, we find the best estimator for using the whitened observation process \( X_n' \) as given by Eq. (10.95). The filter that results from the tandem combination of the whitening filter and the estimation filter is the solution to the Wiener-Hopf equations.

The transfer function of the second filter in Fig. 10.15 is
\[ H_2(f) = \sum_{\beta=0}^{\infty} R_{Z,X'}(m) \cdot \exp(-j \cdot 2\pi \cdot f \cdot m) \]
by Eq. (10.95). To evaluate Eq. (10.97) we need to find
\[ R_{Z,X'}(k) = E[Z_{n+k} X_n'] \]
\[ R_{Z,X'}(k) = \sum_{i=0}^{\infty} w_i \cdot E[Z_{n+k} X_{n-i}] \]
\[ R_{Z,X'}(k) = \sum_{i=0}^{\infty} w_i \cdot R_{Z,X'}(k+i) \]
where \( w_i \) is the impulse response of the whitening filter. The Fourier transform of Eq. (10.98) gives an expression that is easier to work with:
\[ S_{Z,X'}(f) = W(f)^* \cdot S_{Z,X}(f) = \frac{S_{Z,X}(f)}{G(f)^*} \]
The inverse Fourier transform of Eq. (10.99) yields the desired \( R_{Z,X'}(k) \) which can then be substituted into Eq. (10.97) to obtain \( H_2(f) \).

In summary, the optimum filter is found using the following procedure:

1. Factor \( S_X(f) \) as in Eq. (10.96) and obtain a causal whitening filter \( W(f) = 1/G(f) \)
2. Find \( R_{Z,X'}(k) \) from Eq. (10.98) or from Eq. (10.99).
3. \( H_2(f) \) is then given by Eq. (10.97).
4. The optimum filter is then \( H(f) = W(f) \cdot H_2(f) \)

This procedure is valid for the continuous-time version of the optimum causal filter problem, after appropriate changes are made from summations to integrals.
Example 10.25  Wiener Filter

Find the optimum causal filter for estimating a signal \( Z(t) \) from the observation \( X(t) = Z(t) + N(t) \), where \( Z(t) \) and \( N(t) \) are independent random processes, \( N(t) \) is zero-mean white noise density 1, and \( Z(t) \) has power spectral density

\[
S_Z(f) = \frac{2}{1 + 4\pi^2 f^2}.
\]

The optimum filter in this problem is called the **Wiener filter**.

The cross-power spectral density between \( Z(t) \) and \( X(t) \) is

\[
S_{Z,X}(f) = S_Z(f),
\]

since the signal and noise are independent random processes. The power spectral density for the observation process is

\[
S_X(f) = S_Z(f) + S_N(f) = \frac{3 + 4\pi^2 f^2}{1 + 4\pi^2 f^2} = \left( \frac{j2\pi f + \sqrt{3}}{j2\pi f + 1} \right) \left( \frac{-j2\pi f + \sqrt{3}}{-j2\pi f + 1} \right).
\]

If we let

\[
G(f) = \frac{j2\pi f + \sqrt{3}}{j2\pi f + 1},
\]

\[
W(f) = \frac{1}{G(f)} = \frac{j2\pi f + 1}{j2\pi f + \sqrt{3}},
\]

then it is easy to verify that \( W(f) = 1/G(f) \) is the whitening causal filter.

Next we evaluate Eq. (10.99):

\[
S_{Z,X}(f) = \frac{S_{Z,X}(f)}{G^*(f)} = \frac{2}{1 + 4\pi^2 f^2} \frac{1 - j2\pi f}{\sqrt{3} - j2\pi f}
\]

\[
= \frac{2}{(1 + j2\pi f)(\sqrt{3} - j2\pi f)}
\]

\[
= \frac{c}{1 + j2\pi f} + \frac{c}{\sqrt{3} - j2\pi f},
\]

where \( c = 2/(1 + \sqrt{3}) \). If we take the inverse Fourier transform of \( S_{Z,X}(f) \), we obtain

\[
R_{Z,X}(\tau) = \begin{cases} 
  ce^{-\tau} & \tau > 0 \\
  ce^{\sqrt{3}\tau} & \tau < 0.
\end{cases}
\]

Equation (10.97) states that \( H_2(f) \) is given by the Fourier transform of the \( \tau > 0 \) portion of \( R_{Z,X}(\tau) \):

\[
H_2(f) = \mathcal{F}(ce^{-\tau}u(\tau)) = \frac{c}{1 + j2\pi f}.
\]

\[
H(f) = W(f) \cdot H_2(f) = \frac{j2\pi f + 1}{j2\pi f + \sqrt{3}} \cdot \frac{c}{j2\pi f + 1}
\]

The optimum filter is then

\[ H(f) = \frac{1}{G(f)} H_2(f) = \frac{c}{\sqrt{3} + j2\pi f}. \]

The impulse response of this filter is

\[ h(t) = ce^{-\sqrt{3}t} \quad t > 0. \]
10.5 *The Kalman Filter

The optimum linear systems considered in the previous section have two limitations: (1) They assume wide-sense stationary signals; and (2) The number of equations grows with the size of the observation set. In this section, we consider an estimation approach that assumes signals have a certain structure. This assumption keeps the dimensionality of the problem fixed even as the observation set grows. It also allows us to consider certain nonstationary signals.

We will consider the class of signals that can be represented as shown in Fig. 10.16(a):

\[ Z_n = \alpha_{n-a} \cdot Z_{n-1} + W_n \quad n = 1, 2, \ldots \]

where \( Z_0 \) is the random variable at time 0, \( \alpha_n \) is a known sequence of constants, and \( W_n \) is a sequence of zero-mean uncorrelated random variables with possibly time-varying variances \( \{ E[W_n^2] \} \). The resulting process \( Z_n \) is nonstationary in general. We assume that the process \( Z_n \) is not available to us, and that instead, as shown in Fig. 10.16(a), we observe (10.103)

\[ X_n = Z_n + N_n \quad n = 0, 1, 2, \ldots \]

where the observation noise \( N_n \) is a zero-mean, uncorrelated sequence of random variables with possibly time-varying variances \( \{ E[N_n^2] \} \). We assume that \( W_n \) and \( N_n \) are uncorrelated at all times \( n \) and \( n' \). In the special case where \( W_n \) and \( N_n \) are Gaussian random processes, then \( Z_n \) and \( X_n \) will also be Gaussian random processes. We will develop the Kalman filter, which has the structure in Fig. 10.16(b).

Our objective is to find for each time \( n \) the minimum mean square estimate (actually prediction) of based on the observations using a linear estimator that possibly varies with time:

\[ Y_n = \sum_{j=1}^{n} h_j^{(n-1)} \cdot X_{n-j} \]

The orthogonality principle implies that the optimum filter \( \{ h_j^{(n-1)} \} \) satisfies

\[ E[(Z_n - Y_n) \cdot X_l] = E\left[ (Z_n - \sum_{j=1}^{n} h_j^{(n-1)} \cdot X_{n-j}) \cdot X_l \right] = 0 \quad \text{for } l = 0, 1, \ldots, n-1 \]
which leads to a set of \( n \) equations in \( n \) unknowns:

\[
R_{Z,X}(n,l) = \sum_{j=1}^{n} h_j^{(n-l)} \cdot R_{Z,X}(n-j,l) \quad \text{for } l = 0,1,\ldots,n-1
\]

At the next time instant, we need to find

\[
Y_{n+1} = \sum_{j=1}^{n+1} h_j^{(n)} \cdot X_{n+1-j}
\]

by solving a system of \((n+1) \times (n+1)\) equations:

\[
R_{Z,X}(n+1,l) = \sum_{j=1}^{n+1} h_j^{(n)} \cdot R_{Z,X}(n+1-j,l) \quad \text{for } l = 0,1,\ldots,n
\]

Up to this point we have followed the procedure of the previous section and we find that the dimensionality of the problem grows with the number of observations. We now use the signal structure to develop a recursive method for solving Eq. (10.106).

We first need the following two results: For \( t<n \), we have

\[
\begin{align*}
R_{Z,X}(n+1,l) &= E[Z_{n+1} \cdot X_l] = E[(\alpha_n \cdot Z_n + W_n) \cdot X_l] \\
R_{Z,X}(n+1,l) &= \alpha_n \cdot R_{Z,X}(n,l) + E[W_n \cdot X_l] \\
R_{Z,X}(n+1,l) &= \alpha_n \cdot R_{Z,X}(n,l)
\end{align*}
\]

since \( E[W_n \cdot X_l] = E[W_n] \cdot E[X_l] \) that is, \( W_n \) is uncorrelated with the past of the process and the observations prior to time \( n \), as can be seen from Fig. 10.16(a). Also for \( l<n \) we have

\[
R_{Z,X}(n,l) = E[Z_n \cdot X_l] = E[(X_n - N_n) \cdot X_l] \\
R_{Z,X}(n,l) = R_X(n,l) - E[N_n \cdot X_l] = R_X(n,l)
\]

since \( E[N_n \cdot X_l] = E[N_n] \cdot E[X_l] \) that is, the observation noise at time \( n \) is uncorrelated with prior observations.

We now show that the set of equations in Eq. (10.107) can be related to the set in Eq. (10.105). For \( n<l \), we can equate the right-hand sides of Eqs. (10.108) and (10.107):

\[
\begin{align*}
\alpha_n \cdot R_{Z,X}(n,l) &= \sum_{j=1}^{n+1} h_j^{(n)} \cdot R_X(n+1-j,l) \\
\alpha_n \cdot R_{Z,X}(n,l) &= h_1^{(n)} \cdot R_X(n,l) + \sum_{j=2}^{n+1} h_j^{(n)} \cdot R_X(n+1-j,l)
\end{align*}
\]
From Eq. (10.109) we have $R_{Z,X}(n,l) = R_X(n,l)$ so we can replace the first term on the right-hand of Eq. (10.110) and then move the resulting term to the left-hand side:

$$\alpha_n \cdot R_{Z,X}(n,l) - h_1^{(n)} \cdot R_X(n,l) = \sum_{j=2}^{n+1} h_j^{(n)} \cdot R_X(n+1-j,l)$$

$$(\alpha_n - h_1^{(n)}) \cdot R_{Z,X}(n,l) = \sum_{j=2}^{n+1} h_j^{(n)} \cdot R_X(n+1-j,l)$$

Adjusting the index of summation

$$(\alpha_n - h_1^{(n)}) \cdot R_{Z,X}(n,l) = \sum_{j=1}^{n} h_{j+1}^{(n)} \cdot R_X(n-j,l)$$

By dividing both sides, we finally obtain

$$R_{Z,X}(n,l) = \sum_{j=1}^{n} \frac{h_{j+1}^{(n)}}{\alpha_n - h_1^{(n)}} \cdot R_X(n-j,l) \quad \text{for } l = 0, 1, \ldots, n$$

This set of equations is identical to Eq. (10.105) if we set

$$h_j^{(n-1)} = \frac{h_{j+1}^{(n)}}{\alpha_n - h_1^{(n)}} \quad \text{for } j = 0, 1, \ldots, n$$

Therefore, if at step $n$ we have found $h_1^{(n-1)}, h_2^{(n-1)}, \ldots, h_n^{(n-1)}$, and if somehow we have found $h_1^{(n)}$ then we can find the remaining coefficients from

$$h_{j+1}^{(n)} = (\alpha_n - h_1^{(n)}) \cdot h_j^{(n-1)} \quad \text{for } j = 0, 1, \ldots, n$$

Thus the key question is how to find $h_1^{(n)}$.

Suppose we substitute the coefficients in Eq. (10.113b) into Eq. (10.106):

$$Y_{n+1} = \sum_{j=1}^{n+1} h_j^{(n)} \cdot X_{n+1-j}$$

$$Y_{n+1} = h_1^{(n)} \cdot X_n + \sum_{j=2}^{n+1} h_j^{(n)} \cdot X_{n+1-j} = h_1^{(n)} \cdot X_n + \sum_{j=1}^{n} h_{j+1}^{(n)} \cdot X_{n-j}$$

$$Y_{n+1} = h_1^{(n)} \cdot X_n + (\alpha_n - h_1^{(n)}) \cdot \sum_{j=1}^{n} h_j^{(n-1)} \cdot X_{n-j}$$

$$Y_{n+1} = h_1^{(n)} \cdot X_n + (\alpha_n - h_1^{(n)}) \cdot Y_n$$

$$Y_{n+1} = \alpha_n \cdot Y_n + h_1^{(n)} \cdot (X_n - Y_n)$$

where the second equality follows from Eq. (10.104). The above equation has a very pleasing interpretation, as shown in Fig. 10.16(b). Since \( Y_n \) is the prediction for time \( n \), \( \alpha_n \cdot Y_n \) is the prediction for the next time instant, \( n+1 \), based on the “old” information (see Eq. (10.102)).

The term \((X_n - Y_n)\) is called the “innovations,” and it gives the discrepancy between the old prediction and the observation. Finally, the term \( h^{(n)}_1 \) is called the gain, henceforth denoted by \( k_n \) and it indicates the extent to which the innovations should be used to correct \( \alpha_n \cdot Y_n \) to obtain the “new” prediction \( Y_{n+1} \). If we denote the innovations by

\[
I_n = X_n - Y_n
\]

then Eq. (10.114) becomes

\[
Y_{n+1} = \alpha_n \cdot Y_n + k_n \cdot I_n
\]

We still need to determine a means for computing the gain, \( k_n \).

From Eq. (10.115), we have that the innovations satisfy

\[
I_n = X_n - Y_n = Z_n + N_n - Y_n = e_n + N_n
\]

where \( e_n = Z_n - Y_n \) is the prediction error. A recursive equation can be obtained for the prediction error:

\[
e_{n+1} = Z_{n+1} - Y_{n+1} = \alpha_n \cdot Z_n + W_n - \alpha_n \cdot Y_n - k_n \cdot I_n
\]

\[
e_{n+1} = \alpha_n \cdot (Z_n - Y_n) + W_n - k_n \cdot (e_n + N_n)
\]

\[
e_{n+1} = (\alpha_n - k_n) \cdot e_n + W_n - k_n \cdot N_n
\]

with initial condition \( e_0 = Z_0 \). Since \( X_0, W_n, \) and \( N_n \) are zero-mean, it then follows that \( E[e_0] = 0 \) for all \( n \). A recursive equation for the mean square prediction error is obtained from Eq. (10.117):

\[
E[e_{n+1}^2] = (\alpha_n - k_n)^2 \cdot E[e_n^2] + E[W_n^2] + k_n^2 \cdot E[N_n^2]
\]

with initial condition \( E[e_0^2] = E[Z_0^2] \). We are finally ready to obtain an expression for the gain, \( k_n \).

The gain \( k_n \) must minimize the mean square error \( E[e_{n+1}^2] \). Therefore we can differentiate Eq. (10.118) with respect to \( k_n \) and set it equal to zero:

\[
0 = -2 \cdot (\alpha_n - k_n) \cdot E[e_n^2] + 2 \cdot k_n \cdot E[N_n^2]
\]
Then we can solve for $k_n$:

$$k_n = \frac{\alpha_n \cdot E[e_n^2]}{E[e_n^2] + E[N_n^2]}$$

The expression for the mean square prediction error in Eq. (10.118) can be simplified by using Eq. (10.119) (see Problem 10.72):

$$k_n \cdot (E[e_n^2] + E[N_n^2]) = \alpha_n \cdot E[e_n^2]$$

$$k_n^2 \cdot E[N_n^2] = k_n \cdot (\alpha_n - k_n) \cdot E[e_n^2]$$

$$E[e_{n+1}^2] = (\alpha_n - k_n)^2 \cdot E[e_n^2] + E[W_n^2] + k_n \cdot (\alpha_n - k_n) \cdot E[e_n^2]$$

$$E[e_{n+1}^2] = \alpha_n \cdot (\alpha_n - k_n) \cdot E[e_n^2] + E[W_n^2]$$

Equations (10.119), (10.116), and (10.120) when combined yield the recursive procedure that constitutes the Kalman filtering algorithm:

**Kalman filter algorithm:**

**Initialization:**

$$Y_0 = 0 \quad \text{and} \quad E[e_0^2] = E[Z_0^2]$$

For $n = 0, 1, 2, \ldots$

$$k_n = \frac{\alpha_n \cdot E[e_n^2]}{E[e_n^2] + E[N_n^2]}$$

$$Y_{n+1} = \alpha_n \cdot Y_n + k_n \cdot (X_n - Y_n)$$

$$E[e_{n+1}^2] = \alpha_n \cdot (\alpha_n - k_n) \cdot E[e_n^2] + E[W_n^2]$$

Note that the algorithm requires knowledge of the signal structure, i.e., the $\alpha_n$ and the variances $E[N_n^2]$ and $E[W_n^2]$. The algorithm can be implemented easily and has consequently found application in a broad range of detection, estimation, and signal processing problems. The algorithm can be extended in matrix form to accommodate a broader range of processes.
Example 10.26  First-Order Autoregressive Process

Consider a signal defined by

\[ Z_n = aZ_{n-1} + W_n \quad n = 1, 2, \ldots \quad Z_0 = 0, \]

where \( E[W_n^2] = \sigma_W^2 = 0.36 \), and \( a = 0.8 \), and suppose the observations are made in additive white noise

\[ X_n = Z_n + N_n \quad n = 0, 1, 2, \ldots, \]

where \( E[N_n^2] = 1 \). Find the form of the predictor and its mean square error as \( n \to \infty \).

The gain at step \( n \) is given by

\[ k_n = \frac{aE[e_n^2]}{E[e_n^2] + 1}. \]

The mean square error sequence is therefore given by

\[ E[e_0^2] = E[Z_0^2] = 0 \]

\[ E[e_{n+1}^2] = a(a - k_n)E[e_n^2] + \sigma_W^2 \]

\[ = a \left( \frac{a}{1 + E[e_n^2]} \right) E[e_n^2] + \sigma_W^2 \quad \text{for} \ n = 1, 2, \ldots. \]

The steady state mean square error \( e_\infty \) must satisfy

\[ e_\infty = \frac{a^2}{1 + e_\infty} e_\infty + \sigma_W^2. \]

For \( a = 0.8 \) and \( \sigma_W^2 = 0.36 \), the resulting quadratic equation yields \( k_\infty = 0.3 \) and \( e_\infty = 0.6 \).

Thus at steady state the predictor is

\[ Y_{n+1} = 0.8Y_n + 0.3(X_n - Y_n). \]
10.6 *Estimating the Power Spectral Density

Let \( X_0, X_1, \ldots, X_{k-1} \) be \( k \) observations of the discrete-time, zero-mean, wide-sense stationary process \( X_n \). The periodogram estimate for is defined as

\[
\tilde{p}_k(f) = \frac{1}{k} \cdot |\tilde{x}_k(f)|^2
\]

where \( \tilde{x}_k(f) \) is obtained as a Fourier transform of the observation sequence:

\[
\tilde{x}_k(f) = \sum_{m=0}^{k-1} X_m \cdot \exp(-j \cdot 2\pi \cdot f \cdot m)
\]

In Section 10.1 we showed that the expected value of the periodogram estimate is

\[
E[\tilde{p}_k(f)] = \sum_{m=-k+1}^{k-1} \left( 1 - \frac{|m|}{k} \right) \cdot R_X(m') \cdot \exp(-j \cdot 2\pi \cdot f \cdot (m'))
\]

so \( \tilde{p}_k(f) \) is a biased estimator for \( S_X(f) \). However, as \( k \to \infty \)

\[
E[\tilde{p}_k(f)] \to S_X(f)
\]

so the mean of the periodogram estimate approaches \( S_X(f) \).

Before proceeding to find the variance of the periodogram estimate, we note that the periodogram estimate is equivalent to taking the Fourier transform of an estimate for the autocorrelation sequence; that is,

\[
\tilde{p}_k(f) = \sum_{m=-k+1}^{k-1} \hat{r}_k(m) \cdot \exp(-j \cdot 2\pi \cdot f \cdot m)
\]

where the estimate for the autocorrelation is

\[
\hat{r}_k(m) = \frac{1}{k} \cdot \sum_{n=0}^{k-1} X_n \cdot X_{n+m}
\]

We might expect that as we increase the number of samples \( k \), the periodogram estimate converges to \( S_X(f) \). This does not happen. Instead we find that \( \tilde{p}_k(f) \) fluctuates wildly about the true spectral density, and that this random variation does not decrease with increased \( k \) (see Fig. 10.17).

To see why this happens, in the next section we compute the statistics of the periodogram estimate for a white noise Gaussian random process. We find that the estimates given by the periodogram have a variance that does not approach zero as the number of samples is increased. This explains the lack of improvement in the estimate as \( k \) is increased. Furthermore, we show that the periodogram estimates are uncorrelated at uniformly spaced frequencies in the interval \(-1/2 \leq f < 1/2\). This explains the erratic appearance of the periodogram estimate as a function of \( f \). In the final section, we obtain another estimate for \( S_X(f) \) whose variance does approach zero as \( k \) increases.

10.6.1 Variance of Periodogram Estimate

10.6.2 Smoothing of Periodogram Estimate

A fundamental result in probability theory is that the sample mean of a sequence of independent realizations of a random variable approaches the true mean with probability one. We obtain an estimate for that goes to zero with the number of observations $k$ by taking the average of $N$ independent periodograms on samples of size $k$:

$$\langle \tilde{p}_k(f) \rangle_N = \frac{1}{N} \sum_{i=1}^{N} \tilde{p}_{k,i}(f)$$

where $\tilde{p}_{k,i}(f)$ are $N$ independent periodograms computed using separate sets of $k$ samples each.

The choice of $k$ and $N$ is determined by the desired frequency resolution and variance of the estimate. The blocksize $k$ determines the number of frequencies for which the spectral density is computed (i.e., the frequency resolution). The variance of the estimate is controlled by the number of periodograms $N$. The actual choice of $k$ and $N$ depends on the nature of the signal being investigated.
10.7 Numerical Techniques for Processing Random Signals

skipped

10.7.1 FFT Techniques

10.7.2 Filtering Techniques

10.7.3 Generation of Random Sequences
Summary

- The power spectral density of a WSS process is the Fourier transform of its autocorrelation function. The power spectral density of a real-valued random process is a real-valued, nonnegative, even function of frequency.

- The output of a linear, time-invariant system is a WSS random process if its input is a WSS random process that is applied an infinite time in the past.

- The output of a linear, time-invariant system is a Gaussian WSS random process if its input is a Gaussian WSS random process.

- Wide-sense stationary random processes with arbitrary rational power spectral density can be generated by filtering white noise.

- The sampling theorem allows the representation of bandlimited continuous-time processes by the sequence of periodic samples of the process.

- The orthogonality condition can be used to obtain equations for linear systems that minimize mean square error. These systems arise in filtering, smoothing, and prediction problems. Matrix numerical methods are used to find the optimum linear systems.

- The Kalman filter can be used to estimate signals with a structure that keeps the dimensionality of the algorithm fixed even as the size of the observation set increases.

- The variance of the periodogram estimate for the power spectral density does not approach zero as the number of samples is increased. An average of several independent periodograms is required to obtain an estimate whose variance does approach zero as the number of samples is increased.

- The FFT, convolution, and matrix techniques are basic tools for analyzing, simulating, and implementing processing of random signals.
CHECKLIST OF IMPORTANT TERMS

- Amplitude modulation
- ARMA process
- Autoregressive process
- Bandpass signal
- Causal system
- Cross-power spectral density
- Einstein-Wiener-Khinchin theorem
- Filtering
- Impulse response
- Innovations
- Kalman filter
- Linear system
- Long-range dependence
- Moving average process
- Nyquist sampling rate
- Optimum filter
- Orthogonality condition
- Periodogram
- Power spectral density
- Prediction
- Quadrature amplitude modulation
- Sampling theorem
- Smoothed periodogram
- Smoothing
- System
- Time-invariant system
- Transfer function
- Unit-sample response
- White noise
- Wiener filter
- Wiener-Hopf equations
- Yule-Walker equations