Probability and Stochastic Processes

A Friendly Introduction for Electrical and Computer Engineers

Third Edition

SIGNAL PROCESSING SUPPLEMENT (SPS)

Roy D. Yates David J. Goodman

February 4, 2014

This supplement uses a page size matched to the screen of an iPad tablet. If you do print on paper and you have good eyesight, you may wish to print two pages per sheet in landscape mode. On the other hand, a "Fit to Paper" printing option will create "Large Print" output.

Introduction

In designing new equipment and evaluating the performance of existing systems, electrical and computer engineers frequently represent electrical signals as sample functions of wide sense stationary stochastic processes. We use probability density functions and probability mass functions to describe the amplitude characteristics of signals, and we use autocorrelation functions to describe the time-varying nature of the signals. Many important signals – for example, brightness levels of television picture elements – appear as sample functions of random sequences. Others – for example, audio waveforms – appear as sample functions of continuous-time processes. However, practical equipment increasingly uses digital signal processing to perform many operations on continuous-time signals. To do so, the equipment contains an analog-to-digital converter to transform a continuous-time signal to a random sequence. An analog-to-digital converter performs two operations: sampling and quantization. Sampling with a period T_s seconds transforms a continuous-time process $X(t)$ to a random sequence $X_n = X(nT_s)$. Quantization transforms the continuous random variable X_n to a discrete random variable Q_n.

In this chapter we ignore quantization and analyze *linear filtering* of random processes and random sequences resulting from sampling random processes. Linear filtering is a practical technique with many applications. For example, we will see that when we apply the linear estimation methods developed in Chapter 12 to a random process, the result is a linear filter. We will use the Fourier transform of the autocorrelation and cross-correlation functions to develop frequency domain techniques for the analysis of random signals.

1 Linear Filtering of a Continuous-Time Stochastic Process

The output of a linear time-invariant filter is a wide sense stationary stochastic process if the input is a wide sense stationary process.

We begin by describing the relationship of the stochastic process at the output of a linear filter to the stochastic process at the input of the filter. Consider a linear time-invariant (LTI) filter with impulse response $h(t)$. If the input is a deterministic signal $v(t)$, the output, $w(t)$, is the convolution

$$w(t) = \int_{-\infty}^{\infty} h(u)v(t-u)\,du = \int_{-\infty}^{\infty} h(t-u)v(u)\,du. \tag{1}$$

If the possible inputs to the filter are $x(t)$, sample functions of a stochastic process $X(t)$, then the outputs, $y(t)$, are sample functions of another stochastic process, $Y(t)$. Because $y(t)$ is the convolution of $x(t)$ and $h(t)$, we adopt the following notation for the relationship of $Y(t)$ to $X(t)$:

$$Y(t) = \int_{-\infty}^{\infty} h(u)X(t-u)\,du = \int_{-\infty}^{\infty} h(t-u)X(u)\,du. \tag{2}$$

Similarly, the expected value of $Y(t)$ is the convolution of $h(t)$ and $\mathrm{E}[X(t)]$.

Theorem 1

$$\mathrm{E}\left[Y(t)\right] = \mathrm{E}\left[\int_{-\infty}^{\infty} h(u)X(t-u)\,du\right] = \int_{-\infty}^{\infty} h(u)\,\mathrm{E}\left[X(t-u)\right]\,du.$$

When $X(t)$ is wide sense stationary, we use Theorem 1 to derive $R_Y(t,\tau)$ and $R_{XY}(t,\tau)$ as functions of $h(t)$ and $R_X(\tau)$. We will observe that when $X(t)$ is wide sense stationary, $Y(t)$ is also wide sense stationary.

Theorem 2

If the input to an LTI filter with impulse response $h(t)$ is a wide sense stationary process $X(t)$, the output $Y(t)$ has the following properties:

(a) $Y(t)$ is a wide sense stationary process with expected value

$$\mu_Y = \mathrm{E}\left[Y(t)\right] = \mu_X \int_{-\infty}^{\infty} h(u)\,du,$$

and autocorrelation function

$$R_Y(\tau) = \int_{-\infty}^{\infty} h(u) \int_{-\infty}^{\infty} h(v) R_X(\tau + u - v)\, dv\, du.$$

(b) *$X(t)$ and $Y(t)$ are jointly wide sense stationary and have input-output cross-correlation*

$$R_{XY}(\tau) = \int_{-\infty}^{\infty} h(u) R_X(\tau - u)\, du.$$

(c) *The output autocorrelation is related to the input-output cross-correlation by*

$$R_Y(\tau) = \int_{-\infty}^{\infty} h(-w) R_{XY}(\tau - w)\, dw.$$

Proof We recall from Theorem 1 that $\mathrm{E}[Y(t)] = \int_{-\infty}^{\infty} h(u)\, \mathrm{E}[X(t-u)]\, du$. Since $\mathrm{E}[X(t)] = \mu_X$ for all t, $\mathrm{E}[Y(t)] = \int_{-\infty}^{\infty} h(u)\mu_X\, du$, which is independent of t. Next, we observe that Equation (2) implies $Y(t+\tau) = \int_{-\infty}^{\infty} h(v) X(t+\tau-u)\, dv$. Thus

$$\begin{aligned} R_Y(t,\tau) &= \mathrm{E}\left[Y(t)Y(t+\tau)\right] \\ &= \mathrm{E}\left[\int_{-\infty}^{\infty} h(u)X(t-u)\, du \int_{-\infty}^{\infty} h(v)X(t+\tau-v)\, dv\right] \\ &= \int_{-\infty}^{\infty} h(u) \int_{-\infty}^{\infty} h(v)\, \mathrm{E}\left[X(t-u)X(t+\tau-v)\right]\, dv\, du. \end{aligned} \tag{3}$$

Because $X(t)$ is wide sense stationary, $\mathrm{E}[X(t-u)X(t+\tau-v)] = R_X(\tau - v + u)$ so that

$$R_Y(t,\tau) = R_Y(\tau) = \int_{-\infty}^{\infty} h(u) \int_{-\infty}^{\infty} h(v) R_X(\tau - v + u)\, dv\, du. \tag{4}$$

By Definition 13.15, $Y(t)$ is wide sense stationary because $\mathrm{E}[Y(t)]$ is independent of time t, and $R_Y(t,\tau)$ depends only on the time shift τ.

The input-output cross-correlation is

$$R_{XY}(t,\tau) = \mathrm{E}\left[X(t)\int_{-\infty}^{\infty} h(v)X(t+\tau-v)\,dv\right]$$
$$= \int_{-\infty}^{\infty} h(v)\,\mathrm{E}\left[X(t)X(t+\tau-v)\right]\,dv = \int_{-\infty}^{\infty} h(v)R_X(\tau-v)\,dv. \quad (5)$$

Thus $R_{XY}(t,\tau) = R_{XY}(\tau)$, and by Definition 13.18, $X(t)$ and $Y(t)$ are jointly wide sense stationary. From Theorem 2(a) and Theorem 2(b), we observe that

$$R_Y(\tau) = \int_{-\infty}^{\infty} h(u) \underbrace{\int_{-\infty}^{\infty} h(v)R_X(\tau+u-v)\,dv}_{R_{XY}(\tau+u)}\,du = \int_{-\infty}^{\infty} h(u)R_{XY}(\tau+u)\,du. \quad (6)$$

The substitution $w=-u$ yields $R_Y(\tau) = \int_{-\infty}^{\infty} h(-w)R_{XY}(\tau-w)\,dw$, to complete the derivation.

Example 1

$X(t)$, a wide sense stationary stochastic process with expected value $\mu_X = 10$ volts, is the input to a linear time-invariant filter. The filter impulse response is

$$h(t) = \begin{cases} e^{t/0.2} & 0 \le t \le 0.1 \text{ sec}, \\ 0 & \text{otherwise}. \end{cases} \quad (7)$$

What is the expected value of the filter output process $Y(t)$?

Applying Theorem 2, we have

$$\mu_Y = \mu_X \int_{-\infty}^{\infty} h(t)\,dt = 10\int_0^{0.1} e^{t/0.2}\,dt = 2\left(e^{0.5}-1\right) = 1.30 \text{ V}. \quad (8)$$

Example 2

A white Gaussian noise process $W(t)$ with autocorrelation function

$$R_W(\tau) = \eta_0\delta(\tau) \quad (9)$$

is passed through the moving-average filter

$$h(t) = \begin{cases} 1/T & 0 \le t \le T, \\ 0 & \text{otherwise.} \end{cases} \tag{10}$$

For the output $Y(t)$, find the expected value $\mathrm{E}[Y(t)]$, the input-output cross-correlation $R_{WY}(\tau)$ and the autocorrelation $R_Y(\tau)$.

. . . .

We recall from Definition 13.20 that a white noise process $W(t)$ has expected value $\mu_W = 0$. The expected value of the output is $\mathrm{E}[Y(t)] = \mu_W \int_{-\infty}^{\infty} h(t)\,dt = 0$. From Theorem 2(b), the input-output cross-correlation is

$$R_{WY}(\tau) = \frac{\eta_0}{T}\int_0^T \delta(\tau - u)\,du = \begin{cases} \eta_0/T & 0 \le t \le T, \\ 0 & \text{otherwise.} \end{cases} \tag{11}$$

By Theorem 2(c), the output autocorrelation is

$$R_Y(\tau) = \int_{-\infty}^{\infty} h(-v) R_{WY}(\tau - v)\,dv = \frac{1}{T}\int_{-T}^{0} R_{WY}(\tau - v)\,dv. \tag{12}$$

To evaluate this convolution integral, we must express $R_{WY}(\tau - v)$ as a function of v. From Equation (11),

$$R_{WY}(\tau - v) = \begin{cases} \frac{\eta_0}{T} & 0 \le \tau - v \le T \\ 0 & \text{otherwise} \end{cases} = \begin{cases} \frac{\eta_0}{T} & \tau - T \le v \le \tau, \\ 0 & \text{otherwise.} \end{cases} \tag{13}$$

Thus, $R_{WY}(\tau - v)$ is nonzero over the interval $\tau - T \le v \le \tau$. However, in Equation (12), we integrate $R_{WY}(\tau - v)$ over the interval $-T \le v \le 0$. If $\tau < -T$ or $\tau - T > 0$, then these two intervals do not overlap and the integral (12) is zero. Otherwise, for $-T \le \tau \le 0$, it follows from Equation (12) that

$$R_Y(\tau) = \frac{1}{T}\int_{-T}^{\tau} \frac{\eta_0}{T}\,dv = \frac{\eta_0(T + \tau)}{T^2}. \tag{14}$$

Furthermore, for $0 \le \tau \le T$,

$$R_Y(\tau) = \frac{1}{T}\int_{\tau - T}^{0} \frac{\eta_0}{T}\,dv = \frac{\eta_0(T - \tau)}{T^2}. \tag{15}$$

Putting these pieces together, the output autocorrelation is the triangular function

$$R_Y(\tau) = \begin{cases} \eta_0(T - |\tau|)/T^2 & |\tau| \leq T, \\ 0 & \text{otherwise.} \end{cases} \tag{16}$$

In Section 13.11, we observed that a white Gaussian noise $W(t)$ is physically impossible because it has average power $\mathrm{E}[W^2(t)] = \eta_0\delta(0) = \infty$. However, when we filter white noise, the output will be a process with finite power. In Example 2, the output of the moving-average filter $h(t)$ has finite average power $R_Y(0) = \eta_0/T$. This conclusion is generalized in Problem 1.4.

Theorem 2 provides mathematical procedures for deriving the expected value and the autocorrelation function of the stochastic process at the output of a filter from the corresponding functions at the input. In general, however, it is much more complicated to determine the PDF or PMF of the output given the corresponding probability function of the input. One exception occurs when the filter input is a Gaussian stochastic process. The following theorem states that the output is also a Gaussian stochastic process.

Theorem 3

If a stationary Gaussian process $X(t)$ is the input to a LTI filter $h(t)$, the output $Y(t)$ is a stationary Gaussian process with expected value and autocorrelation given by Theorem 2.

Although a proof of Theorem 3 is beyond the scope of this text, we observe that Theorem 3 is analagous to Theorem 8.11 which shows that a linear transformation of jointly Gaussian random variables yields jointly Gaussian random variables. In particular, for a Gaussian input process $X(t)$, the convolution integral of Equation (2) can be viewed as the limiting case of a linear transformation of jointly Gaussian random variables. Theorem 7 proves that if the input to a discrete-time filter is Gaussian, the output is also Gaussian.

Example 3

For the white noise moving-average process $Y(t)$ in Example 2, let $\eta_0 = 10^{-15}$ W/Hz and $T = 10^{-3}$ s. For an arbitrary time t_0, find $\mathrm{P}[Y(t_0) > 4 \cdot 10^{-6}]$.

Applying the given parameter values $\eta = 10^{-15}$ and $T = 10^{-3}$ to Equation (16), we learn that

$$R_Y[\tau] = \begin{cases} 10^{-9}(10^{-3} - |\tau|) & |\tau| \leq 10^{-3} \\ 0 & \text{otherwise.} \end{cases} \tag{17}$$

By Theorem 3, $Y(t)$ is a stationary Gaussian process and thus $Y(t_0)$ is a Gaussian random variable. In particular, since $\mathrm{E}[Y(t_0)] = 0$, we see from Equation (17) that $Y(t_0)$ has variance $\mathrm{Var}[Y(t_0)] = R_Y(0) = 10^{-12}$. This implies

$$\begin{aligned} \mathrm{P}\left[Y(t_0) > 4 \cdot 10^{-6}\right] &= \mathrm{P}\left[\frac{Y(t_0)}{\sqrt{\mathrm{Var}[Y(t_0)]}} > \frac{4 \cdot 10^{-6}}{10^{-6}}\right] \\ &= Q(4) = 3.17 \cdot 10^{-5}. \end{aligned} \tag{18}$$

Quiz 1

Let $h(t)$ be a low-pass filter with impulse response

$$h(t) = \begin{cases} e^{-t} & t \geq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{19}$$

The input to the filter is $X(t)$, a wide sense stationary random process with expected value $\mu_X = 2$ and autocorrelation $R_X(\tau) = \delta(\tau)$. What are the expected value and autocorrelation of the output process $Y(t)$?

2 Linear Filtering of a Random Sequence

The output of a linear time-invariant digital filter is a wide sense stationary random sequence if the input is a wide sense random sequence. The autocorrelation function of the filter output is determined by the autocorrelation function of the input process and the filter impulse response. The relationship is expressed in a double summation.

A strong trend in practical electronics is to use a specialized microcomputer referred to as a *digital signal processor (DSP)* to perform signal processing operations. To use a DSP as a linear filter it is necessary to convert the input signal $x(t)$ to a sequence of samples $x(nT)$, where $n = \cdots, -1, 0, 1, \cdots$ and $1/T$ Hz is the sampling rate. When the continuous-time signal is a sample function of a stochastic process, $X(t)$, the sequence of samples at the input to the digital signal processor is a sample function of a random sequence $X_n = X(nT)$. The autocorrelation function of the random sequence X_n consists of samples of the autocorrelation function of the continuous-time process $X(t)$.

Theorem 4

The random sequence X_n is obtained by sampling the continuous-time process $X(t)$ at a rate of $1/T_s$ samples per second. If $X(t)$ is a wide sense stationary process with expected value $\mathrm{E}[X(t)] = \mu_X$ and autocorrelation $R_X(\tau)$, then X_n is a wide sense stationary random sequence with expected value $\mathrm{E}[X_n] = \mu_X$ and autocorrelation function

$$R_X[k] = R_X(kT_s).$$

Proof Because the sampling rate is $1/T_s$ samples per second, the random variables in X_n are random variables in $X(t)$ occurring at intervals of T_s seconds: $X_n = X(nT_s)$. Therefore, $\mathrm{E}[X_n] = \mathrm{E}[X(nT_s)] = \mu_X$ and

$$R_X[k] = \mathrm{E}[X_n X_{n+k}] = \mathrm{E}[X(nT_s)X([n+k]T_s)] = R_X(kT_s). \tag{20}$$

Example 4

Continuing Example 3, the random sequence Y_n is obtained by sampling the white noise moving-average process $Y(t)$ at a rate of $f_s = 10^4$ samples per second. Derive the autocorrelation function $R_Y[n]$ of Y_n.

. . . .

Given the autocorrelation function $R_Y(\tau)$ in Equation (17), Theorem 4 implies that

$$R_Y[k] = R_Y(k10^{-4}) = \begin{cases} 10^{-9}(10^{-3} - 10^{-4}|k|) & |k10^{-4}| \le 10^{-3} \\ 0 & \text{otherwise} \end{cases}$$

$$= \begin{cases} 10^{-6}(1 - 0.1|k|) & |k| \le 10, \\ 0 & \text{otherwise.} \end{cases} \tag{21}$$

DSP performs *discrete-time linear filtering*. The impulse response of a discrete-time filter is a sequence h_n, $n = \ldots, -1, 0, 1, \ldots$ and the output is a random sequence Y_n, related to the input X_n by the discrete-time convolution,

$$Y_n = \sum_{i=-\infty}^{\infty} h_i X_{n-i}. \tag{22}$$

Corresponding to Theorem 2 for continuous-time processes we have the equivalent theorem for discrete-time processes.

Theorem 5

If the input to a discrete-time linear time-invariant filter with impulse response h_n is a wide sense stationary random sequence, X_n, the output Y_n has the following properties.

(a) Y_n is a wide sense stationary random sequence with expected value

$$\mu_Y = \mathrm{E}[Y_n] = \mu_X \sum_{n=-\infty}^{\infty} h_n,$$

and autocorrelation function

$$R_Y[n] = \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} h_i h_j R_X[n + i - j].$$

(b) Y_n and X_n are jointly wide sense stationary with input-output cross-correlation

$$R_{XY}[n] = \sum_{i=-\infty}^{\infty} h_i R_X[n - i].$$

(c) *The output autocorrelation is related to the input-output cross-correlation by*

$$R_Y[n] = \sum_{i=-\infty}^{\infty} h_{-i} R_{XY}[n-i].$$

Example 5

A wide sense stationary random sequence X_n with $\mu_X = 1$ and autocorrelation function $R_X[n]$ is the input to the order $M-1$ discrete-time moving-average filter h_n where

$$h_n = \begin{cases} 1/M & n = 0, \ldots, M-1 \\ 0 & \text{otherwise,} \end{cases} \quad \text{and} \quad R_X[n] = \begin{cases} 4 & n = 0, \\ 2 & n = \pm 1, \\ 0 & |n| \geq 2. \end{cases} \tag{23}$$

For the case $M = 2$, find the following properties of the output random sequence Y_n: the expected value μ_Y, the autocorrelation $R_Y[n]$, and the variance $\mathrm{Var}[Y_n]$.

For this filter with $M = 2$, we have from Theorem 5,

$$\mu_Y = \mu_X(h_0 + h_1) = \mu_X = 1. \tag{24}$$

By Theorem 5(a), the autocorrelation of the filter output is

$$\begin{aligned} R_Y[n] &= \sum_{i=0}^{1} \sum_{j=0}^{1} (0.25) R_X[n+i-j] \\ &= (0.5) R_X[n] + (0.25) R_X[n-1] + (0.25) R_X[n+1]. \end{aligned} \tag{25}$$

Substituting $R_X[n]$ from the problem statement, we obtain

$$R_Y[n] = \begin{cases} 3 & n = 0, \\ 2 & |n| = 1, \\ 0.5 & |n| = 2, \\ 0 & \text{otherwise.} \end{cases} \tag{26}$$

To obtain $\mathrm{Var}[Y_n]$, we recall from Definition 13.16 that $\mathrm{E}[Y_n^2] = R_Y[0] = 3$. Therefore, $\mathrm{Var}[Y_n] = \mathrm{E}[Y_n^2] - \mu_Y^2 = 2$.

Note in Example 5 that the filter output is the average of the most recent input sample and the previous input sample. Consequently, we could also use Theorem 9.1 and Theorem 9.2 to find the expected value and variance of the output process.

Theorem 5 presents the important properties of discrete-time filters in a general way. In designing and analyzing practical filters, electrical and computer engineers usually confine their attention to *causal* filters with the property that $h_n = 0$ for $n < 0$. They consider filters in two categories:

- *Finite impulse response (FIR)* filters with $h_n = 0$ for $n \geq M$, where M is a positive integer. Note that $M - 1$ is called the *order* of the filter. The input-output convolution is

$$Y_n = \sum_{i=0}^{M-1} h_i X_{n-i}. \tag{27}$$

- *Infinite impulse response (IIR)* filters such that for any positive integer N, $h_n \neq 0$ for at least one value of $n > N$. The practical realization of an IIR filter is recursive. Each filter output is a linear combination of a finite number of input samples and a finite number of previous output samples:

$$Y_n = \sum_{i=0}^{\infty} h_i X_{n-i} = \sum_{i=0}^{M-1} a_i X_{n-i} + \sum_{j=1}^{N} b_j Y_{n-j}. \tag{28}$$

  In this formula, the $a_0, a_2, \ldots, a_{M-1}$ are the coefficients of the *forward section* of the filter and $b_1, b_2, \ldots, b_N$ are the coefficients of the *feedback section*.

Example 6

Why does the index i in the forward section of an IIR filter start at $i = 0$ whereas the index j in the feedback section starts at 1?

In the forward section, the output Y_n depends on the *current* input X_n as well as past inputs. However, in the feedback section, the filter at sample n has access only to *previous* outputs $Y_{n-1}, Y_{n-2}, \ldots$ in producing Y_n.

Example 7

A wide sense stationary random sequence X_n with expected value $\mu_X = 0$ and autocorrelation function $R_X[n] = \sigma^2 \delta_n$ is passed through the order $M-1$ discrete-time moving-average filter

$$h_n = \begin{cases} 1/M & 0 \le n \le M-1, \\ 0 & \text{otherwise.} \end{cases} \tag{29}$$

Find the output autocorrelation $R_Y[n]$.

In Example 5, we found the output autocorrelation directly from Theorem 5(a) because the averaging filter was a simple first-order filter. For a filter of arbitrary order $M-1$, we first find the input-output cross-correlation $R_{XY}[k]$ and then use Theorem 5(c) to find $R_Y[k]$. From Theorem 5(b), the input-output cross-correlation is

$$R_{XY}[k] = \frac{\sigma^2}{M} \sum_{i=0}^{M-1} \delta_{k-i} = \begin{cases} \sigma^2/M & k = 0, 1, \ldots, M-1, \\ 0 & \text{otherwise.} \end{cases} \tag{30}$$

From Theorem 5(c), the output autocorrelation is

$$R_Y[n] = \sum_{i=-\infty}^{\infty} h_{-i} R_{XY}[n-i] = \frac{1}{M} \sum_{i=-(M-1)}^{0} R_{XY}[n-i]. \tag{31}$$

To express $R_{XY}[n-i]$ as a function of i, we use Equation (30) by replacing k in $R_{XY}[k]$ with $n-i$, yielding

$$R_{XY}[n-i] = \begin{cases} \sigma^2/M & i = n-M+1, n-M+2, \ldots, n, \\ 0 & \text{otherwise.} \end{cases} \tag{32}$$

Thus, if $n < -(M-1)$ or $n - M + 1 > 0$, then $R_{XY}[n-i] = 0$ over the interval $-(M-1) \leq i \leq M-1$. For $-(M-1) \leq n \leq 0$,

$$R_Y[n] = \frac{1}{M} \sum_{i=-(M-1)}^{n} \frac{\sigma^2}{M} = \frac{\sigma^2(M+n)}{M^2}. \tag{33}$$

For $0 \leq n \leq M-1$,

$$R_Y[n] = \frac{1}{M} \sum_{i=n-(M-1)}^{0} \frac{\sigma^2}{M} = \frac{\sigma^2(M-n)}{M^2}. \tag{34}$$

Combining these cases, the output autocorrelation is the triangular function

$$R_Y[n] = \begin{cases} \sigma^2(M-|n|)/M^2 & -(M-1) \leq n \leq M-1, \\ 0 & \text{otherwise.} \end{cases} \tag{35}$$

Example 8

A first-order discrete-time integrator with wide sense stationary input sequence X_n has output

$$Y_n = X_n + 0.8Y_{n-1}. \tag{36}$$

What is the filter impulse response h_n?

....

We find the impulse response by repeatedly substituting for Y_k on the right side of Equation (36). To begin, we replace Y_{n-1} with the expression in parentheses in the following formula

$$Y_n = X_n + 0.8(X_{n-1} + 0.8Y_{n-2}) = X_n + 0.8X_{n-1} + 0.8^2 Y_{n-2}. \tag{37}$$

We can continue this procedure, replacing successive values of Y_k. For the third step, replacing Y_{n-2} yields

$$Y_n = X_n + 0.8X_{n-1} + 0.8^2 X_{n-2} + 0.8^3 Y_{n-3}. \tag{38}$$

After k steps of this process, we obtain

$$Y_n = X_n + 0.8X_{n-1} + 0.8^2 X_{n-2} + \cdots + 0.8^k X_{n-k} + 0.8^{k+1} Y_{n-k-1}. \tag{39}$$

Continuing the process indefinitely, we infer that $Y_n = \sum_{k=0}^{\infty} 0.8^k X_{n-k}$. Comparing this formula with Equation (22), we obtain

$$h_n = \begin{cases} 0.8^n & n = 0, 1, 2, \ldots \\ 0 & \text{otherwise.} \end{cases} \tag{40}$$

Example 9

Continuing Example 8, suppose the wide sense stationary input X_n with expected value $\mu_X = 0$ and autocorrelation function

$$R_X[n] = \begin{cases} 1 & n = 0, \\ 0.5 & |n| = 1, \\ 0 & |n| \geq 2, \end{cases} \tag{41}$$

is the input to the first-order integrator h_n. Find the second moment, $\mathrm{E}[Y_n^2]$, of the output.

....

We start by expanding the square of Equation (36):

$$\begin{aligned} \mathrm{E}\left[Y_n^2\right] &= \mathrm{E}\left[X_n^2 + 2(0.8)X_n Y_{n-1} + 0.8^2 Y_{n-1}^2\right] \\ &= \mathrm{E}\left[X_n^2\right] + 2(0.8)\,\mathrm{E}\left[X_n Y_{n-1}\right] + 0.8^2\,\mathrm{E}\left[Y_{n-1}^2\right]. \end{aligned} \tag{42}$$

Because Y_n is wide sense stationary, $\mathrm{E}[Y_{n-1}^2] = \mathrm{E}[Y_n^2]$. Moreover, since

$$\mathrm{E}\left[X_n X_{n-1-i}\right] = R_X[i+1], \tag{43}$$

we see that

$$\begin{aligned} \mathrm{E}\left[X_n Y_{n-1}\right] &= \mathrm{E}\left[X_n(X_{n-1} + 0.8X_{n-2} + 0.8^2 X_{n-3} \ldots)\right] \\ &= \sum_{i=0}^{\infty} 0.8^i R_X[i+1] = R_X[1] = 0.5. \end{aligned} \tag{44}$$

Except for the first term ($i = 0$), all terms in the sum are zero because $R_X[n] = 0$ for $n > 1$. Therefore, from Equation (42),

$$\mathrm{E}\left[Y_n^2\right] = \mathrm{E}\left[X_n^2\right] + 2(0.8)(0.5) + 0.8^2\,\mathrm{E}\left[Y_n^2\right]. \tag{45}$$

With $\mathrm{E}[X_n^2] = R_X[0] = 1$, $\mathrm{E}[Y_n^2] = 1.8/(1 - 0.8^2) = 2.8125$.

Quiz 2

The input to a first-order discrete-time differentiator h_n is X_n, a wide sense stationary Gaussian random sequence with $\mu_X = 0.5$ and autocorrelation function $R_X[k]$. Given

$$R_X[k] = \begin{cases} 1 & k = 0, \\ 0.5 & |k| = 1, \\ 0 & \text{otherwise,} \end{cases} \qquad \text{and} \qquad h_n = \begin{cases} 1 & n = 0, \\ -1 & n = 1, \\ 0 & \text{otherwise,} \end{cases} \tag{46}$$

find the following properties of the output random sequence Y_n: the expected value μ_Y, the autocorrelation $R_Y[n]$, the variance $\mathrm{Var}[Y_n]$.

3 Discrete-Time Linear Filtering: Vectors and Matrices

> Discrete-time signals can be represented as vectors and discrete-time filters as matrices. The vector/matrix notation is more concise than convolutional sums and allows us to understand the properties of discrete-time linear filters in terms of results for random vectors.

In many applications, it is convenient to think of a digital filter in terms of the filter impulse response h_n and the convolution sum of Equation (22). However, in this section, we will observe that we can also represent discrete-time signals as vectors and discrete-time filters as matrices. This can offer some advantages. The vector/matrix notation is generally more concise than convolutional sums. Second, vector notation allows us to understand the properties of discrete-time linear filters in terms of results for random vectors derived in Chapter 8. Lastly, vector notation is the first step to implementing LTI filters in MATLAB.

We can represent a discrete-time input sequence X_n by the L-dimensional vector of samples $\mathbf{X} = \begin{bmatrix} X_0 & \cdots & X_{L-1} \end{bmatrix}'$. From Equation (27), the output at time n of an order $M-1$ FIR filter depends on the M-dimensional vector

$$\mathbf{X}_n = \begin{bmatrix} X_{n-M+1} & \cdots & X_n \end{bmatrix}', \tag{47}$$

which holds the M most recent samples of X_n. The subscript n of the vector $\mathbf{X}_n$ denotes that the most recent observation of the process X_n is at time n. For both $\mathbf{X}$ and $\mathbf{X}_n$, we order the vectors' elements in increasing time index. In the following, we describe properties of $\mathbf{X}_n$, while noting that $\mathbf{X}_n = \mathbf{X}$ when $n = L-1$ and $M = L$.

Theorem 6

If X_n is a wide sense stationary process with expected value μ and autocorrelation function $R_X[k]$, then the vector $\mathbf{X}_n$ has correlation matrix $\mathbf{R}_{\mathbf{X}_n}$ and expected value $\mathrm{E}[\mathbf{X}_n]$ given by

$$\mathbf{R}_{\mathbf{X}_n} = \begin{bmatrix} R_X[0] & R_X[1] & \cdots & R_X[M-1] \\ R_X[1] & R_X[0] & \ddots & \vdots \\ \vdots & \ddots & \ddots & R_X[1] \\ R_X[M-1] & \cdots & R_X[1] & R_X[0] \end{bmatrix}, \qquad \mathrm{E}[\mathbf{X}_n] = \mu \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}.$$

The matrix $\mathbf{R}_{\mathbf{X}_n}$ has a special structure. There are only M different numbers among the M^2 elements of the matrix and each diagonal of $\mathbf{R}_{\mathbf{X}_n}$ consists of identical elements. This matrix is in a category referred to as *Toeplitz forms*. The Toeplitz structure simplifies the computation of the inverse $\mathbf{R}_{\mathbf{X}_n}^{-1}$.

Example 10

The wide sense stationary sequence X_n has autocorrelation $R_X[n]$ as given in Example 5. Find the correlation matrix of $\mathbf{X}_{33} = \begin{bmatrix} X_{30} & X_{31} & X_{32} & X_{33} \end{bmatrix}'$.

From Theorem 6, $\mathbf{X}_{33}$ has length $M = 4$ and Toeplitz correlation matrix

$$\mathbf{R}_{\mathbf{X}_{33}} = \begin{bmatrix} R_X[0] & R_X[1] & R_X[2] & R_X[3] \\ R_X[1] & R_X[0] & R_X[1] & R_X[2] \\ R_X[2] & R_X[1] & R_X[0] & R_X[1] \\ R_X[3] & R_X[2] & R_X[1] & R_X[0] \end{bmatrix} = \begin{bmatrix} 4 & 2 & 0 & 0 \\ 2 & 4 & 2 & 0 \\ 0 & 2 & 4 & 2 \\ 0 & 0 & 2 & 4 \end{bmatrix}. \tag{48}$$

We represent an order $M-1$ LTI FIR filter h_n by the row vector

$$\mathbf{h}' = \begin{bmatrix} h_0 & \cdots & h_{M-1} \end{bmatrix}. \tag{49}$$

With input X_n, the output Y_n at time n, as given by the input-output convolution of Equation (27), can be expressed conveniently in vector notation as

$$Y_n = \overleftarrow{\mathbf{h}}'\mathbf{X}_n \tag{50}$$

where $\overleftarrow{\mathbf{h}'} = \begin{bmatrix} h_{M-1} & \cdots & h_0 \end{bmatrix}$. Because of the equivalence of the filter h_n and the vector $\mathbf{h}'$, it is common terminology to refer to a filter vector $\mathbf{h}'$ as simply an FIR filter.[1] As we did for the signal vector $\mathbf{X}_n$, we represent the FIR filter $\mathbf{h}$ with elements in increasing time index. However, as we observe in Equation (50), discrete-time convolution with the filter h_n employs the filter vector $\mathbf{h}$ with its elements in time-reversed order. In this case, we put a left arrow over $\mathbf{h}$, as in $\overleftarrow{\mathbf{h}}$, for the time-reversed version of $\mathbf{h}$.

Example 11

The order $M-1$ averaging filter h_n given in Example 7 can be represented by the M element vector

$$\mathbf{h} = \frac{1}{M}\begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix}'. \tag{51}$$

[1] In fact, it is convenient to refer to both $\mathbf{h}'$ and its corresponding column vector $\mathbf{h}$ as a filter.

For an FIR filter $\mathbf{h}$, it is often desirable to process a block of inputs represented by the vector $\mathbf{X} = \begin{bmatrix} X_0 & \cdots & X_{L-1} \end{bmatrix}'$. The output vector $\mathbf{Y}$ is the discrete convolution of $\mathbf{h}$ and $\mathbf{X}$. The nth element of $\mathbf{Y}$ is Y_n given by the discrete-time convolution of Equation (27) or, equivalently, Equation (50). In the implementation of discrete convolution of vectors, it is customary to assume $\mathbf{X}$ is padded with zeros such that $X_i = 0$ for $i < 0$ and for $i > N-1$. In this case, the discrete convolution (50) yields

$$Y_0 = h_0 X_0, \tag{52}$$

$$Y_1 = h_1 X_0 + h_0 X_1, \tag{53}$$

and so on. The output vector $\mathbf{Y}$ is related to $\mathbf{X}$ by $\mathbf{Y} = \mathbf{H}\mathbf{X}$ where

$$\underbrace{\begin{bmatrix} Y_0 \\ \vdots \\ Y_{M-1} \\ \vdots \\ \vdots \\ Y_{L-1} \\ \vdots \\ Y_{L+M-2} \end{bmatrix}}_{\mathbf{Y}} = \underbrace{\begin{bmatrix} h_0 & & & & & \\ \vdots & \ddots & & & & \\ h_{M-1} & \cdots & h_0 & & & \\ & \ddots & & \ddots & & \\ & & \ddots & & \ddots & \\ & & & h_{M-1} & \cdots & h_0 \\ & & & & \ddots & \vdots \\ & & & & & h_{M-1} \end{bmatrix}}_{\mathbf{H}} \underbrace{\begin{bmatrix} X_0 \\ \vdots \\ X_{M-1} \\ \vdots \\ \vdots \\ X_{L-M} \\ \vdots \\ X_{L-1} \end{bmatrix}}_{\mathbf{X}}. \tag{54}$$

Just as for continuous-time processes, it is generally difficult to determine the PMF or PDF of either an isolated filter output Y_n or the vector output $\mathbf{Y}$. Just as for continuous-time systems, the special case of a Gaussian input process is a notable exception.

Theorem 7

Let the input to an FIR filter $\mathbf{h} = \begin{bmatrix} h_0 & \cdots & h_{M-1} \end{bmatrix}'$ *be the vector* $\mathbf{X} = \begin{bmatrix} X_0 & \cdots & X_{L-1} \end{bmatrix}'$*, consisting of samples from a stationary Gaussian process* X_n *with expected value* μ *and autocorrelation function* $R_X[k]$*. The output*

$\mathbf{Y} = \mathbf{H}\mathbf{X}$ as given by Equation (54) is a Gaussian random vector with expected value and covariance matrix given by

$$\boldsymbol{\mu}_{\mathbf{Y}} = \mathbf{H}\boldsymbol{\mu}_{\mathbf{X}} \quad \textit{and} \quad \mathbf{C}_{\mathbf{Y}} = \mathbf{H}(\mathbf{R}_{\mathbf{X}} - \boldsymbol{\mu}_{\mathbf{X}}\boldsymbol{\mu}'_{\mathbf{X}})\mathbf{H}'$$

where $\mathbf{R}_{\mathbf{X}}$ and $\boldsymbol{\mu}_{\mathbf{X}}$ are given by $\mathbf{R}_{\mathbf{X}_n}$ and $\mathrm{E}[\mathbf{X}_n]$ in Theorem 6 with $M = L$.

Proof Theorem 8.7 verifies that $\mathbf{X}$ has covariance matrix $\mathbf{C}_{\mathbf{X}} = \mathbf{R}_{\mathbf{X}} - \boldsymbol{\mu}_{\mathbf{X}}\boldsymbol{\mu}'_{\mathbf{X}}$. The present theorem then follows immediately from Theorem 8.11.

In Equation (54), we observe that the middle range outputs $Y_{M-1}, \ldots, Y_{L-1}$ represent a steady-state response. For $M-1 \le i \le L-1$, each Y_i is of the form $Y_i = \overleftarrow{\mathbf{h}}'\mathbf{X}_i$ where

$$\overleftarrow{\mathbf{h}}' = \begin{bmatrix} h_{M-1} & \cdots & h_0 \end{bmatrix}' \tag{55}$$

and

$$\mathbf{X}_i = \begin{bmatrix} X_{i-M+1} & X_{i-M+2} & \cdots & X_i \end{bmatrix}'. \tag{56}$$

On the other hand, the initial outputs $Y_0, \ldots, Y_{M-2}$ represent a transient response resulting from the assumption that $X_i = 0$ for $i < 0$. Similarly, $Y_L, \ldots, Y_{L+M-2}$ are transient outputs that depend on $X_i = 0$ for $i > n$. These transient outputs are simply the result of using a finite-length input for practical computations. Theorem 7 is a general result that accounts for the transient components in in the calculation of the expected value $\mathrm{E}[\mathbf{Y}]$ and covariance matrix $\mathbf{C}_{\mathbf{Y}}$ through the structure of the matrix $\mathbf{H}$.

However, a stationary Gaussian process X_n is defined for all time instances n. In this case, in passing X_n through an FIR filter $h_0, \ldots, h_{M-1}$, each output Y_n is given by Equation (50), which we repeat here: $Y_n = \overleftarrow{\mathbf{h}}'\mathbf{X}_n$. That is, each Y_n combines the M most recent inputs in precisely the same way. When X_n is a stationary process, the vectors $\mathbf{X}_n$ are statistically identical for every n. This observation leads to the following theorem.

Theorem 8

If the input to a discrete-time linear filter h_n is a stationary Gaussian random sequence X_n, the output Y_n is a stationary Gaussian random sequence with expected value and autocorrelation given by Theorem 5.

Proof Since X_n is wide sense stationary, Theorem 5 implies that Y_n is a wide sense stationary sequence. Thus it is sufficient to show that Y_n is a Gaussian random sequence, since Theorem 13.15 will then ensure that Y_n is stationary. We prove this claim only for an FIR filter $\mathbf{h} = [h_0 \quad \cdots \quad h_{M-1}]'$. That is, for every set of time instances $n_1 < n_2 < \cdots < n_k$, we will show that $\mathbf{Y} = [Y_{n_1} \quad Y_{n_2} \quad \cdots \quad Y_{n_k}]'$ is a Gaussian random vector. To start, we define L such that $n_k = n_1 + L - 1$. Next, we define the vectors $\mathbf{X} = [X_{n_1-M+1} \quad X_{n_1-M+2} \quad \cdots \quad X_{n_k}]'$ and $\hat{\mathbf{Y}} = \hat{\mathbf{H}}\mathbf{X}$ where

$$\hat{\mathbf{H}} = \begin{bmatrix} h_{M-1} & & \cdots & h_0 & & \\ & \ddots & & & \ddots & \\ & & h_{M-1} & \cdots & & h_0 \end{bmatrix} \tag{57}$$

is an $L \times (L+M-1)$ Toeplitz matrix. We note that $\hat{Y}_1 = Y_{n_1}, \bar{Y}_2 = Y_{n_1+1}, \ldots, \hat{Y}_L = Y_{n_k}$. Thus the elements of $\mathbf{Y}$ are a subset of the elements of $\hat{\mathbf{Y}}$ and we can define a $k \times L$ selection matrix $\mathbf{B}$ such that $\mathbf{Y} = \mathbf{B}\hat{\mathbf{Y}}$. Note that $B_{ij} = 1$ if $n_i = n_1 + j - 1$; otherwise $B_{ij} = 0$. Thus $\mathbf{Y} = \mathbf{BHX}$ and it follows from Theorem 8.11 that $\mathbf{Y}$ is a Gaussian random vector.

Quiz 3

The stationary Gaussian random sequence $X_0, X_1, \ldots$ with expected value $\mathrm{E}[X_n] = 0$ and covariance function $R_X[n] = \delta_n$ is the input to a the moving-average filter $\mathbf{h} = (1/4)\begin{bmatrix}1 & 1 & 1 & 1\end{bmatrix}'$. The output is Y_n. Find the PDF of $\mathbf{Y} = \begin{bmatrix}Y_{33} & Y_{34} & Y_{35}\end{bmatrix}'$.

4 Discrete-Time Linear Estimation and Prediction Filters

> The optimum linear predictor of a random sequence is the solution to a set of linear equations in which the coefficients are values of the autocorrelation function of the random sequence.

Most cellular telephones contain digital signal processing microprocessors that perform linear prediction as part of a speech compression algorithm. In a linear predictor, a speech waveform is considered to be a sample function of wide sense stationary stochastic process $X(t)$. The waveform is sampled every

T seconds (usually $T = 1/8000$ seconds) to produce the random sequence $X_n = X(nT)$. The prediction problem is to estimate a future speech sample, X_{n+k} using N previous speech samples $X_{n-M+1}, X_{n-M+2}, \ldots, X_n$. The need to minimize the cost, complexity, and power consumption of the predictor makes a DSP-based linear filter an attractive choice.

In the terminology of Theorem 12.6, we wish to form a linear estimate of the random variable $X = X_{n+k}$ using the random observation vector is $\mathbf{Y} = \begin{bmatrix} X_{n-M+1} & \cdots & X_n \end{bmatrix}'$. The general solution to this problem was given in Section 12.4 in the form of Theorems 12.6 and 12.7. When the random variable X has zero expected value, we learned from Theorem 12.6 that the minimum mean square error linear estimator of X given an observation vector $\mathbf{Y}$ is

$$\hat{X}_L(\mathbf{Y}) = \mathbf{R}_{X\mathbf{Y}} \mathbf{R}_{\mathbf{Y}}^{-1} \mathbf{Y}. \tag{58}$$

Equivalently, we can write

$$\hat{X}_L(\mathbf{Y}) = \mathbf{a}'\mathbf{Y} \tag{59}$$

where

$$\mathbf{a}' = \mathbf{R}_{X\mathbf{Y}} \mathbf{R}_{\mathbf{Y}}^{-1}. \tag{60}$$

We emphasize that this is a general solution for random variable X and observation vector $\mathbf{Y}$. In the context of a random sequence X_n, Equations (59) and (60) can solve a wide variety of estimation and prediction problems based on an observation vector $\mathbf{Y}$ derived from the process X_n. When X is a future value of the process X_n, the solution is a linear predictor. When $X = X_n$ and $\mathbf{Y}$ is a collection of noisy observations, the solution is a linear estimator. In the following, we describe some basic examples of prediction and estimation. These examples share the common feature that the linear predictor or estimator can be implemented as a discrete-time linear filter.

Linear Prediction Filters

At time n, a linear prediction filter uses the available observations

$$\mathbf{Y} = \mathbf{X}_n = \begin{bmatrix} X_{n-M+1} & \cdots & X_n \end{bmatrix}' \tag{61}$$

to estimate a future sample $X = X_{n+k}$. We wish to construct an LTI FIR filter h_n with input X_n such that the desired filter output at time n, is the linear minimum mean square error estimate

$$\hat{X}_L(\mathbf{X}_n) = \mathbf{a}'\mathbf{X}_n \tag{62}$$

where $\mathbf{a}' = \mathbf{R}_{X\mathbf{Y}}\mathbf{R}_{\mathbf{Y}}^{-1}$. Following the notation of Equation (50), the filter output at time n will be

$$\hat{X}_L(\mathbf{X}_n) = \overleftarrow{\mathbf{h}}'\mathbf{X}_n. \tag{63}$$

Comparing Equations (60) and (63), we see that the predictor can be implemented in the form the filter $\mathbf{h}$ by choosing $\overleftarrow{\mathbf{h}}' = \mathbf{a}'$, or,

$$\mathbf{h}' = \overleftarrow{\mathbf{a}}'. \tag{64}$$

That is, the optimal filter vector $\mathbf{h}' = \begin{bmatrix} h_0 & \cdots & h_{M-1} \end{bmatrix}$ is simply the time-reversed $\mathbf{a}'$. To complete the derivation of the optimal prediction filter, we observe that $\mathbf{Y} = \mathbf{X}_n$ and that $\mathbf{R}_{\mathbf{Y}} = \mathbf{R}_{\mathbf{X}_n}$, as given by Theorem 6. The cross-correlation matrix $\mathbf{R}_{X\mathbf{Y}}$ is

$$\begin{aligned} \mathbf{R}_{X\mathbf{Y}} = \mathbf{R}_{X_{n+k}\mathbf{X}_n} &= \mathrm{E}\left[X_{n+k}\begin{bmatrix} X_{n-M+1} & \cdots & X_{n-1} & X_n \end{bmatrix}\right] \\ &= \mathrm{E}\left[\begin{bmatrix} X_{n+k}X_{n-M+1} & \cdots & X_{n+k}X_{n-1} & X_{n+k}X_n \end{bmatrix}\right] \\ &= \begin{bmatrix} R_X[M+k-1] & \cdots & R_X[k+1] & R_X[k] \end{bmatrix}. \end{aligned} \tag{65}$$

We summarize this conclusion in the next theorem. An example then follows.

Theorem 9

Let X_n be a wide sense stationary random process with expected value $\mathrm{E}[X_n] = 0$ and autocorrelation function $R_X[k]$. The minimum mean square error linear *filter of order $M-1$ for predicting X_{n+k} at time n is the filter $\mathbf{h}$ such that*

$$\overleftarrow{\mathbf{h}'} = \mathbf{R}_{X_{n+k}\mathbf{X}_n}\mathbf{R}_{\mathbf{X}_n}^{-1},$$

where $\mathbf{R}_{\mathbf{X}_n}$ is given by Theorem 6 and $\mathbf{R}_{\mathbf{X}_n X_{n+k}}$ is given by Equation (65).

Example 12

X_n is a wide sense stationary random sequence with $\mathrm{E}[X_n]=0$ and autocorrelation function $R_X[k]=(-0.9)^{|k|}$. For $M=2$ samples, find $\mathbf{h}=\begin{bmatrix}h_0 & h_1\end{bmatrix}'$, the coefficients of the optimum linear predictor of $X=X_{n+1}$, given $\mathbf{Y}=\begin{bmatrix}X_{n-1} & X_n\end{bmatrix}'$. What is the optimum linear predictor of X_{n+1} given X_{n-1} and X_n?

. . . .

The optimal filter is $\mathbf{h}' = \overleftarrow{\mathbf{a}'}$ where $\mathbf{a}'$ is given by Equation (60). For $M=2$, the vector $\mathbf{a}'=\begin{bmatrix}a_0 & a_1\end{bmatrix}$ must satisfy $\mathbf{a}'\mathbf{R}_{\mathbf{Y}}=\mathbf{R}_{X\mathbf{Y}}$. From Theorem 6 and Equation (65),

$$\begin{bmatrix}a_0 & a_1\end{bmatrix}\begin{bmatrix}R_X[0] & R_X[1]\\ R_X[1] & R_X[0]\end{bmatrix} = \begin{bmatrix}R_X[2] & R_X[1]\end{bmatrix} \tag{66}$$

or

$$\begin{bmatrix}a_0 & a_1\end{bmatrix}\begin{bmatrix}1 & -0.9\\ -0.9 & 1\end{bmatrix} = \begin{bmatrix}0.81 & -0.9\end{bmatrix}. \tag{67}$$

The solution to these equations is $a_0=0$ and $a_1=-0.9$. Therefore, the optimum linear prediction filter is

$$\mathbf{h}'=\begin{bmatrix}h_0 & h_1\end{bmatrix}=\begin{bmatrix}a_1 & a_0\end{bmatrix}=\begin{bmatrix}-0.9 & 0\end{bmatrix}. \tag{68}$$

The optimum linear predictor of X_{n+1} given X_{n-1} and X_n is

$$\hat{X}_{n+1}=\overleftarrow{\mathbf{h}}'\mathbf{Y}=\begin{bmatrix}0 & -0.9\end{bmatrix}\begin{bmatrix}X_{n-1}\\ X_n\end{bmatrix}=-0.9X_n. \tag{69}$$

Example 13

Continuing Example 12, what is the mean square error of the optimal predictor?

. . . .

Since the optimal predictor is $\hat{X}_{n+1} = -0.9X_n$, the mean square error is

$$e_L = \mathrm{E}\left[(X_{n+1} - \hat{X}_{n+1})^2\right] = \mathrm{E}\left[(X_{n+1} + 0.9X_n)^2\right]. \tag{70}$$

By expanding the square and expressing the result in terms of the autocorrelation function $R_X[k]$, we obtain

$$e_L = R_X[0] + 2(0.9)R_X[1] + (0.9)^2 R_X[0] = 1 - (0.9)^2 = 0.19. \tag{71}$$

Examples 12 and 13 are a special case of the following property of random sequences X_n with autocorrelation function of the form $R_X[n] = b^{|n|}R_X[0]$.

Theorem 10

If X_n has autocorrelation function $R_X[n] = b^{|n|}R_X[0]$, the optimum linear predictor of X_{n+k} given the M previous samples $X_{n-M+1}, X_{n-M+2}, \ldots, X_n$ is

$$\hat{X}_{n+k} = b^k X_n$$

and the minimum mean square error is $e_L^* = R_X[0](1 - b^{2k})$.

Proof Since $R_X[n] = b^n$, we observe from Equation (65) that

$$\mathbf{R}_{X\mathbf{Y}} = b^k \begin{bmatrix} R_X[M-1] & \cdots & R_X[1] & R_X[0] \end{bmatrix}. \tag{72}$$

From Theorem 6, we see that $\mathbf{R}_{X\mathbf{Y}}$ is the Mth row of $\mathbf{R}_{\mathbf{Y}}$ scaled by b^k. Thus the solution to $\mathbf{a}'\mathbf{R}_{\mathbf{Y}} = \mathbf{R}_{X\mathbf{Y}}$ is the vector $\mathbf{a}' = \begin{bmatrix} 0 & \cdots & 0 & b^k \end{bmatrix}$. The mean square error is

$$\begin{aligned} e_L^* = \mathrm{E}\left[(X_{n+k} - b^k X_n)^2\right] &= R_X[0] - 2b^k R_X[k] + b^{2k} R_X[0] \\ &= R_X[0] - 2b^{2k} R_X[0] + b^{2k} R_X[0] \\ &= R_X[0](1 - b^{2k}). \end{aligned} \tag{73}$$

When b^k is close to 1, X_{n+k} and X_n are highly correlated. In this case, the sequence varies slowly and the estimate $b^k X_n$ will be close to X_n and also close to X_{n+k}. Since b^{2k} is also close to 1, the mean square error e_L^* will be close to zero. As b^k gets smaller, X_{n+k} and X_n become less correlated. Consequently, the random sequence has much greater variation. The predictor is unable to track this variation and the predicted value approaches $\mathrm{E}[X_{n+k}] = 0$. That is, when the observation tells us little about the future, the predictor must rely on a priori knowledge of the random sequence. In addition, we observe for a fixed value of b that increasing k reduces b^{2k}, and estimates are less accurate. That is, predictions further into the future are less accurate.

Furthermore, Theorem 10 states that for random sequences with autocorrelation functions of the form $R_X[k] = b^{|k|}R_X[0]$, X_n is the only random variable that contributes to the optimum linear prediction of X_{n+1}. Another way of stating this is that at time $n+1$, all of the information about the past history of the sequence is summarized in the value of X_n. Discrete-valued random sequences with this property are referred to as discrete-time Markov chains. The analysis of probability models of discrete-time Markov chains can be found in the Markov Chains Supplement.

Linear Estimation Filters

In the estimation problem, we estimate $X = X_n$ based on the noisy observations $Y_n = X_n + W_n$. In particular, we use the vector $\mathbf{Y} = \mathbf{Y}_n = \begin{bmatrix} Y_{n-M+1} & \cdots & Y_{n-1} & Y_n \end{bmatrix}'$ of the M most recent observations. Our estimates will be the output resulting from passing the sequence Y_n through the LTI FIR filter h_n. We assume that X_n and W_n are independent wide sense stationary sequences with expected values $\mathrm{E}[X_n] = \mathrm{E}[W_n] = 0$ and autocorrelation functions $R_X[n]$ and $R_W[n]$.

From Theorem 12.6, we know that the linear minimum mean square error estimate of X given the observation $\mathbf{Y}_n$ is $\hat{X}_L(\mathbf{Y}_n) = \mathbf{a}'\mathbf{Y}_n$ where $\mathbf{a}' = \mathbf{R}_{X\mathbf{Y}_n}\mathbf{R}_{\mathbf{Y}_n}^{-1}$. The optimal estimation filter is $\mathbf{h}' = \overleftarrow{\mathbf{a}}'$, the time reversal of $\mathbf{a}$. All that remains is to identify $\mathbf{R}_{\mathbf{Y}_n}$ and $\mathbf{R}_{X\mathbf{Y}_n}$. In vector form,

$$\mathbf{X}_n = \begin{bmatrix} X_{n-M+1} & \cdots & X_{n-1} & X_n \end{bmatrix}', \tag{74}$$

$$\mathbf{W}_n = \begin{bmatrix} W_{n-M+1} & \cdots & W_{n-1} & W_n \end{bmatrix}', \tag{75}$$

and

$$\mathbf{Y}_n = \mathbf{X}_n + \mathbf{W}_n. \tag{76}$$

This implies

$$\begin{aligned}\mathbf{R}_{\mathbf{Y}_n} &= \mathrm{E}\left[\mathbf{Y}_n \mathbf{Y'}_n\right] = \mathrm{E}\left[(\mathbf{X}_n + \mathbf{W}_n)(\mathbf{X'}_n + \mathbf{W'}_n)\right] \\ &= \mathrm{E}\left[\mathbf{X}_n \mathbf{X'}_n + \mathbf{X}_n \mathbf{W'}_n + \mathbf{W}_n \mathbf{X'}_n + \mathbf{W}_n \mathbf{W'}_n\right].\end{aligned} \tag{77}$$

Because $\mathbf{X}_n$ and $\mathbf{W}_n$ are independent, $\mathrm{E}[\mathbf{X}_n \mathbf{W'}_n] = \mathrm{E}[\mathbf{X}_n]\,\mathrm{E}[\mathbf{W'}_n] = \mathbf{0}$. Similarly, $\mathrm{E}[\mathbf{W}_n \mathbf{X'}_n] = \mathbf{0}$. This implies

$$\mathbf{R}_{\mathbf{Y}_n} = \mathrm{E}\left[\mathbf{X}_n \mathbf{X'}_n\right] + \mathrm{E}\left[\mathbf{W}_n \mathbf{W'}_n\right] = \mathbf{R}_{\mathbf{X}_n} + \mathbf{R}_{\mathbf{W}_n}. \tag{78}$$

The cross-correlation matrix is

$$\begin{aligned}\mathbf{R}_{X\mathbf{Y}_n} &= \mathrm{E}\left[X_n(\mathbf{X'}_n + \mathbf{W'}_n)\right] \\ &= \mathrm{E}\left[X_n \mathbf{X'}_n\right] + \mathrm{E}\left[X_n \mathbf{W'}_n\right] = \mathrm{E}\left[X_n \mathbf{X'}_n\right]\end{aligned} \tag{79}$$

since $\mathrm{E}[X_n \mathbf{W'}_n] = \mathrm{E}[X_n]\,\mathrm{E}[\mathbf{W}_n] = \mathbf{0'}$. Thus, by Equations (74) and (79),

$$\begin{aligned}\mathbf{R}_{X\mathbf{Y}_n} = \mathbf{R}_{X_n\mathbf{X}_n} &= \mathrm{E}\left[X_n \begin{bmatrix} X_{n-M+1} & \cdots & X_{n-1} & X_n \end{bmatrix}\right] \\ &= \begin{bmatrix} R_X[M-1] & \cdots & R_X[1] & R_X[0] \end{bmatrix}.\end{aligned} \tag{80}$$

These facts are summarized in the following theorem. An example then follows.

Theorem 11

Let X_n and W_n be independent wide sense stationary random processes with expected values $\mathrm{E}[X_n] = \mathrm{E}[W_n] = 0$ and autocorrelation functions $R_X[k]$ and $R_W[k]$. Let $Y_n = X_n + W_n$. The minimum mean square error linear estimation filter of X_n of order $M-1$ given the input Y_n is given by $\mathbf{h}'$ such that

$$\overleftarrow{\mathbf{h}'} = \begin{bmatrix} h_{M-1} & \cdots h_0 \end{bmatrix} = \mathbf{R}_{X_n\mathbf{X}_n}(\mathbf{R}_{\mathbf{X}_n} + \mathbf{R}_{\mathbf{W}_n})^{-1}$$

where $\mathbf{R}_{X_n\mathbf{X}_n}$ is given by Equation (80).

Example 14

The independent random sequences X_n and W_n have zero expected value and autocorrelation functions $R_X[k] = (-0.9)^{|k|}$ and $R_W[k] = (0.2)\delta_k$. Use $M = 2$ samples of the noisy observation sequence $Y_n = X_n + W_n$ to estimate X_n. Find the linear minimum mean square error prediction filter $\mathbf{h} = \begin{bmatrix} h_0 & h_1 \end{bmatrix}'$.

Based on the observation $\mathbf{Y} = \begin{bmatrix} Y_{n-1} & Y_n \end{bmatrix}'$, the linear MMSE estimate of $X = X_n$ is $\mathbf{a}'\mathbf{Y}$ where $\mathbf{a}' = \mathbf{R}_{X\mathbf{Y}}\mathbf{R}_{\mathbf{Y}}^{-1}$. From Equation (80),

$$\mathbf{R}_{X\mathbf{Y}} = \begin{bmatrix} R_X[1] & R_X[0] \end{bmatrix} = \begin{bmatrix} -0.9 & 1 \end{bmatrix}. \tag{81}$$

From Equation (78),

$$\mathbf{R}_{\mathbf{Y}} = \mathbf{R}_{\mathbf{X}_n} + \mathbf{R}_{\mathbf{W}_n} = \begin{bmatrix} 1 & -0.9 \\ -0.9 & 1 \end{bmatrix} + \begin{bmatrix} 0.2 & 0 \\ 0 & 0.2 \end{bmatrix} = \begin{bmatrix} 1.2 & -0.9 \\ -0.9 & 1.2 \end{bmatrix}. \tag{82}$$

This implies

$$\begin{aligned} \mathbf{a}' &= \mathbf{R}_{X\mathbf{Y}}\mathbf{R}_{\mathbf{Y}}^{-1} \\ &= \begin{bmatrix} -0.9 & 1 \end{bmatrix} \begin{bmatrix} 1.2 & -0.9 \\ -0.9 & 1.2 \end{bmatrix}^{-1} = \begin{bmatrix} -0.2857 & 0.6190 \end{bmatrix}. \end{aligned} \tag{83}$$

The optimal filter is $\mathbf{h}' = \overleftarrow{\mathbf{a}'} = \begin{bmatrix} 0.6190 & -0.2857 \end{bmatrix}$.

In Example 14, the estimation filter combines the observations to get the dual benefit of averaging the noise samples as well as exploiting the correlation of X_{n-1} and X_n. A general version of Example 14 is examined in Problem 4.6.

Quiz 4

X_n is a wide sense stationary random sequence with $E[X_n] = 0$ and autocorrelation function

$$R_X[n] = \begin{cases} (0.9)^{|n|} + 0.1 & n = 0, \\ (0.9)^{|n|} & \text{otherwise.} \end{cases} \tag{84}$$

For $M = 2$ samples, find $\mathbf{h} = \begin{bmatrix} h_0 & h_1 \end{bmatrix}'$, the optimum linear prediction filter of X_{n+1}, given X_{n-1} and X_n. What is the mean square error of the optimum linear predictor?

5 Power Spectral Density of a Continuous-Time Process

The power spectral density (PSD) function, which is defined as the Fourier transform of the autocorrelation function, gives a frequency domain interpretation of a wide sense stationary random process.

The autocorrelation function of a continuous-time stochastic process conveys information about the time structure of the process. If $X(t)$ is stationary and $R_X(\tau)$ decreases rapidly with increasing τ, it is likely that a sample function of $X(t)$ will change abruptly in a short time. Conversely, if the decline in $R_X(\tau)$ with increasing τ is gradual, it is likely that a sample function of $X(t)$ will be slowly time-varying. Fourier transforms offer another view of the variability of functions of time. A rapidly varying function of time has a Fourier transform with high magnitudes at high frequencies, and a slowly varying function has a Fourier transform with low magnitudes at high frequencies.

In the study of stochastic processes, the *power spectral density function*, $S_X(f)$, provides a frequency-domain representation of the time structure of $X(t)$. By definition, $S_X(f)$ is the expected value of the squared magnitude of the Fourier transform of a sample function of $X(t)$. To present the definition formally, we first establish our notation for the Fourier transform as a function of the frequency variable f Hz.

Definition 1 — Fourier Transform

Functions $g(t)$ and $G(f)$ are a ***Fourier transform*** *pair if*

$$G(f) = \int_{-\infty}^{\infty} g(t) e^{-j2\pi f t}\, dt, \qquad g(t) = \int_{-\infty}^{\infty} G(f) e^{j2\pi f t}\, df.$$

Tables 1 and 2 provide a list of Fourier transform pairs and Fourier transform properties. Students who have already studied signals and systems may recall that not all functions of time have Fourier transforms. For many functions that extend over infinite time, the time integral in Definition 1 does not exist. Sample functions $x(t)$ of a stationary stochastic process $X(t)$ are

Time function	Fourier Transform
$\delta(\tau)$	1
1	$\delta(f)$
$\delta(\tau-\tau_0)$	$e^{-j2\pi f\tau_0}$
$u(\tau)$	$\frac{1}{2}\delta(f)+\frac{1}{j2\pi f}$
$e^{j2\pi f_0\tau}$	$\delta(f-f_0)$
$\cos 2\pi f_0\tau$	$\frac{1}{2}\delta(f-f_0)+\frac{1}{2}\delta(f+f_0)$
$\sin 2\pi f_0\tau$	$\frac{1}{2j}\delta(f-f_0)-\frac{1}{2j}\delta(f+f_0)$
$ae^{-a\tau}u(\tau)$	$\frac{a}{a+j2\pi f}$
$ae^{-a\lvert\tau\rvert}$	$\frac{2a^2}{a^2+(2\pi f)^2}$
$ae^{-\pi a^2\tau^2}$	$e^{-\pi f^2/a^2}$
$\mathrm{rect}(\tau/T)$	$T\,\mathrm{sinc}(fT)$
$\mathrm{sinc}(2W\tau)$	$\frac{1}{2W}\,\mathrm{rect}(\frac{f}{2W})$

Note that a is a positive constant and that the rectangle and sinc functions are defined as

$$\mathrm{rect}(x)=\begin{cases}1 & |x|<1/2,\\ 0 & \text{otherwise,}\end{cases}$$

$$\mathrm{sinc}(x)=\frac{\sin(\pi x)}{\pi x}.$$

Table 1 Fourier transform pairs of common signals.

Time function	Fourier Transform
$g(\tau - \tau_0)$	$G(f)e^{-j2\pi f\tau_0}$
$g(\tau)e^{j2\pi f_0 \tau}$	$G(f - f_0)$
$g(-\tau)$	$G^*(f)$
$\dfrac{dg(\tau)}{d\tau}$	$j2\pi f G(f)$
$\int_{-\infty}^{\tau} g(v)\,dv$	$\dfrac{G(f)}{j2\pi f} + \dfrac{G(0)}{2}\delta(f)$
$\int_{-\infty}^{\infty} h(v)g(\tau - v)\,dv$	$G(f)H(f)$
$g(t)h(t)$	$\int_{-\infty}^{\infty} H(f')G(f - f')\,df'$

Table 2 Properties of the Fourier transform.

usually of this nature. To work with these functions in the frequency domain, we begin with $x_T(t)$, a truncated version of $x(t)$. It is identical to $x(t)$ for $-T \leq t \leq T$ and 0 elsewhere. We use the notation $X_T(f)$ for the Fourier transform of this function.

$$X_T(f) = \int_{-T}^{T} x(t)e^{-j2\pi f t}\,dt. \tag{85}$$

$|X_T(f)|^2$, the squared magnitude of $X_T(f)$, appears in the definition of $S_X(f)$. If $x(t)$ is an electrical signal measured in volts or amperes, $|X_T(f)|^2$ has units of energy. Its time average, $|X_T(f)|^2/2T$, has units of power. $S_X(f)$ is the limit as the time window goes to infinity of the expected value of this function:

Definition 2 Power Spectral Density

The power spectral density function of the wide sense stationary stochastic *process $X(t)$ is*

$$S_X(f) = \lim_{T\to\infty} \frac{1}{2T} \mathrm{E}\left[|X_T(f)|^2\right] = \lim_{T\to\infty} \frac{1}{2T} \mathrm{E}\left[\left|\int_{-T}^{T} X(t)e^{-j2\pi ft}\,dt\right|^2\right].$$

We refer to $S_X(f)$ as a density function because it can be interpreted as the amount of power in $X(t)$ in the infinitesimal range of frequencies $[f, f+df]$. Physically, $S_X(f)$ has units of watts/Hz = Joules. As stated in Theorem 13(b), the average power of $X(t)$ is the integral of $S_X(f)$ over all frequencies in $-\infty < f < \infty$.

Both the autocorrelation function and the power spectral density function convey information about the time structure of $X(t)$. The Wiener-Khintchine theorem shows that they convey the same information:

Theorem 12 Wiener-Khintchine

If $X(t)$ is a wide sense stationary stochastic process, $R_X(\tau)$ and $S_X(f)$ are the Fourier transform pair

$$S_X(f) = \int_{-\infty}^{\infty} R_X(\tau)e^{-j2\pi f\tau}\,d\tau, \qquad R_X(\tau) = \int_{-\infty}^{\infty} S_X(f)\,e^{j2\pi f\tau}\,df.$$

Proof Since $x(t)$ is real-valued, $X_T^*(f) = \int_{-T}^{T} x(t')e^{j2\pi ft'}\,dt'$. Since $|X_T(f)|^2 = X_T(f)X_T^*(f)$,

$$\begin{aligned}
\mathrm{E}\left[|X_T(f)|^2\right] &= \mathrm{E}\left[\left(\int_{-T}^{T} X(t)e^{-j2\pi ft}\,dt\right)\left(\int_{-T}^{T} x(t')e^{j2\pi ft'}\,dt'\right)\right] \\
&= \int_{-T}^{T}\int_{-T}^{T} \mathrm{E}\left[X(t)X(t')\right] e^{-j2\pi f(t-t')}\,dt\,dt' \\
&= \int_{-T}^{T}\int_{-T}^{T} R_X(t-t')e^{-j2\pi f(t-t')}\,dt\,dt'.
\end{aligned} \tag{86}$$

In Equation (86), we are integrating the deterministic function

$$g(t-t') = R_X(t-t')e^{-j2\pi f(t-t')} \tag{87}$$

over a square box. Making the variable substitution $\tau = t - t'$ and reversing the order of integration, we can show that

$$\int_{-T}^{T} \int_{-T}^{T} g(t-t')\,dt\,dt' = 2T \int_{-2T}^{2T} g(\tau) \left(1 - \frac{|\tau|}{2T}\right) d\tau. \tag{88}$$

With $g(\tau) = R_X(\tau) e^{-j2\pi f\tau}$, it follows from Equations (86) and (88) that

$$\frac{1}{2T} \mathrm{E}\left[|X_T(f)|^2\right] = \int_{-2T}^{2T} R_X(\tau) \left(1 - \frac{|\tau|}{2T}\right) e^{-j2\pi f\tau}\,d\tau. \tag{89}$$

We define the function

$$h_T(\tau) = \begin{cases} 1 - |\tau|/(2T) & |\tau| \le 2T, \\ 0 & \text{otherwise,} \end{cases} \tag{90}$$

and we observe that $\lim_{T\to\infty} h_T(\tau) = 1$ for all finite τ. It follows from Equation (89) that

$$\begin{aligned} \lim_{T\to\infty} \frac{1}{2T} \mathrm{E}\left[|X_T(f)|^2\right] &= \lim_{T\to\infty} \int_{-\infty}^{\infty} h_T(\tau) R_X(\tau) e^{-j2\pi f\tau}\,d\tau \\ &= \int_{-\infty}^{\infty} \left(\lim_{T\to\infty} h_T(\tau)\right) R_X(\tau) e^{-j2\pi f\tau}\,d\tau = S_X(f). \end{aligned} \tag{91}$$

The following theorem states three important properties of $S_X(f)$:

Theorem 13

For a wide sense stationary random process $X(t)$, the power spectral density $S_X(f)$ is a real-valued function with the following properties:

(a) $S_X(f) \ge 0$ *for all* f

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

(c) $S_X(-f) = S_X(f)$

Proof The first property follows from Definition 2, in which $S_X(f)$ is the expected value of the time average of a nonnegative quantity. The second property follows by substituting $\tau = 0$ in Theorem 12 and referring to Definition 1. To prove the third property, we observe that $R_X(\tau) = R_X(-\tau)$ implies

$$S_X(f) = \int_{-\infty}^{\infty} R_X(-\tau) e^{-j2\pi f\tau}\, d\tau. \tag{92}$$

Making the substitution $\tau' = -\tau$ yields

$$S_X(f) = \int_{-\infty}^{\infty} R_X(\tau') e^{-j2\pi(-f)\tau'}\, d\tau' = S_X(-f). \tag{93}$$

Example 15

A wide sense stationary process $X(t)$ has autocorrelation function $R_X(\tau) = Ae^{-b|\tau|}$ where $b > 0$. Derive the power spectral density function $S_X(f)$ and calculate the average power $\mathrm{E}[X^2(t)]$.

. . . .

To find $S_X(f)$, we use Table 1 since $R_X(\tau)$ is of the form $ae^{-a|\tau|}$.

$$S_X(f) = \frac{2Ab}{(2\pi f)^2 + b^2}. \tag{94}$$

The average power is

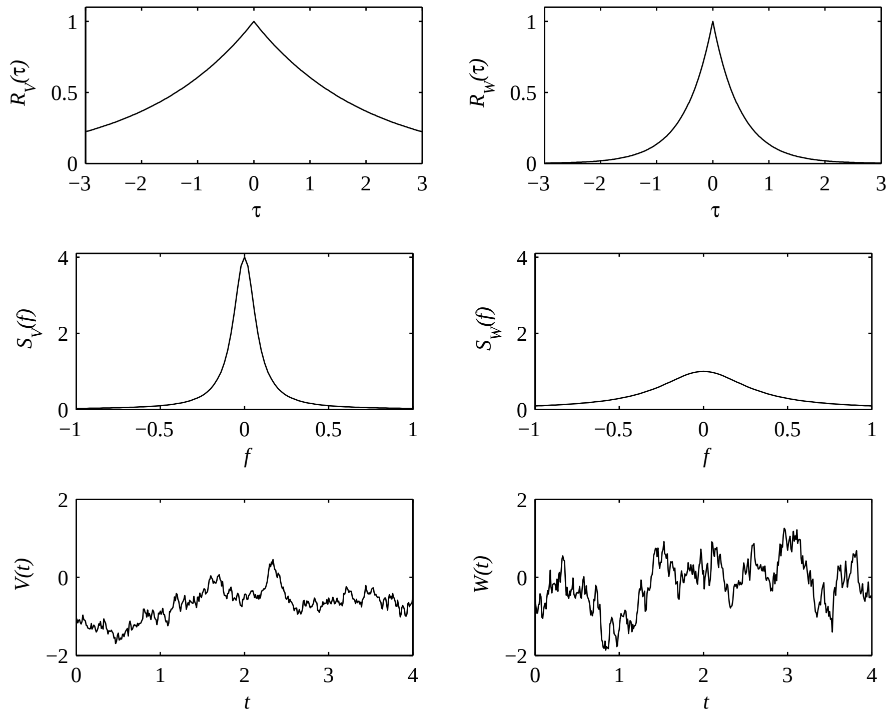
$$\mathrm{E}\left[X^2(t)\right] = R_X(0) = Ae^{-b|0|} = \int_{-\infty}^{\infty} \frac{2Ab}{(2\pi f)^2 + b^2}\, df = A. \tag{95}$$

Figure 1 displays three graphs for each of two stochastic processes in the family studied in Example 15: $V(t)$ with $R_V(\tau) = e^{-0.5|\tau|}$ and $W(t)$ with $R_W(\tau) = e^{-2|\tau|}$. For each process, the three graphs are the autocorrelation function, the power spectral density function, and one sample function. For both processes, the average power is $A = 1$ watt. Note $W(t)$ has a narrower autocorrelation (less dependence between two values of the process with a given time separation) and a wider power spectral density (more power at higher frequencies) than $V(t)$. The sample function $w(t)$ fluctuates more rapidly with time than $v(t)$.

Example 16

For the random process $Y(t)$ of Example 2, find the power spectral density $S_Y(f)$.

. . . .



Random processes $V(t)$ and $W(t)$ with autocorrelation functions $R_V(\tau) = e^{-0.5|\tau|}$ and $R_W(\tau) = e^{-2|\tau|}$ are examples of the process $X(t)$ in Example 15. These graphs show $R_V(\tau)$ and $R_W(\tau)$, the power spectral density functions $S_V(f)$ and $S_W(f)$, and sample paths of $V(t)$ and $W(t)$.

Figure 1 Two examples of the process $X(t)$ in Example 15.

From Definition 2 and the triangular autocorrelation function $R_Y(\tau)$ given in Equation (16), we can write

$$R_Y(\tau) = \frac{\eta_0}{T^2}\left(\int_{-T}^{0}(T+\tau)e^{-j2\pi f\tau}\,d\tau + \int_{0}^{T}(T-\tau)e^{-j2\pi f\tau}\,d\tau\right). \qquad (96)$$

With the substitution $\tau' = -\tau$ in the left integral, we obtain

$$R_Y(\tau) = \frac{\eta_0}{T^2}\left(\int_0^T (T-\tau')e^{j2\pi f\tau'}\,d\tau' + \int_0^T (T-\tau)e^{-j2\pi f\tau}\,d\tau\right)$$
$$= \frac{2\eta_0}{T^2}\int_0^T (T-\tau)\cos(2\pi f\tau)\,d\tau. \tag{97}$$

Using integration by parts (Appendix B, Math Fact B.10), $u = T - \tau$ and $dv = \cos(2\pi f\tau)\,d\tau$, yields

$$R_Y(\tau) = \frac{2\eta_0}{T^2}\left(\left.\frac{(T-\tau)\sin(2\pi f\tau)}{2\pi f}\right|_0^T + \int_0^T \frac{\sin(2\pi f\tau)}{2\pi f}\,d\tau\right)$$
$$= \frac{2\eta_0(1-\cos(2\pi fT))}{(2\pi fT)^2}. \tag{98}$$

Quiz 5

The power spectral density of $X(t)$ is

$$S_X(f) = \frac{5}{W}\operatorname{rect}\left(\frac{f}{2W}\right) = \begin{cases} 5/W & -W \le f \le W, \\ 0 & \text{otherwise.} \end{cases} \tag{99}$$

(a) What is the average power of $X(t)$?

(b) Write a formula for the autocorrelation function of $X(t)$.

(c) Draw graphs of $S_X(f)$ and $R_X(\tau)$ for $W = 1$ kHz and $W = 10$ Hz.

6 Power Spectral Density of a Random Sequence

> The spectral analysis of a random sequence parallels the analysis of a continuous-time process. The power spectral density function of the digital filter output is the product of the power spectral density function of the input and the squared magnitude of the filter transfer function.

A sample function of a random sequence is an ordered list of numbers. Each number in the list is a sample value of a random variable. The *discrete-time Fourier transform (DTFT)* is a spectral representation of an ordered set of numbers.

Definition 3 — Discrete-time Fourier Transform (DTFT)

The sequence $\{\ldots, x_{-1}, x_0, x_1, \ldots\}$ and the function $X(\phi)$ are a discrete-time Fourier transform (DTFT) pair if

$$X(\phi) = \sum_{n=-\infty}^{\infty} x_n e^{-j2\pi\phi n}, \qquad x_n = \int_{-1/2}^{1/2} X(\phi) e^{j2\pi\phi n}\, d\phi.$$

Table 3 provides a table of common discrete-time Fourier transforms.

Note that ϕ is a dimensionless normalized frequency, with range $-1/2 \le \phi \le 1/2$, and $X(\phi)$ is periodic with unit period. This property reflects the fact that a list of numbers in itself has no inherent time scale. When the random sequence $X_n = X(nT_s)$ consists of time samples of a continuous-time process sampled at frequency $f_s = 1/T_s$ Hz, the normalized frequency ϕ corresponds to the frequency $f = \phi f_s$ Hz. The normalized frequency range $-1/2 \le \phi \le 1/2$ reflects the requirement of the Nyquist sampling theorem that sampling a signal at rate f_s allows the sampled signal to describe frequency components in the range $-f_s/2 \le f \le f_s/2$.

Example 17

Calculate the DTFT $H(\phi)$ of the order $M-1$ moving-average filter h_n of Example 5.

Discrete Time function	Discrete Time Fourier Transform
$\delta[n] = \delta_n$	1
1	$\delta(\phi)$
$\delta[n-n_0] = \delta_{n-n_0}$	$e^{-j2\pi\phi n_0}$
$u[n]$	$\dfrac{1}{1-e^{-j2\pi\phi}} + \dfrac{1}{2}\sum_{k=-\infty}^{\infty}\delta(\phi+k)$
$e^{j2\pi\phi_0 n}$	$\sum_{k=-\infty}^{\infty}\delta(\phi-\phi_0-k)$
$\cos 2\pi\phi_0 n$	$\frac{1}{2}\delta(\phi-\phi_0) + \frac{1}{2}\delta(\phi+\phi_0)$
$\sin 2\pi\phi_0 n$	$\frac{1}{2j}\delta(\phi-\phi_0) - \frac{1}{2j}\delta(\phi+\phi_0)$
$a^n u[n]$	$\dfrac{1}{1-ae^{-j2\pi\phi}}$
$a^{\lvert n\rvert}$	$\dfrac{1-a^2}{1+a^2-2a\cos 2\pi\phi}$
g_{n-n_0}	$G(\phi)e^{-j2\pi\phi n_0}$
$g_n e^{j2\pi\phi_0 n}$	$G(\phi-\phi_0)$
g_{-n}	$G^*(\phi)$
$\sum_{k=-\infty}^{\infty} h_k g_{n-k}$	$G(\phi)H(\phi)$
$g_n h_n$	$\int_{-1/2}^{1/2} H(\phi')G(\phi-\phi')\,d\phi'$

Note that $\delta[n]$ is the discrete impulse, $u[n]$ is the discrete unit step, and a is a constant with magnitude $|a| < 1$.

Table 3 Discrete-Time Fourier transform pairs and properties.

Since $h_n = 1/M$ for $n = 0, 1, \ldots, M-1$ and is otherwise zero, we apply Definition 3 to write

$$H(\phi) = \sum_{n=-\infty}^{\infty} h_n e^{-j2\pi\phi n} = \frac{1}{M} \sum_{n=0}^{M-1} e^{-j2\pi\phi n}. \tag{100}$$

Using Math Fact B.4 with $\alpha = e^{-j2\pi\phi}$ and $n = M-1$, we obtain

$$H(\phi) = \frac{1}{M} \left(\frac{1 - e^{-j2\pi\phi M}}{1 - e^{-j2\pi\phi}} \right). \tag{101}$$

We noted earlier that many time functions do not have Fourier transforms. Similarly, the sum in Definition 3 does not converge for sample functions of many random sequences. To work with these sample functions, our analysis is similar to that of continuous-time processes. We define a truncated sample function that is identical to x_n for $-L \leq n \leq L$ and 0 elsewhere. We use the notation $X_L(\phi)$ for the discrete-time Fourier transform of this sequence:

$$X_L(\phi) = \sum_{n=-L}^{L} x_n e^{-j2\pi\phi n}. \tag{102}$$

The power spectral density function of the random sequence X_n is the expected value $|X_L(\phi)|^2/(2N+1)$.

Definition 4 Power Spectral Density of a Random Sequence

The power spectral density function of the wide sense stationary random sequence X_n is

$$S_X(\phi) = \lim_{L\to\infty} \frac{1}{2L+1} \mathrm{E}\left[\left| \sum_{n=-L}^{L} X_n e^{-j2\pi\phi n} \right|^2 \right].$$

The Wiener-Khintchine Theorem also holds for random sequences.

Theorem 14 Discrete-Time Wiener-Khintchine

If X_n is a wide sense stationary stochastic process, $R_X[k]$ and $S_X(\phi)$ are a discrete-time Fourier transform pair:

$$S_X(\phi) = \sum_{k=-\infty}^{\infty} R_X[k] \, e^{-j2\pi\phi k}, \qquad R_X[k] = \int_{-1/2}^{1/2} S_X(\phi) \, e^{j2\pi\phi k} \, d\phi.$$

The properties of the power spectral density function of a random sequence are similar to the properties of the power spectral density function of a continuous-time stochastic process. The following theorem corresponds to Theorem 13 and, in addition, describes the periodicity of $S_X(\phi)$ for a random sequence.

Theorem 15

For a wide sense stationary random sequence X_n, the power spectral density $S_X(\phi)$ has the following properties:

(a) $S_X(\phi) \geq 0$ *for all* f,

(b) $\int_{-1/2}^{1/2} S_X(\phi) \, d\phi = \mathrm{E}[X_n^2] = R_X[0]$,

(c) $S_X(-\phi) = S_X(\phi)$,

(d) for any integer n, $S_X(\phi + n) = S_X(\phi)$.

Example 18

The wide sense stationary random sequence X_n has zero expected value and autocorrelation function

$$R_X[k] = \begin{cases} \sigma^2(2 - |n|)/4 & n = -1, 0, 1, \\ 0 & \text{otherwise.} \end{cases} \tag{103}$$

Derive the power spectral density function of X_n.

Applying Theorem 14 and Definition 3, we have

$$S_X(\phi) = \sum_{n=-1}^{1} R_X[n] e^{-j2\pi n\phi}$$
$$= \sigma^2 \left[\frac{(2-1)}{4} e^{j2\pi\phi} + \frac{2}{4} + \frac{(2-1)}{4} e^{-j2\pi\phi} \right]$$
$$= \frac{\sigma^2}{2} + \frac{\sigma^2}{2}\cos(2\pi\phi). \tag{104}$$

Example 19

The wide sense stationary random sequence X_n has zero expected value and power spectral density

$$S_X(\phi) = \frac{1}{2}\delta(\phi - \phi_0) + \frac{1}{2}\delta(\phi + \phi_0) \tag{105}$$

where $0 < \phi_0 < 1/2$. What is the autocorrelation $R_X[k]$?

From Theorem 14, we have

$$R_X[k] = \int_{-1/2}^{1/2} (\delta(\phi - \phi_0) + \delta(\phi + \phi_0)) e^{j2\pi\phi k} \, d\phi. \tag{106}$$

By the sifting property of the continuous-time delta function,

$$R_X[k] = \frac{1}{2} e^{j2\pi\phi_0 k} + \frac{1}{2} e^{-j2\pi\phi_0 k} = \cos(2\pi\phi_0 k). \tag{107}$$

Quiz 6

The random sequence X_n has power spectral density function $S_X(\phi) = 10$. Derive $R_X[k]$, the autocorrelation function of X_n.

7 Cross Spectral Density

The cross spectral density of jointly wide sense stationary signals is the Fourier transform of the cross correlation function.

When two processes are jointly wide sense stationary, we can study the cross-correlation in the frequency domain.

Definition 5 — Cross Spectral Density

For jointly wide sense stationary random processes $X(t)$ and $Y(t)$, the Fourier transform of the cross-correlation yields the ***cross spectral density***

$$S_{XY}(f) = \int_{-\infty}^{\infty} R_{XY}(\tau) e^{-j2\pi f\tau}\, d\tau.$$

. . . .

For jointly wide sense stationary random sequences X_n and Y_n, the discrete-time Fourier transform of the cross-correlation yields the cross spectral density

$$S_{XY}(\phi) = \sum_{k=-\infty}^{\infty} R_{XY}[k]\, e^{-j2\pi\phi k}.$$

We encounter cross-correlations in experiments that involve noisy observations of a wide sense stationary random process $X(t)$.

Example 20

In Example 13.24, we were interested in $X(t)$ but we could observe only $Y(t) = X(t) + N(t)$ where $N(t)$ is a wide sense stationary noise process with $\mu_N = 0$. In this case, when $X(t)$ and $N(t)$ are jointly wide sense stationary, we found that

$$R_Y(\tau) = R_X(\tau) + R_{XN}(\tau) + R_{NX}(\tau) + R_N(\tau). \tag{108}$$

Find the power spectral density of the output $Y(t)$.

. . . .

By taking a Fourier transform of both sides, we obtain the power spectral density of the observation $Y(t)$.

$$S_Y(f) = S_X(f) + S_{XN}(f) + S_{NX}(f) + S_N(f). \tag{109}$$

Example 21

Continuing Example 20, suppose $N(t)$ and $X(t)$ are independent. Find the autocorrelation and power spectral density of the observation $Y(t)$.

In this case,

$$R_{XN}(\tau) = \mathrm{E}[X(t)N(t+\tau)] = \mathrm{E}[X(t)]\,\mathrm{E}[N(t+\tau)] = 0. \tag{110}$$

Similarly, $R_{NX}(\tau) = 0$. This implies

$$R_Y(\tau) = R_X(\tau) + R_N(\tau), \tag{111}$$

and

$$S_Y(f) = S_X(f) + S_N(f). \tag{112}$$

Quiz 7

Random process $Y(t) = X(t - t_0)$ is a time delayed version of the wide sense stationary process $X(t)$. Find the cross spectral density $S_{XY}(\tau)$.

8 Frequency Domain Filter Relationships

When a wide sense stationary process is the input to a linear filter, the PSD of the output is the product input PSD and the squared magnitude of the filter frequency response.

Electrical and computer engineers are well acquainted with frequency domain representations of filtering operations. If a linear filter has impulse response $h(t)$, $H(f)$, the Fourier transform of $h(t)$, is referred to as the *frequency response* of the filter. The Fourier transform of the filter output $W(f)$ is related to the transform of the input $V(f)$ and the filter frequency response $H(f)$ by

$$W(f) = H(f)V(f). \tag{113}$$

Equation (113) is the frequency domain representation of the fundamental input-output property of a continuous-time filter. For discrete-time signals W_n and V_n and the discrete-time filter with impulse response h_n, the corresponding relationship is

$$W(\phi) = H(\phi)V(\phi) \tag{114}$$

where the three functions of frequency are discrete-time Fourier transforms defined in Definition 3. When the discrete-time filter contains forward and feedback sections as in Equation (28), the transfer function is

$$H(\phi) = A(\phi)/(1 - B(\phi)), \tag{115}$$

where $A(\phi)$ and $B(\phi)$ are the discrete-time Fourier transforms of the forward section and feedback section, respectively.

In studying stochastic processes, our principal frequency domain representation of the continuous-time process $X(t)$ is the power spectral density function $S_X(f)$ in Definition 2, and our principal representation of the random sequence X_n is $S_X(\phi)$ in Definition 4. When $X(t)$ or X_n is the input to a linear filter with transfer function $H(f)$ or $H(\phi)$, the following theorem presents the relationship of the power spectral density function of the output of the filter to the corresponding function of the input process.

Theorem 16

When a wide sense stationary stochastic process $X(t)$ is the input to a linear time-invariant filter with transfer function $H(f)$, the power spectral density

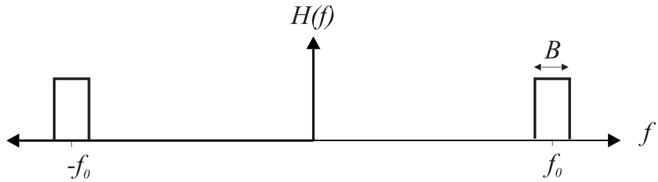


Figure 2 The ideal bandpass filter $H(f)$ with center frequency f_0 and bandwidth B Hz.

of the output $Y(t)$ is

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

. . . .

When a wide sense stationary random sequence X_n is the input to a linear time-invariant filter with transfer function $H(\phi)$, the power spectral density of the output Y_n is

$$S_Y(\phi) = |H(\phi)|^2 \, S_X(\phi).$$

Proof We refer to Theorem 12 to recall that $S_Y(f)$ is the Fourier transform of $R_Y(\tau)$. We then substitute Theorem 2(a) in the definition of the Fourier transform to obtain

$$S_Y(f) = \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(u)h(v)R_X(\tau + v - u)\, du\, dv \right) e^{-j2\pi f\tau}\, d\tau. \tag{116}$$

Substituting $\tau' = \tau + v - u$ yields

$$S_Y(f) = \underbrace{\int_{-\infty}^{\infty} h(u)e^{-j2\pi fu}\, du}_{H(f)} \underbrace{\int_{-\infty}^{\infty} h(v)e^{j2\pi fv}\, dv}_{H^*(f)} \underbrace{\int_{-\infty}^{\infty} R_X(\tau')e^{-j2\pi f\tau'}\, d\tau'}_{S_X(f)} \tag{117}$$

Thus $S_Y(f) = H(f)H^*(f)S_X(f) = |H(f)|^2 S_X(f)$. The proof is similar for random sequences.

Now we are ready to justify the interpretation of $S_X(f)$ as a density function. As shown in Figure 2, suppose $H(f)$ is a narrow, ideal bandpass filter of bandwidth B centered at frequency f_0. That is,

$$H(f) = \begin{cases} 1 & |f \pm f_0| \leq B/2, \\ 0 & \text{otherwise.} \end{cases} \tag{118}$$

In this case, passing the random process $X(t)$ through the filter $H(f)$ always produces an output waveform $Y(t)$ that is in the passband of the filter $H(f)$. As we have shown, the power spectral density of the filter output is

$$S_Y(f) = |H(f)|^2 S_X(f). \tag{119}$$

Moreover, Theorem 13(b) implies that the average power of $Y(t)$ is

$$\begin{aligned} \mathrm{E}\left[Y^2(t)\right] &= \int_{-\infty}^{\infty} S_Y(f)\, df \\ &= \int_{-f_0-B/2}^{-f_0+B/2} S_X(f)\, df + \int_{f_0-B/2}^{f_0+B/2} S_X(f)\, df. \end{aligned} \tag{120}$$

Since $S_X(f) = S_X(-f)$, when B is very small, we have

$$\mathrm{E}\left[Y^2(t)\right] \approx 2BS_X(f_0). \tag{121}$$

We see that the average power of the filter output is approximately the power spectral density of the input at the center frequency of the filter times the bandwidth of the filter. The approximation becomes exact as the bandwidth approaches zero. Therefore, $S_X(f_0)$ is the power per unit frequency (the definition of a density function) of $X(t)$ at frequencies near f_0.

Example 22

A wide sense stationary process $X(t)$ with autocorrelation function $R_X(\tau) = e^{-b|\tau|}$ is the input to an RC filter with impulse response

$$h(t) = \begin{cases} (1/RC)e^{-t/RC} & t \geq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{122}$$

Assuming $b > 0$ and $b \neq 1/RC$, find $S_Y(f)$ and $R_Y(\tau)$, the power spectral density and autocorrelation of the filter output $Y(t)$. What is the average power of the output stochastic process?

For convenience, let $a = 1/RC$. Since the impulse response has the form $h(t) = ae^{-at}u(t)$ and $R_X(\tau) = e^{-b|\tau|}$, Table 1 tells us that

$$H(f) = \frac{a}{a + j2\pi f}, \qquad S_X(f) = \frac{2b}{(2\pi f)^2 + b^2}. \tag{123}$$

Therefore,

$$|H(f)|^2 = \frac{a^2}{(2\pi f)^2 + a^2} \tag{124}$$

and, by Theorem 16, $S_Y(f) = |H(f)|^2 S_X(f)$. We use the method of partial fractions to write

$$\begin{aligned} S_Y(f) &= \frac{2ba^2}{[(2\pi f)^2 + a^2][(2\pi f)^2 + b^2]} \\ &= \frac{2ba^2/(b^2 - a^2)}{(2\pi f)^2 + a^2} - \frac{2ba^2/(b^2 - a^2)}{(2\pi f)^2 + b^2}. \end{aligned} \tag{125}$$

Recognizing that for any constant $c > 0$, $e^{-c|\tau|}$ and $2c/((2\pi f)^2 + c^2)$ are Fourier transform pairs, we obtain the output autocorrelation

$$R_Y(\tau) = \frac{ba}{b^2 - a^2} e^{-a|\tau|} - \frac{a^2}{b^2 - a^2} e^{-b|\tau|}. \tag{126}$$

From Theorem 13, we obtain the average power

$$\mathrm{E}\left[Y^2(t)\right] = R_Y(0) = \frac{ba - a^2}{b^2 - a^2} = \frac{a}{(b + a)}. \tag{127}$$

Example 23

The random sequence X_n has power spectral density

$$S_X(\phi) = 2 + 2\cos(2\pi\phi). \tag{128}$$

This sequence is the input to the filter in Quiz 2 with impulse response

$$h_n = \begin{cases} 1 & n = 0, \\ -1 & n = -1, 1, \\ 0 & \text{otherwise.} \end{cases} \tag{129}$$

Derive $S_Y(\phi)$, the power spectral density function of the output sequence Y_n. What is $\mathrm{E}[Y_n^2]$?

The discrete Fourier transform of h_n is

$$H(\phi) = 1 - e^{j2\pi\phi} - e^{-j2\pi\phi} = 1 - 2\cos(2\pi\phi). \tag{130}$$

Therefore, from Theorem 16 we have

$$\begin{aligned} S_Y(\phi) = |H(\phi)|^2 S_X(\phi) &= [1 - 2\cos(2\pi\phi)]^2 [2 + 2\cos(2\pi\phi)] \\ &= 2 - 6\cos(2\pi\phi) + 8\cos^3(2\pi\phi). \end{aligned} \tag{131}$$

Applying the identity $\cos^3(x) = 0.75\cos(x) + 0.25\cos(3x)$, we can simplify this formula to obtain

$$S_Y(\phi) = 2 + 2\cos(6\pi\phi). \tag{132}$$

We obtain the mean square value from Theorem 15(b):

$$\mathrm{E}\left[Y_n^2\right] = \int_{-1/2}^{1/2} [2 + 2\cos(6\pi\phi)]\, d\phi = 2. \tag{133}$$

Example 24

We recall that in Example 7 that the wide sense stationary random sequence X_n with expected value $\mu_X = 0$ and autocorrelation function $R_X[n] = \sigma^2\delta_n$ is passed through the order $M - 1$ discrete-time moving-average filter

$$h_n = \begin{cases} 1/M & 0 \le n \le M - 1, \\ 0 & \text{otherwise.} \end{cases} \tag{134}$$

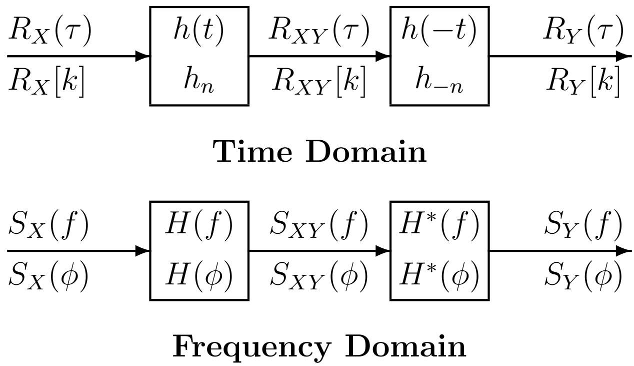


Figure 3 Input-Output Correlation and Spectral Density Functions

Find the power spectral density $S_Y(\phi)$ for the discrete-time moving-average filter output Y_n.

. . . .

By using Theorem 16, $S_Y(\phi) = |H(\phi)|^2 S_X(\phi)$. We note that $S_X(\phi) = \sigma^2$. In Example 17, we found that

$$H(\phi) = \frac{1}{M}\left(\frac{1 - e^{-j2\pi\phi M}}{1 - e^{-j2\pi\phi}}\right). \tag{135}$$

Since $|H(\phi)|^2 = H(\phi)H^*(\phi)$, it follows that

$$\begin{aligned} S_Y(\phi) = H(\phi)H^*(\phi)S_X(\phi) &= \frac{\sigma^2}{M^2}\left(\frac{1 - e^{-j2\pi\phi M}}{1 - e^{-j2\pi\phi}}\right)\left(\frac{1 - e^{j2\pi\phi M}}{1 - e^{j2\pi\phi}}\right) \\ &= \frac{\sigma^2}{M^2}\left(\frac{1 - \cos(2\pi M\phi)}{1 - \cos(2\pi\phi)}\right). \end{aligned} \tag{136}$$

We next consider the cross power spectral density function of the input and output of a linear filter. Theorem 2(b) can be interpreted as follows: If a deterministic waveform $R_X(\tau)$ is the input to a filter with impulse response $h(t)$, the output is the waveform $R_{XY}(\tau)$. Similarly, Theorem 2(c) states that if the waveform $R_{XY}(\tau)$ is the input to a filter with impulse response $h(-t)$, the output is the waveform $R_Y(\tau)$. The upper half of Figure 3 illustrates the relationships between $R_X(\tau)$, $R_{XY}(\tau)$, and $R_Y(\tau)$, as well as the corresponding relationships for discrete-time systems. The lower half of the figure demonstrates the same relationships in the frequency domain. It uses the fact that the Fourier transform of $h(-t)$, the time-reversed impulse response, is $H^*(f)$, the complex conjugate of the Fourier transform of $h(t)$. The following theorem states these relationships mathematically.

Theorem 17

If the wide sense stationary process $X(t)$ is the input to a linear time-invariant filter with transfer function $H(f)$, and $Y(t)$ is the filter output, the input-output cross power spectral density function and the output power spectral density function are

$$S_{XY}(f) = H(f)S_X(f), \qquad S_Y(f) = H^*(f)S_{XY}(f).$$

. . . .

If the wide sense stationary random sequence X_n is the input to a linear time-invariant filter with transfer function $H(\phi)$, and Y_n is the filter output, the input-output cross power spectral density function and the output power spectral density function are

$$S_{XY}(\phi) = H(\phi)S_X(\phi), \qquad S_Y(\phi) = H^*(\phi)S_{XY}(\phi).$$

Quiz 8

A wide sense stationary stochastic process $X(t)$ with expected value $\mu_X = 0$ and autocorrelation function $R_X(\tau) = e^{-5000|\tau|}$ is the input to an RC filter with time constant $RC = 100\mu$s. The filter output is the stochastic process $Y(t)$.

(a) Derive $S_{XY}(f)$, the cross power spectral density function of $X(t)$ and $Y(t)$.

(b) Derive $S_Y(f)$, the power spectral density function of $Y(t)$.

(c) What is $\mathrm{E}[Y^2(t)]$, the average power of the filter output?

9 Linear Estimation of Continuous-Time Processes

> The transfer function of the optimum linear estimator of a continuous-time stochastic process is the ratio of the cross spectral density function of the observed and estimated process and the power spectral density function of the observed process.

In this section, we observe a sample function of a wide sense stationary continuous-time stochastic process $Y(t)$ and design a linear filter to estimate a sample function of another wide sense stationary process $X(t)$. The linear filter that minimizes the mean square error is referred to as a *Wiener filter*. The properties of a Wiener filter are best represented in the frequency domain.

Theorem 18

$X(t)$ and $Y(t)$ are wide sense stationary stochastic processes with power spectral density functions $S_X(f)$ and $S_Y(f)$, and cross spectral density function $S_{XY}(f)$. $\hat{X}(t)$ is the output of a linear filter with input $Y(t)$ and transfer function $H(f)$. The transfer function that minimizes the mean square error $\mathrm{E}[(X(t)-\hat{X}(t))^2]$ is

$$\hat{H}(f) = \begin{cases} \dfrac{S_{XY}(f)}{S_Y(f)} & S_Y(f) > 0, \\ 0 & \textit{otherwise}. \end{cases}$$

The minimum mean square error is

$$e_L^* = \int_{-\infty}^{\infty} \left(S_X(f) - \frac{|S_{XY}(f)|^2}{S_Y(f)} \right) df.$$

We omit the proof of this theorem.

In practice, one of the most common estimation procedures is separating a signal from additive noise. In this application, the observed stochastic process, $Y(t) = X(t) + N(t)$, is the sum of the signal of interest and an extraneous stochastic process $N(t)$ referred to as additive noise. Usually $X(t)$ and $N(t)$ are independent stochastic processes. It follows that the power spectral function of $Y(t)$ is $S_Y(f) = S_X(f) + S_N(f)$, and the cross spectral density function is $S_{XY}(f) = S_X(f)$. By Theorem 18, the transfer function of the optimum estimation filter is

$$\hat{H}(f) = \frac{S_X(f)}{S_X(f) + S_N(f)}, \tag{137}$$

and the minimum mean square error is

$$\begin{aligned} e_L^* &= \int_{-\infty}^{\infty} \left(S_X(f) - \frac{|S_X(f)|^2}{S_X(f) + S_N(f)} \right) df \\ &= \int_{-\infty}^{\infty} \frac{S_X(f)\, S_N(f)}{S_X(f) + S_N(f)}\, df. \end{aligned} \tag{138}$$

Example 25

$X(t)$ is a wide sense stationary stochastic process with $\mu_X = 0$ and autocorrelation function

$$R_X(\tau) = \frac{\sin(2\pi 5000\tau)}{2\pi 5000\tau}. \tag{139}$$

Observe $Y(t) = X(t) + N(t)$, where $N(t)$ is a wide sense stationary process with power spectral density function $S_N(f) = 10^{-5}$. $X(t)$ and $N(t)$ are mutually independent.

(a) What is the transfer function of the optimum linear filter for estimating $X(t)$ given $Y(t)$?

(b) What is the mean square error of the optimum estimation filter?

. . . .

(a) From Table 1, we deduce the power spectral density function of $X(t)$ is

$$R_X(\tau) = 10^{-4}\,\mathrm{rect}(f/10^4) = \begin{cases} 10^{-4} & |f| \leq 5000, \\ 0 & \text{otherwise.} \end{cases} \tag{140}$$

(b) From Equation (137) we obtain $\hat{H}(f) = 1/1.1$, $|f| \leq 5000$, $\hat{H}(f) = 0$ otherwise. The mean square error in Equation (138) is

$$e_L^* = \int_{-5000}^{5000} \frac{10^{-5}10^{-4}}{10^{-4}+10^{-5}}\, df = 0.1/1.1 = 0.0909. \tag{141}$$

In this example, the optimum estimation filter is an ideal lowpass filter with an impulse response that has nonzero values for all τ from $-\infty$ to ∞. This filter cannot be implemented in practical hardware. The design of filters that minimize the mean square estimation error under practical constraints is covered in more advanced texts. Equation (138) is a lower bound on the estimation error of any practical filter.

Quiz 9

In a spread spectrum communications system, $X(t)$ is an information signal modeled as wide sense stationary stochastic process with $\mu_X = 0$ and autocorrelation function

$$R_X(\tau) = \sin(2\pi 5000\tau)/(2\pi 5000\tau). \tag{142}$$

A radio receiver obtains $Y(t) = X(t) + N(t)$, where the interference $N(t)$ is a wide sense stationary process with $\mu_N = 0$ and $\mathrm{Var}[N] = 1$ watt. The power spectral density function of $N(t)$ is $S_N(f) = N_0$ for $|f| \leq B$ and $S_N(f) = 0$ for $|f| > B$. $X(t)$ and $N(t)$ are independent.

(a) What is the relationship between the noise power spectral density N_0 and the interference bandwidth B?

(b) What is the transfer function of the optimum estimation filter that processes $Y(t)$ in order to estimate $X(t)$?

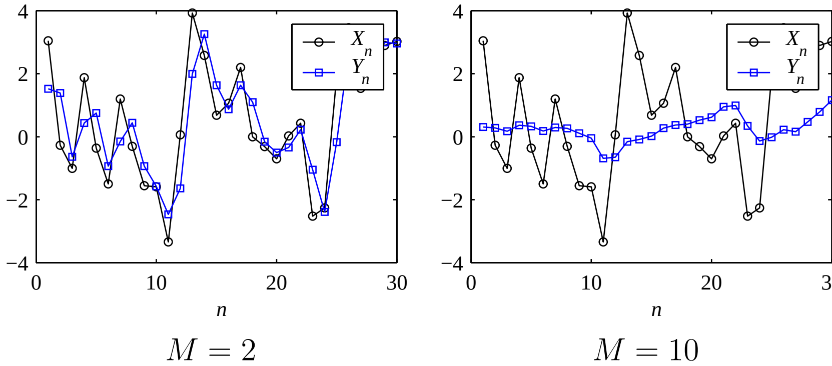


Figure 4 Sample paths of the input X_n and output Y_n of the order $M-1$ moving-average filter of Example 26.

(c) What is the minimum value of the interference bandwidth B Hz that results in a minimum mean square error $e_L^* \leq 0.05$?

10 MATLAB

We have seen in prior chapters that MATLAB makes vector and matrix processing simple. In the context of random signals, efficient use of MATLAB requires discrete-time random sequences. These may be the result either of a discrete-time process or of sampling a continuous-time process at a rate not less than the Nyquist sampling rate. In addition, MATLAB processes only finite-length vectors corrresponding to either finite-duration discrete-time signals or to length L blocks of an infinite-duration signal. In either case, we assume our objective is to process a discrete-time signal, a length L random sequence, represented by the vector

$$\mathbf{x} = \begin{bmatrix} x_0 & \cdots & x_{L-1} \end{bmatrix}'. \tag{143}$$

Note that $\mathbf{x}$ may be deterministic or random. For the moment, we drop our usual convention of denoting random quanitities by uppercase letters since in MATLAB code, we will use uppercase letters for vectors in frequency domain. The vector $\mathbf{x}$ is often called the *data* so as to distinguish it from discrete-time filters that are also represented by vectors.

Time-Domain Processing for Finite-Length Data

Time-domain methods, such as the linear filtering for a random sequence introduced in Section 2, can be implemented directly using MATLAB matrix and vector processing techniques. In MATLAB, block processing is especially common.

Example 26

For the order $M-1$ moving-average filter h_n and random processes X_n and Y_n of Example 5, write the MATLAB function `smoothfilter(L,M)` that generates a length L sample path $\mathbf{X} = \begin{bmatrix} X_0 & \cdots & X_{L-1} \end{bmatrix}'$ under the assumption that X_n is a Gaussian process. In addition, generate the length $L+M-1$ output process $\mathbf{Y} = \begin{bmatrix} Y_0 & \cdots & Y_{L+M-2} \end{bmatrix}'$.

```
function y=smoothfilter(L,M);
rx=[4 2 zeros(1,L-2)];
cx=rx-ones(size(rx));
x=gaussvector(1,cx,1);
h= ones(1,M)/M;
y=conv(h,x);
plot(1:L,x,1:L,y(1:L),':');
xlabel('\it  n');
legend('\it X_n','\it Y_n');
```

The function `smoothfilter` generates the vectors $\mathbf{X}$ and $\mathbf{Y}$. The vector `rx` is the vector of autocorrelations padded with zeros so that all needed autocorrelation terms are specified. Implementing $C_X[k] = R_X[k] - \mu_X^2$, the vector `cx` is the corresponding autocovariance function.

Next, the vector `x` of samples $X_0, \ldots, X_{L-1}$ is generated by `gaussvector`. We recall from Chapter 13 that we extended `gaussvector` such that if the second argument is an $N \times 1$ vector, then `gaussvector` internally generates the symmetric Toeplitz covariance matrix. Finally, `conv` implements the discrete convolution of X_n and the filter h_n.

For a single sample path X_n, Figure 4 shows X_n and the output of the smoothing filter for $M = 2$ and $M = 10$. For $M = 2$, the sharpest peaks of X_n are attenuated in the output sequence Y_n. However, the difference between input and output is small because the first-order filter is not very effective. For $M = 10$, much more of the high-frequency content of the signal is removed.

Given an autocorrelation function $R_X[k]$ and filter h_n, MATLAB can also use convolution and Theorem 5 to calculate the output autocorrelation $R_Y[k]$ and the input-output cross-correlation $R_{XY}[k]$.

Example 27

For random process X_n and smoothing filter h_n from Example 5, use MATLAB to calculate $R_Y[k]$ and $R_{XY}[k]$.

The solution, given by `smoothcorrelation.m`, and corresponding output are:

```
%smoothcorrelation.m
rx=[2 4 2];
h=[0.5 0.5];
rxy=conv(h,rx)
ry=conv(fliplr(h),rxy)
```

```
>> smoothcorrelation
rxy =
   1.00  3.00  3.00  1.00
ry =
   0.50  2.00  3.00  2.00  0.50
```

The functions $R_X[k]$, $R_{XY}[k]$, and $R_Y[k]$ are represented by the vectors `rx`, `rxy`, and `ry`. By Theorem 5(c), the output autocorrelation `ry` is the convolution of $R_{XY}[k]$ with the time-reversed filter h_{-n}. In MATLAB, `fliplr(h)` time reverses, or flips, the row vector `h`. Note that the vectors `rx`, `rxy`, and `ry` do not record the starting time of the corresponding functions. Because `rx` and `ry` describe autocorrelation functions, we can infer that those functions are centered around the origin. However, without knowledge of the filter h_n and autocorrelation $R_X[k]$, there is no way to deduce that `rxy(1)` $= R_{XY}[-1]$.

MATLAB makes it easy to implement linear estimation and prediction filters. In the following example, we generalize the solution of Example 12 to find the order N linear predictor for X_{n+1} given M prior samples.

Example 28

For the stationary random sequence X_n with expected value $\mu_X = 0$, write a MATLAB program `lmsepredictor.m` to calculate the order $M-1$ LMSE predictor filter $\mathbf{h}$ for $X = X_{n+1}$ given the observation $\mathbf{X}_n = \begin{bmatrix} X_{n-M+1} & \cdots & X_n \end{bmatrix}'$.

From Theorem 9, the filter vector $\mathbf{h}$ expressed in time-reversed form, is $\overleftarrow{\mathbf{h}} = \mathbf{R}_{\mathbf{X}_n}^{-1}\mathbf{R}_{\mathbf{X}_n X_{n+1}}$. In this solution, $\mathbf{R}_{\mathbf{X}_n}$ is given in Theorem 6 and $\mathbf{R}_{\mathbf{X}_n X_{n+1}}$ appears in Equation (65) with $k=1$. The solution requires that we know the set of autocorrelation values $\{R_X(0), \ldots, R_X(M)\}$. There are several ways to implement a MATLAB solution. One way would be for `lmsepredictor` to require as an argument a handle for a MATLAB function that calculates $R_X(k)$. We choose a simpler approach and pass a vector `rx` equal to $\begin{bmatrix} R_X(0) & \cdots & R_X(m-1) \end{bmatrix}'$. If $M \geq m$, then we pad `rx` with zeros.

```
function h=lmsepredictor(r,M);
m=length(r);
rx=[r(:);zeros(M-m+1,1)];
    %append zeros if needed
RY=toeplitz(r(1:M));
RYX=r(M+1:-1:2);
a=RY\RYX;
h=a(M:-1:1);
```

MATLAB provides the `toeplitz` function to generate the matrix $\mathbf{R}_{\mathbf{X}_n}$ (denoted by `RY` in the code). Finally, we find $\mathbf{a}$ such that $\mathbf{a}'\mathbf{R}_{\mathbf{X}_n} = \mathbf{R}_{X_{n+1}\mathbf{X}_n}$. Note that the code, as written, equivalently solves $\mathbf{R}_{\mathbf{X}_n}\mathbf{a} = \mathbf{R}_{\mathbf{X}_n X_{n+1}}$. The output is $\mathbf{h} = \overleftarrow{\mathbf{a}}$, the time-reversed $\mathbf{a}$.

The function `lmsepredictor.m` assumes $R_X(k) = 0$ for $k \geq m$. If $R_X(k)$ has an infinite-length tail ($R_X(k) = (0.9)^{|k|}$ for example) then the user must make sure that $m > M$; otherwise, the results will be innaccurate. This is examined in Problem 10.5.

Frequency Domain Processing of Finite-Length Data

Our frequency domain analysis of random sequences has employed the discrete-time Fourier transform (DTFT). However, the DTFT is a continuous function of the normalized frequency variable ϕ. For frequency domain methods in MATLAB, we employ the Discrete Fourier Transform (DFT).

Definition 6 Discrete Fourier Transform (DFT)

For a finite-duration signal represented by the vector $\mathbf{x} = \begin{bmatrix} x_0 & \cdots & x_{L-1} \end{bmatrix}'$, *the* N ***point DFT*** *is the length* N *vector* $\tilde{\mathbf{X}} = \begin{bmatrix} \tilde{X}_0 & \cdots & \tilde{X}_{N-1} \end{bmatrix}'$, *where*

$$\tilde{X}_n = \sum_{k=0}^{L-1} x_k e^{-j2\pi(n/N)k}, \qquad n = 0, 1, \ldots, N-1.$$

By comparing Definition 6 for the DFT and Definition 3 for the DTFT, we see that for a finite-length signal $\mathbf{x}$, the DFT evaluates the DTFT $X(\phi)$ at N equally spaced frequencies,

$$\phi_n = \frac{n}{N}, \qquad n = 0, 1, \ldots, N-1, \tag{144}$$

over the full range of normalized frequencies $0 \le \phi \le 1$. Recall that the DTFT is symmetric around $\phi = 0$ and periodic with period 1. It is customary to observe the DTFT over the interval $-1/2 \le \phi \le 1/2$. By contrast MATLAB produces values of the DFT over $0 \le \phi \le 1$. When L=N, the two transforms coincide over $0 \le \phi \le 1/2$ while the image of the DFT over $1/2 < \phi \le 1$ corresponds to samples of the DTFT over $-1/2 < \phi \le 0$. That is, $\tilde{X}_{N/2} = X(-1/2)$, $\tilde{X}_{N/2+1} = X(-1/2 + 1/N)$, and so on.

Because of the special structure of the DFT transformation, the DFT can be implemented with great efficiency as the "Fast Fourier Transform" or FFT. In MATLAB, if `x` is a length L vector, then `X=fft(x)` produces the L point DFT of `x`. In addition, `X=fft(x,N)` produces the N point DFT of `x`. In this case, if $L < N$, then MATLAB pads the vector `x` with $N - L$ zeros. However, if $L > N$, then MATLAB returns the DFT of the truncation of `x` to its first N elements. Note that this is inconsistent with Definition 6 of the DFT. In general, we assume $L = N$ when we refer to an N point DFT without reference to the data length L. In the form of the FFT, the DFT is both efficient and useful in representing a random process in the frequency domain.

Example 29

Let h_n denote the order $M - 1$ smoothing filter of Example 5. For $M = 2$ and $M = 10$, compute the $N = 32$ point DFT. For each M, plot the magnitude of

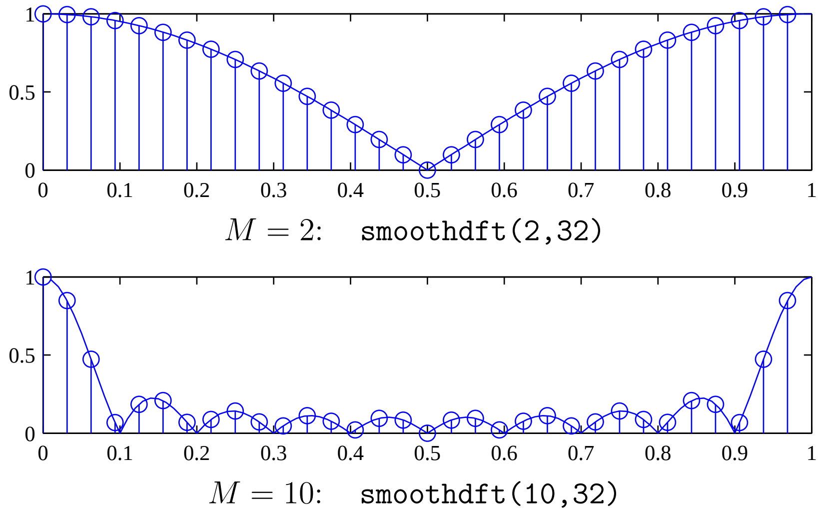


Figure 5 The magnitude of the DTFT and the 32-point DFT for the order $M-1$ moving-average filter h_n of Example 29.

each DFT and the magnitude of the corresponding DTFT $|H(\phi)|$.

. . . .

We calculate the DFT of the M-element vector $\mathbf{h} = \begin{bmatrix} 1/M & \cdots & 1/M \end{bmatrix}'$ for each M. In Example 17, we found that the DTFT of h_n is

$$H(\phi) = \frac{1}{M}\left(\frac{1-e^{-j2\pi\phi M}}{1-e^{-j2\pi\phi}}\right). \tag{145}$$

```
function smoothdft(M,N);
phi=0.01:0.01:1;
hdtft=(1-exp(-j*2*pi*phi*M))./ ...
    (M*(1-exp(-j*2*pi*phi)));
h=ones(1,M)/M;
hdft=fft(h,N);
n=(0:(N-1))/N;
stem(n,abs(hdft));
hold on;
plot(phi,abs(hdtft));
hold off;
```

The program `smoothdft.m` calculates $H(\phi)$ and the N point DFT and generates the graphs shown in Figure 5. The magnitude $|H(\phi)|$ appears in the figure as the solid curve. The DFT magnitude is displayed as a stem plot for normalized frequencies n/N to show the correspondence with $|H(\phi)|$ at frequencies $\phi = n/N$.

In Figures 5, 7, and 8, we see the mirror image symmetry of both the DTFT and the DFT around frequency $\phi = 1/2$. This is a general property of the DFT for real-valued signals. We explain this property for the smoothing filter h_n of Example 29. When the elements of h_n are real, $H(-\phi) = H^*(\phi)$. Thus $|H(-\phi)| = |H(\phi)|$, corresponding to mirror image symmetry for the magnitude spectrum about $\phi = 0$. Moreover, because the DTFT $H(\phi)$ is periodic with unit period, $H(\phi) = H(\phi - 1)$. Thus $H(\phi)$ over the interval $1/2 \leq \phi < 1$ is identical to $H(\phi)$ over the interval $-1/2 \leq \phi < 0$. That is, in a graph of either the DTFT or DFT, the right side of the graph corresponding to normalized frequencies $1/2 \leq \phi < 1$ is the same as the image for negative frequencies $-1/2 \leq \phi < 1$. Combining this observation with the mirror image symmetry of $H(\phi)$ about $\phi = 0$ yields mirror image symmetry around $\phi = 1/2$. Thus, when interpreting a DTFT or DFT graph, low frequencies are represented by the points near the left edge and right edge of the graph. High-frequency components are described by points near the middle of the graph.

When $L = N$, we can express the DFT as the $N \times N$ matrix transformation

$$\tilde{\mathbf{X}} = \mathrm{DFT}(\mathbf{x}) = \tilde{\mathbf{F}}\mathbf{x} \tag{146}$$

where the DFT matrix $\tilde{\mathbf{F}}$ has n, k element

$$\tilde{\mathbf{F}}_{nk} = e^{-j2\pi(n/N)k} = \left[e^{-j2\pi/N}\right]^{nk}. \tag{147}$$

Because $\tilde{\mathbf{F}}$ is an $N \times N$ square matrix with a special structure, its inverse exists and is given by

$$\tilde{\mathbf{F}}^{-1} = \frac{1}{N}\tilde{\mathbf{F}}^*, \tag{148}$$

where $\tilde{\mathbf{F}}^*$ is obtained by taking the complex conjugate of each entry of $\tilde{\mathbf{F}}$. Thus, given the DFT $\tilde{\mathbf{X}} = \tilde{\mathbf{F}}\mathbf{x}$, we can recover the original signal $\mathbf{x}$ by the inverse DFT

$$\mathbf{x} = \mathrm{IDFT}(\tilde{\mathbf{X}}) = \frac{1}{N}\tilde{\mathbf{F}}^*\tilde{\mathbf{X}}. \tag{149}$$

Example 30

Write a function `F=dftmat(N)` that returns the N-point DFT matrix. Find the DFT matrix F for $N = 4$ and show numerically that $F^{-1} = (1/N)F^*$.

```
function F = dftmat(N);
Usage: F=dftmat(N)
%F is the N by N DFT matrix
n=(0:N-1)';
F=exp((-1.0j)*2*pi*(n*(n'))/N);
```

The function `dftmat.m` is a direct implementation of Equation (147). In the code, `n` is the column vector $\begin{bmatrix}0 & 1 & \cdots N-1\end{bmatrix}'$ and `n*(n')` produces the $N \times N$ matrix with n, k element equal to nk.

As shown in Figure 6, executing `F=dftmat(4)` produces the 4×4 DFT matrix

$$F = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -j & -1 & j \\ 1 & -1 & 1 & -1 \\ 1 & j & -1 & -j \end{bmatrix}. \tag{150}$$

In the figure, we also verify that $F^{-1} = (1/4)F^*$.

As a consequence, the structure of $\tilde{\mathbf{F}}$ allows us to show in the next theorem that an algorithm for the DFT can also be used as an algorithm for the IDFT.

Theorem 19

$$\mathrm{IDFT}(\tilde{\mathbf{X}}) = \frac{1}{N}\left(\mathrm{DFT}(\tilde{\mathbf{X}}^*)\right)^*.$$

```
>> F=dftmat(4)
F =
   1.00             1.00             1.00             1.00
   1.00             0.00 - 1.00i    -1.00 - 0.00i    -0.00 + 1.00i
   1.00            -1.00 - 0.00i     1.00 + 0.00i    -1.00 - 0.00i
   1.00            -0.00 + 1.00i    -1.00 - 0.00i     0.00 - 1.00i
>> (1/4)*F*(conj(F))
ans =
   1.00            -0.00 + 0.00i     0.00 + 0.00i     0.00 + 0.00i
  -0.00 - 0.00i     1.00            -0.00             0.00 + 0.00i
   0.00 - 0.00i    -0.00 - 0.00i     1.00 + 0.00i    -0.00 + 0.00i
   0.00 - 0.00i     0.00 - 0.00i    -0.00 - 0.00i     1.00
>>
```

Figure 6 The 4-point DFT matrix for Example 30 as generated by `dftmat.m`

Proof By taking complex conjugates of Equation (149), we obtain

$$\mathbf{x}^* = \frac{1}{N}\tilde{\mathbf{F}}\tilde{\mathbf{X}}^* = \frac{1}{N}\,\mathrm{DFT}(\tilde{\mathbf{X}}^*). \tag{151}$$

By conjugating once again, we obtain

$$\mathbf{x} = \frac{1}{N}\left(\mathrm{DFT}(\tilde{\mathbf{X}}^*)\right)^*. \tag{152}$$

Observing that $\mathbf{x} = \mathrm{IDFT}(\tilde{\mathbf{X}})$ completes the proof.

Thus, if $\mathbf{x}$ is a length L data vector with $L > N$, then the N-point DFT produces a vector $\tilde{\mathbf{X}} = \mathrm{DFT}(\mathbf{x})$ of length N. For MATLAB, `X=fft(x,N)` produces the DFT of the truncation of `x` to its first N values. In either case, the IDFT defined by the $N \times N$ matrix $\tilde{\mathbf{F}}^{-1}$, or equivalently `ifft`, also returns a length N output. If $L > N$, the IDFT output cannot be the same as the length L original input `x`. In the case of MATLAB, `ifft` returns the truncated version of the original signal. If $L < N$, then the inverse DFT returns the original input $\mathbf{x}$ padded with zeros to length N.

The DFT for the Power Spectral Density

The DFT can also be used to transform the discrete autocorrelation $R_X[n]$ into the power spectral density $S_X(\phi)$. We assume that $R_X[n]$ is the length $2L-1$ signal given by the vector

$$\mathbf{r} = \begin{bmatrix} R_X\left[-(L-1)\right] & R_X\left[-L+2\right] & \cdots & R_X\left[L-1\right] \end{bmatrix}'. \tag{153}$$

The corresponding PSD is

$$S_X\left(\phi\right) = \sum_{k=-(L-1)}^{L-1} R_X\left[k\right] e^{-j2\pi\phi k}. \tag{154}$$

Given a vector $\mathbf{r}$, we would like to generate the vector $\mathbf{S} = \begin{bmatrix} S_0 & \cdots & S_{N-1} \end{bmatrix}'$ such that $S_n = S_X(n/N)$. A complication is that the PSD given by Equation (154) is a double-sided transform while the DFT in the form of Definition 6, or the MATLAB function `fft`, is a single-sided transform. For single-sided signals, the two transforms are the same. However, for a double-sided autocorrelation function $R_X[n]$, the difference matters.

Although the vector $\mathbf{r}$ represents an autocorrelation function, it can be viewed as simply a sequence of numbers, the same as any other discrete-time signal. Viewed this way, MATLAB can use `fft` to calculate the DFT $\tilde{\mathbf{R}} = \mathrm{DFT}(\mathbf{r})$ where

$$\tilde{R}_n = \sum_{l=0}^{2L-2} R_X\left[l-(L-1)\right] e^{-j2\pi(n/N)l}. \tag{155}$$

We make the substitution $k = l-(L-1)$ to write

$$\begin{aligned}\tilde{R}_n &= \sum_{k=-(L-1)}^{L-1} R_X\left[k\right] e^{-j2\pi(n/N)(k+L-1)} \\ &= e^{-j2\pi(n/N)(L-1)} S_X\left(n/N\right). \end{aligned} \tag{156}$$

Equivalently, we can write

$$S_n = S_X\left(n/N\right) = \tilde{\mathbf{R}}_n e^{j2\pi(n/N)(L-1)}. \tag{157}$$

Thus the difference between the one-sided DFT and the double-sided DTFT results in the elements of $\tilde{\mathbf{R}}$ being linearly phase shifted from the values of $S_X(n/N)$ that we wished for. If all we are interested in is the magnitude of the frequency response, this phase can be ignored because $|\tilde{\mathbf{R}}_n| = |S_X(n/N)|$. However, in many instances, the correct phase of the PSD is needed.

```
function S=fftc(varargin);
%DFT for a signal r
%centered at the origin
%Usage:
%  fftc(r,N): N point DFT of r
%  fftc(r): length(r) DFT of r
r=varargin{1};
L=1+floor(length(r)/2);
if (nargin>1)
    N=varargin{2}(1);
else
    N=(2*L)-1;
end
R=fft(r,N);
n=reshape(0:(N-1),size(R));
phase=2*pi*(n/N)*(L-1);
S=R.*exp((1.0j)*phase);
```

In this case, the progam `fftc.m` provides a DFT for signals `r` that are *centered* at the origin. Given the simple task of `fftc`, the code may seem unusually long. In fact, the actual phase shift correction occurs only in the very last line line of the program. The additional code employs MATLAB's `varargin` function to support the same variable argument calling conventions as `fft`; `fftc(r)` returns the n-point DFT where n is the length of `r`, while `fftc(r,N)` returns the `N`-point DFT.

Circular Convolution

Given a discrete-time signal x_n as the input to a filter h_n, we know that the output y_n is the convolution of x_n and h_n. Equivalently, convolution in the time domain yields the multiplication $Y(\phi) = H(\phi)X(\phi)$ in the frequency domain.

When x_n is a length L input given by the vector $\mathbf{x} = \begin{bmatrix} x_0 & \cdots & x_{L-1} \end{bmatrix}'$, and h_n is the order $M-1$ causal filter $\mathbf{h} = \begin{bmatrix} h_0 & \cdots & h_{M-1} \end{bmatrix}'$, the corresponding output $\mathbf{y}$ has length $L+M-1$. If we choose $N \geq L+M-1$, then the N-point DFTs of $\mathbf{h}$, $\mathbf{x}$, and $\mathbf{y}$ have the property that

$$\tilde{\mathbf{Y}}_k = Y(k/N), \quad \tilde{\mathbf{H}}_k = H(k/n), \quad \tilde{\mathbf{X}}_k = X(k/N). \tag{158}$$

It follows that at frequencies k/N, we have that $\tilde{\mathbf{Y}}_k = \tilde{\mathbf{H}}_k \tilde{\mathbf{X}}_k$. Moreover, IDFT($\tilde{\mathbf{Y}}$) will recover the time-domain output signal y_n.

Infinite-duration Random Sequences

We have observed that for finite-duration signals and finite-order filters, the DFT can be used almost as a substitute for the DTFT. The only noticeable differences are phase shifts that result from the `fft` function assuming that all signal vectors $\mathbf{r}$ are indexed to start at time $n = 0$. For sequences of length $L \leq N$, the DFT provides both an invertible frequency domain transformation as well as providing samples of the DTFT at frequencies $\phi_n = n/N$.

On the other hand, there are many applications involving infinite duration random sequences for which we would like to obtain a frequency domain characterization. In this case, we can only apply the DFT to a finite-length data sample $\mathbf{x} = \begin{bmatrix} x_i & x_{i+1} & \cdots & x_{i+L-1} \end{bmatrix}'$. In this case, the DFT should be used with some care as an approximation or substitute for the DTFT. First, as we see in the following example, the DFT frequency samples may not exactly match the peak frequencies of $X(\phi)$.

Example 31

Calculate the $N = 20$ point and $N = 64$ point DFT of the length $L = N$ sequence $x_k = \cos(2\pi(0.15)k)$. Plot the magnitude of the DFT as a function of the normalized frequency.

. . . .

```
function X=fftexample1(L,N)
k=0:L-1;
x=cos(2*pi*(0.15)*k);
X=fft(x,N);
phi=(0:N-1)/N;
stem(phi,abs(X));
```

The program `fftexample1.m` generates and plots the magnitude of the N-point DFT. The plots for $N = 20$ and $N = 64$ appear in Figure 7. We see that the DFT does not precisely describe the DTFT unless the sinusoid x_k is at frequency n/N for an integer n.

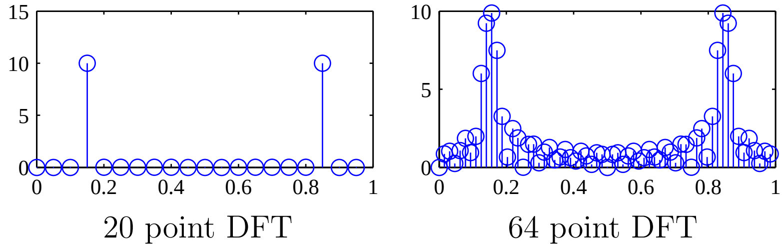


Figure 7 20 point and 64 point DFTs of $x_k = \cos(2\pi(0.15)k)$

Even if N is chosen to be large, the DFT is likely to be a poor approximation to the DTFT if the data length L is chosen too small. To understand this, we view $\mathbf{x}$ as a vector representation of the signal $x_k w_L[k]$ where $w_L[k]$ is a windowing function satisfying

$$w_L[k] = \begin{cases} 1 & k = 0, \ldots, L-1, \\ 0 & \text{otherwise.} \end{cases} \tag{159}$$

From Equation (159), we see that $w_L[k]$ is just a scaled version of the order $M-1$ moving-average filter h_n in Example 26. In particular, for $M = L$, $w_L[k] = Lh_k$ and the DTFT of $w_L[k]$ is $W_L(\phi) = LH(\phi)$. Thus, for $L = M = 10$, $W_L(\phi)$ will resemble $H(\phi)$ in the lower graph of Figure 5. The DTFT of the time domain multiplication $x_k w_L[k]$ results in the frequency domain convolution of $X(\phi)$ and $W_L(\phi)$. Thus the windowing introduces spurious high-frequency components associated with the sharp transitions of the function $w_L[k]$. In particular, if $X(\phi)$ has sinusoidal components at frequencies ϕ_1 and ϕ_2, those signal components will not be distinguishable if $|\phi_1 - \phi_2| < 1/L$, no matter how large we choose N.

Example 32

Consider the discrete-time function

$$x_k = \cos(2\pi(0.20)k) + \cos(2\pi(0.25)k), \qquad k = 0, 1, \ldots, L-1. \tag{160}$$

For $L = 10$ and $L = 20$, plot the magnitude of the $N = 20$ point DFT.

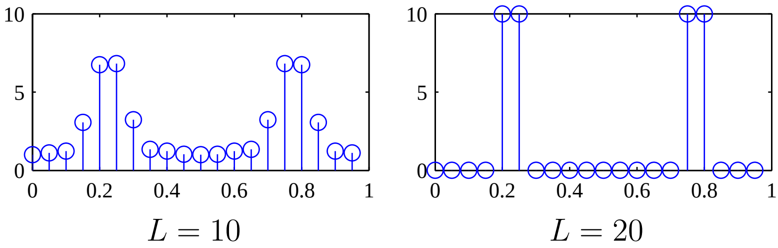


Figure 8 The $N = 20$ point DFT of $x_k = \cos(2\pi(0.20)k) + \cos(2\pi(0.25)k)$ for data length L.

```
function X=fftexample2(L,N)
k=0:L-1;
x=cos(2*pi*(0.20)*k) ...
      +cos(2*pi*(0.25)*k);
X=fft(x,N);
stem((0:N-1)/N,abs(X));
```

The program `fftexample2.m` plots the magnitude of the N-point DFT. Results for $L = 10, 20$ appear in Figure 8. The graph for $L = 10$ shows how the short data length introduce spurious frequency components even though the $N = 20$ point DFT should be sufficient to identify the two sinusoidal components. For data length $L = 20$, we are able to precisely identify the two signal components.

We have observed that we can specify the number of frequency points N of the DFT separately from the data length L. However, if x_n is an infinite-length signal, then the ability of the DFT to resolve frequency components is limited by the data length L. Thus, if we are going to apply an N point DFT to a length L sequence extracted from an infinite-duration data signal, there is little reason to extract fewer than N samples.

On the other hand, if $L > N$, we can generate a length N *wrapped signal* with the same DFT. In particular, if $\mathbf{x}$ has length $L > N$, we express $\mathbf{x}$ as the concatenation of length N segments

$$\mathbf{x} = [\mathbf{x}_1 \ \mathbf{x}_2 \ \cdots \ \mathbf{x}_i]. \tag{161}$$

where the last segment $\mathbf{x}_i$ may be padded with zeroes to have length N. In this case, we define the length N wrapped signal as $\hat{\mathbf{x}} = \mathbf{x}_1 + \mathbf{x}_2 + \cdots + \mathbf{x}_i$. Because $e^{-j2\pi(n/N)k}$ has period N, it can be shown that $\mathbf{x}$ and $\hat{\mathbf{x}}$ have the same DFT. If $\tilde{\mathbf{X}} = \mathrm{DFT}(\mathbf{x}) = \mathrm{DFT}(\hat{\mathbf{x}})$, then $\mathrm{IDFT}(\tilde{\mathbf{X}})$ will return the wrapped signal $\hat{\mathbf{x}}$. Thus, to preserve the invertibility of the DFT, it is desirable to use data record of length L equal to N, the number of points in the DFT.

Quiz 10

The wide sense stationary process X_n of Example 5 is passed through the order $M-1$ moving-average filter. The output is Y_n. Use a 32 point DFT to plot the power spectral density functions

(a) $S_X(\phi)$,

(b) $S_Y(\phi)$ for $M = 2$,

(c) $S_Y(\phi)$ for $M = 10$.

Use these graphs to interpret the results of Example 26.

Further Reading: Probability theory, stochastic processes, and digital signal processing (DSP) intersect in the study of random signal processing. [MSY98] and [Orf96] are accessible entry points to the study of DSP with the help of MATLAB. [BH92] covers the theory of random signal processing in depth and concludes with a chapter devoted to the Global Positioning System. [Hay96] presents basic principles of random signal processing in the context of a few important practical applications.

BH92. R. G. Brown and P. Y. C. Hwang. *Introduction to Random Signals.* John Wiley & Sons, 2nd edition, 1992.

Hay96. M. H. Hayes. *Statistical Digital Signal Processing and Modeling.* 1996.

MSY98. J. H. McClellan, R.W. Shafer, and M.A. Yoder. *DSP First: A Multimedia Approach.* Prentice Hall, 1998.

Orf96. S. J. Orfanidis. *Introduction to Signal Processing.* Prentice Hall, 1996.

Problems

Difficulty: ● Easy ■ Moderate ♦ Difficult ♦♦ Experts Only

1.1● Let $X(t)$ denote a wide sense stationary process with $\mu_X = 0$ and autocorrelation $R_X(\tau)$. Let $Y(t) = 2 + X(t)$. What is $R_Y(t, \tau)$? Is $Y(t)$ wide sense stationary?

1.2● $X(t)$, the input to a linear time-invariant filter is a wide sense stationary random process with expected value $\mu_X = -3$ volts. The impulse response of the filter is

$$h(t) = \begin{cases} 1 - 10^6 t^2 & 0 \le t \le 10^{-3} \text{ sec}, \\ 0 & \text{otherwise.} \end{cases}$$

What is the expected value of the output process $Y(t)$?

1.3● $X(t)$, the input to a linear time-invariant filter is a wide sense stationary stochastic process with expected value $\mu_X = 4$ volts. The filter output $Y(t)$ is a wide sense stationary stochastic process with expected $\mu_Y = 1$ volt. The filter impulse response is

$$h(t) = \begin{cases} e^{-t/a} & t \ge 0, \\ 0 & \text{otherwise.} \end{cases}$$

What is the value of the time constant a?

1.4■ A white Gaussian noise signal $W(t)$ with autocorrelation function $R_W(\tau) = \eta_0 \delta(\tau)$ is passed through an LTI filter $h(t)$. Prove that the output $Y(t)$ has average power

$$\mathrm{E}\left[Y^2(t)\right] = \eta_0 \int_{-\infty}^{\infty} h^2(u)\, du.$$

2.1● The random sequence X_n is the input to a discrete-time filter. The output is

$$Y_n = \frac{X_{n+1} + X_n + X_{n-1}}{3}.$$

(a) What is the impulse response h_n?

(b) Find the autocorrelation of the output Y_n when X_n is a wide sense stationary random sequence with $\mu_X = 0$ and autocorrelation

$$R_X[n] = \begin{cases} 1 & n = 0, \\ 0 & \text{otherwise.} \end{cases}$$

2.2● $X(t)$ is a wide sense stationary process with autocorrelation function

$$R_X(\tau) = 10 \frac{\sin(2000\pi t) + \sin(1000\pi t)}{2000\pi t}.$$

The process $X(t)$ is sampled at rate $1/T_s = 4{,}000$ Hz, yielding the discrete-time process X_n. What is the autocorrelation function $R_X[k]$ of X_n?

2.3■ The output Y_n of the smoothing filter in Example 5 has $\mu_Y = 1$ and autocorrelation function

$$R_Y[n] = \begin{cases} 3 & n = 0, \\ 2 & |n| = 1, \\ 0.5 & |n| = 2, \\ 0 & \text{otherwise.} \end{cases}$$

Y_n is the input to another smoothing filter with impulse response

$$h_n = \begin{cases} 1 & n = 0, 1, \\ 0 & \text{otherwise.} \end{cases}$$

The output of this second smoothing filter is W_n.

(a) What is μ_W, the expected value of the output of the second filter?

(b) What is the autocorrelation function of W_n?

(c) What is the variance of W_n?

(d) W_n is a linear function of the original input process X_n in Example 5:

$$W_n = \sum_{i=0}^{M-1} g_i X_{n-i}.$$

What is the impulse response g_i of the combined filter?

2.4■ Let the random sequence Y_n in Problem 2.3 be the input to the differentiator with impulse response,

$$h_n = \begin{cases} 1 & n = 0, \\ -1 & n = 1, \\ 0 & \text{otherwise.} \end{cases}$$

The output is the random sequence V_n.

(a) What is the expected value of the output μ_V?

(b) What is the autocorrelation function of the output $R_V[n]$?

(c) What is the variance of the output $\text{Var}[V_n]$?

(d) V_n is a linear function of the original input process X_n in Example 5:

$$V_n = \sum_{i=0}^{M-1} f_i X_{n-i}.$$

What is the impulse response f_i of the combined filter?

2.5■ Example 5 describes a discrete-time smoothing filter with impulse response

$$h_n = \begin{cases} 1 & n = 0, 1, \\ 0 & \text{otherwise.} \end{cases}$$

For a particular wide sense stationary input process X_n, the output process Y_n is a wide sense stationary random sequence, with $\mu_Y = 0$ and autocorrelation function

$$R_Y[n] = \begin{cases} 2 & n = 0, \\ 1 & |n| = 1, \\ 0 & \text{otherwise.} \end{cases}$$

What is the autocorrelation $R_X[n]$ of the input X_n?

2.6♦ The input $\ldots, X_{-1}, X_0, X_1, \ldots$ and output $\ldots, Y_{-1}, Y_0, Y_1, \ldots$ of a digital filter obey

$$Y_n = \frac{1}{2}\left(X_n + Y_{n-1}\right).$$

Let the inputs be a sequence of iid random variables with $\text{E}[X_i] = \mu_{X_i} = 0$ and $\text{Var}[X_i] = \sigma^2$. Find the following properties of the output sequence: $\text{E}[Y_i]$, $\text{Var}[Y_i]$, $\text{Cov}[Y_{i+1}, Y_i]$, and ρ_{Y_{i+1},Y_i}.

2.7♦ $Z_0, Z_1, \ldots$ is an iid random sequence with $\text{E}[Z_n] = 0$ and $\text{Var}[Z_n] =$

$\hat{\sigma}^2$. The random sequence $X_0, X_1, \ldots$ obeys the recursive equation

$$X_n = cX_{n-1} + Z_{n-1}, \quad n = 1, 2, \ldots,$$

where c is a constant satisfying $|c| < 1$. Find $\hat{\sigma}^2$ such that $X_0, X_1, \ldots$ is a random sequence with $\mathrm{E}[X_n] = 0$ such that for $n \geq 0$ and $n + k \geq 0$,

$$R_X[n, k] = \sigma^2 c^{|k|}.$$

2.8◆◆ The iid random sequence $X_0, X_1, \ldots$ of standard normal random variables is the input to a digital filter. For a constant a satisfying $|a| < 1$, the filter output is the random sequence $Y_0, Y_1, \ldots$ such that $Y_0 = 0$ and for $n > 1$,

$$Y_n = a\left(X_n + Y_{n-1}\right).$$

Find $\mathrm{E}[Y_i]$ and the autocorrelation $R_Y[m, k]$ of the output sequence. Is the random sequence Y_n wide sense stationary?

3.1● X_n is a stationary Gaussian sequence with expected value $\mathrm{E}[X_n] = 0$ and autocorrelation function $R_X[k] = 2^{-|k|}$. Find the PDF of $\mathbf{X} = [X_1 \quad X_2 \quad X_3]'$.

3.2● X_n is a sequence of independent random variables such that $X_n = 0$ for $n < 0$ while for $n \geq 0$, each X_n is a Gaussian $(0, 1)$ random variable. Passing X_n through the filter $\mathbf{h} = [1 \quad -1 \quad 1]'$ yields the output Y_n. Find the PDFs of:

(a) $\mathbf{Y}_3 = [Y_1 \quad Y_2 \quad Y_3]'$,

(b) $\mathbf{Y}_2 = [Y_1 \quad Y_2]'$.

3.3■ The stationary Gaussian process X_n with expected value $\mathrm{E}[X_n] = 0$ and autocorrelation function $R_X[k] = 2^{-|k|}$ is passed through the linear filter $\mathbf{h} = [1 \quad -1 \quad 1]'$, yielding the output Y_n. Find the PDF of $\mathbf{Y} = [Y_3 \quad Y_4 \quad Y_5]'$.

3.4■ The stationary Gaussian process X_n with expected value $\mathrm{E}[X_n] = 0$ and autocorrelation function $R_X[k] = 2^{-|k|}$ is passed through the linear filter $\mathbf{h} = [1 \quad 0 \quad -1]'$. For the filter output Y_n, find the PDF of $\mathbf{Y} = [Y_3 \quad Y_4 \quad Y_5]'$.

4.1● Let X_n be a wide sense stationary Gaussian random sequence with expected value $\mathrm{E}[X_n] = 0$ and autocorrelation function

$$R_X[k] = \mathrm{E}\left[X_n X_{n+k}\right] = 2^{-|k|}.$$

We observe the noisy random sequence $Y_n = X_n + Z_n$ where Z_n is an iid Gaussian noise sequence, independent of X_n, with $\mathrm{E}[Z_n] = 0$ and $\mathrm{Var}[Z_n] = 1/2$.

(a) Find the LMSE estimate $\hat{X}_n$ of X_n given *only* the observation Y_n. Note that $\hat{X}_n$ cannot use prior observations $Y_{n-1}, Y_{n-2}, \ldots$ That is, $\hat{X}_n$ is the output of an order 0 filter.

(b) What is the mean square error e_0 of the estimator $\hat{X}_n$?

(c) What is the PDF $f_{\hat{X}_n}(x)$ of $\hat{X}_n$?

(d) Find the conditional expectation $\mathrm{E}[X_n | Y_n]$.

4.2● X_n is a wide sense stationary random sequence with $\mu_X = 0$ and autocorrelation function

$$R_X[k] = \begin{cases} 1 - 0.25\,|k| & |k| \leq 4, \\ 0 & \text{otherwise.} \end{cases}$$

For $M = 2$ samples, find $\mathbf{h} = [h_0 \quad h_1]'$, the coefficients of the optimum linear prediction filter of X_{n+1}, given X_{n-1} and X_n. What is the mean square error of the optimum linear predictor?

4.3■ A communication system, is used to transmit a sequence of binary random variables $X_1, X_2, \ldots$ such that $X_n \in \{-1, 1\}$ form a wide sense stationary process with $\mathrm{E}[X_n] = 0$ and autocovariance function:

$$C_X[k] = \begin{cases} 1 & k = 0, \\ \alpha & k = 1, \\ 0 & \text{otherwise.} \end{cases}$$

For each transmitted X_n, the receiver observes $Y_n = X_n + W_n$ where the W_i are iid Gaussian $(0, 1)$ noises, independent of all X_j.

(a) What values of α are possible?

(b) Based on the observation Y_n, find $\hat{X}_n$, the linear MMSE estimate of X_n given Y_n.

(c) Based on the observation $\mathbf{Y}_n = [Y_n, Y_{n-1}]'$, find $\tilde{X}_n$, the linear MMSE estimate of X_n given $\mathbf{Y}_n$.

4.4● X_n is a wide sense stationary random sequence with $\mu_X = 0$ and autocorrelation function

$$R_X[k] = \begin{cases} 1.1 & k = 0, \\ 1 - 0.25\,|k| & 1 \leq |k| \leq 4, \\ 0 & \text{otherwise.} \end{cases}$$

For $M = 2$ samples, find $\mathbf{h} = [h_0 \quad h_1]'$, the coefficients of the optimum linear predictor of X_{n+1}, given X_{n-1} and X_n. What is the mean square error of the optimum linear predictor?

4.5● Let X_n be a wide sense stationary random sequence with expected value $\mathrm{E}[X_n] = 0$ and autocorrelation function

$$R_X[k] = \mathrm{E}[X_n X_{n+k}] = c^{|k|},$$

where $|c| < 1$. We observe the random sequence

$$Y_n = X_n + Z_n,$$

where Z_n is an iid noise sequence, independent of X_n, with $\mathrm{E}[Z_n] = 0$ and $\mathrm{Var}[Z_n] = \eta^2$. Find the LMSE estimate of X_n given Y_n.

4.6■ Continuing Problem 4.5, find the mean square error of the optimal linear filter $\mathbf{h} = [h_0 \quad h_1]'$ based on $M = 2$ samples of Y_n. What value of c minimizes the mean square estimation error?

4.7◆ Suppose X_n is a random sequence satisfying

$$X_n = cX_{n-1} + Z_{n-1},$$

where $Z_1, Z_2, \ldots$ is an iid random sequence with $\mathrm{E}[Z_n] = 0$ and $\mathrm{Var}[Z_n] = \sigma^2$ and c is a constant satisfying $|c| < 1$. In addition, for convenience, we assume $\mathrm{E}[X_0] = 0$ and $\mathrm{Var}[X_0] = \sigma^2/(1 - c^2)$. We make the following noisy measurement

$$Y_{n-1} = dX_{n-1} + W_{n-1},$$

where $W_1, W_2, \ldots$ is an iid measurement noise sequence with $\mathrm{E}[W_n] = 0$ and $\mathrm{Var}[W_n] = \eta^2$ that is independent of X_n and Z_n.

(a) Find the optimal linear predictor, $\hat{X}_n(Y_{n-1})$, of X_n using the noisy observation Y_{n-1}.

(b) Find the mean square estimation error

$$e_L^*(n) = \mathrm{E}\left[\left(X_n - \hat{X}_n(Y_{n-1})\right)^2\right].$$

5.1● $X(t)$ is a wide sense stationary process with autocorrelation function

$$R_X(\tau) = 10\frac{\sin(2000\pi\tau) + \sin(1000\pi\tau)}{2000\pi\tau}.$$

What is the power spectral density of $X(t)$?

5.2■ $X(t)$ is a wide sense stationary process with $\mu_X = 0$ and $Y(t) = X(\alpha t)$ where α is a nonzero constant. Find $R_Y(\tau)$ in terms of $R_X(\tau)$. Is $Y(t)$ wide sense stationary? If so, find the power spectral density $S_Y(f)$.

6.1● X_n is a wide sense stationary discrete-time random sequence with autocorrelation function $R_X[k]$ such that

$$R_X[k] = \delta[k] + (0.1)^{|k|}$$

for $k = 0, \pm1, \pm2, \ldots$, and $R_X[k] = 0$ otherwise. Find the power spectral density $S_X(f)$.

7.1● For jointly wide sense stationary processes $X(t)$ and $Y(t)$, prove that the cross spectral density satisfies

$$S_{YX}(f) = S_{XY}(-f) = [S_{XY}(f)]^*.$$

8.1● A wide sense stationary process $X(t)$ with autocorrelation function $R_X(\tau) = 100e^{-100|\tau|}$ is the input to an RC filter with impulse response

$$h(t) = \begin{cases} e^{-t/RC} & t \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$

The filter output process has average power $\mathrm{E}[Y^2(t)] = 100$.

(a) Find the output autocorrelation $R_Y(\tau)$.

(b) What is the value of RC?

8.2● Let $W(t)$ denote a wide sense stationary Gaussian noise process with $\mu_W = 0$ and power spectral density $S_W(f) = 1$.

(a) What is $R_W(\tau)$, the autocorrelation of $W(t)$?

(b) $W(t)$ is the input to a linear time-invariant filter with impulse response

$$H(f) = \begin{cases} 1 & |f| \leq B/2 \\ 0 & \text{otherwise.} \end{cases}$$

The filter output is $Y(t)$. What is the power spectral density function of $Y(t)$?

(c) What is the average power of $Y(t)$?

(d) What is the expected value of the filter output?

8.3■ The wide sense stationary process $X(t)$ with autocorrelation function $R_X(\tau)$ and power spectral density $S_X(f)$ is the input to a tapped delay line filter

$$H(f) = a_1 e^{-j2\pi f t_1} + a_2 e^{-j2\pi f t_2}.$$

Find the output power spectral density $S_Y(f)$ and the output autocorrelation $R_Y(\tau)$.

8.4■ A wide sense stationary process $X(t)$ with autocorrelation function

$$R_X(\tau) = e^{-4\pi\tau^2}.$$

is the input to a filter with transfer function

$$H(f) = \begin{cases} 1 & 0 \le |f| \le 2, \\ 0 & \text{otherwise.} \end{cases}$$

Find

(a) The average power of the input $X(t)$
(b) The output power spectral density $S_Y(f)$
(c) The average power of the output $Y(t)$

8.5■ A wide sense stationary process $X(t)$ with power spectral density

$$S_X(f) = \begin{cases} 10^{-4} & |f| \le 100, \\ 0 & \text{otherwise,} \end{cases}$$

is the input to an RC filter with frequency response

$$H(f) = \frac{1}{100\pi + j2\pi f}.$$

The filter output is the stochastic process $Y(t)$.

(a) What is $E[X^2(t)]$?
(b) What is $S_{XY}(f)$?
(c) What is $S_{YX}(f)$?
(d) What is $S_Y(f)$?
(e) What is $E[Y^2(t)]$?

8.6■ A wide sense stationary stochastic process $X(t)$ with autocorrelation function $R_X(\tau) = e^{-4|\tau|}$ is the input to a linear time-invariant filter with impulse response

$$h(t) = \begin{cases} e^{-7t} & t \ge 0, \\ 0 & \text{otherwise.} \end{cases}$$

The filter output is $Y(t)$.

(a) Find the cross spectral density $S_{XY}(f)$.
(b) Find cross-correlation $R_{XY}(\tau)$.

8.7■ A white Gaussian noise process $N(t)$ with power spectral density of 10^{-15} W/Hz is the input to the lowpass filter $H(f) = 10^6 e^{-10^6|f|}$. Find the following properties of the output $Y(t)$:

(a) The expected value μ_Y
(b) The power spectral density $S_Y(f)$
(c) The average power $E[Y^2(t)]$
(d) $P[Y(t) > 0.01]$

8.8♦ In Problem 13.11.3, we found that passing a stationary white noise process through an integrator produced a *non-stationary* output process $Y(t)$. Does this example violate Theorem 2?

8.9♦ Let $M(t)$ be a wide sense stationary random process with average power $E[M^2(t)] = q$ and power spectral density $S_M(f)$. The Hilbert transform of $M(t)$ is $\hat{M}(t)$, a signal obtained by passing $M(t)$ through a linear time-invariant filter with frequency response

$$H(f) = -j\operatorname{sgn}(f) = \begin{cases} -j & f \ge 0, \\ j & f < 0. \end{cases}$$

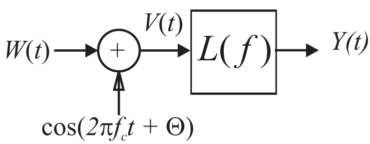
(a) Find the power spectral density $S_{\hat{M}}(f)$ and the average power $\hat{q} = E[\hat{M}^2(t)]$.

(b) In a single sideband communications system, the upper sideband signal is

$$U(t) = M(t)\cos(2\pi f_c t + \Theta) - \hat{M}(t)\sin(2\pi f_c t + \Theta),$$

where Θ has a uniform PDF over $[0, 2\pi)$, independent of $M(t)$ and $\hat{M}(t)$. What is the average power $\mathrm{E}[U^2(t)]$?

8.10♦ As depicted below, a white noise Gaussian process $W(t)$ with power spectral density $S_W(f) = 10^{-15}$ W/Hz is passed through a random phase modulator to produce $V(t)$. The process $V(t)$ is then low-pass filtered ($L(f)$ is an ideal unity gain low-pass filter of bandwidth B) to create $Y(t)$.



The random phase Θ is assumed to be independent of $W(t)$ and to have a uniform distribution over $[0, 2\pi]$.

(a) What is the autocorrelation $R_W(\tau)$?

(b) What is $\mathrm{E}[V(t)]$?

(c) What is the autocorrelation $R_V(\tau)$? Simplify as much as possible.

(d) Is $Y(t)$ a wide sense stationary process?

(e) What is the average power of $Y(t)$?

9.1● $X(t)$ is a wide sense stationary stochastic process with $\mu_X = 0$ and autocorrelation function

$$R_X(\tau) = \sin(2\pi W\tau)/(2\pi W\tau).$$

Observe $Y(t) = X(t) + N(t)$, where $N(t)$ is a wide sense stationary process with power spectral density function $S_N(f) = 10^{-5}$ such that $X(t)$ and $N(t)$ are mutually independent.

(a) What is the transfer function of the optimum linear filter for estimating $X(t)$ given $Y(t)$?

(b) What is the maximum signal bandwidth W Hz that produces a minimum mean square error $e_L^* \leq 0.04$?

9.2■ $X(t)$ is a wide sense stationary stochastic process with $\mu_X = 0$ and autocorrelation function

$$R_X(\tau) = e^{-5000|\tau|}.$$

Observe $Y(t) = X(t) + N(t)$, where $N(t)$ is a wide sense stationary process with power spectral density function $S_N(f) = 10^{-5}$. $X(t)$ and $N(t)$ are mutually independent.

(a) What is the transfer function of the optimum linear filter for estimating $X(t)$ given $Y(t)$?

(b) What is the mean square error of the optimum estimation filter?

10.1● For the digital filter of Problem 2.6, generate 100 sample paths $Y_0, \ldots, Y_{500}$ assuming the X_i are iid Gaussian $(0,1)$ random variables and $Y(0) = 0$. Estimate the expected value and variance of Y_n for $n = 5$, $n = 50$ and $n = 500$.

10.2● In the program `fftc.m`, the vector `n` simply holds the elements $0, 1, \ldots, N-1$. Why is it defined as `n=reshape(0:(N-1),size(R))` rather than just `n=0:N-1`?

10.3● The stationary Gaussian random sequence X_n has expected value $E[X_n] = 0$ and autocorrelation function $R_X[k] = \cos(0.04\pi k)$. Use `gaussvector` to generate and plot 10 sample paths of the form $X_0, \ldots, X_{99}$. What do you appear to observe? Confirm your suspicions by calculating the 100 point DFT on each sample path.

10.4● The LMSE predictor is $\mathbf{a} = \mathbf{R}_{\mathbf{Y}}^{-1}\mathbf{R}_{\mathbf{Y}X}$, However, the next to last line of `lmsepredictor.m` isn't `a=inv(RY)*RYX`. Why?

10.5● X_n is a wide sense stationary random sequence with expected value $\mu_X = 0$ and autocorrelation function $R_X[k] = (-0.9)^{|k|}$. Suppose

```
rx=(-0.9).^(0:5)
```

For what values of N does

```
a=lmsepredictor(rx,N)
```

produce the correct coefficient vector $\hat{\mathbf{a}}$?

10.6● For the discrete-time process X_n in Problem 2.2, calculate an approximation to the power spectral density by finding the DFT of the truncated autocorrelation function

$$\mathbf{r} = \begin{bmatrix} r_{-100} & r_{-99} & \cdots & r_{100} \end{bmatrix}',$$

where $r_k = R_X[k]$. Compare your DFT output against the DTFT $S_X(\phi)$.

10.7■ For the random sequence X_n defined in Problem 4.2, find the filter $\mathbf{h} = \begin{bmatrix} h_0 & \cdots & h_{M-1} \end{bmatrix}'$ of the optimum linear predictor of X_{n+1}, given $X_{n-M+1}, \ldots, X_{n-1}, \ldots, X_n$. What is the mean square error $e_L^*(M)$ of the M-tap optimum linear predictor? Graph $e_L^*(M)$ as a function for M for $M = 1, 2, \ldots, 10$.

10.8■ For a wide sense stationary sequence X_n with zero expected value, extend `lmsepredictor.m` to a function

```
function h = kpredictor(rx,M,k)
```

which produces the filter vector $\mathbf{h}$ of the optimal k step linear predictor of X_{n+k} given the observation

$$\mathbf{Y}_n = \begin{bmatrix} X_{n-M+1} & \cdots & X_{n-1} & X_n \end{bmatrix}'.$$

10.9♦ Continuing Problem 4.7 of the noisy predictor, generate sample paths of X_n and $\hat{X}_n$ for $n = 0, 1, \ldots, 50$ with the following parameters:

(a) $c = 0.9$, $d = 10$

(b) $c = 0.9$, $d = 1$

(c) $c = 0.9$, $d = 0.1$

(d) $c = 0.6$, $d = 10$

(e) $c = 0.6$, $d = 1$

(f) $c = 0.6$, $d = 0.1$

In each experiment, use $\eta = \sigma = 1$. Use the analysis of Problem 4.7 to interpret your results.

Quiz Solutions

Quiz 1 Solution

By Theorem 2,

$$\mu_Y = \mu_X \int_{-\infty}^{\infty} h(t)dt = 2\int_0^{\infty} e^{-t}\,dt = 2. \tag{1}$$

Since $R_X(\tau) = \delta(\tau)$, the autocorrelation function of the output is

$$\begin{aligned} R_Y(\tau) &= \int_{-\infty}^{\infty} h(u) \int_{-\infty}^{\infty} h(v)\delta(\tau + u - v)\,dv\,du \\ &= \int_{-\infty}^{\infty} h(u)h(\tau + u)\,du. \end{aligned} \tag{2}$$

For $\tau > 0$, we have

$$R_Y(\tau) = \int_0^{\infty} e^{-u}e^{-\tau-u}\,du = e^{-\tau}\int_0^{\infty} e^{-2u}\,du = \frac{1}{2}e^{-\tau}. \tag{3}$$

For $\tau < 0$, we can deduce that $R_Y(\tau) = \frac{1}{2}e^{-|\tau|}$ by symmetry. Just to be safe though, we can double check. For $\tau < 0$,

$$R_Y(\tau) = \int_{-\tau}^{\infty} h(u)h(\tau + u)\,du = \int_{-\tau}^{\infty} e^{-u}e^{-\tau-u}\,du = \frac{1}{2}e^{\tau}. \tag{4}$$

Hence,

$$R_Y(\tau) = \frac{1}{2}e^{-|\tau|}. \tag{5}$$

Quiz 2 Solution

The expected value of the output is

$$\mu_Y = \mu_X \sum_{n=-\infty}^{\infty} h_n = 0.5(1 + -1) = 0. \tag{1}$$

The autocorrelation of the output is

$$\begin{aligned} R_Y[n] &= \sum_{i=0}^{1}\sum_{j=0}^{1} h_i h_j R_X[n+i-j] \\ &= 2R_X[n] - R_X[n-1] - R_X[n+1] \\ &= \begin{cases} 1 & n = 0, \\ 0 & \text{otherwise.} \end{cases} \end{aligned} \tag{2}$$

Since $\mu_Y = 0$, The variance of Y_n is $\mathrm{Var}[Y_n] = \mathrm{E}[Y_n^2] = R_Y[0] = 1$.

Quiz 3 Solution

By Theorem 8, $\mathbf{Y} = \begin{bmatrix} Y_{33} & Y_{34} & Y_{35} \end{bmatrix}'$ is a Gaussian random vector since X_n is a Gaussian random process. Moreover, by Theorem 5, each Y_n has expected value $\mathrm{E}[Y_n] = \mu_X \sum_{n=-\infty}^{\infty} h_n = 0$. Thus $\mathrm{E}[\mathbf{Y}] = \mathbf{0}$. Fo find the PDF of the Gaussian vector $\mathbf{Y}$, we need to find the covariance matrix $\mathbf{C_Y}$, which equals the correlation matrix $\mathbf{R_Y}$ since $\mathbf{Y}$ has zero expected value. One way to find the $\mathbf{R_Y}$ is to observe that $\mathbf{R_Y}$ has the Toeplitz structure of Theorem 6 and to use Theorem 5 to find the autocorrelation function

$$R_Y[n] = \sum_{i=-\infty}^{\infty}\sum_{j=-\infty}^{\infty} h_i h_j R_X[n+i-j]. \tag{1}$$

Despite the fact that $R_X[k]$ is an impulse, using Equation (1) is surprisingly tedious because we still need to sum over all i and j such that $n+i-j=0$.

In this problem, it is simpler to observe that $\mathbf{Y} = \mathbf{HX}$ where

$$\mathbf{X} = \begin{bmatrix} X_{30} & X_{31} & X_{32} & X_{33} & X_{34} & X_{35} \end{bmatrix}' \tag{2}$$

and

$$\mathbf{H} = \frac{1}{4}\begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix}. \tag{3}$$

In this case, following Theorem 7, or by directly applying Theorem 8.8 with $\boldsymbol{\mu}_{\mathbf{X}} = \mathbf{0}$ and $\mathbf{A} = \mathbf{H}$, we obtain $\mathbf{R}_{\mathbf{Y}} = \mathbf{H}\mathbf{R}_{\mathbf{X}}\mathbf{H}'$. Since $R_X[n] = \delta_n$, $\mathbf{R}_{\mathbf{X}} = \mathbf{I}$, the identity matrix. Thus

$$\mathbf{C}_{\mathbf{Y}} = \mathbf{R}_{\mathbf{Y}} = \mathbf{H}\mathbf{H}' = \frac{1}{16}\begin{bmatrix} 4 & 3 & 2 \\ 3 & 4 & 3 \\ 2 & 3 & 4 \end{bmatrix}. \tag{4}$$

It follows (very quickly if you use MATLAB for 3×3 matrix inversion) that

$$\mathbf{C}_{\mathbf{Y}}^{-1} = 16\begin{bmatrix} 7/12 & -1/2 & 1/12 \\ -1/2 & 1 & -1/2 \\ 1/12 & -1/2 & 7/12 \end{bmatrix}. \tag{5}$$

Thus, the PDF of $\mathbf{Y}$ is

$$f_{\mathbf{Y}}(\mathbf{y}) = \frac{1}{(2\pi)^{3/2}[\det(\mathbf{C}_{\mathbf{Y}})]^{1/2}} \exp\left(-\frac{1}{2}\mathbf{y}'\mathbf{C}_{\mathbf{Y}}^{-1}\mathbf{y}\right). \tag{6}$$

A disagreeable amount of algebra will show $\det(\mathbf{C}_{\mathbf{Y}}) = 3/1024$ and that the PDF can be "simplified" to

$$f_{\mathbf{Y}}(\mathbf{y}) = \frac{16\exp\left[-8\left(\frac{7}{12}y_{33}^2 + y_{34}^2 + \frac{7}{12}y_{35}^2 - y_{33}y_{34} + \frac{1}{6}y_{33}y_{35} - y_{34}y_{35}\right)\right]}{\sqrt{6\pi^3}}. \tag{7}$$

Equation (7) shows that one of the nicest features of the multivariate Gaussian distribution is that $\mathbf{y}'\mathbf{C}_{\mathbf{Y}}^{-1}\mathbf{y}$ is a very concise representation of the cross-terms in the exponent of $f_{\mathbf{Y}}(\mathbf{y})$.

Quiz 4 Solution

This quiz is solved using Theorem 9 for the case of $k = 1$ and $M = 2$. In this case, $\mathbf{X}_n = \begin{bmatrix} X_{n-1} & X_n \end{bmatrix}'$ and

$$\mathbf{R}_{\mathbf{X}_n} = \mathrm{E}\left[\mathbf{X}_n\mathbf{X}'_n\right] = \begin{bmatrix} R_X[0] & R_X[1] \\ R_X[1] & R_X[0] \end{bmatrix} = \begin{bmatrix} 1.1 & 0.9 \\ 0.9 & 1.1 \end{bmatrix} \tag{1}$$

and

$$\begin{aligned}\mathbf{R}_{X_{n+1}\mathbf{X}_n} &= \mathrm{E}\left[X_{n+1}\begin{bmatrix}X_{n-1} & X_n\end{bmatrix}\right] \\ &= \begin{bmatrix}R_X[2] & R_X[1]\end{bmatrix} = \begin{bmatrix}0.81 & 0.9\end{bmatrix}. \end{aligned} \tag{2}$$

The MMSE linear first order filter for predicting X_{n+1} at time n is the filter $\mathbf{h}'$ such that

$$\begin{aligned}\overleftarrow{\mathbf{h}'} &= \mathbf{R}_{X_{n+1}\mathbf{X}_n}\mathbf{R}_{\mathbf{X}_n}^{-1} \\ &= \begin{bmatrix}0.81 & 0.9\end{bmatrix}\begin{bmatrix}1.1 & 0.9\\ 0.9 & 1.1\end{bmatrix}^{-1} = \begin{bmatrix}\frac{81}{400} & \frac{261}{400}\end{bmatrix}. \end{aligned} \tag{3}$$

It follows that the filter is $\mathbf{h} = \begin{bmatrix}261/400 & 81/400\end{bmatrix}'$ and the MMSE linear predictor is

$$\hat{X}_{n+1} = \frac{81}{400}X_{n-1} + \frac{261}{400}X_n. \tag{4}$$

To find the mean square error, one approach is to follow the method of Example 13 and to directly calculate

$$e_L^* = \mathrm{E}\left[(X_{n+1} - \hat{X}_{n+1})^2\right]. \tag{5}$$

This method is workable for this simple problem but becomes increasingly tedious for higher order filters. Instead, we can derive the mean square error for an arbitary prediction filter $\mathbf{h}$. Since $\hat{X}_{n+1} = \overleftarrow{\mathbf{h}}'\mathbf{X}_n$,

$$\begin{aligned}e_L^* &= \mathrm{E}\left[\left(X_{n+1} - \overleftarrow{\mathbf{h}}'\mathbf{X}_n\right)^2\right] \\ &= \mathrm{E}\left[(X_{n+1} - \overleftarrow{\mathbf{h}}'\mathbf{X}_n)(X_{n+1} - \overleftarrow{\mathbf{h}}'\mathbf{X}_n)'\right] \\ &= \mathrm{E}\left[(X_{n+1} - \overleftarrow{\mathbf{h}}'\mathbf{X}_n)(X_{n+1} - \mathbf{X}_n'\overleftarrow{\mathbf{h}})\right] \end{aligned} \tag{6}$$

After a bit of algebra, we obtain

$$e_L^* = R_X[0] - 2\overleftarrow{\mathbf{h}}'\mathbf{R}_{\mathbf{X}_n X_{n+1}} + \overleftarrow{\mathbf{h}}'\mathbf{R}_{\mathbf{X}_n}\overleftarrow{\mathbf{h}}. \tag{7}$$

With the substitution

$$\overleftarrow{\mathbf{h}} = \mathbf{R}_{\mathbf{X}_n}^{-1} \mathbf{R}_{\mathbf{X}_n X_{n+1}}, \tag{8}$$

we obtain

$$\begin{aligned} e_L^* &= R_X[0] - \mathbf{R}'_{\mathbf{X}_n X_{n+1}} \mathbf{R}_{\mathbf{X}_n}^{-1} \mathbf{R}_{\mathbf{X}_n X_{n+1}} \\ &= R_X[0] - \overleftarrow{\mathbf{h}}' \mathbf{R}_{\mathbf{X}_n X_{n+1}}. \end{aligned} \tag{9}$$

Note that this is essentially the same result as Theorem 12.6 with $\mathbf{Y} = \mathbf{X}_n$, $X = X_{n+1}$ and $\hat{\mathbf{a}}' = \overleftarrow{\mathbf{h}}'$. It is noteworthy that the result is derived in a much simpler way in the proof of Theorem 12.6 by using the orthoginality property of the LMSE estimator.

In any case, the mean square error is

$$\begin{aligned} e_L^* &= R_X[0] - \overleftarrow{\mathbf{h}}' \mathbf{R}_{\mathbf{X}_n X_{n+1}} \\ &= 1.1 - \frac{1}{400} \begin{bmatrix} 81 & 261 \end{bmatrix} \begin{bmatrix} 0.81 \\ 0.9 \end{bmatrix} = \frac{506}{1451} = 0.3487. \end{aligned} \tag{10}$$

Keep in mind that the blind estimate would yield a mean square error of $\text{Var}[X] = 1.1$, we see that observing X_{n-1} and X_n improves the accuracy of our prediction of X_{n+1}.

Quiz 5 Solution

(a) By Theorem 13(b), the average power of $X(t)$ is

$$\text{E}\left[X^2(t)\right] = \int_{-\infty}^{\infty} S_X(f)\, df = \int_{-W}^{W} \frac{5}{W}\, df = 10 \text{ Watts}. \tag{1}$$

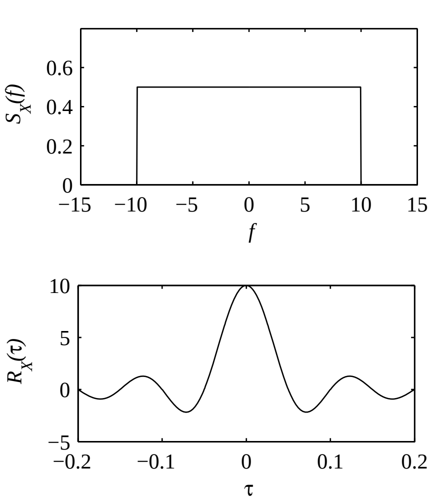
(b) The autocorrelation function is the inverse Fourier transform of $S_X(f)$. Consulting Table 1, we note that

$$S_X(f) = 10 \frac{1}{2W} \text{rect}\left(\frac{f}{2W}\right). \tag{2}$$

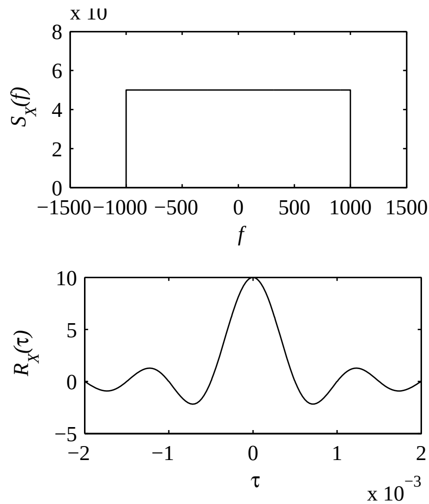
It follows that the inverse transform of $S_X(f)$ is

$$R_X(\tau) = 10\,\mathrm{sinc}(2W\tau) = 10\frac{\sin(2\pi W\tau)}{2\pi W\tau}. \tag{3}$$

(c) For $W = 10$ Hz and $W = 1$ kHZ, here are the graphs of $S_X(f)$ and $R_X(\tau)$:



(a) $W = 10$



(b) $W = 1000$

Quiz 6 Solution

In a sampled system, the discrete time impulse $\delta[n]$ has a flat discrete Fourier transform. That is, if $R_X[n] = 10\delta[n]$, then

$$S_X(\phi) = \sum_{n=-\infty}^{\infty} 10\delta[n]e^{-j2\pi\phi n} = 10. \tag{1}$$

Thus, $R_X[n] = 10\delta[n]$. (This quiz is really lame!)

Quiz 7 Solution

Since $Y(t) = X(t - t_0)$,

$$\begin{aligned} R_{XY}(t,\tau) &= \mathrm{E}\left[X(t)Y(t+\tau)\right] \\ &= \mathrm{E}\left[X(t)X(t+\tau-t_0)\right] \\ &= R_X(\tau - t_0). \end{aligned} \tag{1}$$

We see that $R_{XY}(t,\tau) = R_{XY}(\tau) = R_X(\tau - t_0)$. From Table 1, we recall the property that $g(\tau - \tau_0)$ has Fourier transform $G(f)e^{-j2\pi f\tau_0}$. Thus the Fourier transform of

$$R_{XY}(\tau) = R_X(\tau - t_0) = g(\tau - t_0) \tag{2}$$

is

$$S_{XY}(f) = S_X(f)\, e^{-j2\pi f t_0}. \tag{3}$$

Quiz 8 Solution

We solve this quiz using Theorem 17. First we need some preliminary facts. Let $a_0 = 5{,}000$ so that

$$R_X(\tau) = \frac{1}{a_0} a_0 e^{-a_0|\tau|}. \tag{1}$$

Consulting with the Fourier transforms in Table 1, we see that

$$S_X(f) = \frac{1}{a_0}\frac{2a_0^2}{a_0^2 + (2\pi f)^2} = \frac{2a_0}{a_0^2 + (2\pi f)^2}. \tag{2}$$

The RC filter has impulse response $h(t) = a_1 e^{-a_1 t}u(t)$, where $u(t)$ is the unit step function and $a_1 = 1/RC$ where $RC = 10^{-4}$ is the filter time constant. From Table 1,

$$H(f) = \frac{a_1}{a_1 + j2\pi f}. \tag{3}$$

(a) By Theorem 17,

$$S_{XY}(f) = H(f)S_X(f) = \frac{2a_0a_1}{[a_1 + j2\pi f]\,[a_0^2 + (2\pi f)^2]}. \tag{4}$$

(b) Again by Theorem 17,

$$S_Y(f) = H^*(f)S_{XY}(f) = |H(f)|^2\, S_X(f)\,. \tag{5}$$

Note that

$$\begin{aligned} |H(f)|^2 &= H(f)H^*(f) \\ &= \frac{a_1}{(a_1 + j2\pi f)}\frac{a_1}{(a_1 - j2\pi f)} \\ &= \frac{a_1^2}{a_1^2 + (2\pi f)^2}. \end{aligned} \tag{6}$$

Thus,

$$\begin{aligned} S_Y(f) &= |H(f)|^2\, S_X(f) \\ &= \frac{2a_0a_1^2}{[a_1^2 + (2\pi f)^2]\,[a_0^2 + (2\pi f)^2]}. \end{aligned} \tag{7}$$

(c) To find the average power at the filter output, we can either use basic calculus and calculate $\int_{-\infty}^{\infty} S_Y(f)\,df$ directly or we can find $R_Y(\tau)$ as an inverse transform of $S_Y(f)$. Using partial fractions and the Fourier transform table, the latter method is actually less algebra. In particular, some algebra will show that

$$S_Y(f) = \frac{K_0}{a_0^2 + (2\pi f)^2} + \frac{K_1}{a_1 + (2\pi f)^2} \tag{8}$$

where

$$K_0 = \frac{2a_0a_1^2}{a_1^2 - a_0^2}, \qquad K_1 = \frac{-2a_0a_1^2}{a_1^2 - a_0^2}. \tag{9}$$

Thus,

$$S_Y(f) = \frac{K_0}{2a_0^2}\frac{2a_0^2}{a_0^2 + (2\pi f)^2} + \frac{K_1}{2a_1^2}\frac{2a_1^2}{a_1 + (2\pi f)^2}. \tag{10}$$

Consulting with Table 1, we see that

$$R_Y(\tau) = \frac{K_0}{2a_0^2} a_0 e^{-a_0|\tau|} + \frac{K_1}{2a_1^2} a_1 e^{-a_1|\tau|} \tag{11}$$

Substituting the values of K_0 and K_1, we obtain

$$R_Y(\tau) = \frac{a_1^2 e^{-a_0|\tau|} - a_0 a_1 e^{-a_1|\tau|}}{a_1^2 - a_0^2}. \tag{12}$$

The average power of the $Y(t)$ process is

$$R_Y(0) = \frac{a_1}{a_1 + a_0} = \frac{2}{3}. \tag{13}$$

Note that the input signal has average power $R_X(0) = 1$. Since the RC filter has a 3dB bandwidth of 10,000 rad/sec and the signal $X(t)$ has most of its its signal energy below 5,000 rad/sec, the output signal has almost as much power as the input.

Quiz 9 Solution

This quiz implements an example of Equations (137) and (138) for a system in which we filter $Y(t) = X(t) + N(t)$ to produce an optimal linear estimate of $X(t)$. The solution to this quiz is just to find the filter $\hat{H}(f)$ using Equation (137) and to calculate the mean square error e_L* using Equation (138).

Comment: Since the text omitted the derivations of Equations (137) and (138), we note that Example 13.24 showed that

$$R_Y(\tau) = R_X(\tau) + R_N(\tau), \qquad R_{YX}(\tau) = R_X(\tau). \tag{1}$$

Taking Fourier transforms, it follows that

$$S_Y(f) = S_X(f) + S_N(f), \qquad S_{YX}(f) = S_X(f). \tag{2}$$

Now we can go on to the quiz, at peace with the derivations.

(a) Since $\mu_N = 0$, $R_N(0) = \text{Var}[N] = 1$. This implies

$$R_N(0) = \int_{-\infty}^{\infty} S_N(f)\, df = \int_{-B}^{B} N_0\, df = 2N_0 B. \tag{3}$$

Thus $N_0 = 1/(2B)$. Because the noise process $N(t)$ has constant power $R_N(0) = 1$, decreasing the single-sided bandwidth B increases the power spectral density of the noise over frequencies $|f| < B$.

(b) Since $R_X(\tau) = \text{sinc}(2W\tau)$, where $W = 5{,}000$ Hz, we see from Table 1 that

$$S_X(f) = \frac{1}{10^4} \text{rect}\left(\frac{f}{10^4}\right). \tag{4}$$

The noise power spectral density can be written as

$$S_N(f) = N_0 \,\text{rect}\left(\frac{f}{2B}\right) = \frac{1}{2B} \text{rect}\left(\frac{f}{2B}\right), \tag{5}$$

From Equation (137), the optimal filter is

$$\hat{H}(f) = \frac{S_X(f)}{S_X(f) + S_N(f)} = \frac{\frac{1}{10^4} \text{rect}\left(\frac{f}{10^4}\right)}{\frac{1}{10^4} \text{rect}\left(\frac{f}{10^4}\right) + \frac{1}{2B} \text{rect}\left(\frac{f}{2B}\right)}. \tag{6}$$

(c) We produce the output $\hat{X}(t)$ by passing the noisy signal $Y(t)$ through the filter $\hat{H}(f)$. From Equation (138), the mean square error of the estimate is

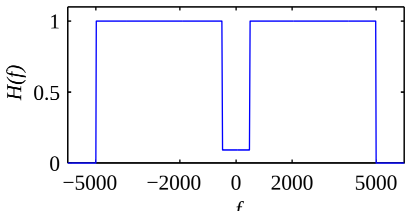
$$\begin{aligned} e_L^* &= \int_{-\infty}^{\infty} \frac{S_X(f)\, S_N(f)}{S_X(f) + S_N(f)}\, df \\ &= \int_{-\infty}^{\infty} \frac{\frac{1}{10^4} \text{rect}\left(\frac{f}{10^4}\right) \frac{1}{2B} \text{rect}\left(\frac{f}{2B}\right)}{\frac{1}{10^4} \text{rect}\left(\frac{f}{10^4}\right) + \frac{1}{2B} \text{rect}\left(\frac{f}{2B}\right)}\, df. \end{aligned} \tag{7}$$

To evaluate the MSE e_L^*, we need to whether $B \le W$. Since the problem asks us to find the largest possible B, let's suppose $B \le W$. We can go back and consider the case $B > W$ later. When $B \le W$, the MSE is

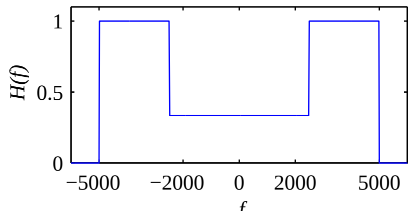
$$e_L^* = \int_{-B}^{B} \frac{\frac{1}{10^4}\frac{1}{2B}}{\frac{1}{10^4} + \frac{1}{2B}}\, df = \frac{\frac{1}{10^4}}{\frac{1}{10^4} + \frac{1}{2B}} = \frac{1}{1 + \frac{5{,}000}{B}} \tag{8}$$

To obtain MSE $e_L^* \le 0.05$ requires $B \le 5{,}000/19 = 263.16$ Hz.

Although this completes the solution to the quiz, what is happening may not be obvious. The noise power is always $\text{Var}[N] = 1$ Watt, for all values of B. As B is decreased, the PSD $S_N(f)$ becomes increasingly tall, but only over a bandwidth B that is decreasing. Thus as B descreases, the filter $\hat{H}(f)$ makes an increasingly deep and narrow notch at frequencies $|f| \le B$. Here are two examples of the filter $\hat{H}(f)$:



$B = 500$



$B = 2500$

As B shrinks, the filter suppresses less of the signal of $X(t)$. The result is that the MSE goes down. Finally, we note that we can choose B very large and also achieve MSE $e_L^* = 0.05$. In particular, when $B > W = 5000$, $S_N(f) = 1/2B$ over frequencies $|f| < W$. In this case, the Wiener filter $\hat{H}(f)$ is an ideal (flat) lowpass filter

$$\hat{H}(f) = \begin{cases} \frac{\frac{1}{10^4}}{\frac{1}{10^4} + \frac{1}{2B}} & |f| < 5{,}000, \\ 0 & \text{otherwise.} \end{cases} \tag{9}$$

Thus increasing B spreads the constant 1 watt of power of $N(t)$ over more bandwidth. The Wiener filter removes the noise that is outside the band of the desired signal. The mean square error is

$$e_L^* = \int_{-5000}^{5000} \frac{\frac{1}{10^4}\frac{1}{2B}}{\frac{1}{10^4} + \frac{1}{2B}}\, df = \frac{\frac{1}{2B}}{\frac{1}{10^4} + \frac{1}{2B}} = \frac{1}{\frac{B}{5000} + 1}. \tag{10}$$

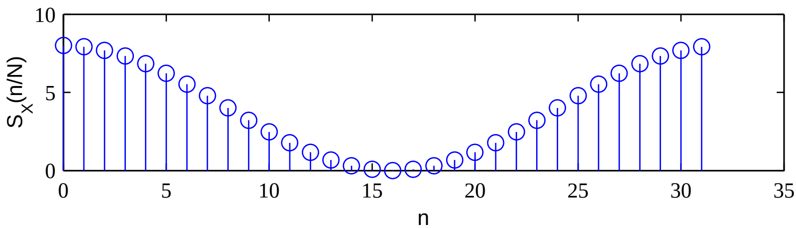
In this case, $B \ge 9.5 \times 10^4$ guarantees $e_L^* \le 0.05$.

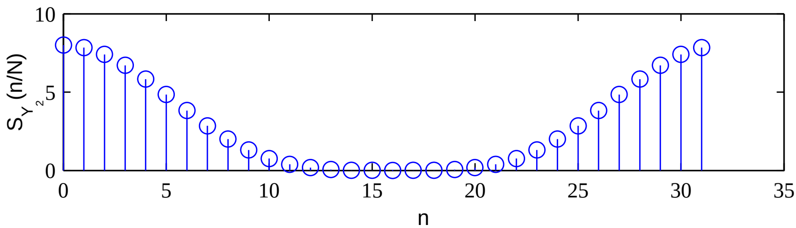
Quiz 10 Solution

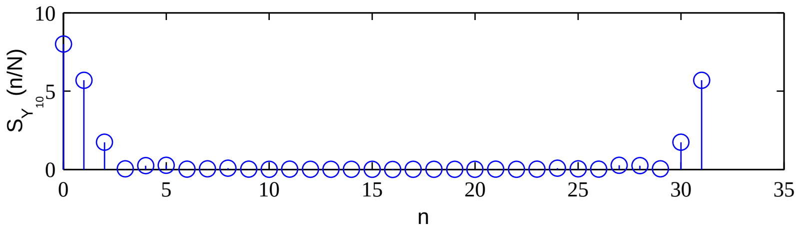
It is fairly straightforward to find $S_X(\phi)$ and $S_Y(\phi)$. The only thing to keep in mind is to use `fftc` to transform the autocorrelation $R_X[f]$ into the power spectral density $S_X(\phi)$. The following MATLAB program

```
%mquiz11.m
N=32;
%autocorrelation and PSD
rx=[2 4 2]; SX=fftc(rx,N);
stem(0:N-1,abs(sx));
xlabel('n');ylabel('S_X(n/N)');
%impulse/filter response: M=2
h2=0.5*[1 1]; H2=fft(h2,N);
SY2=SX.* ((abs(H2)).^2);
figure; stem(0:N-1,abs(SY2)); %PSD of Y for M=2
xlabel('n');ylabel('S_{Y_2}(n/N)');
%impulse/filter response: M=10
h10=0.1*ones(1,10); H10=fft(h10,N);
SY10=sx.*((abs(H10)).^2);
figure; stem(0:N-1,abs(SY10));
xlabel('n');ylabel('S_{Y_{10}}(n/N)');
```

generates and plots graphs of $S_X(\phi)$, $S_Y(n/N)$ for $M = 2$, and $S_\phi(n/N)$ for $M = 10$ using an $N = 32$ point DFT.:







Relative to $M = 2$, when $M = 10$, the filter $H(\phi)$ filters out almost all of the high frequency components of $X(t)$. In the context of Example 26, the low pass moving average filter for $M = 10$ removes the high frquency components and results in a filter output that varies very slowly.

As an aside, note that the vectors `SX`, `SY2` and `SY10` in `mquiz11` should all be real-valued vectors. However, the finite numerical precision of MATLAB results in tiny imaginary parts. Although these imaginary parts have no computational significance, they tend to confuse the `stem` function. Hence, we generate stem plots of the magnitude of each power spectral density.

Problem Solutions

Problem 1.1 Solution

For this problem, it is easiest to work with the expectation operator. The mean function of the output is

$$\mathrm{E}[Y(t)] = 2 + \mathrm{E}[X(t)] = 2. \tag{1}$$

The autocorrelation of the output is

$$\begin{aligned} R_Y(t,\tau) &= \mathrm{E}[(2+X(t))(2+X(t+\tau))] \\ &= \mathrm{E}[4 + 2X(t) + 2X(t+\tau) + X(t)X(t+\tau)] \\ &= 4 + 2\,\mathrm{E}[X(t)] + 2\,\mathrm{E}[X(t+\tau)] + \mathrm{E}[X(t)X(t+\tau)] \\ &= 4 + R_X(\tau). \end{aligned} \tag{2}$$

We see that $R_Y(t,\tau)$ only depends on the time difference τ. Thus $Y(t)$ is wide sense stationary.

Problem 1.3 Solution

By Theorem 2, the mean of the output is

$$\mu_Y = \mu_X \int_{-\infty}^{\infty} h(t)\,dt = 4\int_0^{\infty} e^{-t/a}\,dt = -4ae^{-t/a}\Big|_0^{\infty} = 4a. \tag{1}$$

Since $\mu_Y = 1 = 4a$, we must have $a = 1/4$.

Problem 2.1 Solution

(a) Note that

$$Y_i = \sum_{n=-\infty}^{\infty} h_n X_{i-n} = \frac{1}{3}X_{i+1} + \frac{1}{3}X_i + \frac{1}{3}X_{i-1}. \tag{1}$$

By matching coefficients, we see that

$$h_n = \begin{cases} 1/3 & n = -1, 0, 1, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

(b) By Theorem 5, the output autocorrelation is

$$\begin{aligned} R_Y[n] &= \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} h_i h_j R_X[n+i-j] \\ &= \frac{1}{9} \sum_{i=-1}^{1} \sum_{j=-1}^{1} R_X[n+i-j] \\ &= \frac{1}{9} \Big(R_X[n+2] + 2R_X[n+1] + 3R_X[n] \\ &\qquad + 2R_X[n-1] + R_X[n-2] \Big). \end{aligned} \tag{3}$$

Substituting in $R_X[n]$ yields

$$R_Y[n] = \begin{cases} 1/3 & n = 0, \\ 2/9 & |n| = 1, \\ 1/9 & |n| = 2, \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

Problem 2.3 Solution

(a) By Theorem 5, the expected value of the output is

$$\mu_W = \mu_Y \sum_{n=-\infty}^{\infty} h_n = 2\mu_Y = 2. \tag{1}$$

(b) Theorem 5 also says that the output autocorrelation is

$$\begin{aligned} R_W[n] &= \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} h_i h_j R_Y[n+i-j] \\ &= \sum_{i=0}^{1} \sum_{j=0}^{1} R_Y[n+i-j] \\ &= R_Y[n-1] + 2R_Y[n] + R_Y[n+1]. \end{aligned} \tag{2}$$

For $n = -3$,

$$R_W[-3] = R_Y[-4] + 2R_Y[-3] + R_Y[-2] = R_Y[-2] = 0.5. \tag{3}$$

Following the same procedure, its easy to show that $R_W[n]$ is nonzero for $|n| = 0, 1, 2$. Specifically,

$$R_W[n] = \begin{cases} 0.5 & |n| = 3, \\ 3 & |n| = 2, \\ 7.5 & |n| = 1, \\ 10 & n = 0, \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

(c) The second moment of the output is $\mathrm{E}[W_n^2] = R_W[0] = 10$. The variance of W_n is

$$\mathrm{Var}[W_n] = \mathrm{E}\left[W_n^2\right] - (\mathrm{E}[W_n])^2 = 10 - 2^2 = 6. \tag{5}$$

(d) This part doesn't require any probability. It just checks your knowledge of linear systems and convolution. There is a bit of confusion because h_n is used to denote both the filter that transforms X_n to Y_n as well as the filter that transforms Y_n to W_n. To avoid confusion, we will use $\hat{h}_n$ to denote the filter that transforms X_n to Y_n. Using Equation (22) for discrete-time convolution, we can write

$$W_n = \sum_{j=-\infty}^{\infty} h_j Y_{n-j}, \qquad Y_{n-j} = \sum_{i=-\infty}^{\infty} \hat{h}_i X_{n-j-i}. \tag{6}$$

Combining these equations yields

$$W_n = \sum_{j=-\infty}^{\infty} h_j \sum_{i=-\infty}^{\infty} \hat{h}_i X_{n-j-i}. \tag{7}$$

For each j, we make the substitution $k = i + j$, and then reverse the order of summation to obtain

$$W_n = \sum_{j=-\infty}^{\infty} h_j \sum_{k=-\infty}^{\infty} \hat{h}_{k-j} X_{n-k} = \sum_{k=-\infty}^{\infty} \left(\sum_{j=-\infty}^{\infty} h_j \hat{h}_{k-j} \right) X_{n-k}. \tag{8}$$

Thus we see that

$$g_k = \sum_{j=-\infty}^{\infty} h_j \hat{h}_{k-j}. \tag{9}$$

That is, the filter g_n is the convolution of the filters $\hat{h}_n$ and h_n.
In the context of our particular problem, the filter $\hat{h}_n$ that transforms X_n to Y_n is given in Example 5. The two filters are

$$\hat{\mathbf{h}} = \begin{bmatrix} \hat{h}_0 & \hat{h}_1 \end{bmatrix}' = \begin{bmatrix} 1/2 & 1/2 \end{bmatrix}', \qquad \mathbf{h} = \begin{bmatrix} h_0 & h_1 \end{bmatrix}' = \begin{bmatrix} 1 & 1 \end{bmatrix}'. \tag{10}$$

Keep in mind that $h_n = \hat{h}_n = 0$ for $n < 0$ or $n > 1$. From Equation (9),

$$g_k = \hat{h}_k + \hat{h}_{k-1} = \begin{cases} 1/2 & k = 0, \\ 1 & k = 1, \\ 1/2 & k = 2, \\ 0 & \text{otherwise.} \end{cases} \tag{11}$$

Problem 2.5 Solution

We start with Theorem 5:

$$\begin{aligned} R_Y[n] &= \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} h_i h_j R_X[n+i-j] \\ &= R_X[n-1] + 2R_X[n] + R_X[n+1]. \end{aligned} \tag{1}$$

First we observe that for $n \leq -2$ or $n \geq 2$,

$$R_Y[n] = R_X[n-1] + 2R_X[n] + R_X[n+1] = 0. \tag{2}$$

This suggests that $R_X[n] = 0$ for $|n| > 1$. In addition, we have the following facts:

$$R_Y[0] = R_X[-1] + 2R_X[0] + R_X[1] = 2, \tag{3}$$

$$R_Y[-1] = R_X[-2] + 2R_X[-1] + R_X[0] = 1, \tag{4}$$

$$R_Y[1] = R_X[0] + 2R_X[1] + R_X[2] = 1. \tag{5}$$

A simple solution to this set of equations is $R_X[0] = 1$ and $R_X[n] = 0$ for $n \neq 0$.

Problem 2.7 Solution

There is a technical difficulty with this problem since X_n is not defined for $n < 0$. This implies $C_X[n,k]$ is not defined for $k < -n$ and thus $C_X[n,k]$ cannot be completely independent of k. When n is large, corresponding to a process that has been running for a long time, this is a technical issue, and not a practical concern. Instead, we will find $\bar{\sigma}^2$ such that $C_X[n,k] = C_X[k]$ for all n and k for which the covariance function is defined. To do so, we need to express X_n in terms of $Z_0, Z_1, \ldots, Z_{n_1}$. We do this in the following way:

$$\begin{aligned}
X_n &= cX_{n-1} + Z_{n-1} \\
&= c[cX_{n-2} + Z_{n-2}] + Z_{n-1} \\
&= c^2[cX_{n-3} + Z_{n-3}] + cZ_{n-2} + Z_{n-1} \\
&\vdots \\
&= c^n X_0 + c^{n-1} Z_0 + c^{n-2} Z_2 + \cdots + Z_{n-1} \\
&= c^n X_0 + \sum_{i=0}^{n-1} c^{n-1-i} Z_i.
\end{aligned} \tag{1}$$

Since $\mathrm{E}[Z_i] = 0$, the mean function of the X_n process is

$$\mathrm{E}\left[X_n\right] = c^n \, \mathrm{E}\left[X_0\right] + \sum_{i=0}^{n-1} c^{n-1-i} \, \mathrm{E}\left[Z_i\right] = \mathrm{E}\left[X_0\right]. \tag{2}$$

Thus, for X_n to be a zero mean process, we require that $\mathrm{E}[X_0] = 0$. The autocorrelation function can be written as

$$\begin{aligned} &R_X\left[n,k\right] \\ &\quad = \mathrm{E}\left[X_n X_{n+k}\right] \\ &\quad = \mathrm{E}\left[\left(c^n X_0 + \sum_{i=0}^{n-1} c^{n-1-i} Z_i\right)\left(c^{n+k} X_0 + \sum_{j=0}^{n+k-1} c^{n+k-1-j} Z_j\right)\right]. \end{aligned} \tag{3}$$

Although it was unstated in the problem, we will assume that X_0 is independent of $Z_0, Z_1, \ldots$ so that $\mathrm{E}[X_0 Z_i] = 0$. Since $\mathrm{E}[Z_i] = 0$ and $\mathrm{E}[Z_i Z_j] = 0$ for $i \neq j$, most of the cross terms will drop out. For $k \geq 0$, autocorrelation simplifies to

$$\begin{aligned} R_X\left[n,k\right] &= c^{2n+k} \, \mathrm{Var}[X_0] + \sum_{i=0}^{n-1} c^{2(n-1)+k-2i} \bar{\sigma}^2 \\ &= c^{2n+k} \, \mathrm{Var}[X_0] + \bar{\sigma}^2 c^k \frac{1-c^{2n}}{1-c^2}. \end{aligned} \tag{4}$$

Since $\mathrm{E}[X_n] = 0$, $\mathrm{Var}[X_0] = R_X[n,0] = \sigma^2$ and we can write for $k \geq 0$,

$$R_X\left[n,k\right] = \bar{\sigma}^2 \frac{c^k}{1-c^2} + c^{2n+k}\left(\sigma^2 - \frac{\bar{\sigma}^2}{1-c^2}\right). \tag{5}$$

For $k < 0$, we have

$$R_X[n,k] = \mathrm{E}\left[\left(c^n X_0 + \sum_{i=0}^{n-1} c^{n-1-i} Z_i\right)\left(c^{n+k} X_0 + \sum_{j=0}^{n+k-1} c^{n+k-1-j} Z_j\right)\right]$$

$$= c^{2n+k}\,\mathrm{Var}[X_0] + c^{-k}\sum_{j=0}^{n+k-1} c^{2(n+k-1-j)}\bar{\sigma}^2$$

$$= c^{2n+k}\sigma^2 + \bar{\sigma}^2 c^{-k}\frac{1-c^{2(n+k)}}{1-c^2}$$

$$= \frac{\bar{\sigma}^2}{1-c^2}c^{-k} + c^{2n+k}\left(\sigma^2 - \frac{\bar{\sigma}^2}{1-c^2}\right). \tag{6}$$

We see that $R_X[n,k] = \sigma^2 c^{|k|}$ by choosing

$$\bar{\sigma}^2 = (1-c^2)\sigma^2. \tag{7}$$

Problem 3.1 Solution

Since the process X_n has expected value $\mathrm{E}[X_n] = 0$, we know that

$$C_X(k) = R_X(k) = 2^{-|k|}. \tag{1}$$

Thus $\mathbf{X} = \begin{bmatrix} X_1 & X_2 & X_3 \end{bmatrix}'$ has covariance matrix

$$\mathbf{C_X} = \begin{bmatrix} 2^0 & 2^{-1} & 2^{-2} \\ 2^{-1} & 2^0 & 2^{-1} \\ 2^{-2} & 2^{-1} & 2^0 \end{bmatrix} = \begin{bmatrix} 1 & 1/2 & 1/4 \\ 1/2 & 1 & 1/2 \\ 1/4 & 1/2 & 1 \end{bmatrix}. \tag{2}$$

From Definition 8.12, the PDF of $\mathbf{X}$ is

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}[\det(\mathbf{C_X})]^{1/2}} \exp\left(-\frac{1}{2}\mathbf{x}'\mathbf{C}_{\mathbf{X}}^{-1}\mathbf{x}\right). \tag{3}$$

If we are using MATLAB for calculations, it is best to decalre the problem solved at this point. However, if you like algebra, we can write out the PDF in terms of the variables x_1, x_2 and x_3. To do so we find that the inverse covariance matrix is

$$\mathbf{C}_{\mathbf{X}}^{-1} = \begin{bmatrix} 4/3 & -2/3 & 0 \\ -2/3 & 5/3 & -2/3 \\ 0 & -2/3 & 4/3 \end{bmatrix}. \tag{4}$$

A little bit of algebra will show that $\det(\mathbf{C}_{\mathbf{X}}) = 9/16$ and that

$$\frac{1}{2}\mathbf{x}'\mathbf{C}_{\mathbf{X}}^{-1}\mathbf{x} = \frac{2x_1^2}{3} + \frac{5x_2^2}{6} + \frac{2x_3^2}{3} - \frac{2x_1x_2}{3} - \frac{2x_2x_3}{3}. \tag{5}$$

It follows that

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{4}{3(2\pi)^{3/2}} \exp\left(-\frac{2x_1^2}{3} - \frac{5x_2^2}{6} - \frac{2x_3^2}{3} + \frac{2x_1x_2}{3} + \frac{2x_2x_3}{3}\right). \tag{6}$$

Problem 3.3 Solution

The sequence X_n is passed through the filter

$$\mathbf{h} = \begin{bmatrix} h_0 & h_1 & h_2 \end{bmatrix}' = \begin{bmatrix} 1 & -1 & 1 \end{bmatrix}'. \tag{1}$$

The output sequence is Y_n. Following the approach of Equation (54), we can write the output $\mathbf{Y} = \begin{bmatrix} Y_1 & Y_2 & Y_3 \end{bmatrix}'$ as

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix} = \begin{bmatrix} h_2 & h_1 & h_0 & 0 & 0 \\ 0 & h_2 & h_1 & h_0 & 0 \\ 0 & 0 & h_2 & h_1 & h_0 \end{bmatrix} \begin{bmatrix} X_{-1} \\ X_0 \\ X_1 \\ X_2 \\ X_3 \end{bmatrix} \tag{2}$$

Substituting in the filter values $\mathbf{h} = \begin{bmatrix} h_0 & h_1 & h_2 \end{bmatrix}' = \begin{bmatrix} 1 & -1 & 1 \end{bmatrix}'$, we have

$$\mathbf{Y} = \underbrace{\begin{bmatrix} 1 & -1 & 1 & 0 & 0 \\ 0 & 1 & -1 & 1 & 0 \\ 0 & 0 & 1 & -1 & 1 \end{bmatrix}}_{\mathbf{H}} \underbrace{\begin{bmatrix} X_{-1} \\ X_0 \\ X_1 \\ X_2 \\ X_3 \end{bmatrix}}_{\mathbf{X}}. \tag{3}$$

Since X_n has autocovariance function $C_X(k) = 2^{-|k|}$, $\mathbf{X}$ has covariance matrix

$$\mathbf{C_X} = \begin{bmatrix} 1 & 1/2 & 1/4 & 1/8 & 1/16 \\ 1/2 & 1 & 1/2 & 1/4 & 1/8 \\ 1/4 & 1/2 & 1 & 1/2 & 1/4 \\ 1/8 & 1/4 & 1/2 & 1 & 1/2 \\ 1/16 & 1/8 & 1/4 & 1/2 & 1 \end{bmatrix}. \tag{4}$$

Since $\mathbf{Y} = \mathbf{HX}$,

$$\mathbf{C_Y} = \mathbf{H C_X H'} = \begin{bmatrix} 3/2 & -3/8 & 9/16 \\ -3/8 & 3/2 & -3/8 \\ 9/16 & -3/8 & 3/2 \end{bmatrix}. \tag{5}$$

Some calculation (preferably by MATLAB) will show that $\det(\mathbf{C_Y}) = 675/256$ and that

$$\mathbf{C_Y^{-1}} = \frac{1}{15} \begin{bmatrix} 12 & 2 & -4 \\ 2 & 11 & 2 \\ -4 & 2 & 12 \end{bmatrix}. \tag{6}$$

Some algebra will show that

$$\mathbf{y}' \mathbf{C_Y^{-1}} \mathbf{y} = \frac{12y_1^2 + 11y_2^2 + 12y_3^2 + 4y_1y_2 + -8y_1y_3 + 4y_2y_3}{15}. \tag{7}$$

This implies $\mathbf{Y}$ has PDF

$$f_{\mathbf{Y}}(\mathbf{y}) = \frac{1}{(2\pi)^{3/2}[\det(\mathbf{C}_{\mathbf{Y}})]^{1/2}} \exp\left(-\frac{1}{2}\mathbf{y}'\mathbf{C}_{\mathbf{Y}}^{-1}\mathbf{y}\right)$$
$$= \frac{16\exp\left(-\frac{12y_1^2+11y_2^2+12y_3^2+4y_1y_2+-8y_1y_3+4y_2y_3}{30}\right)}{(2\pi)^{3/2}15\sqrt{3}}. \tag{8}$$

This solution is another demonstration of why the PDF of a Gaussian random vector should be left in vector form.

Comment: We know from Theorem 5 that Y_n is a stationary Gaussian process. As a result, the random variables Y_1, Y_2 and Y_3 are identically distributed and $\mathbf{C}_{\mathbf{Y}}$ is a symmetric Toeplitz matrix. This might make on think that the PDF $f_{\mathbf{Y}}(\mathbf{y})$ should be symmetric in the variables y_1, y_2 and y_3. However, because Y_2 is in the middle of Y_1 and Y_3, the information provided by Y_1 and Y_3 about Y_2 is different than the information Y_1 and Y_2 convey about Y_3. This fact appears as asymmetry in $f_{\mathbf{Y}}(\mathbf{y})$.

Problem 4.1 Solution

(a) Since both X_n and Z_n have zero expected value $\hat{X}_n = aY_n$ for the optimal choice of a. We could derive the optimal a from first principles, or we could use apply our LMSE estimator formula

$$\hat{X}_n = R_{X_nY_n}R_{Y_n}^{-1}Y_n. \tag{1}$$

In this case, all terms are simply scalars. In particular,

$$R_{Y_n} = \text{Var}[Y_n] = \text{Var}[X_n] + \text{Var}[Z_n] = 1 + \frac{1}{2} = \frac{3}{2} \tag{2}$$

and

$$R_{X_nY_n} = \text{E}[X_nY_n] = \text{E}[X_n(X_n+Z_n)] = \text{E}\left[X_n^2\right] = 1. \tag{3}$$

It follows that

$$\hat{X}_n = \frac{2}{3} Y_n. \tag{4}$$

(b) From first principles, the mean squared error is

$$\begin{aligned} e_0 &= \mathrm{E}\left[(\hat{X}_n - X_n)^2\right] \\ &= \mathrm{E}\left[\left(\frac{2}{3}Y_n - X_n\right)^2\right] = \mathrm{E}\left[\left(\frac{2}{3}(X_n + Z_n) - X_n\right)^2\right]. \end{aligned} \tag{5}$$

Simplifying and then expanding the square, we find that

$$\begin{aligned} e_0 &= \mathrm{E}\left[\left(-\frac{1}{3}X_n + \frac{2}{3}Z_n\right)^2\right] \\ &= \mathrm{E}\left[\frac{1}{9}X_n^2 - \frac{4}{9}X_n Z_n + \frac{4}{9}Z_n^2\right] = \frac{1}{9} + \frac{4}{9}\left(\frac{1}{2}\right) = \frac{1}{3}. \end{aligned} \tag{6}$$

(c) Since $\hat{X}_n = (2/3)X_n + (2/3)Z_n$ is the sum of independent Gaussian random variables, we know that $\hat{X}_n$ is Gaussian. Moreover,

$$\mathrm{E}\left[\hat{X}_n\right] = \frac{2}{3}(\mathrm{E}[X_n] + \mathrm{E}[Z_n]) = 0, \tag{7}$$

$$\mathrm{Var}[\hat{X}_n] = \frac{4}{9}\,\mathrm{Var}[X_n] + \frac{4}{9}\,\mathrm{Var}[Z_n] = \frac{2}{3}. \tag{8}$$

Since $\hat{X}_n$ is Gaussian, it has PDF

$$f_{\hat{X}_n}(x) = \frac{1}{\sqrt{2\pi(2/3)}} e^{-x^2/[2(2/3)]} = \sqrt{\frac{3}{4\pi}} e^{-3x^2/4}. \tag{9}$$

(d) Because X_n, Z_n, Y_n and $\hat{X}_n$ are all jointly Gaussian, the LMSE estimate $\hat{X}_n$ equals the conditional expected value $\mathrm{E}[X_n|Y_n]$. That is $\mathrm{E}[X_n|Y_n] = \hat{X}_n = 2Y_n/3$.

Problem 4.3 Solution

(a)

$$\begin{aligned}\alpha = C_X[1] &= \mathrm{Cov}\left[X_n, X_{n+1}\right]\\ &= \rho_{X_n,X_{n+1}} \sqrt{\mathrm{Var}\left[X_n\right] \mathrm{Var}\left[X_{n+1}\right]}\\ &= \rho_{X_n,X_{n+1}} C_X(0) = \rho_{X_n,X_{n+1}}.\end{aligned} \tag{1}$$

Thus $|\alpha| \leq 1$.

(b) Here we can use the linear MMSE estimator formula with the scalar input Y_n:

$$\hat{X}_n = \mathbf{R}_{X_n Y_n} \mathbf{R}_{Y_n}^{-1} Y_n. \tag{2}$$

Since $Y_n = X_n + W_n$,

$$\begin{aligned}\mathbf{R}_{X_n Y_n} &= \mathrm{E}\left[X_n Y_n\right] = \mathrm{E}\left[X_n (X_n + W_n)\right]\\ &= \mathrm{E}\left[X_n^2\right] + \mathrm{E}\left[X_n\right]\mathrm{E}\left[W_n\right] = C_X\left[0\right] = 1\end{aligned} \tag{3}$$

and

$$\begin{aligned}\mathbf{R}_{Y_n} = \mathrm{E}\left[Y_n^2\right] &= \mathrm{E}\left[(X_n + W_n)^2\right]\\ &= \mathrm{E}\left[X_n^2\right] + 2\,\mathrm{E}\left[X_n W_n\right] + \mathrm{E}\left[W_n^2\right]\\ &= C_X[0] + 0 + 1 = 2.\end{aligned} \tag{4}$$

This implies $\hat{X}_n = Y_n/2$.

(c) Here we use the linear MMSE estimator formula with vector input $\mathbf{Y}_n$:

$$\hat{X}_n = \mathbf{R}_{X_n \mathbf{Y}_n} \mathbf{R}_{\mathbf{Y}_n}^{-1} \mathbf{Y}_n. \tag{5}$$

Since $Y_n = X_n + W_n$ and $\mathbf{Y}_n = \begin{bmatrix} Y_n & Y_{n-1} \end{bmatrix}'$,

$$\begin{aligned}
\mathbf{R}_{X_n\mathbf{Y}_n} &= \mathrm{E}\left[X_n \mathbf{Y}'_n\right] \\
&= \mathrm{E}\left[X_n \begin{bmatrix} Y_n & Y_{n-1} \end{bmatrix}\right] \\
&= \begin{bmatrix} \mathrm{E}\left[X_n Y_n\right] & \mathrm{E}\left[X_n Y_{n-1}\right] \end{bmatrix} \\
&= \begin{bmatrix} (\mathrm{E}\left[X_n^2\right] + \mathrm{E}\left[X_n\right]\mathrm{E}\left[W_n\right]) & (\mathrm{E}\left[X_n X_{n-1}\right] + \mathrm{E}\left[X_n W_{n-1}\right]) \end{bmatrix} \\
&= \begin{bmatrix} R_X\left[0\right] & R_X\left[1\right] \end{bmatrix} \\
&= \begin{bmatrix} 1 & \alpha \end{bmatrix}.
\end{aligned} \tag{6}$$

In addition,

$$\begin{aligned}
\mathbf{R}_{\mathbf{Y}_n} &= \mathrm{E}\left[\mathbf{Y}_n \mathbf{Y}'_n\right] \\
&= \mathrm{E}\left[\begin{bmatrix} X_n + W_n \\ X_{n-1} + W_{n-1} \end{bmatrix} \begin{bmatrix} (X_n + W_n) & (X_{n-1} + W_{n-1}) \end{bmatrix}\right] \\
&= \begin{bmatrix} \mathrm{E}\left[X_n^2\right] + \mathrm{E}\left[W_n^2\right] & \mathrm{E}\left[X_n X_{n-1}\right] \\ \mathrm{E}\left[X_{n-1} X_n\right] & \mathrm{E}\left[X_{n-1}^2\right] + \mathrm{E}\left[W_{n-1}^2\right] \end{bmatrix} \\
&= \begin{bmatrix} 2 & \alpha \\ \alpha & 2 \end{bmatrix}.
\end{aligned} \tag{7}$$

Defining the vectors $\mathbf{X}_n = \begin{bmatrix} X_n & X_{n-1} \end{bmatrix}'$ and $\mathbf{W}_n = \begin{bmatrix} W_n & W_{n-1} \end{bmatrix}'$, we can simplify the calculation in (7). Since $\mathbf{Y}_n = \mathbf{X}_n + \mathbf{W}_n$ and $\mathbf{X}_n$ and $\mathbf{W}_n$ are independent and have zero expected value, we have a situation identical to that in Equation (76). Following the steps that follow Equation (76), we obtain

$$\begin{aligned}
\mathbf{R}_{\mathbf{Y}_n} &= \mathbf{R}_{\mathbf{X}_n} + \mathbf{R}_{\mathbf{W}_n} \\
&= \begin{bmatrix} 1 & \alpha \\ \alpha & 1 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & \alpha \\ \alpha & 2 \end{bmatrix}.
\end{aligned} \tag{8}$$

No matter which way you calculate $\mathbf{R}_{\mathbf{Y}_n}$, this implies

$$\begin{aligned}\hat{X}_n &= \mathbf{R}_{X_n\mathbf{Y}_n}\mathbf{R}_{\mathbf{Y}_n}^{-1}\mathbf{Y}_n \\ &= \begin{bmatrix}1 & \alpha\end{bmatrix}\frac{1}{4-\alpha^2}\begin{bmatrix}2 & -\alpha \\ -\alpha & 2\end{bmatrix}\begin{bmatrix}Y_n \\ Y_{n-1}\end{bmatrix} \\ &= \frac{2-\alpha^2}{4-\alpha^2}Y_n + \frac{\alpha}{4-\alpha^2}Y_{n-1}. \qquad (9)\end{aligned}$$

Note that as α approaches zero, the contribution of the previous sample Y_{n-1} goes to zero.

Problem 4.5 Solution

This problem generalizes Example 14 in that -0.9 is replaced by the parameter c and the noise variance 0.2 is replaced by η^2. Because we are only finding the first order filter $\mathbf{h} = \begin{bmatrix}h_0 & h_1\end{bmatrix}'$, it is relatively simple to generalize the solution of Example 14 to the parameter values c and η^2.

Based on the observation $\mathbf{Y} = \begin{bmatrix}Y_{n-1} & Y_n\end{bmatrix}'$, Theorem 11 states that the linear MMSE estimate of $X = X_n$ is $\overleftarrow{\mathbf{h}}'\mathbf{Y}$ where

$$\overleftarrow{\mathbf{h}}' = \mathbf{R}_{X_n\mathbf{Y}}\mathbf{R}_{\mathbf{Y}}^{-1} = \mathbf{R}_{X_n\mathbf{X}_n}(\mathbf{R}_{\mathbf{X}_n} + \mathbf{R}_{\mathbf{W}_n})^{-1}. \qquad (1)$$

From Equation (80),

$$\mathbf{R}_{X_n\mathbf{X}_n} = \begin{bmatrix}R_X[1] & R_X[0]\end{bmatrix} = \begin{bmatrix}c & 1\end{bmatrix}. \qquad (2)$$

From the problem statement,

$$\mathbf{R}_{\mathbf{X}_n} + \mathbf{R}_{\mathbf{W}_n} = \begin{bmatrix}1 & c \\ c & 1\end{bmatrix} + \begin{bmatrix}\eta^2 & 0 \\ 0 & \eta^2\end{bmatrix} = \begin{bmatrix}1+\eta^2 & c \\ c & 1+\eta^2\end{bmatrix}. \qquad (3)$$

This implies

$$
\begin{aligned}
\overleftarrow{\mathbf{h}}' &= \begin{bmatrix} c & 1 \end{bmatrix} \begin{bmatrix} 1+\eta^2 & c \\ c & 1+\eta^2 \end{bmatrix}^{-1} \\
&= \begin{bmatrix} c & 1 \end{bmatrix} \frac{1}{(1+\eta^2)^2 - c^2} \begin{bmatrix} 1+\eta^2 & -c \\ -c & 1+\eta^2 \end{bmatrix} \\
&= \begin{bmatrix} \dfrac{c\eta^2}{(1+\eta^2)^2 - c^2} & \dfrac{1+\eta^2 - c^2}{(1+\eta^2)^2 - c^2} \end{bmatrix}. \qquad (4)
\end{aligned}
$$

To find the mean square error of this predictor, we recall that Theorem 11 is just Theorem 12.6 expressed in the terminology of filters. Expressing part (c) of Theorem 12.6 in terms of the linear estimation filter $\mathbf{h}$, the mean square error of the estimator is

$$
\begin{aligned}
e_L^* &= \operatorname{Var}[X_n] - \overleftarrow{\mathbf{h}}'\mathbf{R}_{\mathbf{Y}_n X_n} \\
&= \operatorname{Var}[X_n] - \overleftarrow{\mathbf{h}}'\mathbf{R}_{\mathbf{X}_n X_n} \\
&= R_X[0] - \overleftarrow{\mathbf{h}}' \begin{bmatrix} c \\ 1 \end{bmatrix} \\
&= 1 - \frac{c^2\eta^2 + \eta^2 + 1 - c^2}{(1+\eta^2)^2 - c^2}. \qquad (5)
\end{aligned}
$$

Note that we always find that $e_L^* < \operatorname{Var}[X_n] = 1$ simply because the optimal estimator cannot be worse than the blind estimator that ignores the observation $\mathbf{Y}_n$.

Problem 4.7 Solution

This problem involves both linear estimation and prediction because we are using Y_{n-1}, the noisy observation of X_{n-1} to estimate the future value X_n. Thus we can't follow the cookbook recipes of Theorem 9 or Theorem 11. Instead we go back to Theorem 12.3 to find the minimum mean square error linear estimator. In that theorem, X_n and Y_{n-1} play the roles of X and Y.

That is, our estimate $\hat{X}_n$ of X_n is

$$\begin{aligned}\hat{X}_n &= \hat{X}_L(Y_{n-1}) \\ &= \rho_{X_n,Y_{n-1}} \left(\frac{\mathrm{Var}[X_n]}{\mathrm{Var}[Y_{n-1}]}\right)^{1/2} (Y_{n-1} - \mathrm{E}\left[Y_{n-1}\right]) + \mathrm{E}\left[X_n\right]. \qquad (1)\end{aligned}$$

By recursive application of $X_n = cX_{n-1} + Z_{n-1}$, we obtain

$$X_n = a^n X_0 + \sum_{j=1}^{n} a^{j-1} Z_{n-j}. \qquad (2)$$

The expected value of X_n is

$$\mathrm{E}\left[X_n\right] = a^n \,\mathrm{E}\left[X_0\right] + \sum_{j=1}^{n} a^{j-1}\,\mathrm{E}\left[Z_{n-j}\right] = 0. \qquad (3)$$

The variance of X_n is

$$\begin{aligned}\mathrm{Var}[X_n] &= a^{2n}\,\mathrm{Var}[X_0] + \sum_{j=1}^{n} [a^{j-1}]^2\,\mathrm{Var}[Z_{n-j}] \\ &= a^{2n}\,\mathrm{Var}[X_0] + \sigma^2 \sum_{j=1}^{n} [a^2]^{j-1}. \qquad (4)\end{aligned}$$

Since $\mathrm{Var}[X_0] = \sigma^2/(1-c^2)$, we obtain

$$\mathrm{Var}[X_n] = \frac{c^{2n}\sigma^2}{1-c^2} + \frac{\sigma^2(1-c^{2n})}{1-c^2} = \frac{\sigma^2}{1-c^2}. \qquad (5)$$

Note that $\mathrm{E}[Y_{n-1}] = d\,\mathrm{E}[X_{n-1}] + \mathrm{E}[W_n] = 0$. The variance of Y_{n-1} is

$$\begin{aligned}\mathrm{Var}[Y_{n-1}] &= d^2\,\mathrm{Var}[X_{n-1}] + \mathrm{Var}[W_n] \\ &= \frac{d^2\sigma^2}{1-c^2} + \eta^2. \qquad (6)\end{aligned}$$

Since X_n and Y_{n-1} have zero mean, the covariance of X_n and Y_{n-1} is

$$\begin{aligned}\mathrm{Cov}\left[X_n, Y_{n-1}\right] &= \mathrm{E}\left[X_n Y_{n-1}\right] \\ &= \mathrm{E}\left[(cX_{n-1} + Z_{n-1})\,(dX_{n-1} + W_{n-1})\right]. \qquad (7)\end{aligned}$$

From the problem statement, we learn that

$$\mathrm{E}[X_{n-1}W_{n-1}] = 0 \qquad \mathrm{E}[X_{n-1}]\,\mathrm{E}[W_{n-1}] = 0,$$
$$\mathrm{E}[Z_{n-1}X_{n-1}] = 0 \qquad \mathrm{E}[Z_{n-1}W_{n-1}] = 0.$$

Hence, the covariance of X_n and Y_{n-1} is

$$\mathrm{Cov}\left[X_n, Y_{n-1}\right] = cd\,\mathrm{Var}[X_{n-1}]. \tag{8}$$

The correlation coefficient of X_n and Y_{n-1} is

$$\rho_{X_n,Y_{n-1}} = \frac{\mathrm{Cov}\left[X_n, Y_{n-1}\right]}{\sqrt{\mathrm{Var}[X_n]\,\mathrm{Var}[Y_{n-1}]}}. \tag{9}$$

Since $\mathrm{E}[Y_{n-1}]$ and $\mathrm{E}[X_n]$ are zero, the linear predictor for X_n becomes

$$\begin{aligned}\hat{X}_n &= \rho_{X_n,Y_{n-1}}\left(\frac{\mathrm{Var}[X_n]}{\mathrm{Var}[Y_{n-1}]}\right)^{1/2} Y_{n-1} \\ &= \frac{\mathrm{Cov}\left[X_n, Y_{n-1}\right]}{\mathrm{Var}[Y_{n-1}]}Y_{n-1} \\ &= \frac{cd\,\mathrm{Var}[X_{n-1}]}{\mathrm{Var}[Y_{n-1}]}Y_{n-1}. \end{aligned} \tag{10}$$

Substituting the above result for $\mathrm{Var}[X_n]$, we obtain the optimal linear predictor of X_n given Y_{n-1}.

$$\hat{X}_n = \frac{c}{d}\frac{1}{1+\beta^2(1-c^2)}Y_{n-1}, \tag{11}$$

where $\beta^2 = \eta^2/(d^2\sigma^2)$. From Theorem 12.3, the mean square estimation error at step n

$$\begin{aligned} e_L^*(n) &= \mathrm{E}\left[(X_n - \hat{X}_n)^2\right] \\ &= \mathrm{Var}[X_n]\left(1-\rho^2_{X_n,Y_{n-1}}\right) = \sigma^2\frac{1+\beta^2}{1+\beta^2(1-c^2)}. \end{aligned} \tag{12}$$

We see that mean square estimation error $e_L^*(n) = e_L^*$, a constant for all n. In addition, e_L^* is an increasing function β.

Problem 5.1 Solution

To use Table 1, we write $R_X(\tau)$ in terms of the autocorrelation

$$\operatorname{sinc}(x) = \frac{\sin(\pi x)}{\pi x}. \tag{1}$$

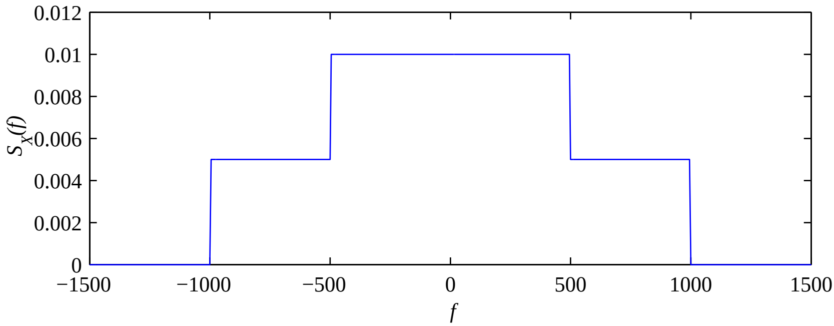
In terms of the sinc(·) function, we obtain

$$R_X(\tau) = 10\operatorname{sinc}(2000\tau) + 5\operatorname{sinc}(1000\tau). \tag{2}$$

From Table 1,

$$S_X(f) = \frac{10}{2{,}000}\operatorname{rect}\left(\frac{f}{2000}\right) + \frac{5}{1{,}000}\operatorname{rect}\left(\frac{f}{1{,}000}\right). \tag{3}$$

Here is a graph of the PSD:



Problem 6.1 Solution

Since the random sequence X_n has autocorrelation function

$$R_X[k] = \delta_k + (0.1)^{|k|}, \tag{1}$$

We can find the PSD directly from Table 3 with $0.1^{|k|}$ corresponding to $a^{|k|}$. The table yields

$$S_X(\phi) = 1 + \frac{1-(0.1)^2}{1+(0.1)^2 - 2(0.1)\cos 2\pi\phi} = \frac{2 - 0.2\cos 2\pi\phi}{1.01 - 0.2\cos 2\pi\phi}. \tag{2}$$

Problem 7.1 Solution

First we show that $S_{YX}(f) = S_{XY}(-f)$. From the definition of the cross spectral density,

$$S_{YX}(f) = \int_{-\infty}^{\infty} R_{YX}(\tau) e^{-j2\pi f\tau}\, d\tau. \tag{1}$$

Making the subsitution $\tau' = -\tau$ yields

$$S_{YX}(f) = \int_{-\infty}^{\infty} R_{YX}(-\tau') e^{j2\pi f\tau'}\, d\tau'. \tag{2}$$

By Theorem 13.14, $R_{YX}(-\tau') = R_{XY}(\tau')$. This implies

$$S_{YX}(f) = \int_{-\infty}^{\infty} R_{XY}(\tau') e^{-j2\pi(-f)\tau'}\, d\tau' = S_{XY}(-f). \tag{3}$$

To complete the problem, we need to show that $S_{XY}(-f) = [S_{XY}(f)]^*$. First we note that since $R_{XY}(\tau)$ is real valued, $[R_{XY}(\tau)]^* = R_{XY}(\tau)$. This implies

$$\begin{aligned} [S_{XY}(f)]^* &= \int_{-\infty}^{\infty} [R_{XY}(\tau)]^* [e^{-j2\pi f\tau}]^*\, d\tau \\ &= \int_{-\infty}^{\infty} R_{XY}(\tau) e^{-j2\pi(-f)\tau}\, d\tau \\ &= S_{XY}(-f). \end{aligned} \tag{4}$$

Problem 8.1 Solution

Let $a = 1/RC$. The solution to this problem parallels Example 22.

(a) From Table 1, we observe that

$$S_X(f) = \frac{2 \cdot 10^4}{(2\pi f)^2 + 10^4}, \qquad H(f) = \frac{1}{a + j2\pi f}. \tag{1}$$

By Theorem 16,

$$S_Y(f) = |H(f)|^2 S_X(f) = \frac{2 \cdot 10^4}{[(2\pi f)^2 + a^2][(2\pi f)^2 + 10^4]}. \tag{2}$$

To find $R_Y(\tau)$, we use a form of partial fractions expansion to write

$$S_Y(f) = \frac{A}{(2\pi f)^2 + a^2} + \frac{B}{(2\pi f)^2 + 10^4}. \tag{3}$$

Note that this method will work only if $a \neq 100$. This same method was also used in Example 22. The values of A and B can be found by

$$A = \left.\frac{2 \cdot 10^4}{(2\pi f)^2 + 10^4}\right|_{f=\frac{ja}{2\pi}} = \frac{-2 \cdot 10^4}{a^2 - 10^4}, \tag{4}$$

$$B = \left.\frac{2 \cdot 10^4}{a^2 + 10^4}\right|_{f=\frac{j100}{2\pi}} = \frac{2 \cdot 10^4}{a^2 - 10^4}. \tag{5}$$

This implies the output power spectral density is

$$S_Y(f) = \frac{-10^4/a}{a^2 - 10^4} \frac{2a}{(2\pi f)^2 + a^2} + \frac{1}{a^2 - 10^4} \frac{200}{(2\pi f)^2 + 10^4}. \tag{6}$$

Since $e^{-c|\tau|}$ and $2c/((2\pi f)^2 + c^2)$ are Fourier transform pairs for any constant $c > 0$, we see that

$$R_Y(\tau) = \frac{-10^4/a}{a^2 - 10^4} e^{-a|\tau|} + \frac{100}{a^2 - 10^4} e^{-100|\tau|}. \tag{7}$$

(b) To find $a = 1/(RC)$, we use the fact that

$$\mathrm{E}\left[Y^2(t)\right] = 100 = R_Y(0) = \frac{-10^4/a}{a^2 - 10^4} + \frac{100}{a^2 - 10^4}. \tag{8}$$

Rearranging, we find that a must satisfy

$$a^3 - (10^4 + 1)a + 100 = 0. \tag{9}$$

This cubic polynomial has three roots:

$$a = 100, \qquad a = -50 + \sqrt{2501}, \qquad a = -50 - \sqrt{2501}. \tag{10}$$

Recall that $a = 100$ is not a valid solution because our expansion of $S_Y(f)$ was not valid for $a = 100$. Also, we require $a > 0$ in order to take the inverse transform of $S_Y(f)$. Thus $a = -50 + \sqrt{2501} \approx 0.01$ and $RC \approx 100$.

Problem 8.3 Solution

Since $S_Y(f) = |H(f)|^2 S_X(f)$, we first find

$$\begin{aligned} |H(f)|^2 &= H(f)H^*(f) \\ &= \left(a_1 e^{-j2\pi f t_1} + a_2 e^{-j2\pi f t_2}\right)\left(a_1 e^{j2\pi f t_1} + a_2 e^{j2\pi f t_2}\right) \\ &= a_1^2 + a_2^2 + a_1 a_2 \left(e^{-j2\pi f(t_2 - t_1)} + e^{-j2\pi f(t_1 - t_2)}\right). \end{aligned} \tag{1}$$

It follows that the output power spectral density is

$$\begin{aligned} S_Y(f) = (a_1^2 + a_2^2) S_X(f) \\ + a_1 a_2 S_X(f) e^{-j2\pi f(t_2 - t_1)} + a_1 a_2 S_X(f) e^{-j2\pi f(t_1 - t_2)}. \end{aligned} \tag{2}$$

Using Table 1, the autocorrelation of the output is

$$\begin{aligned} R_Y(\tau) = (a_1^2 + a_2^2) R_X(\tau) \\ + a_1 a_2 \left(R_X(\tau - (t_1 - t_2)) + R_X(\tau + (t_1 - t_2))\right). \end{aligned} \tag{3}$$

Problem 8.5 Solution

(a) From Theorem 13(b),

$$\mathrm{E}\left[X^2(t)\right] = \int_{-\infty}^{\infty} S_X(f)\, df \int_{-100}^{100} 10^{-4}\, df = 0.02. \tag{1}$$

(b) From Theorem 17

$$S_{XY}(f) = H(f)S_X(f) = \begin{cases} 10^{-4}H(f) & |f| \le 100, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

(c) From Theorem 13.14,

$$R_{YX}(\tau) = R_{XY}(-\tau). \tag{3}$$

From Table 1, if $g(\tau)$ and $G(f)$ are a Fourier transform pair, then $g(-\tau)$ and $G^*(f)$ are a Fourier transform pair. This implies

$$S_{YX}(f) = S_{XY}^*(f) = \begin{cases} 10^{-4}H^*(f) & |f| \le 100, \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

(d) By Theorem 17,

$$\begin{aligned} S_Y(f) &= H^*(f)S_{XY}(f) \\ &= |H(f)|^2 S_X(f) \\ &= \begin{cases} 10^{-4}/[10^4\pi^2 + (2\pi f)^2] & |f| \le 100, \\ 0 & \text{otherwise.} \end{cases} \end{aligned} \tag{5}$$

(e) By Theorem 13,

$$\begin{aligned} \mathrm{E}\left[Y^2(t)\right] &= \int_{-\infty}^{\infty} S_Y(f)\, df \\ &= \int_{-100}^{100} \frac{10^{-4}}{10^4\pi^2 + 4\pi^2 f^2}\, df \\ &= \frac{2}{10^8\pi^2} \int_0^{100} \frac{df}{1 + (f/50)^2}. \end{aligned} \tag{6}$$

By making the substitution, $f = 50\tan\theta$, we have $df = 50\sec^2\theta\,d\theta$. Using the identity $1+\tan^2\theta = \sec^2\theta$, we have

$$\begin{aligned} \mathrm{E}\left[Y^2(t)\right] &= \frac{100}{10^8\pi^2}\int_0^{\tan^{-1}(2)} d\theta \\ &= \frac{\tan^{-1}(2)}{10^6\pi^2} = 1.12\times 10^{-7}. \end{aligned} \tag{7}$$

Problem 8.7 Solution

(a) Since $\mathrm{E}[N(t)] = \mu_N = 0$, the expected value of the output is $\mu_Y = \mu_N H(0) = 0$.

(b) The output power spectral density is

$$S_Y(f) = |H(f)|^2 S_N(f) = 10^{-3} e^{-2\times 10^6 |f|}. \tag{1}$$

(c) The average power is

$$\begin{aligned} \mathrm{E}\left[Y^2(t)\right] = \int_{-\infty}^{\infty} S_Y(f)\,df &= \int_{-\infty}^{\infty} 10^{-3} e^{-2\times 10^6|f|}\,df \\ &= 2\times 10^{-3}\int_0^{\infty} e^{-2\times 10^6 f}\,df \\ &= 10^{-3}. \end{aligned} \tag{2}$$

(d) Since $N(t)$ is a Gaussian process, Theorem 3 says $Y(t)$ is a Gaussian process. Thus the random variable $Y(t)$ is Gaussian with

$$\mathrm{E}[Y(t)] = 0, \qquad \mathrm{Var}[Y(t)] = \mathrm{E}\left[Y^2(t)\right] = 10^{-3}. \tag{3}$$

Thus we can use Table 4.2 to calculate

$$\begin{aligned}
\mathrm{P}\left[Y(t) > 0.01\right] &= \mathrm{P}\left[\frac{Y(t)}{\sqrt{\mathrm{Var}[Y(t)]}} > \frac{0.01}{\sqrt{\mathrm{Var}[Y(t)]}}\right] \\
&= 1 - \Phi\left(\frac{0.01}{\sqrt{0.001}}\right) \\
&= 1 - \Phi(0.32) = 0.3745. \qquad (4)
\end{aligned}$$

Problem 8.9 Solution

(a) Note that $|H(f)| = 1$. This implies $S_{\hat{M}}(f) = S_M(f)$. Thus the average power of $\hat{M}(t)$ is

$$\hat{q} = \int_{-\infty}^{\infty} S_{\hat{M}}(f)\, df = \int_{-\infty}^{\infty} S_M(f)\, df = q. \qquad (1)$$

(b) The average power of the upper sideband signal is

$$\begin{aligned}
\mathrm{E}\left[U^2(t)\right] &= \mathrm{E}\left[M^2(t)\cos^2(2\pi f_c t + \Theta)\right] \\
&\quad - \mathrm{E}\left[2M(t)\hat{M}(t)\cos(2\pi f_c t + \Theta)\sin(2\pi f_c t + \Theta)\right] \\
&\quad + \mathrm{E}\left[\hat{M}^2(t)\sin^2(2\pi f_c t + \Theta)\right]. \qquad (2)
\end{aligned}$$

To find the expected value of the random phase cosine, for an integer $n \neq 0$, we evaluate

$$\begin{aligned}
\mathrm{E}\left[\cos(2\pi f_c t + n\Theta)\right] &= \int_0^{2\pi} \cos(2\pi f_c t + n\theta)\frac{1}{2\pi}\, d\theta \\
&= \frac{1}{2n\pi}\sin(2\pi f_c t + n\theta)\Big|_0^{2\pi} \\
&= \frac{1}{2n\pi}\left(\sin(2\pi f_c t + 2n\pi) - \sin(2\pi f_c t)\right) = 0. \qquad (3)
\end{aligned}$$

Similar steps will show that for any integer $n \neq 0$, the random phase sine also has expected value

$$\mathrm{E}\left[\sin(2\pi f_c t + n\Theta)\right] = 0. \tag{4}$$

Using the trigonometric identity $\cos^2\phi = (1+\cos 2\phi)/2$, we can show

$$\mathrm{E}\left[\cos^2(2\pi f_c t + \Theta)\right] = \mathrm{E}\left[\frac{1}{2}\left(1+\cos(2\pi(2f_c)t + 2\Theta)\right)\right] = 1/2. \tag{5}$$

Similarly,

$$\mathrm{E}\left[\sin^2(2\pi f_c t + \Theta)\right] = \mathrm{E}\left[\frac{1}{2}\left(1-\cos(2\pi(2f_c)t + 2\Theta)\right)\right] = 1/2. \tag{6}$$

In addition, the identity $2\sin\phi\cos\phi = \sin 2\phi$ implies

$$\mathrm{E}\left[2\sin(2\pi f_c t + \Theta)\cos(2\pi f_c t + \Theta)\right] = \mathrm{E}\left[\cos(4\pi f_c t + 2\Theta)\right] = 0. \tag{7}$$

Since $M(t)$ and $\hat{M}(t)$ are independent of Θ, the average power of the upper sideband signal is

$$\begin{aligned}
\mathrm{E}\left[U^2(t)\right] &= \mathrm{E}\left[M^2(t)\right]\mathrm{E}\left[\cos^2(2\pi f_c t + \Theta)\right] \\
&\qquad + \mathrm{E}\left[\hat{M}^2(t)\right]\mathrm{E}\left[\sin^2(2\pi f_c t + \Theta)\right] \\
&\qquad\qquad - \mathrm{E}\left[M(t)\hat{M}(t)\right]\mathrm{E}\left[2\cos(2\pi f_c t + \Theta)\sin(2\pi f_c t + \Theta)\right] \\
&= q/2 + q/2 + 0 = q.
\end{aligned} \tag{8}$$

Problem 9.1 Solution

The system described in this problem corresponds exactly to the system in the text that yielded Equation (137).

(a) From Equation (137), the optimal linear filter is

$$\hat{H}(f) = \frac{S_X(f)}{S_X(f) + S_N(f)}. \tag{1}$$

In this problem, $R_X(\tau) = \mathrm{sinc}(2W\tau)$ so that

$$S_X(f) = \frac{1}{2W}\,\mathrm{rect}\left(\frac{f}{2W}\right). \tag{2}$$

It follows that the optimal filter is

$$\hat{H}(f) = \frac{\frac{1}{2W}\,\mathrm{rect}\left(\frac{f}{2W}\right)}{\frac{1}{2W}\,\mathrm{rect}\left(\frac{f}{2W}\right) + 10^{-5}} = \frac{10^5}{10^5 + 2W}\,\mathrm{rect}\left(\frac{f}{2W}\right). \tag{3}$$

(b) From Equation (138), the minimum mean square error is

$$\begin{aligned} e_L^* = \int_{-\infty}^{\infty} \frac{S_X(f)\,S_N(f)}{S_X(f) + S_N(f)}\,df &= \int_{-\infty}^{\infty} \hat{H}(f) S_N(f)\,df \\ &= \frac{10^5}{10^5 + 2W} \int_{-W}^{W} 10^{-5}\,df \\ &= \frac{2W}{10^5 + 2W}. \end{aligned} \tag{4}$$

It follows that the mean square error satisfies $e_L^* \leq 0.04$ if and only if $W \leq 2{,}083.3$ Hz. What is occurring in this problem is the optimal filter is simply an ideal lowpass filter of bandwidth W. Increasing W increases the bandwidth of the signal and the bandwidth of the filter $\hat{H}(f)$. This allows more noise to pass through the filter and decreases the quality of our estimator.

Problem 10.1 Solution

Although it is straightforward to calculate sample paths of Y_n using the filter response $Y_n = \frac{1}{2}Y_{n-1} + \frac{1}{2}X_n$ directly, the necessary loops makes for a slow program. A solution using vectors and matrices tends to run faster. From the filter response, we can write

$$Y_1 = \frac{1}{2}X_1, \tag{1}$$
$$Y_2 = \frac{1}{4}X_1 + \frac{1}{2}X_2, \tag{2}$$
$$Y_3 = \frac{1}{8}X_1 + \frac{1}{4}X_2 + \frac{1}{2}X_3, \tag{3}$$
$$\vdots$$
$$Y_n = \frac{1}{2^n}X_1 + \frac{1}{2^{n-1}}X_2 + \cdots + \frac{1}{2}X_n. \tag{4}$$

In vector notation, these equations become

$$\underbrace{\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}}_{\mathbf{Y}} = \underbrace{\begin{bmatrix} 1/2 & 0 & \cdots & 0 \\ 1/4 & 1/2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 1/2^n & \cdots & 1/4 & 1/2 \end{bmatrix}}_{\mathbf{H}} \underbrace{\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix}}_{\mathbf{X}}. \tag{5}$$

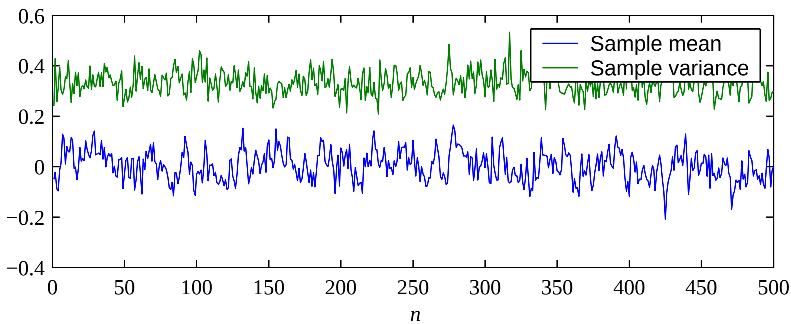
When $\mathbf{X}$ is a column of iid Gaussian $(0,1)$ random variables, the column vector $\mathbf{Y} = \mathbf{HX}$ is a single sample path of $Y_1, \ldots, Y_n$. When $\mathbf{X}$ is an $n \times m$ matrix of iid Gaussian $(0,1)$ random variables, each column of $\mathbf{Y} = \mathbf{HX}$ is a sample path of $Y_1, \ldots, Y_n$. In this case, let matrix entry $Y_{i,j}$ denote a sample Y_i of the jth sample path. The samples $Y_{i,1}, Y_{i,2}, \ldots, Y_{i,m}$ are iid samples of Y_i. We can estimate the mean and variance of Y_i using the sample mean $M_n(Y_i)$ and sample variance $V_m(Y_i)$ of Section 10.4. These estimates are

$$M_n(Y_i) = \frac{1}{m}\sum_{j=1}^{m} Y_{i,j}, \qquad V(Y_i) = \frac{1}{m-1}\sum_{j=1}^{m} \left(Y_{i,j} - M_n(Y_i)\right)^2. \tag{6}$$

This is the approach of the following program.

```
function ymv=yfilter(m);
%ymv(i) is the mean and var (over m paths) of y(i),
%the filter output of 11.2.6 and 11.10.1
X=randn(500,m);
H=toeplitz([(0.5).^(1:500)],[0.5 zeros(1,499)]);
Y=H*X;
yav=sum(Y,2)/m;
yavmat=yav*ones(1,m);
yvar=sum((Y-yavmat).^2,2)/(m-1);
ymv=[yav yvar];
```

The commands `ymv=yfilter(100);plot(ymv)` will generate a plot similar to this:



We see that each sample mean is small, on the other order of 0.1. Note that $\mathrm{E}[Y_i] = 0$. For $m = 100$ samples, the sample mean has variance $1/m = 0.01$ and standard deviation 0.1. Thus it is to be expected that we observe sample mean values around 0.1.

Also, it can be shown (in the solution to Problem 2.6 for example) that as i becomes large, $\mathrm{Var}[Y_i]$ converges to 1/3. Thus our sample variance results are also not surprising.

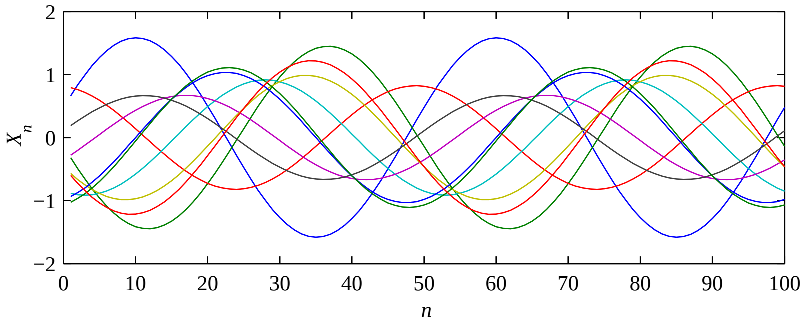
Comment: Although within each sample path, Y_i and Y_{i+1} are quite correlated, the sample means of Y_i and Y_{i+1} are not very correlated when a large number of sample paths are averaged. Exact calculation of the covariance of the sample means of Y_i and Y_{i+1} might be an interesting exercise. The same observations apply to the sample variance as well.

Problem 10.3 Solution

The program `cospaths.m` generates Gaussian sample paths with the desired autocorrelation function $R_X(k) = \cos(0.04 * pi * k)$. Here is the code:

```
function x=cospaths(n,m);
%Generate m  sample paths of length n of a
%Gaussian process with ACF R[k]=cos(0.04*pi*k)
k=0:n-1;
rx=cos(0.04*pi*k)';
x=gaussvector(0,rx,m);
```

The program is simple because if the second input parameter to `gaussvector` is a length m vector `rx`, then `rx` is assumed to be the first row of a symmetric Toeplitz covariance matrix. The commands `x=cospaths(100,10);plot(x)` will produce a graph like this one:

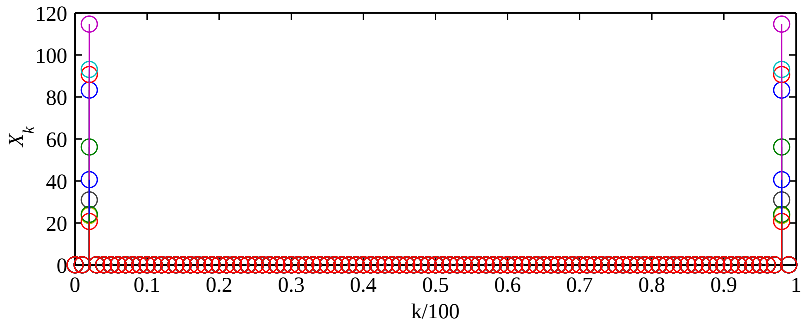


We note that every sample path of the process is Gaussian random sequence. However, it would also appear from the graph that every sample path is a perfect sinusoid. this may seem strange if you are used to seeing Gaussian processes simply as noisy processes or fluctating Brownian motion processes. However, in this case, the amplitude and phase of each sample path is random such that over the ensemble of sinusoidal sample functions, each sample X_n is a Gaussian $(0, 1)$ random variable.

Finally, to confirm that that each sample path is a perfect sinusoid, rather than just resembling a sinusoid, we calculate the DFT of each sample path. The commands

```
>> x=cospaths(100,10);
>> X=fft(x);
>> stem((0:99)/100,abs(X));
```

will produce a plot similar to this:



The above plot consists of ten overlaid 100-point DFT magnitude stem plots, one for each Gaussian sample function. Each plot has exactly two nonzero components at frequencies $k/100 = 0.02$ and $(100 - k)/100 = 0.98$ corresponding to each sample path sinusoid having frequency 0.02. Note that the magnitude of each 0.02 frequency component depends on the magnitude of the corresponding sinusoidal sample path.

Problem 10.5 Solution

The function `lmsepredictor.m` is designed so that if the sequence X_n has a finite duration autocorrelation function such that $R_X[k] = 0$ for $|k| \geq m$, but the LMSE filter of order $M - 1$ for $M \geq m$ is supposed to be returned, then the `lmsepredictor` automatically pads the autocorrelation vector `rx` with a sufficient number of zeros so that the output is the order $M - 1$ filter. Conversely, if `rx` specifies more values $R_X[k]$ than are needed, then the operation `rx(1:M)` extracts the M values $R_X[0], \ldots, R_X[M-1]$ that are needed.

However, in this problem $R_X[k] = (-0.9)^{|k|}$ has infinite duration. When we call `lmsepredictor(rx,N)` for $N \geq 6$ and pass the truncated representation `rx` of length 6 as an argument, the result is that `rx` is incorrectly padded with zeros. The resulting filter output will be the LMSE filter for the filter response

$$R_X[k] = \begin{cases} (-0.9)^{|k|} & |k| \le 5, \\ 0 & \text{otherwise}, \end{cases} \tag{1}$$

rather than the LMSE filter for the true autocorrelation function. Put another way, we obtain the correct output for $N \le 5$.

Problem 10.7 Solution

In this problem, we generalize the solution of Problem 4.2 using Theorem 9 with $k = 1$ for filter order $M > 2$. The optimum linear predictor filter

$$\mathbf{h} = \begin{bmatrix} h_0 & h_1 & \cdots & h_{M-1} \end{bmatrix}' \tag{1}$$

of X_{n+1} given

$$\mathbf{X}_n = \begin{bmatrix} X_{n-(M-1)} & \cdots & X_{n-1} & X_n \end{bmatrix}' \tag{2}$$

is given by

$$\overleftarrow{\mathbf{h}} = \begin{bmatrix} h_{M-1} \\ \vdots \\ h_0 \end{bmatrix} = \mathbf{R}_{\mathbf{X}_n}^{-1} \mathbf{R}_{\mathbf{X}_n X_{n+1}}, \tag{3}$$

where $\mathbf{R}_{\mathbf{X}_n}$ is given by Theorem 6 and $\mathbf{R}_{\mathbf{X}_n X_{n+1}}$ is given by Equation (65). In this problem,

$$\mathbf{R}_{\mathbf{X}_n X_{n+1}} = \mathrm{E}\left[\begin{bmatrix} X_{n-M+1} \\ \vdots \\ X_n \end{bmatrix} X_{n+1} \right] = \begin{bmatrix} R_X[M] \\ \vdots \\ R_X[1] \end{bmatrix}. \tag{4}$$

Going back to Theorem 12.7, the mean square error is

$$e_L^* = \mathrm{E}\left[(X_{n+1} - \overleftarrow{h}'\mathbf{X}_n)^2 \right] = \mathrm{Var}[X_{n+1}] - \overleftarrow{\mathbf{h}}'\mathbf{R}_{\mathbf{X}_n X_{n+1}}. \tag{5}$$

For $M > 2$, we use the MATLAB function `onesteppredictor(r,M)` to perform the calculations.

```
function [h,e]=onesteppredictor(r,M);
%usage: h=onesteppredictor(r,M);
%input: r=[R_X(0) R_X(1) .. R_X(m-1)]
%assumes R_X(n)==0 for n >=m
%output=vector h for lmse predictor
% xx=h'[X(n),X(n-1),..,X(n-M+1)] for X(n+1)
m=length(r);
r=[r(:);zeros(M-m+1,1)];%append zeros if needed
RY=toeplitz(r(1:M));
RYX=r(M+1:-1:2);
h=flipud(RY\RYX);
e=r(1)-(flipud(h))'*RYX;
```

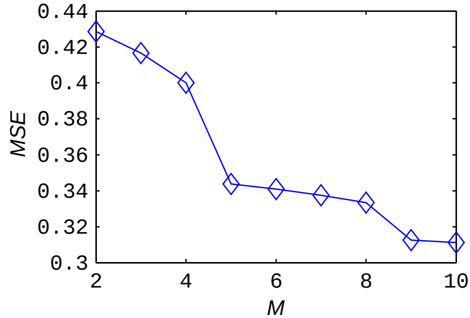
The code is pretty straightforward. Here are two examples just to show it works.

```
>> [h2,e2]=onesteppredictor(r,2)
h2 =
    0.8571
   -0.1429
e2 =
    0.4286

>> [h4,e4]=onesteppredictor(r,4)
h4 =
    0.8000
    0.0000
   -0.0000
   -0.2000
e4 =
    0.4000
```

The problem also requested that we calculate the mean square error as a function of the filter order M. Here is a script and the resulting plot of the MSE.

```
%onestepmse.m
r=1-0.25*(0:3);
ee=[ ];
for M=2:10,
  [h,e]=onesteppredictor(r,M);
  ee=[ee,e];
end
plot(2:10,ee,'-d');
xlabel('\itM');
ylabel('\it MSE');
```



Problem 10.9 Solution

To generate the sample paths X_n and Y_n is relatively straightforward. For $\sigma = \eta = 1$, the solution to Problem 4.7 showed that the optimal linear predictor $\hat{X}_n(Y_{n-1})$ of X_n given Y_{n-1} is

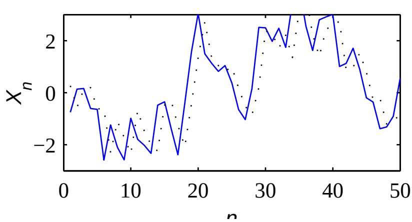
$$\hat{X}_n = \frac{cd}{d^2 + (1 - c^2)} Y_{n-1}. \tag{1}$$

In the following, we plot sample paths of X_n and $\hat{X}_n$. To show how sample paths are similar for different values of c and d, we construct all samples paths using the same sequence of noise samples W_n and Z_n. In addition, given c, we choose $X_0 = X_{00}/\sqrt{1-c^2}$ where X_{00} is a Gaussian $(0,1)$ random variable that is the same for all sample paths. To do this, we write a function `xpathplot(c,d,x00,z,w)` which constructs a sample path given the parameters c and d, and the noise vectors `z` and `w`. A simple script `predictpaths.m` calls `xpathplot` to generate the requested sample paths. Here are the two programs:

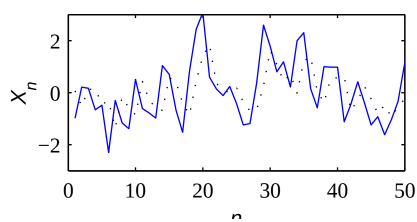
```
function [x,xhat]=xpaths(c,d,x00,z,w)
n=length(z);
n0=(0:(n-1))'; n1=(1:n)';
vx=1/(1-c^2);
x0=sqrt(vx)*x00;
x=(c.^(n1)*x0)+...
     toeplitz(c.^(n0),eye(1,n))*z;
y=d*[x0;x(1:n-1)]+w;
vy=((d^2)*vx) + 1;
xhat=((c*d*vx)/vy)*y;
plot(n1,x,'b-',n1,xhat,'k:');
axis([0 50 -3 3]);
xlabel('\itn'); ylabel('\itX_n');
```

```
%predictpaths.m
w=randn(50,1);z=randn(50,1);
x00=randn(1);
xpaths(0.9,10,x00,z,w);
pause;
xpaths(0.9,1,x00,z,w);
pause;
xpaths(0.9,0.1,x00,z,w);
pause;
xpaths(0.6,10,x00,z,w);
pause;
xpaths(0.6,1,x00,z,w);
pause;
xpaths(0.6,0.1,x00,z,w);
```

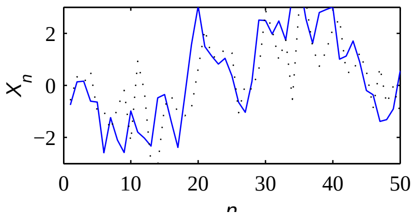
Some sample paths for X_n and $\hat{X}_n$ for the requested parameters are shown below. In each pair, the one-step prediction $\hat{X}_n$ is marked by dots.



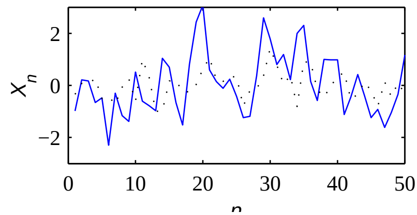
(a) $c = 0.9$, $d = 10$



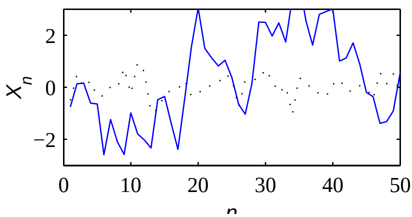
(d) $c = 0.6$, $d = 10$



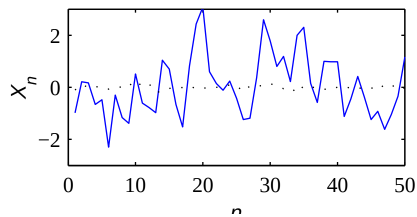
(b) $c = 0.9$, $d = 1$



(e) $c = 0.6$, $d = 1$



(c) $c = 0.9$, $d = 0.1$



(f) $c = 0.6$, $d = 0.1$

The mean square estimation error at step n was found to be

$$e_L^*(n) = e_L^* = \sigma^2 \frac{d^2 + 1}{d^2 + (1 - c^2)}. \tag{2}$$

We see that the mean square estimation error is $e_L^*(n) = e_L^*$, a constant for all n. In addition, e_L^* is a decreasing function of d. In graphs (a) through (c), we see that the predictor tracks X_n less well as d decreases because decreasing d corresponds to decreasing the contribution of X_{n-1} to the measurement Y_{n-1}. Effectively, the impact of the measurement noise is increased. As d decreases, the predictor places less emphasis on the measurement Y_n and instead makes predictions closer to $\mathrm{E}[X] = 0$. That is, when d is small in graphs (c) and (f), the predictor stays close to zero. With respect to c, the performance of the predictor is less easy to understand. In Equation (12), the mean square error e_L^* is the product of

$$\mathrm{Var}[X_n] = \frac{\sigma^2}{1 - c^2} \tag{3}$$

and

$$1 - \rho^2_{X_n, Y_{n-1}} = \frac{(d^2 + 1)(1 - c^2)}{d^2 + (1 - c^2)}. \tag{4}$$

As a function of increasing c^2, $\text{Var}[X_n]$ increases while $1-\rho^2_{X_n,Y_{n-1}}$ decreases. Overall, the mean square error e^*_L is an increasing function of c^2. However, $\text{Var}[X]$ is the mean square error obtained using a blind estimator that always predicts $\text{E}[X]$ while $1-\rho^2_{X_n,Y_{n-1}}$ characterizes the extent to which the optimal linear predictor is better than the blind predictor. When we compare graphs (a)-(c) with $c = 0.9$ to graphs (d)-(f) with $c = 0.6$, we see greater variation in X_n for larger a but in both cases, the predictor worked well when d was large.