Chapter 16

Basic Random Processes

16.1 Introduction

So far we have studied the probabilistic description of a *finite* number of random variables. This is useful for random phenomena that have definite beginning and end times. Many physical phenomena, however, are more appropriately modeled as ongoing in time. Such is the case for the annual summer rainfall in Rhode Island as shown in Figure 1.1 and repeated for convenience in Figure 16.1. This physical

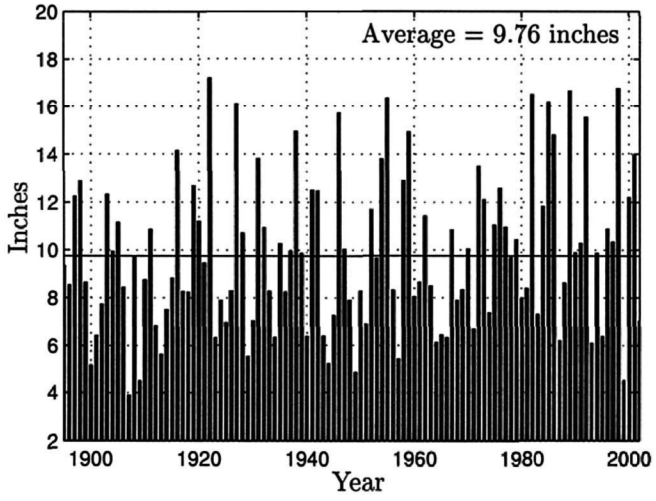


Figure 16.1: Annual summer rainfall in Rhode Island from 1895 to 2002.

process has been ongoing for all time and will undoubtedly continue into the future. It is only our limited ability to measure the rainfall over several lifetimes that has produced the data shown in Figure 16.1. It therefore seems more reasonable to attempt to study the probabilistic characteristics of the annual summer rainfall in

Rhode Island for *all time*. To do so let $X[n]$ be a random variable that denotes the annual summer rainfall for year n. Then, we will be interested in the behavior of the *infinite* tuple of random variables $(\ldots, X[-1], X[0], X[1], \ldots)$, where the corresponding year for $n = 0$ can be chosen for convenience (maybe according to the Christian or Hebrew calendars, as examples). Note that we cannot employ our previous probabilistic methods directly since the number of random variables is not finite or N-dimensional.

Given our interest in the annual summer rainfall, what types of questions are pertinent? A meterologist might wish to determine if the rainfall totals are increasing with time. Hence, he may question if the *average* rainfall is really constant. If it is not constant with time, then our estimate of the average, obtained by taking the sample mean of the values shown in Figure 16.1, is meaningless. As an example, we would also have obtained an average of 9.76 inches if the rainfall totals were increasing linearly with time, starting at 7.76 inches and ending at 11.76 inches. The meterologist might argue that due to global warming the rainfall totals should be increasing. We will return to this question in Section 16.8. Another question might be to assess the probability that the following year the rainfall will be 12 inches or more if we know the entire past history of rainfall totals. This is the problem of prediction, which is a fundamental problem in many scientific disciplines.

A second example of a random process, which is of intense interest, is a man-made one: the Dow-Jones industrial average (DJIA) for stocks. At the end of each trading day the average of the prices of a representative group of stocks is computed to give an indication of the health of the U.S. stock market. Its usefulness is that this value also gives an indication of the overall health of the U.S. economy. Some recent weekly values are shown in Figure 16.2. The overall trend beginning at week 10 is upward until about week 60, at which point it fluctuates up and down. Some questions of interest are whether the index will go back up again after week 92 and to what degree is it possible to predict the movement of the stock market, of which the DJIA is an indicator. The financial industry and in fact the health of the U.S. economy depends in a large degree upon the answers to these questions! In the remaining chapters we will describe the theory and application of random processes. As always, the theory will serve as a foundation upon which we will be able to analyze random processes. In any practical situation, however, the ideal theoretical analysis must be tempered with the constraints and additional complexities of the real world.

16.2 Summary

A random process is defined in Section 16.3. Four different types of random processes are described in Section 16.4. They are classified according to whether they are defined for all time or only for uniformly spaced time samples, and also according to their possible values as being discrete or continuous. Figure 16.5 illustrates the various types. A stationary random process is one for which its probabilistic

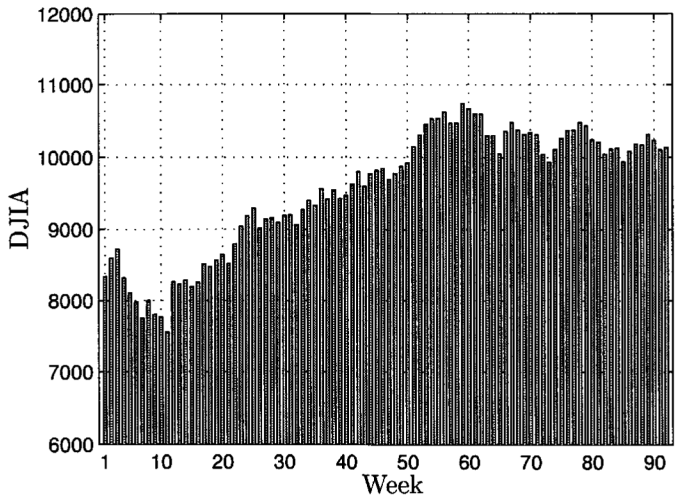


Figure 16.2: Dow-Jones industrial average at the end of each week from January 8, 2003 to September 29, 2004 [DowJones.com 2004].

description does not change with the chosen time origin, which is expressed mathematically by (16.3). An IID random process is stationary as shown in Example 16.3. The concept of a random process having stationary and independent increments is described in Section 16.5 with an illustration given in Example 16.5. Some more examples of random processes are given in Section 16.6. The most useful moments of a random process, the mean sequence and the covariance sequence, are defined by (16.5) and (16.7), respectively. Finally, in Section 16.8 an application of the estimation of the mean sequence to predicting average rainfall totals is described. The least squares estimator of the slope and intercept of a straight line is found using (16.9) and is commonly used in data analysis problems.

16.3 What Is a Random Process?

To define the concept of a random process we will begin by considering our usual example of a coin tossing experiment. Assume that at some start time we toss a coin and then repeat this subexperiment at one second intervals for all time. Letting n denote the time in seconds, we therefore generate successive outcomes at times $n = 0, 1, \ldots$. The experiment continues indefinitely. Since there are two possible outcomes for each coin toss and we will assume that the tosses are independent, we have an infinite sequence of Bernoulli trials. This is termed a *Bernoulli random process* and extends the finite Bernoulli set of random variables first introduced in Section 4.6.2, in which a finite number of trials were carried out. As usual, we let the probability of a head ($X = 1$) be p and the prob-

ability of a tail ($X = 0$) be $1 - p$ for each trial. With this setup, a random process can be defined as a mapping from the original experimental sample space $\mathcal{S} = \{(H, H, T, \ldots), (H, T, H, \ldots), (T, T, H, \ldots), \ldots\}$ to the numerical sample space $\mathcal{S}_X = \{(1, 1, 0, \ldots), (1, 0, 1, \ldots), (0, 0, 1, \ldots), \ldots\}$. Note that each simple event or element of $\mathcal{S}$ is an infinite sequence of H's and T's which is then mapped into an infinite sequence of 1's and 0's, which is the corresponding simple event in $\mathcal{S}_X$. One may picture a random process as being generated by the "random process generator" shown in Figure 16.3. The random process is composed of the infinite (but

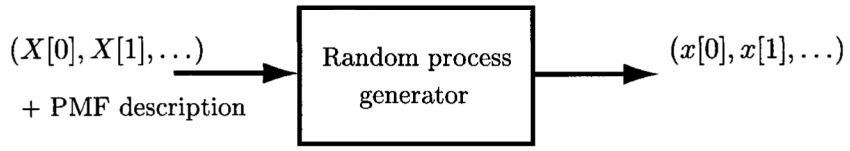


Figure 16.3: A conceptual random process generator. The input is an infinite sequence of random variables with their probabilistic description and the output is an infinite sequence of numbers.

countable) "vector" of random variables $(X[0], X[1], \ldots)$, each of which is a Bernoulli random variable, and each outcome of the random process is given by the infinite sequence of numerical values $(x[0], x[1], \ldots)$. As usual, uppercase letters are used for the random variables and lowercase letters for the values they take on. Some typical outcomes of the Bernoulli random process are shown in Figure 16.4. They were generated in MATLAB using `x=floor(rand(31,1)+0.5)` for each outcome. Each sequence in Figure 16.4 is called an *outcome* or by its synonyms of *realization* or *sample sequence*. We will prefer the use of the term "realization". Each realization is an infinite sequence of numbers. Hence, the random process is a mapping from $\mathcal{S}$, which is a set of infinite sequential experimental outcomes, to $\mathcal{S}_X$, which is a set of infinite sequences of 1's and 0's or realizations. The total number of realizations is not countable (see Problem 16.3). The set of all realizations is sometimes referred to as the *ensemble of realizations*. Just as for the case of a single random variable, which is a mapping from $\mathcal{S}$ to $\mathcal{S}_X$ and therefore is represented as the set function $X(s)$, a similar notation is used for random processes. Now, however, we will use $X[n, s]$ to represent the mapping from an element of $\mathcal{S}$ to a realization $x[n]$. In Figure 16.4 we see the result of the mapping for $s = s_1$, which is $X[n, s_1] = x_1[n]$, as well as others. It is important to note that if we fix n at $n = 18$, for example, then $X[18, s]$ is a *random variable* that has a Bernoulli PMF. Three of its outcomes are shown highlighted in Figure 16.4 with dashed boxes. Hence, all the methods developed for a single random variable are applicable. Likewise, if we fix two samples at $n = 20$ and $n = 22$, then $X[20, s]$ and $X[22, s]$ becomes a bivariate random vector. Again all our previous methods for two-dimensional random vectors apply.

To summarize, a random process is defined to be an infinite sequence of random

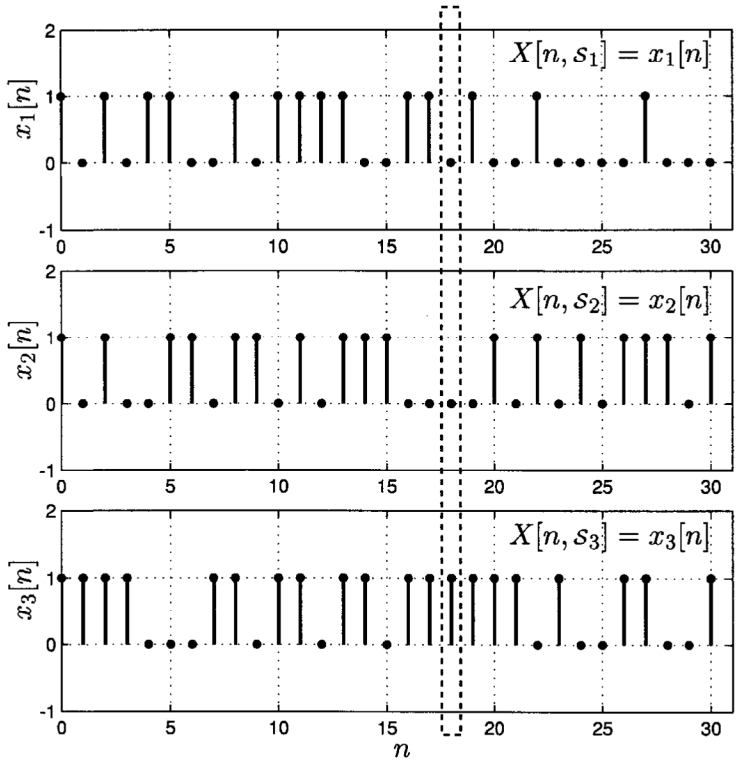


Figure 16.4: Typical outcomes of Bernoulli random process with $p = 0.5$. The realization starts at $n = 0$ and continues indefinitely. The dashed box indicates the realizations of the *random variable* $X[18, s]$.

variables $(X(0), X(1), \ldots)$, with one random variable for each time instant, and each realization of the random process takes on a value that is represented as an infinite sequence of numbers or $(x[0], x[1], \ldots)$. We will denote the random process more succinctly by $X[n]$ and the realization by $x[n]$ but *it is understood that the* n *denotes the values* $n = 0, 1, \ldots$. If we wish to indicate the random process at a *fixed time instant*, then we will use $n = n_0$ or $n = n_1$, etc. so that $X[n_0]$ is the random process at $n = n_0$ (which is just a random variable) and its realization at that time is $x[n_0]$ (which is a number). Finally, we have used the $[\cdot]$ notation to remind us that $X[n]$ is defined only for *discrete integer* times. This type of random process is known as a *discrete-time* random process. In the next section the *continuous-time* random process will be discussed. Before continuing, however, we look at a typical probability calculation for a random process.

Example 16.1 – Bernoulli random process

For the infinite coin tossing example, we might ask for the probability of the first

5 tosses coming up all heads. Thus, we wish to evaluate

$$P[X[0] = 1, X[1] = 1, X[2] = 1, X[3] = 1, X[4] = 1, X[5] = 0 \text{ or } 1, X[6] = 0 \text{ or } 1, \ldots].$$

It would seem that since we don't care what the outcomes of $X[n]$ for $n = 5, 6, \ldots$ are, then the probability expression could be replaced by

$$P[X[0] = 1, X[1] = 1, X[2] = 1, X[3] = 1, X[4] = 1]$$

and indeed this is the case, although it is not so easy to prove [Billingsley 1986]. Then, by using the assumption of independence of a Bernoulli random process we have

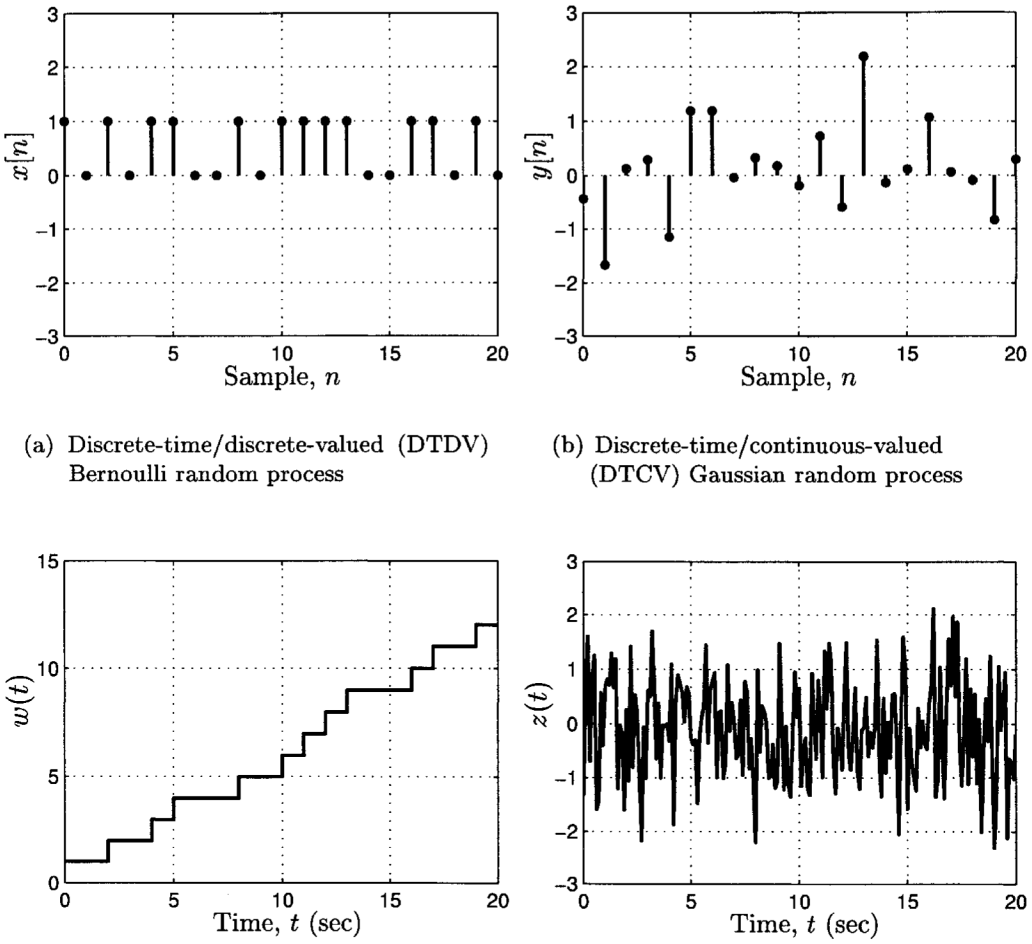
$$P[X[0] = 1, X[1] = 1, X[2] = 1, X[3] = 1, X[4] = 1] = \prod_{n=0}^{4} P[X[n] = 1] = p^5.$$

A related question is to determine the probability that we will *ever* observe 5 ones in a row. Intuitively, we expect this probability to be 1, but how do we prove this? It is not easy! Such is the difficulty encountered when we make the leap from a random vector, having a finite number of random variables, to a random process, having an infinite number of random variables.

$\diamondsuit$

16.4 Types of Random Processes

The previous example of an infinite number of coin tosses produced a random process $X[n]$ for $n = 0, 1, \ldots$. In some cases, however, we wish to think of the random process as having started sometime in the infinite past. If $X[n]$ is defined for $n = \ldots, -1, 0, 1, \ldots$ or equivalently $-\infty < n < \infty$, where it is assumed that n is an integer, then $X[n]$ is called an *infinite* random process. In contrast, the previous example is referred to as a *semi-infinite* random process. Another categorization of random processes involves whether the times at which the random variables are defined and the values that they take on are either discrete or continuous. The infinite coin toss example is a *discrete-time* random process, since it is defined for $n = 0, 1, \ldots$, and is a *discrete-valued* random process, since it takes on values 0 and 1 only. It is referred to as a *discrete-time/discrete valued* (DTDV) random process. Other types of random processes are discrete-time/continuous-valued (DTCV), continuous-time/discrete-valued (CTDV), and continuous-time/continuous-valued (CTCV). A realization of each type is shown in Figure 16.5. In Figure 16.5a a realization of the Bernoulli random process, as previously described, is shown while in Figure 16.5b a realization of a Gaussian random process with $Y[n] \sim \mathcal{N}(0, 1)$ is shown. The Bernoulli random process is defined for $n = 0, 1, \ldots$ (semi-infinite) while the Gaussian random process is defined for $-\infty < n < \infty$ and n an integer (infinite).



(a) Discrete-time/discrete-valued (DTDV) Bernoulli random process

(b) Discrete-time/continuous-valued (DTCV) Gaussian random process

(c) Continuous-time/discrete-valued (CTDV) binomial random process

(d) Continuous-time/continuous-valued (CTCV) Gaussian random process

Figure 16.5: Typical realizations of different types of random processes.

Both these random processes are discrete-time with the first one taking on only the values 0 and 1 and the second one taking on all real values. In Figure 16.5c is shown a random process, also known as a continuous-time binomial random process, which is defined as $W(t) = \sum_{n=0}^{[t]} X[n]$, where $X[n]$ is a Bernoulli random process and $[t]$ denotes the largest integer less than or equal to t. This process effectively counts the number of successes or ones of the Bernoulli random process (compare Figure 16.5c with Figure 16.5a). It is defined for all time; hence, it is a continuous-time random process, and it takes on only integer values in the range $\{0, 1, \ldots\}$; hence, it is discrete-valued. Finally, in Figure 16.5d is shown a realization of another

Gaussian process but with $Z(t) \sim \mathcal{N}(0,1)$ for all time t. This is a continuous-time random process that takes on all real values; hence, it is continuous-valued. We will generally use a discrete-time random process, with either discrete or continuous values, to introduce new concepts. This is because a continuous-time random process introduces a host of mathematical subtleties which in many cases are beyond the scope of this text. When possible, however, we will quote the analogous results for continuous-time random processes. Note finally that a realization of $X[n]$, which is $x[n]$, is also called a *sample sequence*, while a realization of $X(t)$, which is $x(t)$, is also called a *sample function*. We will, however, reserve the use of the word *sample* to refer to a *time sample* of the random process. Hence, a time sample will refer to either the random variable $X[n_0]$ $(X(t_0))$ or the realization $x[n_0]$ $(x(t_0))$ of the random process, with the meaning determined by the context of the discussion. We next revisit the random walk of Example 9.5.

Example 16.2 – Random walk (continued from Example 9.5)

Recall that

$$X_n = \sum_{i=1}^{n} U_i \qquad n = 1, 2, \ldots$$

where

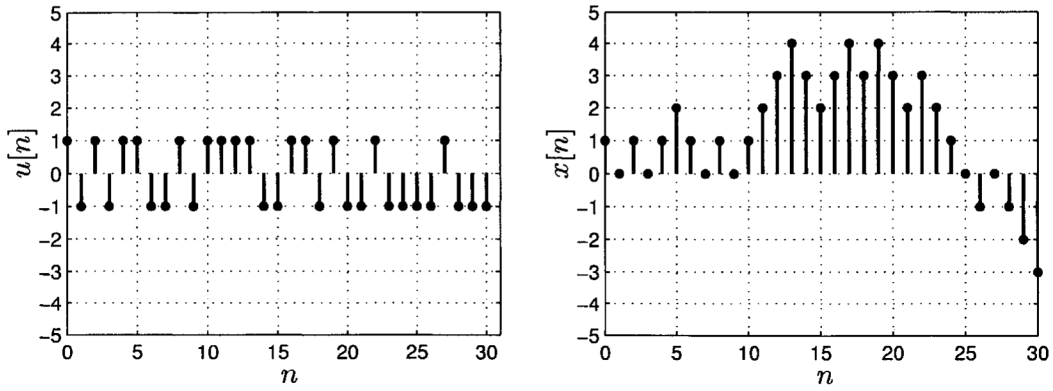
$$p_U[k] = \begin{cases} \frac{1}{2} & k = -1 \\ \frac{1}{2} & k = 1 \end{cases} \tag{16.1}$$

and the U_i's are IID. The random walk is a random process so that rewriting the definition in our new notation, we have

$$X[n] = \sum_{i=0}^{n} U[i] \qquad n = 0, 1, \ldots$$

where the $U[i]$'s are IID random variables having the PMF of (16.1). We also assume that the random walk starts at time $n = 0$. The $U[i]$'s comprise the random variables of a Bernoulli random process but with values of ± 1, instead of the usual 0 and 1. As such, we can view the $U[i]$'s as comprising a Bernoulli random process $U[n]$ for $n = 0, 1, \ldots$. Realizations of $U[n]$ and $X[n]$ are shown in Figure 16.6. One question that comes to mind is the behavior of the random walk for large n. For example, we might be interested in the PDF of $X[n]$ for large n. Relying on the central limit theorem (see Chapter 15), we can assert that the PDF is Gaussian, and therefore we need only determine the mean and variance. This easily follows from the definition of the random walk as

$$\begin{aligned} E[X[n]] &= \sum_{i=0}^{n} E[U[i]] = (n+1)E[U[0]] = 0 \\ \text{var}(X[n]) &= \sum_{i=0}^{n} \text{var}(U[i]) = (n+1)\text{var}(U[0]) = n+1 \end{aligned}$$



(a) Realization of Bernoulli random process $U[n]$

(b) Realization of random walk $X[n]$

Figure 16.6: Typical realization of a random walk.

since $E[U[i]] = 0$ and $\text{var}(U[i]) = 1$. (Note that since the $U[i]$'s are identically distributed, they all have the same mean and variance. We have arbitrarily chosen $U[0]$ in the expression for the mean and variance of a single sample.) Hence, for large n we have approximately that $X[n] \sim \mathcal{N}(0, n+1)$. Does this appear to explain the behavior of $x[n]$ shown in Figure 16.6b?

$\Diamond$

16.5 The Important Property of Stationarity

The simplest type of random process is an IID random process. The Bernoulli random process is an example of this. Each random variable $X[n_0]$ is independent of all the others and each random variable has the same marginal PMF. As such, the joint PMF of any finite number of samples can immediately be written as

$$p_{X[n_1],X[n_2],\ldots,X[n_N]}[x_1, x_2, \ldots, x_N] = \prod_{i=1}^{N} p_{X[n_i]}[x_i] \tag{16.2}$$

and used for probability calculations. For example, for a Bernoulli random process with values $0, 1$ the probability of the first 10 samples being $1, 0, 1, 0, 1, 0, 1, 0, 1, 0$ is $p^5(1-p)^5$. Note that we are able to specify the joint PMF for *any* finite number of sample times. This is sometimes referred to as being able to specify the *finite dimensional distribution* (FDD). It is the most complete probabilistic description that we can manage for a random process and reduces the analysis of a random process to the analysis of a *finite* but arbitrary set of random variables.

A generalization of the IID random process is a random process for which the FDD does not change with the time origin. This is to say that the PMF or PDF of the samples $\{X[n_1], X[n_2], \ldots, X[n_N]\}$ is the same as for $\{X[n_1 + n_0], X[n_2 + n_0], \ldots, X[n_N + n_0]\}$, where n_0 is an arbitrary integer. Alternatively, the set of samples can be shifted in time, with each one being shifted the same amount, without affecting the joint PMF or joint PDF. Mathematically, for the FDD not to change with the time origin, we must have that

$$p_{X[n_1+n_0],X[n_2+n_0],\ldots,X[n_N+n_0]} = p_{X[n_1],X[n_2],\ldots,X[n_N]} \tag{16.3}$$

for all n_0, and for any arbitrary choice of N and $n_1, n_2, \ldots, n_N$. Such a random process is said to be *stationary*. It is implicit from (16.3) that all joint and marginal PMFs or PDFs must have probabilities that do not depend on the time origin. For example, by letting $N = 1$ in (16.3) we have that $p_{X[n_1+n_0]} = p_{X[n_1]}$ and setting $n_1 = 0$, we have that $p_{X[n_0]} = p_{X[0]}$ for all n_0. This says that the marginal PMF or PDF is the same for every sample in a stationary random process. We next prove that an IID random process is stationary.

Example 16.3 – IID random process is stationary.

To prove that the IID random process is a special case of a stationary random process we must show that (16.3) is satisfied. This follows from

$$\begin{aligned} p_{X[n_1+n_0],X[n_2+n_0],\ldots,X[n_N+n_0]} &= \prod_{i=1}^{N} p_{X[n_i+n_0]} && \text{(by independence)} \\ &= \prod_{i=1}^{N} p_{X[n_i]} && \text{(by identically distributed)} \\ &= p_{X[n_1],X[n_2],\ldots,X[n_N]} && \text{(by independence).} \end{aligned}$$

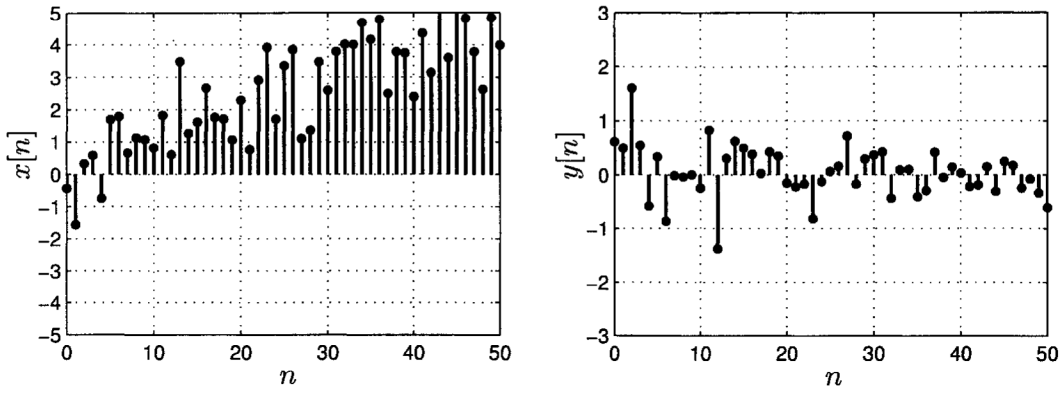
$\diamondsuit$

If a random process is stationary, then all its joint moments and more generally all expected values of functions of the random process, must also be stationary since

$$E_{X[n_1+n_0],\ldots,X[n_N+n_0]}[\cdot] = E_{X[n_1],\ldots,X[n_N]}[\cdot]$$

which follows from (16.3). Examples then of random processes that are not stationary are ones whose means and/or variances change in time, which implies that the marginal PMF or PDF change with time. In Figure 16.7 we show typical realizations of random processes whose mean in Figure 16.7a and whose variance in Figure 16.7b change with time. They were generated using the MATLAB code:

```
randn('state',0)
N=51;
x=randn(N,1)+0.1*[0:N-1]'; % for Figure 16.7a
y=sqrt(0.95.^[0:50]').*randn(N,1); % for Figure 16.7b
```



(a) Mean increasing with n

(b) Variance decreasing with n

Figure 16.7: Random processes that are not stationary.

In Figure 16.7a the true mean increases linearly from 0 to 5 while in Figure 16.7b the variance decreases exponentially as 0.95^n. It is clear then that the samples all have different moments and therefore $p_{X[n_1+n_0]} \neq p_{X[n_1]}$ which violates the condition for stationarity.

 **It is impossible to determine if a random process is stationary from a single realization.**

A realization of a random process is a *single outcome* of the random process. This is analogous to observing a single outcome of a coin toss. We cannot determine if the coin is fair by observing that the outcome was a head. What is required are multiple realizations of the coin tossing experiment. So it is with random processes. In Figure 16.7b, although we generated the realization using a variance that decreased with time, and hence the random process is not stationary, the realization shown *could have been generated with a constant variance.* Then, the values of the realization near $n = 50$ just happen to be smaller than the ones near $n = 0$, which is possible, although maybe not very probable. To better discern whether a random process is stationary we require *multiple realizations.*

Another example of a random process that is not stationary follows.

Example 16.4 – Sum random process

A sum random process is a slight generalization of the random walk process of

Example 16.2. As before, $X[n] = \sum_{i=0}^{n} U[i]$, where the $U[i]$'s are IID but for the general sum process, the $U[i]$'s can have any, although the same, PMF or PDF. Thus, the sum random process is not stationary since

$$\begin{aligned} E[X[n]] &= (n+1)E_U[U[0]] \\ \text{var}(X[n]) &= (n+1)\text{var}(U[0]) \end{aligned}$$

both of which change with n. Hence, it violates the condition for stationarity.

$\diamondsuit$

A random process that is not stationary is said to be *nonstationary*. In light of the fact that an IID random process lends itself to simple probability calculations, it is advantageous, if possible, to transform a nonstationary random process into a stationary one (see Problem 16.12 on transforming the random processes of Figure 16.7 into stationary ones). As an example, for the sum random process this can be done by "reversing" the summing operation. Specifically, we difference the random process. Then $X[n] - X[n-1] = U[n]$ for $n \geq 0$, where we define $X[-1] = 0$. This is an IID random process. The differences or *increment* random variables $U[n]$ are independent and identically distributed. More generally, for the sum random process any two *increments* of the form

$$\begin{aligned} X[n_2] - X[n_1] &= \sum_{i=n_1+1}^{n_2} U[i] \\ X[n_4] - X[n_3] &= \sum_{i=n_3+1}^{n_4} U[i] \end{aligned}$$

are independent if $n_4 > n_3 \geq n_2 > n_1$. Thus, nonoverlapping increments for a sum random process are independent. (Recall that functions of independent random variables are themselves independent.) If furthermore, $n_4 - n_3 = n_2 - n_1$, then they also have the same PMF or PDF since they are composed of the same number of IID random variables. It is then said that for the sum random process, the increments are independent and stationary (equivalent to being identically distributed) or that it has *stationary independent increments*. The reader may wish to ponder whether a random process can have independent but nonstationary increments (see Problem 16.13). Many random processes (an example of which follows) that we will encounter have this property and it allows us to more easily analyze the probabilistic behavior.

Example 16.5 – Binomial counting random process

Consider the repeated coin tossing experiment where we are interested in the number of heads that occurs. Letting $U[n]$ be a Bernoulli random process with $U[n] = 1$ with probability p and $U[n] = 0$ with probability $1-p$, the number of heads is given

by the *binomial counting* or sum process

$$X[n] = \sum_{i=0}^{n} U[i] \qquad n = 0, 1, \ldots$$

or equivalently

$$X[n] = \begin{cases} U[0] & n = 0 \\ X[n-1] + U[n] & n \geq 1 \,. \end{cases}$$

A typical realization is shown in Figure 16.8. The random process has stationary

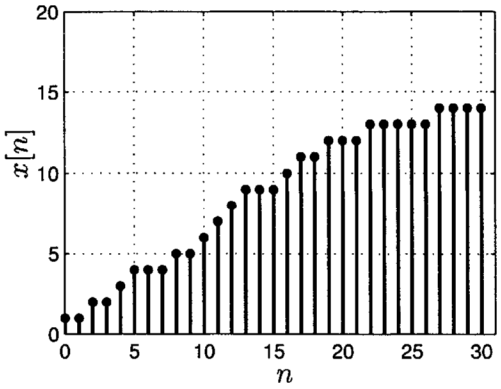


Figure 16.8: Typical realization of binomial counting random process with $p = 0.5$.

and independent increments since the changes over two nonoverlapping intervals are composed of different sets of identically distributed $U[i]$'s. We can use this property to more easily determine probabilities of events. For example, to determine $p_{X[1],X[2]}[1,2] = P[X[1] = 1, X[2] = 2]$, we can note that the event $X[1] = 1, X[2] = 2$ is equivalent to the event $Y_1 = X[1] - X[-1] = 1$, $Y_2 = X[2] - X[1] = 1$, where $X[-1]$ is defined to be identically zero. But Y_1 and Y_2 are nonoverlapping increments (but of unequal length), making them independent random variables. Thus,

$$\begin{aligned} P[X[1] = 1, X[2] = 2] &= P[Y_1 = 1, Y_2 = 1] = P[Y_1 = 1]P[Y_2 = 1] \\ &= P[\underbrace{U[0] + U[1]}_{\text{bin}(2,p)} = 1]P[U[2] = 1] \\ &= \binom{2}{1} p^1(1-p)^1 \cdot p \\ &= 2p^2(1-p). \end{aligned}$$

$\Diamond$

16.6 Some More Examples

We continue our discussion by examining some random processes of practical interest.

Example 16.6 – White Gaussian noise

A common model for physical noise, such as resistor noise due to electron motion fluctuations in an electric field, is termed *white Gaussian noise* (WGN). It is assumed that the noise has been sampled in time to yield a DTCV random process $X[n]$. The WGN random process is defined to be an IID one whose marginal PDF is Gaussian so that $X[n] \sim \mathcal{N}(0, \sigma^2)$ for $-\infty < n < \infty$. Each random variable $X[n_0]$ has a mean of zero, consistent with our notion of a noise process, and the same variance or because the mean is zero, the same power $E[X^2[n_0]]$. A typical realization is shown in Figure 16.5b for $\sigma^2 = 1$. The WGN random process is stationary since it is an IID random process. Its joint PDF is

$$\begin{aligned} p_{X[n_1],X[n_2],\ldots,X[n_N]}(x_1, x_2, \ldots, x_N) &= \prod_{i=1}^{N} p_{X[n_i]}(x_i) \\ &= \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{1}{2\sigma^2} x_i^2\right) \\ &= \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{N} x_i^2\right). \end{aligned} \quad (16.4)$$

Note that the joint PDF is $\mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$, which is a special form of the multivariate Gaussian PDF (see Problem 16.15). The terminology of "white" derives from the property that such a random process may be synthesized from a sum of different frequency random sinusoids each having the same power, much the same as white light is composed of equal contributions of each visible wavelength of light. We will justify this property in Chapter 17 when we discuss the power spectral density.

$\diamondsuit$

Example 16.7 – Moving average random process

The moving average (MA) random process is a DTCV random process defined as

$$X[n] = \tfrac{1}{2}(U[n] + U[n-1]) \qquad -\infty < n < \infty$$

where $U[n]$ is a WGN random process with variance σ_U^2. (To avoid confusion with the variance of other random variables we will sometimes use a subscript on σ^2, in this case σ_U^2, to refer to the variance of the $U[n_0]$ random variable.) The terminology of moving average refers to the averaging of the current random variable $U[n]$ with the previous random variable $U[n-1]$ to form the current moving average random

variable. Also, this averaging "moves" in time, as for example,

$$\begin{aligned} X[0] &= \tfrac{1}{2}(U[0]+U[-1]) \\ X[1] &= \tfrac{1}{2}(U[1]+U[0]) \\ X[2] &= \tfrac{1}{2}(U[2]+U[1]) \\ &\text{etc.} \end{aligned}$$

A typical realization of $X[n]$ is shown in Figure 16.9 and should be compared to the realization of $U[n]$ shown in Figure 16.5b. It is seen that the moving average random process is "smoother" than the WGN random process, from which it was obtained. Further smoothing is possible by averaging more WGN samples together (see Problem 16.17). The MATLAB code shown below was used to generate the realization.

```
randn('state',0)
u=randn(21,1);
for i=1:21
  if i==1
    x(i,1)=0.5*(u(1)+randn(1,1)); % needed to initialize sequence
  else
    x(i,1)=0.5*(u(i)+u(i-1));
  end
end
```

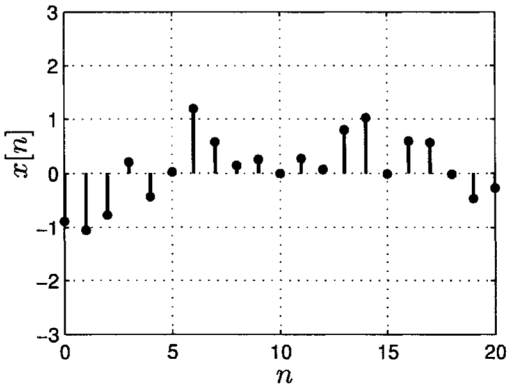


Figure 16.9: Typical realization of moving average random process. The realization of the $U[n]$ random process is shown in Figure 16.5b.

The joint PDF of $X[n]$ can be determined by observing that it is a linearly transformed version of $U[n]$. As an example, to determine the joint PDF of the random

vector $[X[0]\,X[1]]^T$, we have from the definition of the MA random process

$$\begin{bmatrix} X[0] \\ X[1] \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} U[-1] \\ U[0] \\ U[1] \end{bmatrix}$$

or in matrix/vector notation $\mathbf{X} = \mathbf{GU}$. Now recalling that $\mathbf{U}$ is a Gaussian random vector (see (16.4)) and that a linear transformation of a Gaussian random vector produces another Gaussian random vector, we have from Example 14.3 that

$$\mathbf{X} \sim \mathcal{N}(\mathbf{G}E[\mathbf{U}], \mathbf{G}\mathbf{C}_U\mathbf{G}^T).$$

Explicitly, since each sample of $U[n]$ is zero mean with variance σ_U^2 and all samples are independent, we have that $E[\mathbf{U}] = \mathbf{0}$ and $\mathbf{C}_U = \sigma_U^2\mathbf{I}$. This results in

$$\mathbf{X} = \begin{bmatrix} X[0] \\ X[1] \end{bmatrix} \sim \mathcal{N}(\mathbf{0}, \sigma_U^2\mathbf{G}\mathbf{G}^T)$$

where

$$\mathbf{G}\mathbf{G}^T = \begin{bmatrix} \frac{1}{2} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{2} \end{bmatrix}.$$

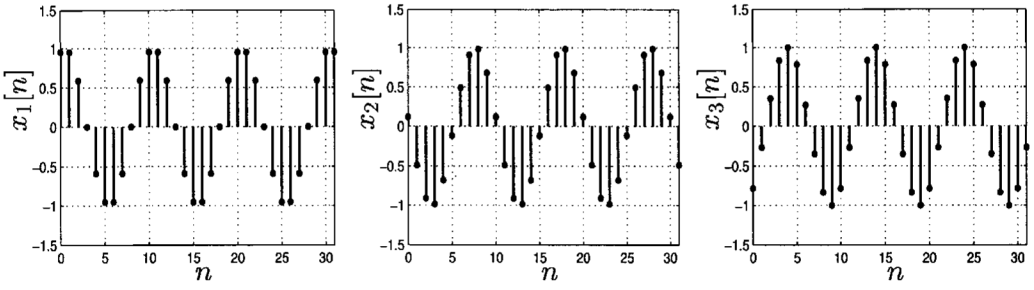
It can furthermore be shown that the MA random process is stationary (see Example 20.2 and Property 20.2).

$\diamondsuit$

Example 16.8 – Randomly phased sinusoid (or sine wave)
Consider the DTCV random process given as

$$X[n] = \cos(2\pi(0.1)n + \Theta) \qquad -\infty < n < \infty$$

where $\Theta \sim \mathcal{U}(0, 2\pi)$. Some typical realizations are shown in Figure 16.10. The MATLAB statements `n=[0:31]'` and `x=cos(2*pi*0.1*n+2*pi*rand(1,1))` can be used to generate each realization. This random process is frequently used to model an analog sinusoid whose phase is unknown and that has been sampled by an analog-to-digital convertor. It is nearly a deterministic signal, except for the phase uncertainty, and is therefore perfectly predictable. This is to say that once we observe two successive samples, then all the remaining ones are known (see Problem 16.20). This is in contrast to the WGN random process, for which regardless of how many samples we observe, we cannot predict any of the remaining ones due to the independence of the samples. Because of the predictability of the randomly phased sinusoidal process, the joint PDF can only be represented using impulsive functions. As an example, you might try to find the PDF of (X, Y) if (X, Y) has the bivariate Gaussian



(a) $\theta = 5.9698$ (b) $\theta = 1.4523$ (c) $\theta = 3.8129$

Figure 16.10: Typical realizations for randomly phased sinusoid.

PDF with $\rho = 1$. We will not pursue this further. However, we can determine the marginal PDF $p_{X[n]}$. To do so we use the transformation formula of (10.30), where the Y random variable is $X[n_0]$ (considering the random process at a fixed time) and the X random variable is Θ. The transformation is shown in Figure 16.11 for $n_0 = 0$. Note that there are two solutions for any given $x[n_0] = y$ (except for the

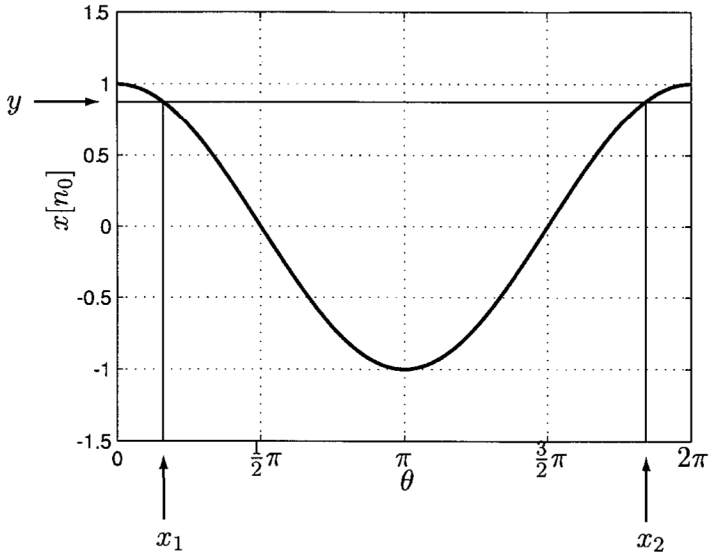


Figure 16.11: Function transforming Θ into $X[n_0]$ for the value $n_0 = 0$, where $X[n_0] = \cos(2\pi(0.1)n_0 + \Theta)$.

point at $\theta = \pi$, which has probability zero). We denote the solutions as $\theta = x_1, x_2$. Using our previous notation of $y = g(x)$ for a transformation of a single random

variable we have that

$$y = \cos(2\pi(0.1)n_0 + x)$$

so that the solutions are

$$\begin{aligned} x_1 &= \arccos(y) - 2\pi(0.1)n_0 = g_1^{-1}(y) \\ x_2 &= 2\pi - [\arccos(y) - 2\pi(0.1)n_0] = g_2^{-1}(y) \end{aligned}$$

for $-1 < y < 1$ and thus $0 < \arccos(y) < \pi$. Using $d\arccos(y)/dy = 1/\sqrt{1-y^2}$, we have

$$\begin{aligned} p_Y(y) &= p_X(g_1^{-1}(y)) \left| \frac{dg_1^{-1}(y)}{dy} \right| + p_X(g_2^{-1}(y)) \left| \frac{dg_2^{-1}(y)}{dy} \right| \\ &= \frac{1}{2\pi} \left| \frac{1}{\sqrt{1-y^2}} \right| + \frac{1}{2\pi} \left| -\frac{1}{\sqrt{1-y^2}} \right| \\ &= \frac{1}{\pi\sqrt{1-y^2}}. \end{aligned}$$

Finally, in our original notation we have the marginal PDF for $X[n]$ for any n

$$p_{X[n]}(x) = \begin{cases} \frac{1}{\pi\sqrt{1-x^2}} & -1 < x < 1 \\ 0 & \text{otherwise.} \end{cases}$$

This PDF is shown in Figure 16.12. Note that the values of $X[n]$ that are most probable are near $x = \pm 1$. Can you explain why? (Hint: Determine the values of θ for which $0.9 < \cos\theta < 1$ and also $0 < \cos\theta < 0.1$ in Figure 16.11.)

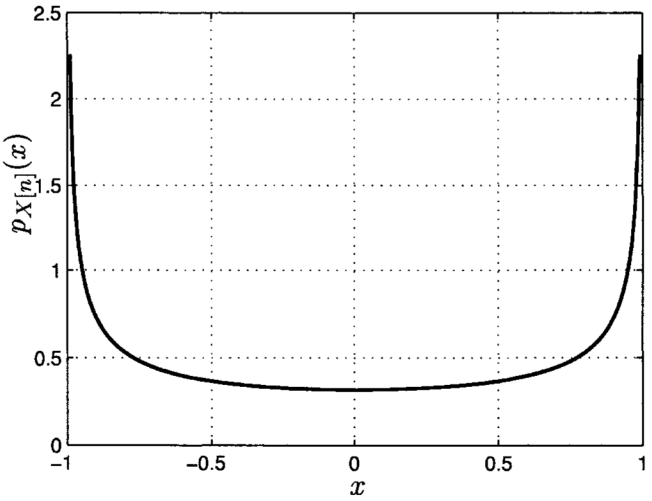


Figure 16.12: Marginal PDF for randomly phased sinusoid.

16.7 Joint Moments

The first and second moments or equivalently the mean and variance of a random process at a given sample time are of great practical importance since they are easily determined. Also, the covariance between two samples of the random process at two different times is easily found. At worst, the first and second moments can always be estimated in practice. This is in contrast to the joint PMF or joint PDF, which in practice may be difficult to determine. Hence, we next define and give some examples of the mean, variance, and covariance sequences for a DTCV random process. The mean sequence is defined as

$$\mu_X[n] = E[X[n]] \qquad -\infty < n < \infty \tag{16.5}$$

while the variance sequence is defined as

$$\sigma_X^2[n] = \text{var}(X[n]) \qquad -\infty < n < \infty \tag{16.6}$$

and finally the covariance sequence is defined as

$$\begin{aligned} c_X[n_1, n_2] &= \text{cov}(X[n_1], X[n_2]) \\ &= E[(X[n_1] - \mu_X[n_1])(X[n_2] - \mu_X[n_2])] \qquad \begin{matrix} -\infty < n_1 < \infty \\ -\infty < n_2 < \infty . \end{matrix} \end{aligned} \tag{16.7}$$

The expectations for the mean and variance are taken with respect to the PMF or PDF $p_{X[n]}$ for a particular value of n. Similarly, the expectation needed for the evaluation of the covariance is with respect to the joint PMF or PDF $p_{X[n_1],X[n_2]}$ for particular values of n_1 and n_2. Since the required PMF or PDF should be clear from the context, we henceforth do not subscript the expectation operator as we have done so previously. Note that the usual symmetry property of the covariance holds, which results in $c_X[n_2, n_1] = c_X[n_1, n_2]$. Also, it follows from the definition of the covariance sequence that $c_X[n, n] = \sigma_X^2[n]$. The actual evaluation of the moments proceeds exactly the same as for random variables.

If the random process is a continuous-time one, then the corresponding definitions are

$$\begin{aligned} \mu_X(t) &= E[X(t)] \\ \sigma_X^2(t) &= \text{var}(X(t)) \\ c_X(t_1, t_2) &= E[(X(t_1) - \mu_X(t_1))(X(t_2) - \mu_X(t_2))]. \end{aligned}$$

These are called the *mean function*, *variance function*, and *covariance function*, respectively. We next examine the moments for the examples of the previous section. Noting that the variance is just the covariance sequence evaluated at $n_1 = n_2 = n$, we need only determine the mean and covariance sequences.

Example 16.9 – White Gaussian noise
Since $X[n] \sim \mathcal{N}(0, \sigma^2)$ for all n, we have that

$$\begin{aligned} \mu_X[n] &= 0 \qquad -\infty < n < \infty \\ \sigma_X^2[n] &= \sigma^2 \qquad -\infty < n < \infty. \end{aligned}$$

The covariance sequence for $n_1 \neq n_2$ must be zero since the random variables are all independent. Recalling that the covariance between $X[n]$ and itself is just the variance, we have that

$$c_X[n_1, n_2] = \begin{cases} 0 & n_1 \neq n_2 \\ \sigma^2 & n_1 = n_2 \,. \end{cases}$$

This can be written in more succinct form by using the discrete delta function as

$$c_X[n_1, n_2] = \sigma^2 \delta[n_2 - n_1].$$

In summary, for a WGN random process we have that $\mu_X[n] = 0$ for all n and $c_X[n_1, n_2] = \sigma^2 \delta[n_2 - n_1]$.

$\Diamond$

Example 16.10 – Moving average random process
The mean sequence is

$$\mu_X[n] = E[X[n]] = E[\tfrac{1}{2}(U[n] + U[n-1])] = 0 \qquad -\infty < n < \infty$$

since $U[n]$ is white Gaussian noise, which has a zero mean for all n. To find the covariance sequence using $X[n] = (U[n] + U[n-1])/2$, we have

$$\begin{aligned} c_X[n_1, n_2] &= E[(X[n_1] - \mu_X[n_1])(X[n_2] - \mu_X[n_2])] \\ &= E[X[n_1]X[n_2]] \\ &= \frac{1}{4} E[(U[n_1] + U[n_1 - 1])(U[n_2] + U[n_2 - 1])] \\ &= \frac{1}{4} \left(E[U[n_1]U[n_2]] + E[U[n_1]U[n_2 - 1]] \right. \\ &\quad \left. + E[U[n_1 - 1]U[n_2]] + E[U[n_1 - 1]U[n_2 - 1]] \right). \end{aligned}$$

But $E[U[k]U[l]] = \sigma_U^2 \delta[l - k]$ since $U[n]$ is WGN, and as a result

$$\begin{aligned} c_X[n_1, n_2] &= \frac{1}{4} \left(\sigma_U^2 \delta[n_2 - n_1] + \sigma_U^2 \delta[n_2 - 1 - n_1] + \sigma_U^2 \delta[n_2 - n_1 + 1] + \sigma_U^2 \delta[n_2 - n_1] \right) \\ &= \frac{\sigma_U^2}{2} \delta[n_2 - n_1] + \frac{\sigma_U^2}{4} \delta[n_2 - n_1 - 1] + \frac{\sigma_U^2}{4} \delta[n_2 - n_1 + 1]. \end{aligned}$$

This is plotted in Figure 16.13 versus $\Delta n = n_2 - n_1$. It is seen that the covariance sequence is zero unless the two samples are at most one unit apart or $\Delta n = n_2 - n_1 =$

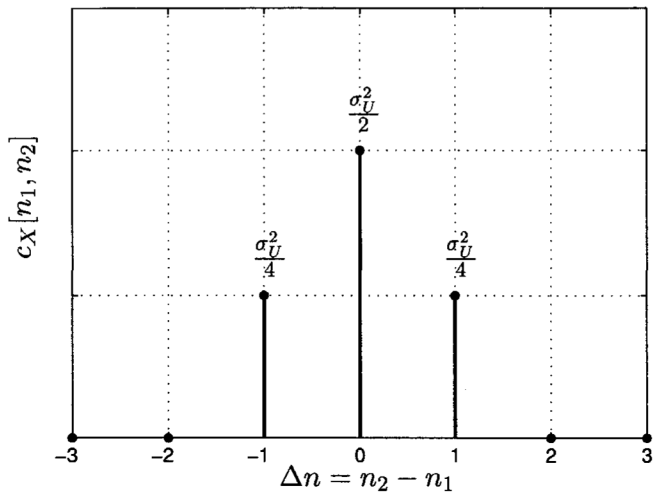


Figure 16.13: Covariance sequence for moving average random process.

± 1. Note that the covariance between any two samples spaced one unit apart is the same. Thus, for example, $X[1]$ and $X[2]$ have the covariance $c_X[1,2] = \sigma_U^2/4$, as do $X[9]$ and $X[10]$ since $c_X[9,10] = \sigma_U^2/4$, and as do $X[-3]$ and $X[-2]$ since $c_X[-3,-2] = \sigma_U^2/4$ (see Figure 16.13). Any samples that are spaced *more than* one unit apart are *uncorrelated*. This is because for $|n_2 - n_1| > 1$, $X[n_1]$ and $X[n_2]$ are independent, being composed of two sets of *different* WGN samples (recall that functions of independent random variables are independent). In summary, we have that

$$\begin{aligned} \mu_X[n] &= 0 \\ c_X[n_1, n_2] &= \begin{cases} \frac{\sigma_U^2}{2} & n_1 = n_2 \\ \frac{\sigma_U^2}{4} & |n_2 - n_1| = 1 \\ 0 & |n_2 - n_1| > 1 \,. \end{cases} \end{aligned}$$

and the variance is $c_X[n,n] = \sigma_U^2/2$ for all n. Also, note from Figure 16.13 that the covariance sequence is symmetric about $\Delta n = 0$.

$\diamondsuit$

Example 16.11 – Randomly phased sinusoid

Recalling that the phase is uniformly distributed on $(0, 2\pi)$ we have that the mean

sequence is

$$\begin{aligned}\mu_X[n] &= E[X[n]] = E[\cos(2\pi(0.1)n + \Theta)] \\ &= \int_0^{2\pi} \cos(2\pi(0.1)n + \theta)\frac{1}{2\pi}d\theta \qquad \text{(use (11.10))} \\ &= \frac{1}{2\pi}\sin(2\pi(0.1)n + \theta)\Big|_0^{2\pi} = 0\end{aligned}$$

for all n. Noting that the mean sequence is zero, the covariance sequence becomes

$$\begin{aligned}c_X[n_1, n_2] &= E[X[n_1]X[n_2]] \\ &= \int_0^{2\pi} [\cos(2\pi(0.1)n_1 + \theta)\cos(2\pi(0.1)n_2 + \theta)]\frac{1}{2\pi}d\theta \\ &= \int_0^{2\pi} \left[\frac{1}{2}\cos[2\pi(0.1)(n_2 - n_1)] + \frac{1}{2}\cos[2\pi(0.1)(n_1 + n_2) + 2\theta]\right]\frac{1}{2\pi}d\theta \\ &= \frac{1}{2}\cos[2\pi(0.1)(n_2 - n_1)] + \frac{1}{8\pi}\sin[2\pi(0.1)(n_1 + n_2) + 2\theta]\Big|_0^{2\pi} \\ &= \frac{1}{2}\cos[2\pi(0.1)(n_2 - n_1)].\end{aligned}$$

Once again the covariance sequence depends only on the spacing between the two

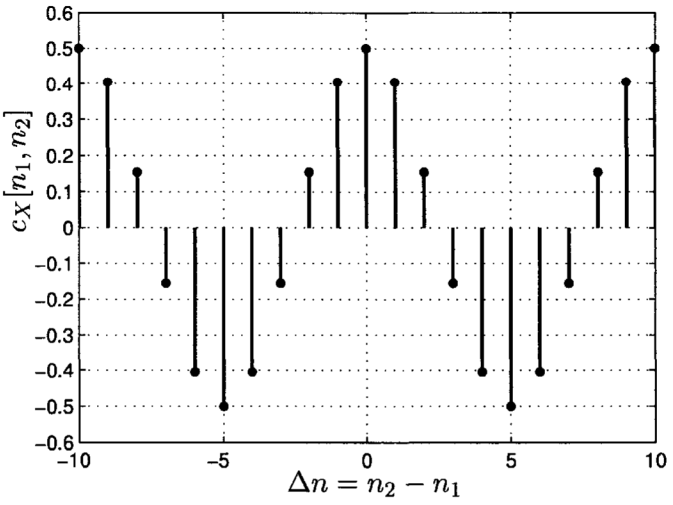


Figure 16.14: Covariance sequence for randomly phased sinusoid.

samples or on $n_2 - n_1$. The covariance sequence is shown in Figure 16.14. The reader should note the symmetry of the covariance sequence about $\Delta n = 0$. Also, the variance follows as $\sigma_X^2[n] = c_X[n, n] = 1/2$ for all n. It is interesting to observe that in this example the fact that the mean sequence is zero makes intuitive sense.

To see this we have plotted 50 realizations of the random process in an overlaid fashion in Figure 16.15. This representation is called a *scatter diagram.* Also is

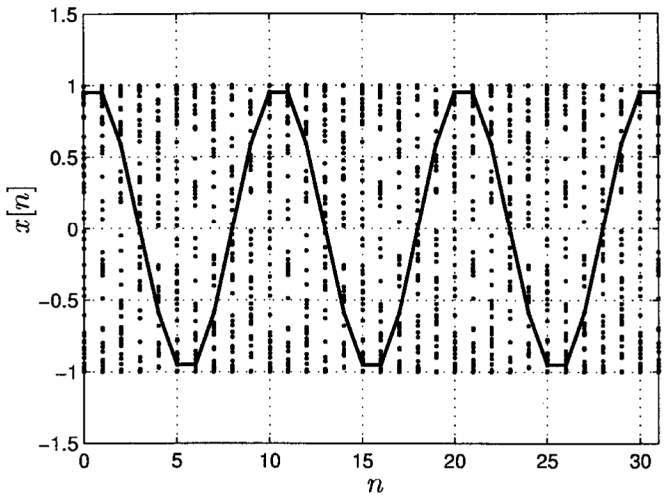


Figure 16.15: Fifty realizations of randomly phased sinusoid plotted in an overlaid format with one realization shown with its points connected by straight lines.

plotted the first realization with the values connected by straight lines for easier viewing. The difference in the realizations is due to the different values of phase realized. It is seen that for a given time instant the values are nearly symmetric about zero, as is predicted by the PDF shown in Figure 16.12 and that the majority of the values are near ± 1, again in agreement with the PDF. The MATLAB code used to generate Figure 16.15 (but omitting the solid curve) is given below.

```
clear all
rand('state',0)
n=[0:31]';
nreal=50;
for i=1:nreal
   x(:,i)=cos(2*pi*0.1*n+2*pi*rand(1,1));
end
plot(n,x(:,1),'.')
grid
hold on
for i=2:nreal
   plot(n,x(:,i),'.')
end
axis([0 31 -1.5 1.5])
```

In these three examples the covariance sequence only depends on $|n_2 - n_1|$. This is not always the case, as is illustrated in Problem 16.26. Also, another counterexample is the random process whose realization is shown in Figure 16.7b. This random process has $\text{var}(X[n]) = c_X[n,n]$ which is not a function of $n_2 - n_1 = n - n = 0$ since otherwise its variance would be a constant for all n.

16.8 Real-World Example – Statistical Data Analysis

It was mentioned in the introduction that some meterologists argue that the annual summer rainfall totals are increasing due to global warming. Referring to Figure 16.1 this supposition asserts that if $X[n]$ is the annual summer rainfall total for year n, then $\mu_X[n_2] > \mu_X[n_1]$ for $n_2 > n_1$. One way to attempt to confirm or dispute this supposition is to assume that $\mu_X[n] = an + b$ and then determine if $a > 0$, as would be the case if the mean were increasing. From the data shown in Figure 16.1 we can estimate a. To do so we let the year 1895, which is the beginning of our data set, be indexed as $n = 0$ and note that $an + b$ when plotted versus n is a straight line. We estimate a by fitting a straight line to the data set using a *least squares* procedure [Kay 1993]. The least squares estimate chooses as estimates of a and b the values that minimize the *least squares error*

$$J(a,b) = \sum_{n=0}^{N-1} (x[n] - (an + b))^2 \tag{16.8}$$

where $N = 108$ for our data set. This approach can be shown to be an optimal one under the condition that the random process is actually given by $X[n] = an + b + U[n]$, where $U[n]$ is a WGN random process [Kay 1993]. Note that if we did not suspect that the mean rainfall totals were changing, then we might assume that $\mu_X[n] = b$ and the least squares estimate of b would result from minimizing

$$J(b) = \sum_{n=0}^{N-1} (x[n] - b)^2.$$

If we differentiate $J(b)$ with respect to b, set the derivative equal to zero, and solve for b, we obtain (see Problem 16.32)

$$\hat{b} = \frac{1}{N} \sum_{n=0}^{N-1} x[n]$$

or $\hat{b} = \bar{x}$, where $\bar{x}$ is the sample mean, which for our data set is 9.76. Now, however, we obtain the least squares estimates of a and b by differentiating (16.8) with respect

to b and a to yield

$$\begin{aligned}\frac{\partial J}{\partial b} &= -2\sum_{n=0}^{N-1}(x[n]-an-b)=0\\ \frac{\partial J}{\partial a} &= -2\sum_{n=0}^{N-1}(x[n]-an-b)n=0.\end{aligned}$$

This results in two simultaneous linear equations

$$\begin{aligned}bN+a\sum_{n=0}^{N-1}n &= \sum_{n=0}^{N-1}x[n]\\ b\sum_{n=0}^{N-1}n+a\sum_{n=0}^{N-1}n^2 &= \sum_{n=0}^{N-1}nx[n].\end{aligned}$$

In vector/matrix form this is

$$\begin{bmatrix} N & \sum_{n=0}^{N-1}n \\ \sum_{n=0}^{N-1}n & \sum_{n=0}^{N-1}n^2 \end{bmatrix}\begin{bmatrix} b \\ a \end{bmatrix}=\begin{bmatrix} \sum_{n=0}^{N-1}x[n] \\ \sum_{n=0}^{N-1}nx[n] \end{bmatrix} \tag{16.9}$$

which is easily solved to yield the estimates $\hat{b}$ and $\hat{a}$. For the data of Figure 16.1 the estimates are $\hat{a}=0.0173$ and $\hat{b}=8.8336$. The data along with the estimated mean sequence $\hat{\mu}_X[n]=0.0173n+8.8336$ are shown in Figure 16.16. Note that the

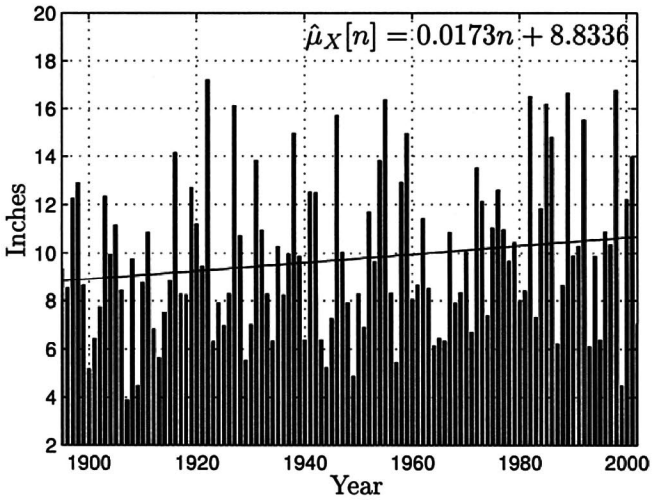


Figure 16.16: Annual summer rainfall in Rhode Island and the estimated mean sequence, $\hat{\mu}_X[n]=0.0173n+8.8336$, where $n=0$ corresponds to the year 1895.

mean indeed appears to be increasing with time. The least squares error sequence,

which is defined as $e[n] = x[n] - (\hat{a}n + \hat{b})$, is shown in Figure 16.17. It is sometimes referred to as the *fitting error*. Note that the error can be quite large. In fact, we

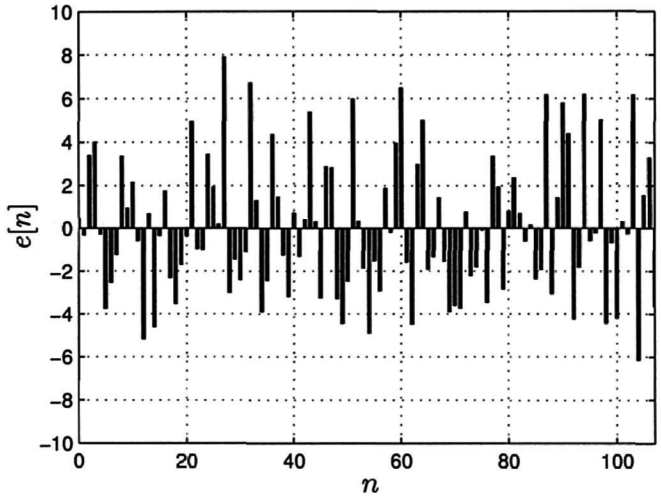


Figure 16.17: Least squares error sequence for annual summer rainfall in Rhode Island fitted with a straight line.

have that $(1/N)\sum_{n=0}^{N-1} e^2[n] = 10.05$.

Now the real question is whether the estimated mean increase in rainfall is significant. The increase is $\hat{a} = 0.0173$ *per year* for a total increase of about 1.85 inches over the course of 108 years. Is it possible that the true mean rainfall has not changed, or that it is really $\mu_X[n] = b$ with the true value of a being zero? In effect, is the value of $\hat{a} = 0.0173$ only due to estimation error? One way to answer this question is to hypothesize that $a = 0$ and then determine the probability density function of $\hat{a}$ as obtained from (16.9). This can be done analytically by assuming $X[n] = b + U[n]$, where $U[n]$ is white Gaussian noise (see Problem 16.33). However, we can gain some quick insight into the problem by resorting to a computer simulation. To do so we assume that the true model for the rainfall data is $X[n] = b + U[n] = 9.76 + U[n]$, where $U[n]$ is white Gaussian noise with variance σ^2. Since we do not know the value of σ^2, we estimate it by using the results shown in Figure 16.17. The least squares error sequence $e[n]$, which is the original data with its estimated mean sequence subtracted, should then be an estimate of $U[n]$. Therefore, we use $\hat{\sigma^2} = (1/N)\sum_{n=0}^{N-1} e^2[n] = 10.05$ in our simulation. In summary, we generate 20 realizations of the random process $X[n] = 9.76 + U[n]$, where $U[n]$ is WGN with $\sigma^2 = 10.05$. Then, we use (16.9) to estimate a and b and finally we plot our mean sequence estimate, which is $\hat{\mu}_X[n] = \hat{a}n + \hat{b}$ for each realization. Using the MATLAB code shown at the end of this section, the results are shown in Figure 16.18. It is seen that even though the true value of a is zero, the estimated value will take on nonzero values with a high probability. Since some of the lines are decreasing, some

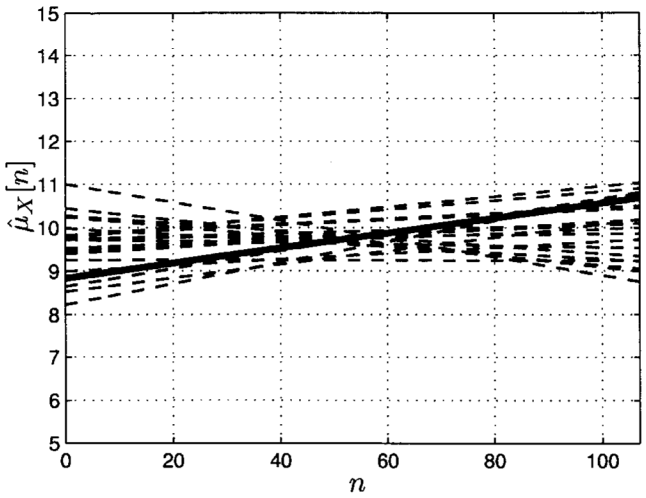


Figure 16.18: Twenty realizations of the estimated mean sequence $\hat{\mu}_X[n] = \hat{a}n + \hat{b}$ based on the random process $X[n] = 9.76 + U[n]$ with $U[n]$ being WGN with $\sigma^2 = 10.05$. The realizations are shown as dashed lines. The estimated mean sequence from Figure 16.16 is shown as the solid line.

of the estimated values of a are even negative. Hence, we would be hard pressed to say that the mean rainfall totals are indeed increasing. Such is the quandry that scientists must deal with on an everyday basis. The only way out of this dilemma is to accumulate more data so that hopefully our estimate of a will be more accurate (see also Problem 16.34).

```
clear all
randn('state',0)
years=[1895:2002]';
N=length(years);
n=[0:N-1]';
A=[N sum(n);sum(n) sum(n.^2)]; % precompute matrix (see (16.9))
B=inv(A); % invert matrix
for i=1:20
  xn=9.76+sqrt(10.05)*randn(N,1); % generate realizations
  baest=B*[sum(xn);sum(n.*xn)]; % estimate a and b using (16.9)
  aest=baest(2);best=baest(1);
  meanest(:,i)=aest*n+best; % determine mean sequence estimate
end
figure % plot mean sequence estimates and overlay
plot(n,meanest(:,1))
grid
xlabel('n')
```

```
ylabel('Estimated mean')
axis([0 107 5 15])
hold on
for i=2:20
  plot(n,meanest(:,i))
end
```

References

Billingsley, P., *Probability and Measure*, John Wiley & Sons, New York, 1986.

DowJones.com, "DowJones Averages," http://averages.dowjones.com/jsp/uiHistoricalIndexRep.jsp, 2004.

Kay, S., *Fundamentals of Statistical Signal Processing: Estimation Theory*, Prentice-Hall, Englewood Cliffs, NJ, 1993.

Problems

16.1 (‿) **(w)** Describe a random process that you are likely to encounter in the following situations:

a. listening to the daily weather forecast

b. paying the monthly telephone bill

c. leaving for work in the morning

Why is each process a random one?

16.2 (w) A single die is tossed repeatedly. What are $\mathcal{S}$ and $\mathcal{S}_X$? Also, can you determine the joint PMF for any N sample times?

16.3 (t) An infinite sequence of 0's and 1's, denoted as $b_1, b_2, \ldots$, can be used to represent any number x in the interval $[0, 1]$ using the binary representation formula

$$x = \sum_{i=1}^{\infty} b_i 2^{-i}.$$

For example, we can represent 3/4 as $0.b_1b_2\ldots = 0.11000\ldots$ and 1/16 as $0.b_1b_2\ldots = 0.0001000\ldots$. Find the representations for 7/8 and 5/8. Is the total number of infinite sequences of 0's and 1's countable?

16.4 (‿) **(w)** For a Bernoulli random process determine the probability that we will observe an alternating sequence of 1's and 0's for the first 100 samples with the first sample being a 1. What is the probability that we will observe an alternating sequence of 1's and 0's for all n?

16.5 (w) Classify the following random processes as either DTDV, DTCV, CTDV, or CTCV:

a. temperature in Rhode Island

b. outcomes for continued spins of a roulette wheel

c. daily weight of person

d. number of cars stopped at an intersection

16.6 (c) Simulate a realization of the random walk process described in Example 16.2 on a computer. What happens as n becomes large?

16.7 (☺) (c,f) A *biased* random walk process is defined as $X[n] = \sum_{i=0}^{n} U[i]$, where $U[i]$ is a Bernoulli random process with

$$p_U[k] = \begin{cases} \frac{1}{4} & k = -1 \\ \frac{3}{4} & k = 1\,. \end{cases}$$

What is $E[X[n]]$ and $\text{var}(X[n])$ as a function of n? Next, simulate on a computer a realization of this random process. What happens as $n \to \infty$ and why?

16.8 (w) A random process $X[n]$ is stationary. If it is known that $E[X[10]] = 10$ and $\text{var}(X[10]) = 1$, then determine $E[X[100]]$ and $\text{var}(X[100])$.

16.9 (☺) (f) The IID random process $X[n]$ has the marginal PDF $p_X(x) = \exp(-x)u(x)$. What is the probability that $X[0], X[1], X[2]$ will all be greater than 1?

16.10 (w) If an IID random process $X[n]$ is transformed to the random process $Y[n] = X^2[n]$, is the transformed random process also IID?

16.11 (w) A Bernoulli random process $X[n]$ that takes on values 0 or 1, each with probability of $p = 1/2$, is transformed using $Y[n] = (-1)^n X[n]$. Is the random process $Y[n]$ IID?

16.12 (w,f) A nonstationary random process is defined as $X[n] = a^{|n|}U[n]$, where $0 < a < 1$ and $U[n]$ is WGN with variance σ_U^2. Find the mean and covariance sequences of $X[n]$. Can you transform the $X[n]$ random process to make it stationary?

16.13 (☺) (w) Consider the random process $X[n] = \sum_{i=0}^{n} U[i]$, which is defined for $n \geq 0$. The $U[n]$ random process consists of independent Gaussian random variables with marginal PDF $U[n] \sim \mathcal{N}(0, (1/2)^n)$. Are the increments independent? Are the increments stationary?

16.14 (c) Plot 50 realizations of a WGN random process $X[n]$ with $\sigma^2 = 1$ for $n = 0, 1, \ldots, 49$ using a scatter diagram (see Figure 16.15 for an example). Use the MATLAB commands `plot(x,y,'.')` and `hold on` to plot each realization as dots and to overlay the realizations on the same graph, respectively. For a fixed n can you explain the observed distribution of the dots?

16.15 (f) Prove that

$$\frac{1}{(2\pi)^{N/2}\det^{1/2}(\mathbf{C})}\exp\left(-\tfrac{1}{2}\mathbf{x}^T\mathbf{C}^{-1}\mathbf{x}\right)$$

where $\mathbf{x} = [x_1\, x_2 \ldots x_N]^T$ and $\mathbf{C} = \sigma^2\mathbf{I}$ for $\mathbf{I}$ an $N \times N$ identity matrix, reduces to (16.4).

16.16 (☺) (f) A "white" uniform random process is defined to be an IID random process with $X[n] \sim \mathcal{U}(-\sqrt{3}, \sqrt{3})$ for all n. Determine the mean and covariance sequences for this random process and compare them to those of the WGN random process. Explain your results.

16.17 (w) A moving average random process can be defined more generally as one for which N samples of WGN are averaged, instead of only $N = 2$ samples as in Example 16.7. It is given by $X[n] = (1/N)\sum_{i=0}^{N-1} U[n-i]$ for all n, where $U[n]$ is a WGN random process with variance σ_U^2. Determine the correlation coefficient for $X[0]$ and $X[1]$. What happens as N increases?

16.18 (☺) (f) For the moving average random process defined in Example 16.7 determine $P[X[n] > 3]$ and compare it to $P[U[n] > 3]$. Explain the difference in terms of "smoothing".

16.19 (c) For the randomly phased sinusoid defined in Example 16.8 determine the mean sequence using a computer simulation.

16.20 (t) For the randomly phased sinusoid of Example 16.8 assume that the realization $x[n] = \cos(2\pi(0.1)n + 0)$ is generated. Prove that if we observe only the samples $x[0] = 1$ and $x[1] = \cos(2\pi(0.1)) = 0.8090$, then all the future samples can be found by using the recursive formula $x[n] = 2\cos(2\pi(0.1))x[n-1] - x[n-2]$ for $n \geq 2$. Could you also find the past samples or $x[n]$ for $n \leq -1$? See also Problem 18.25 for prediction of a sinusoidal random process.

16.21 (c) Verify the PDF of the randomly phased sinusoid given in Figure 16.12 by using a computer simulation.

16.22 (☺) (f,c) A continuous-time random process known as the random *amplitude* sinusoid is defined as $X(t) = A\cos(2\pi t)$ for $-\infty < t < \infty$ and $A \sim \mathcal{N}(0,1)$. Find the mean and covariance functions. Then, plot some realizations of $X(t)$ in an overlaid fashion.

16.23 (f) A random process is the sum of WGN and a deterministic sinusoid and is given as $X[n] = U[n] + \sin(2\pi f_0 n)$ for all n, where $U[n]$ is WGN with variance σ_U^2. Determine the mean and covariance sequences.

16.24 (‿) (w) A random process is IID with samples $X[n] \sim \mathcal{N}(\mu, 1)$. It is desired to remove the mean of the random process by forming the new random process $Y[n] = X[n] - X[n-1]$. First determine the mean sequence of $Y[n]$. Next find $\text{cov}(Y[0], Y[1])$. Is $Y[n]$ an IID random process with a zero mean sequence?

16.25 (f) If a random process is defined as $X[n] = h[0]U[n] + h[1]U[n-1]$, where $h[0]$ and $h[1]$ are constants and $U[n]$ is WGN with variance σ_U^2, find the covariance for $X[0]$ and $X[1]$. Repeat for $X[9]$ and $X[10]$. How do they compare?

16.26 (‿) (f) If a sum random process is defined as $X[n] = \sum_{i=0}^{n} U[i]$ for $n \geq 0$, where $E[U[i]] = 0$ and $\text{var}(U[i]) = \sigma_U^2$ for $i \geq 0$ and the $U[i]$ are IID, find the mean and covariance sequences of $X[n]$.

16.27 (‿) (c) For the MA random process defined in Example 16.7 find $c_X[1,1]$, $c_X[1,2]$ and $c_X[1,3]$ if $\sigma_U^2 = 1$. Next simulate on a computer $M = 10,000$ realizations of the random process $X[n]$ for $n = 0, 1, \ldots, 10$. Estimate the previous covariance sequence samples using $\hat{c}_X[n_1, n_2] = (1/M) \sum_{i=1}^{M} x_i[n_1] x_i[n_2]$, where $x_i[n]$ is the ith realization of $X[n]$. Note that since $X[n]$ is zero mean, $c_X[n_1, n_2] = E[X[n_1]X[n_2]]$.

16.28 (w) For the randomly phased sinusoid described in Example 16.11 determine the minimum mean square estimate of $X[10]$ based on observing $x[0]$. How accurate do you think this prediction will be?

16.29 (f) For a random process $X[n]$ the mean sequence $\mu_X[n]$ and covariance sequence $c_X[n_1, n_2]$ are known. It is desired to predict k samples into the future. If $x[n_0]$ is observed, find the minimum mean square estimate of $X[n_0 + k]$. Next assume that $\mu_X[n] = \cos(2\pi f_0 n)$ and $c_X[n_1, n_2] = 0.9^{|n_2 - n_1|}$ and evaluate the estimate. Finally, what happens to your prediction as $k \to \infty$ and why?

16.30 (f) A random process is defined as $X[n] = As[n]$ for all n, where $A \sim \mathcal{N}(0, 1)$ and $s[n]$ is a deterministic signal. Find the mean and covariance sequences.

16.31 (‿) (f) A random process is defined as $X[n] = AU[n]$ for all n, where $A \sim \mathcal{N}(0, \sigma_A^2)$ and $U[n]$ is WGN with variance σ_U^2, and A is independent of $U[n]$ for all n. Find the mean and covariance sequences. What type of random process is $X[n]$?

16.32 (f) Verify that by differentiating $\sum_{n=0}^{N-1} (x[n] - b)^2$ with respect to b, setting the derivative equal to zero, and solving for b, we obtain the sample mean.

16.33 (t) In this problem we show how to obtain the variance of $\hat{a}$ as obtained by solving (16.9). The variance of $\hat{a}$ is derived under the assumption that $X[n] = b + U[n]$, where $U[n]$ is WGN with variance σ^2. This says that we assume the true value of a is zero. The steps are as follows:

a. Let

$$\mathbf{H} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ \vdots & \vdots \\ 1 & N-1 \end{bmatrix} \qquad \mathbf{X} = \begin{bmatrix} X[0] \\ X[1] \\ X[2] \\ \vdots \\ X[N-1] \end{bmatrix}$$

where $\mathbf{H}$ is an $N \times 2$ matrix and $\mathbf{X}$ is an $N \times 1$ random vector. Now show that that the equations of (16.9) can be written as

$$\mathbf{H}^T\mathbf{H} \begin{bmatrix} b \\ a \end{bmatrix} = \mathbf{H}^T\mathbf{X}.$$

b. The solution for b and a can now be written symbolically as

$$\begin{bmatrix} \hat{b} \\ \hat{a} \end{bmatrix} = \underbrace{(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T}_{\mathbf{G}}\mathbf{X}$$

Since $\mathbf{X}$ is a Gaussian random vector, show that $[\hat{b}\,\hat{a}]^T$ is also a Gaussian random vector with mean $[b\,0]^T$ and covariance matrix $\sigma^2(\mathbf{H}^T\mathbf{H})^{-1}$.

c. As a result we can assert that the marginal PDF of $\hat{a}$ is Gaussian with mean zero and variance equal to the $(2,2)$ element of $\sigma^2(\mathbf{H}^T\mathbf{H})^{-1}$. Show then that $\hat{a} \sim \mathcal{N}(0, \text{var}(\hat{a}))$, where

$$\text{var}(\hat{a}) = \frac{\sigma^2}{\sum_{n=0}^{N-1} n^2 - \frac{1}{N}(\sum_{n=0}^{N-1} n)^2}.$$

Next assume that $\sigma^2 = 10.05$, $N = 108$ and find the probability that $\hat{a} > 0.0173$. Can we assert that the estimated mean sequence shown in Figure 16.16 is not just due to estimation error?

16.34 (☺) (f) Using the results of Problem 16.33 determine the required value of N so that the probability that $\hat{a} > 0.0173$ is less than 10^{-6}.

Chapter 17

Wide Sense Stationary Random Processes

17.1 Introduction

Having introduced the concept of a random process in the previous chapter, we now wish to explore an important subclass of stationary random processes. This is motivated by the very restrictive nature of the stationarity condition, which although mathematically expedient, is almost never satisfied in practice. A somewhat weaker type of stationarity is based on requiring the mean to be a constant in time and the covariance sequence to depend only on the separation in time between the two samples. We have already encountered these types of random processes in Examples 16.9–16.11. Such a random process is said to be stationary in the wide sense or *wide sense stationary* (WSS). It is also termed a *weakly stationary* random process to distinguish it from a stationary process, which is said to be *strictly stationary.* We will use the former terminology to refer to such a process as a WSS random process. In addition, as we will see in Chapter 19, if the random process is Gaussian, then wide sense stationarity implies stationarity. For this reason alone, it makes sense to explore WSS random processes since the use of Gaussian random processes for modeling is ubiquitous.

Once we have discussed the concept of a WSS random process, we will be able to define an extremely important measure of the WSS random process—the *power spectral density* (PSD). This function extends the idea of analyzing the behavior of a deterministic signal by decomposing it into a sum of sinusoids of different frequencies to that of a random process. The difference now is that the amplitudes and phases of the sinusoids will be random variables and so it will be convenient to quantify the *average power* of the various sinusoids. This description of a random phenomenon is important in nearly every scientific field that is concerned with the analysis of time series data such as systems control [Box and Jenkins 1970], signal processing [Schwartz and Shaw 1975], economics [Harvey 1989], geophysics [Robinson 1967],

vibration testing [McConnell 1995], financial analysis [Taylor 1986], and others. As an example, in Figure 17.1 the Wolfer sunspot data [Tong 1990] is shown, with the data points connected by straight lines for easier viewing. It measures the average number of sunspots visually observed through a telescope each year. The importance of the sunspot number is that as it increases, an increase in solar flares occurs. This has the effect of disrupting all radio communications as the solar flare particles reach the earth. Clearly from the data we see a periodic type property. The estimated PSD of this data set is shown in Figure 17.2. We see that the distribution of power versus frequency is highest at a frequency of about 0.09 cycles per year. This means that the random process exhibits a large periodic component with a period of about $1/0.09 \approx 11$ years per cycle, as is also evident from Figure 17.1. This is a powerful prediction tool and therefore is of great interest. How the PSD is actually estimated will be discussed in this chapter, but before doing so, we will need to lay some groundwork.

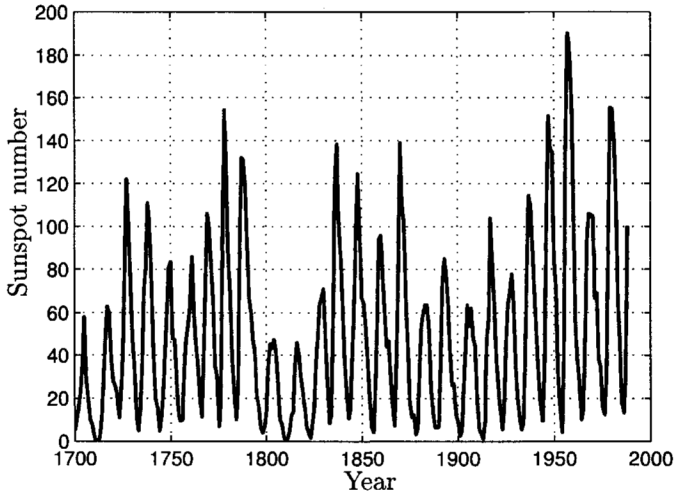


Figure 17.1: Annual number of sunspots – Wolfer sunspot data.

17.2 Summary

A less restrictive form of stationarity, termed wide sense stationarity, is defined by (17.4) and (17.5). The conditions require the mean to be the same for all n and the covariance sequence to depend only on the time difference between the samples. A random process that is stationary is also wide sense stationary as shown in Section 17.3. The autocorrelation sequence is defined by (17.9) with n being arbitrary. It is the covariance between two samples separated by k units for a zero mean WSS random process. Some of its properties are summarized by Properties 17.1–17.4. Under certain conditions the mean of a WSS random process can be found by using

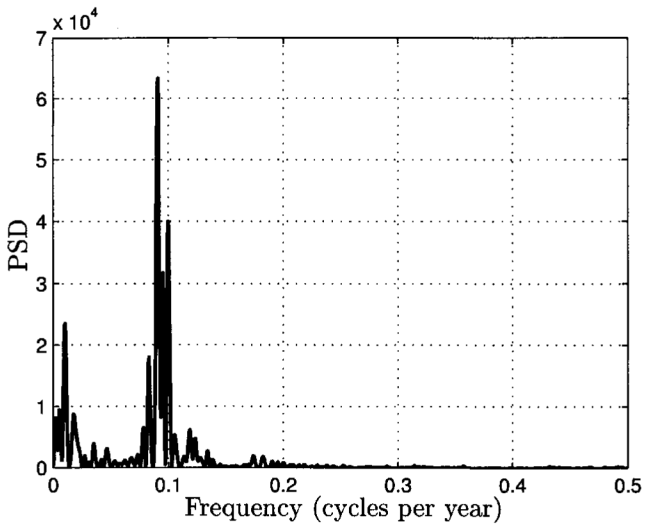


Figure 17.2: Estimated power spectral density for Wolfer sunspot data of Figure 17.1. The sample mean has been computed and removed from the data prior to estimation of the PSD.

the temporal average of (17.25). Such a process is said to be ergodic in the mean. For this to be true the variance of the temporal average given by (17.28) must converge to zero as the number of samples averaged becomes large. The power spectral density (PSD) of a WSS random process is defined by (17.30) and can be evaluated more simply using (17.34). The latter relationship says that the PSD is the Fourier transform of the autocorrelation sequence. It measures the amount of average power per unit frequency or the distribution of average power with frequency. Some of its properties are summarized in Properties 17.7–17.12. From a finite segment of a realization of the random process the autocorrelation sequence can be estimated using (17.43) and the PSD can be estimated by using the averaged periodogram estimate of (17.44) and (17.45). The analogous definitions for a continuous-time WSS random process are given in Section 17.8. Also, an important example is described that relates sampled continuous-time white Gaussian noise to discrete-time white Gaussian noise. Finally, an application of the use of PSDs to random vibration testing is given in Section 17.9.

17.3 Definition of WSS Random Process

Consider a discrete-time random process $X[n]$, which is defined for $-\infty < n < \infty$ with n an integer. Previously, we defined the mean and covariance sequences of

$X[n]$ to be

$$\mu_X[n] = E[X[n]] \quad -\infty < n < \infty \tag{17.1}$$

$$c_X[n_1, n_2] = E[(X[n_1] - \mu_X[n_1])(X[n_2] - \mu_X[n_2])] \quad \begin{matrix} -\infty < n_1 < \infty \\ -\infty < n_2 < \infty \end{matrix} \tag{17.2}$$

where n_1, n_2 are integers. Having knowledge of these sequences allows us to assess important characteristics of the random process such as the mean level and the correlation between samples. In fact, based on only this information we are able to predict $X[n_2]$ based on observing $X[n_1] = x[n_1]$ as

$$\hat{X}[n_2] = \mu_X[n_2] + \frac{c_X[n_1, n_2]}{c_X[n_1, n_1]}(x[n_1] - \mu_X[n_1]) \tag{17.3}$$

which is just the usual linear prediction formula of (7.41) with x replaced by $x[n_1]$ and Y replaced by $X[n_2]$, and which makes use of the mean and covariance sequences defined in (17.1) and (17.2), respectively. However, since in general the mean and covariance change with time, i.e., they are nonstationary, it would be exceedingly difficult to estimate them in practice. To extend the practical utility we would like the mean not to depend on time and the covariance only to depend on the separation between samples or on $|n_2 - n_1|$. This will allow us to estimate these quantities as described later. Thus, we are led to a weaker form of stationarity known as *wide sense stationarity.* A random process is defined to be WSS if

$$\mu_X[n] = \mu \quad \text{(a constant)} \qquad -\infty < n < \infty \tag{17.4}$$

$$c_X[n_1, n_2] = g(|n_2 - n_1|) \qquad -\infty < n_1 < \infty, -\infty < n_2 < \infty \tag{17.5}$$

for some function g. Note that since

$$c_X[n_1, n_2] = E[X[n_1]X[n_2]] - E[X[n_1]]E[X[n_2]]$$

these conditions are equivalent to requiring that $X[n]$ satisfy

$$E[X[n]] = \mu \qquad -\infty < n < \infty$$

$$E[X[n_1]X[n_2]] = h(|n_2 - n_1|) \qquad -\infty < n_1 < \infty, -\infty < n_2 < \infty$$

for some function h. *The mean should not depend on time and the average value of the product of two samples should depend only upon the time interval between the samples.* Some examples of WSS random processes have already been given in Examples 16.9–16.11. For the MA process of Example 16.10 we showed that

$$\mu_X[n] = 0 \quad -\infty < n < \infty$$

$$c_X[n_1, n_2] = \begin{cases} \frac{1}{2}\sigma_U^2 & |n_2 - n_1| = 0 \\ \frac{1}{4}\sigma_U^2 & |n_2 - n_1| = 1 \\ 0 & |n_2 - n_1| > 1. \end{cases}$$

It is seen that every random variable $X[n]$ for $-\infty < n < \infty$ has a mean of zero and the covariance for two samples depends only on the time interval between the samples, which is $|n_2 - n_1|$. Also, this implies that the variance does not depend on time since $\text{var}(X[n]) = c_X[n,n] = \sigma_U^2/2$ for $-\infty < n < \infty$. In contrast to this behavior consider the random processes for which typical realizations are shown in Figure 16.7. In Figure 16.7a the mean changes with time (with the variance being constant) and in Figure 16.7b the variance changes with time (with the mean being constant). Clearly, these random processes are not WSS.

A WSS random process is a special case of a stationary random process. To see this recall that if $X[n]$ is stationary, then from (16.3) with $N = 1$ and $n_1 = n$, we have

$$p_{X[n+n_0]} = p_{X[n]} \qquad \text{for all } n \text{ and for all } n_0.$$

As a consequence, if we let $n = 0$, then

$$p_{X[n_0]} = p_{X[0]} \qquad \text{for all } n_0$$

and since the PDF does not depend on the particular time n_0, the mean must not depend on time. Thus,

$$\mu_X[n] = \mu \qquad -\infty < n < \infty. \tag{17.6}$$

Next, using (16.3) with $N = 2$, we have

$$p_{X[n_1+n_0],X[n_2+n_0]} = p_{X[n_1],X[n_2]} \qquad \text{for all } n_1, n_2 \text{ and } n_0. \tag{17.7}$$

Now if $n_0 = -n_1$ we have from (17.7)

$$p_{X[0],X[n_2-n_1]} = p_{X[n_1],X[n_2]}$$

and if $n_0 = -n_2$, we have

$$p_{X[n_1-n_2],X[0]} = p_{X[n_1],X[n_2]}.$$

This results in

$$\begin{aligned} p_{X[n_1],X[n_2]} &= p_{X[0],X[n_2-n_1]} \\ p_{X[n_1],X[n_2]} &= p_{X[n_1-n_2],X[0]} \end{aligned}$$

which leads to

$$\begin{aligned} E[X[n_1]X[n_2]] &= E[X[0]X[n_2 - n_1]] \\ E[X[n_1]X[n_2]] &= E[X[n_1 - n_2]X[0]] = E[X[0]X[n_1 - n_2]]. \end{aligned}$$

Finally, these two conditions combine to give

$$E[X[n_1]X[n_2]] = E[X[0]X[|n_2 - n_1|]] \tag{17.8}$$

which along with the mean being constant with time yields the second condition for wide sense stationarity of (17.5) that

$$c_X[n_1, n_2] = E[X[n_1]X[n_2]] - E[X[n_1]E[X[n_2]] = E[X[0]X[|n_2 - n_1|] - \mu^2.$$

This proves the assertion that a stationary random process is WSS but the converse is not generally true (see Problem 17.5).

17.4 Autocorrelation Sequence

If $X[n]$ is WSS, then as we have seen $E[X[n_1]X[n_2]]$ depends only on the separation in time between the samples. We can therefore define a new joint moment by letting $n_1 = n$ and $n_2 = n + k$ to yield

$$r_X[k] = E[X[n]X[n + k]] \tag{17.9}$$

which is called the *autocorrelation sequence* (ACS). It depends only on the time difference between samples which is $|n_2 - n_1| = |(n + k) - n| = |k|$ so that *the value of n used in the definition is arbitrary.* It is termed the *auto*correlation sequence (ACS) since it measures the correlation between two samples of the *same* random process. Later we will have occasion to define correlation between two different random processes (see Section 19.3). Note that the time interval between samples is also called the *lag.* An example of the computation of the ACS is given next.

Example 17.1 – A Differencer

Define a random process as $X[n] = U[n] - U[n - 1]$, where $U[n]$ is an IID random process with mean μ and variance σ_U^2. A realization of this random process for which $U[n]$ is a Gaussian random variable for all n is shown in Figure 17.3. Although $U[n]$ was chosen here to be a sequence of Gaussian random variables for the sake of displaying the realization in Figure 17.3, the ACS to be found will be the same regardless of the PDF of $U[n]$. This is because it relies on *only the first two moments of $U[n]$ and not its PDF.* The ACS is found as

$$\begin{aligned} r_X[k] &= E[X[n]X[n + k]] \\ &= E[(U[n] - U[n - 1])(U[n + k] - U[n + k - 1])] \\ &= E[U[n]U[n + k]] - E[U[n]U[n + k - 1]] \\ &\quad - E[U[n - 1]U[n + k]] + E[U[n - 1]U[n + k - 1]]. \end{aligned}$$

But for $n_1 \neq n_2$

$$\begin{aligned} E[U[n_1]U[n_2]] &= E[U[n_1]]E[U[n_2]] \quad \text{(independence)} \\ &= \mu^2 \end{aligned}$$

and for $n_1 = n_2 = n$

$$E[U[n_1]U[n_2]] = E[U^2[n]] = E[U^2[0]] = \sigma_U^2 + \mu^2 \quad \text{(identically distributed).}$$

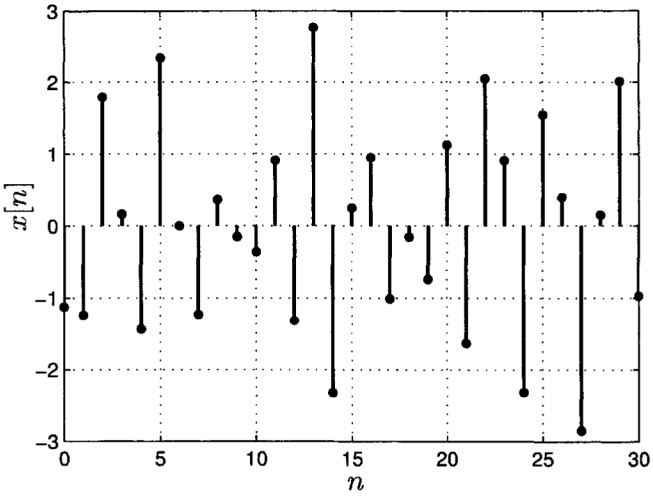


Figure 17.3: Typical realization of a differenced IID Gaussian random process with $U[n] \sim \mathcal{N}(1,1)$.

Combining these results we have that

$$E[U[n_1]U[n_2]] = \mu^2 + \sigma_U^2 \delta[n_2 - n_1]$$

and therefore the ACS becomes

$$r_X[k] = 2\sigma_U^2 \delta[k] - \sigma_U^2 \delta[k-1] - \sigma_U^2 \delta[k+1]. \tag{17.10}$$

This is shown in Figure 17.4. Several observations can be made. The only nonzero correlation is between adjacent samples and this correlation is negative. This accounts for the observation that the realization shown in Figure 17.3 exhibits many adjacent samples that are opposite in sign. Some other observations are that $r_X[0] > 0$, $|r_X[k]| \leq r_X[0]$ for all k, and finally $r_X[-k] = r_X[k]$. In words, the ACS has a maximum at $k = 0$, which is positive, and is a symmetric sequence about $k = 0$ (also called an *even sequence*). These properties hold in general as we now prove.

$\diamondsuit$

Property 17.1 – ACS is positive for the zero lag or $r_X[0] > 0$.
Proof:

$$r_X[k] = E[X[n]X[n+k]] \qquad \text{(definition)}$$

so that with $k = 0$ we have $r_X[0] = E[X^2[n]] > 0$.

$\square$

Note that $r_X[0]$ is the *average power* of the random process at all sample times n. One can view $X[n]$ as the voltage across a 1 ohm resistor and hence $x^2[n]/1$

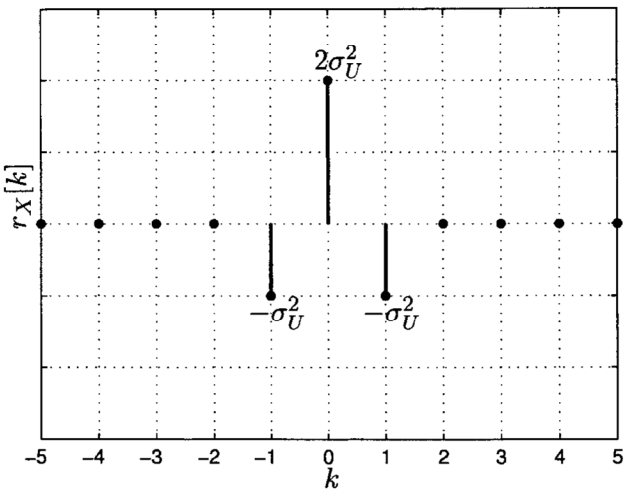


Figure 17.4: Autocorrelation sequence for differenced random process.

is the power for any particular realization of $X[n]$ at time n. The *average* power $E[X^2[n]] = r_X[0]$ does not change with time.

Property 17.2 – ACS is an even sequence or $r_X[-k] = r_X[k]$.

Proof:

$$\begin{aligned} r_X[k] &= E[X[n]X[n+k]] \qquad \text{(definition)} \\ r_X[-k] &= E[X[n]X[n-k]] \end{aligned}$$

and letting $n = m + k$ since the choice of n in the definition of the ACS is arbitrary, we have

$$\begin{aligned} r_X[-k] &= E[X[m+k]X[m]] \\ &= E[X[m]X[m+k]] \\ &= E[X[n]X[n+k]] \qquad \text{(ACS not dependent on } n\text{)} \\ &= r_X[k]. \end{aligned}$$

□

Property 17.3 – Maximum value of ACS is at $k = 0$ or $|r_X[k]| \leq r_X[0]$.

Note that it is possible for some values of $r_X[k]$ for $k \neq 0$ to also equal $r_X[0]$. As an example, for the randomly phased sinusoid of Example 16.11 we had $c_X[n_1, n_2] = \frac{1}{2}\cos[2\pi(0.1)(n_2 - n_1)]$ with a mean of zero. Thus, $r_X[k] = \frac{1}{2}\cos[2\pi(0.1)k]$ and therefore $r_X[10] = r_X[0]$. Hence, the property says that no value of the ACS can *exceed* $r_X[0]$, although there may be multiple values of the ACS that are equal to $r_X[0]$.

Proof: The proof is based on the Cauchy-Schwarz inequality, which from Appendix 7A is

$$|E_{V,W}[VW]| \leq \sqrt{E_V[V^2]}\sqrt{E_W[W^2]}$$

with equality holding if and only if $W = cV$ for c a constant. Letting $V = X[n]$ and $W = X[n+k]$, we have

$$|E[X[n]X[n+k]]| \leq \sqrt{E[X^2[n]]}\sqrt{E[X^2[n+k]]}$$

from which it follows that

$$|r_X[k]| \leq \sqrt{r_X[0]}\sqrt{r_X[0]} = |r_X[0]| = r_X[0] \qquad (\text{since } r_X[0] > 0).$$

Note that equality holds if and only if $X[n+k] = cX[n]$ for all n. This implies perfect predictability of a sample based on the realization of another sample spaced k units ahead or behind in time (see Problem 17.10 for an example involving periodic random processes).

□

Property 17.4 – ACS measures the predictability of a random process. The correlation coefficient for two samples of a zero mean WSS random process is

$$\rho_{X[n],X[n+k]} = \frac{r_X[k]}{r_X[0]} \tag{17.11}$$

For a nonzero mean the expression is easily modified (see Problem 17.11). Proof: Recall that the correlation coefficient for two random variables V and W is defined as

$$\rho_{V,W} = \frac{\text{cov}(V,W)}{\sqrt{\text{var}(V)\text{var}(W)}}.$$

Assuming that V and W are zero mean, this becomes

$$\rho_{V,W} = \frac{E_{V,W}[VW]}{\sqrt{E_V[V^2]E_W[W^2]}}$$

and letting $V = X[n]$ and $W = X[n+k]$, we have

$$\begin{aligned}
\rho_{X[n],X[n+k]} &= \frac{E[X[n]X[n+k]]}{\sqrt{E[X^2[n]]E[X^2[n+k]]}} \\
&= \frac{r_X[k]}{\sqrt{r_X[0]r_X[0]}} \\
&= \frac{r_X[k]}{|r_X[0]|} \\
&= \frac{r_X[k]}{r_X[0]} \qquad (\text{from Property 17.1}).
\end{aligned}$$

□

As an example, for the differencer of Example 17.1 we have from Figure 17.4

$$\rho_{X[n],X[n+k]} = \begin{cases} 1 & k = 0 \\ -\frac{1}{2} & k = \pm 1 \\ 0 & \text{otherwise.} \end{cases}$$

As mentioned previously, the adjacent samples are negatively correlated and the magnitude of the correlation coefficient is now seen to be 1/2.

We next give some more examples of the computation of the ACS.

Example 17.2 – White noise

White noise is defined as *a WSS random process with zero mean, identical variance* σ^2*, and uncorrelated samples.* It is a more general case of the white noise random process first described in Example 16.9. There we assumed the stronger condition of zero mean *IID* samples (hence they must have the same variance due to the identically distributed assumption and also be uncorrelated due to the independence assumption). In addition, it was assumed there that each sample had a *Gaussian PDF*. Note, however, that the definition given above for white noise does not specify a particular PDF. To find the ACS we note that from the definition of the white noise random process

$$\begin{aligned} r_X[k] &= E[X[n]X[n+k]] \\ &= E[X[n]]E[X[n+k]] = 0 \quad k \neq 0 \qquad \text{(uncorrelated and zero mean samples)} \\ &= E[X^2[n]] = \sigma^2 \quad k = 0 \qquad \text{(equal variance samples).} \end{aligned}$$

Therefore, we have that

$$r_X[k] = \sigma^2 \delta[k]. \tag{17.12}$$

Could you predict $X[1]$ from a realization of $X[0]$?

$\Diamond$

As an aside, for WSS random processes, we can find the covariance sequence from the ACS and the mean since

$$\begin{aligned} c_X[n_1, n_2] &= E[X[n_1]X[n_2]] - \mu_X[n_1]\mu_X[n_2] \\ &= r_X[n_2 - n_1] - \mu^2. \end{aligned} \tag{17.13}$$

Another property of the ACS that is evident from (17.13) concerns the behavior of the ACS as $k \to \infty$. Letting $n_1 = n$ and $n_2 = n + k$, we have that

$$r_X[k] = c_X[n, n+k] + \mu^2. \tag{17.14}$$

If two samples becomes uncorrelated or $c_X[n, n+k] \to 0$ as $k \to \infty$, then we see that $r_X[k] \to \mu^2$ as $k \to \infty$. Thus, as another property of the ACS we have the following.

Property 17.5 – ACS approaches μ^2 as $k \to \infty$
This assumes that the samples become uncorrelated for large lags, which is usually the case.

□

If the mean is zero, then from (17.14)

$$r_X[k] = c_X[n, n+k] \tag{17.15}$$

and the ACS approaches zero as the lag increases. We continue with some more examples.

◊

Example 17.3 – MA random process
This random process was shown in Example 16.10 to have a zero mean and a covariance sequence

$$c_X[n_1, n_2] = \begin{cases} \frac{\sigma_U^2}{2} & n_1 = n_2 \\ \frac{\sigma_U^2}{4} & |n_2 - n_1| = 1 \\ 0 & \text{otherwise.} \end{cases} \tag{17.16}$$

Since the covariance sequence depends only on $|n_2 - n_1|$, $X[n]$ is WSS from (17.15). Specifically, the ACS follows from (17.15) and (17.16) with $k = n_2 - n_1$ as

$$r_X[k] = \begin{cases} \frac{\sigma_U^2}{2} & k = 0 \\ \frac{\sigma_U^2}{4} & k = \pm 1 \\ 0 & \text{otherwise.} \end{cases}$$

See Figure 16.13 for a plot of the ACS (replace Δn with k.) Could you predict $X[1]$ from a realization of $X[0]$?

◊

Example 17.4 – Randomly phased sinusoid
This random process was shown in Example 16.11 to have a zero mean and a covariance sequence $c_X[n_1, n_2] = \frac{1}{2}\cos[2\pi(0.1)(n_2 - n_1)]$. Since the covariance sequence depends only on $|n_2 - n_1|$, $X[n]$ is WSS. Hence, from (17.15) we have that

$$r_X[k] = \frac{1}{2}\cos[2\pi(0.1)k].$$

See Figure 16.14 for a plot of the ACS (replace Δn with k.) Could you predict $X[1]$ from a realization of $X[0]$?

◊

In determining predictability of a WSS random process, it is convenient to consider the linear predictor, which depends only on the first two moments. Then, the MMSE linear prediction of $X[n_0+k]$ given $x[n_0]$ is from (17.3) and (17.13) with $n_1 = n_0$ and $n_2 = n_0 + k$

$$\hat{X}[n_0+k] = \mu + \frac{r_X[k] - \mu^2}{r_X[0] - \mu^2}(x[n_0] - \mu) \quad \text{for all } k \text{ and } n_0.$$

For a zero mean random process this becomes

$$\begin{aligned} \hat{X}[n_0+k] &= \frac{r_X[k]}{r_X[0]} x[n_0] \\ &= \rho_{X[n_0],X[n_0+k]} x[n_0] \quad \text{for all } k \text{ and } n_0. \end{aligned}$$

One last example is the autoregressive random process which we will use to illustrate several new concepts for WSS random processes.

Example 17.5 – Autoregressive random process

An autoregressive (AR) random process $X[n]$ is defined to be a WSS random process with a zero mean that evolves according to the recursive difference equation

$$X[n] = aX[n-1] + U[n] \qquad -\infty < n < \infty \tag{17.17}$$

where $|a| < 1$ and $U[n]$ is WGN. The WGN random process $U[n]$ (see Example 16.6), has a zero mean and variance σ_U^2 for all n and its samples are all independent with a Gaussian PDF. The name *autoregressive* is due to the *regression* of $X[n]$ upon $X[n-1]$, which is another sample of the same random process, hence, the prefix *auto*. The evolution of $X[n]$ proceeds, for example, as

$$\begin{aligned} &\vdots \\ X[0] &= aX[-1] + U[0] \\ X[1] &= aX[0] + U[1] \\ X[2] &= aX[1] + U[2] \\ &\vdots \end{aligned}$$

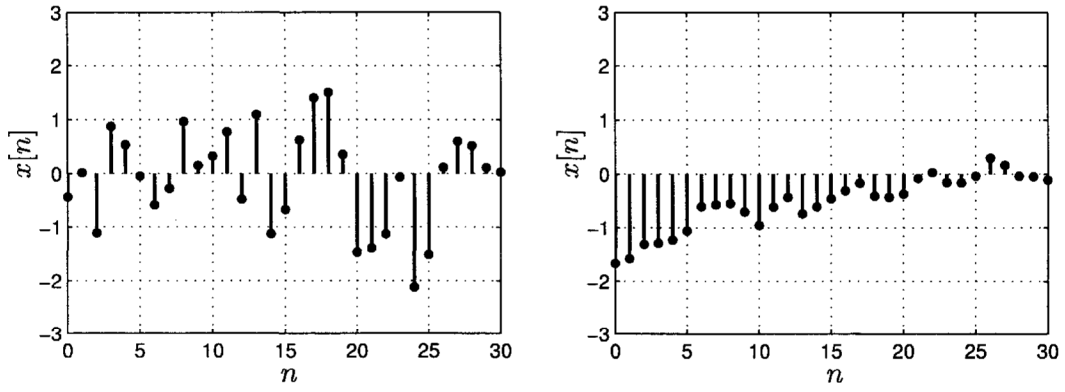
Note that $X[n]$ depends only upon the present and past values of $U[n]$ since for example

$$\begin{aligned} X[2] &= aX[1] + U[2] = a(aX[0] + U[1]) + U[2] = a^2X[0] + aU[1] + U[2] \\ &= a^2(aX[-1] + U[0]) + aU[1] + U[2] = a^3X[-1] + a^2U[0] + aU[1] + U[2] \\ &\vdots \\ &= \sum_{k=0}^{\infty} a^k U[2-k] \end{aligned} \tag{17.18}$$

where the term involving $a^k U[2-k]$ decays to zero as $k \to \infty$ since $|a| < 1$. We see that $X[2]$ depends only on $\{U[2], U[1], \ldots\}$ and it is therefore uncorrelated with $\{U[3], U[4], \ldots\}$. More generally, it can be shown that (see also Problem 19.6)

$$E[X[n]U[n+k]] = 0 \qquad k \geq 1. \tag{17.19}$$

It is seen from (17.18) that in order for the recursion to be stable and hence $X[n]$ to be WSS it is required that $|a| < 1$. The AR random process can be used to model a wide variety of physical random processes with various ACSs, depending upon the choice of the parameters a and σ_U^2. Some typical realizations of the AR random process for different values of a are shown in Figure 17.5. The WGN random process $U[n]$ has been chosen to have a variance $\sigma_U^2 = 1 - a^2$. We will soon see that this choice of variance results in $r_X[0] = 1$ for both AR processes shown in Figure 17.5. The MATLAB code used to generate the realizations shown is given below.



(a) $a = 0.25$, $\sigma_U^2 = 1 - a^2$ (b) $a = 0.98$, $\sigma_U^2 = 1 - a^2$

Figure 17.5: Typical realizations of autoregressive random process with different parameters.

```
clear all
randn('state',0)
a1=0.25;a2=0.98;
varu1=1-a1^2;varu2=1-a2^2;
varx1=varu1/(1-a1^2);varx2=varu2/(1-a2^2); % this is r_X[0]
x1(1,1)=sqrt(varx1)*randn(1,1); % set initial condition X[-1]
                                % see Problems 17.17, 17.18
x2(1,1)=sqrt(varx2)*randn(1,1);
for n=2:31
   x1(n,1)=a1*x1(n-1)+sqrt(varu1)*randn(1,1);
   x2(n,1)=a2*x2(n-1)+sqrt(varu2)*randn(1,1);
end
```

We next derive the ACS. In Chapter 18 we will see how to alternatively obtain the ACS using results from linear systems theory. Using (17.17) we have for $k \geq 1$

$$\begin{aligned} r_X[k] &= E[X[n]X[n+k]] \\ &= E[X[n](aX[n+k-1]+U[n+k])] \\ &= aE[X[n]X[n+k-1]] \qquad \text{(using (17.19))} \\ &= ar_X[k-1]. \end{aligned} \tag{17.20}$$

The solution of this recursive linear difference equation is readily seen to be $r_X[k] = ca^k$, for c any constant and for $k \geq 1$. For $k = 1$ we have that $r_X[1] = ca$ and so from (17.20) $r_X[1] = ar_X[0]$, which implies $c = r_X[0]$. In Problem 17.15 it is shown that

$$r_X[0] = \frac{\sigma_U^2}{1-a^2}$$

so that for all $k \geq 0$, $r_X[k] = r_X[0]a^k$ becomes

$$r_X[k] = \frac{\sigma_U^2}{1-a^2}a^k.$$

Finally, noting that $r_X[-k] = r_X[k]$ from Property 17.2, we obtain the ACS as

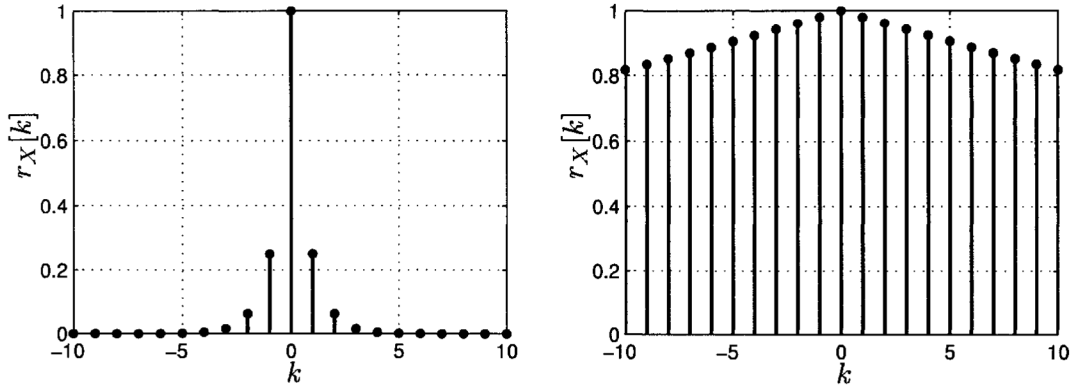
$$r_X[k] = \frac{\sigma_U^2}{1-a^2}a^{|k|} \qquad -\infty < k < \infty. \tag{17.21}$$

(See also Problem 17.16 for an alternative derivation of the ACS.) The ACS is plotted in Figure 17.6 for $a = 0.25$ and $a = 0.98$ and $\sigma_U^2 = 1-a^2$. For both values of a the value of σ_U^2 has been chosen to ensure that $r_X[0] = 1$. Note that for $a = 0.25$ the ACS dies off very rapidly which means that the random process samples quickly become uncorrelated as the separation between them increases. This is consistent with the typical realization shown in Figure 17.5a. For $a = 0.98$ the ACS decays very slowly, indicating a strong positive correlation between samples, and again being consistent with the typical realization shown in Figure 17.5b. In either case the samples become uncorrelated as $k \to \infty$ since $|a| < 1$ and therefore, $r_X[k] \to 0$ as $k \to \infty$ in accordance with Property 17.5. However, the random process with the slower decaying ACS is more predictable.

$\diamondsuit$

One last property that is necessary for a sequence to be a valid ACS is the property of positive definiteness. As its name implies, it is related to the positive definite property of the covariance matrix. As an example, consider the random vector $\mathbf{X} = [X[0]\, X[1]]^T$. Then we know from the proof of Property 9.2 (covariance matrix is positive semidefinite) that if $Y = a_0X[0] + a_1X[1]$ cannot be made equal to a constant by any choice of a_0 and a_1, then

$$\text{var}(Y) = \underbrace{\begin{bmatrix} a_0 & a_1 \end{bmatrix}}_{\mathbf{a}^T} \underbrace{\begin{bmatrix} \text{cov}(X[0],X[0]) & \text{cov}(X[0],X[1]) \\ \text{cov}(X[1],X[0]) & \text{cov}(X[1],X[1]) \end{bmatrix}}_{\mathbf{C}_X} \underbrace{\begin{bmatrix} a_0 \\ a_1 \end{bmatrix}}_{\mathbf{a}} > 0.$$



(a) $a = 0.25$, $\sigma_U^2 = 1 - a^2$ (b) $a = 0.98$, $\sigma_U^2 = 1 - a^2$

Figure 17.6: The autocorrelation sequence for autoregressive random processes with different parameters.

Since this holds for all $\mathbf{a} \neq \mathbf{0}$, the covariance matrix $\mathbf{C}_X$ is by definition positive definite (see Appendix C). (If it were possible to choose a_0 and a_1 so that $Y = c$, for c a constant, then $X[1]$ would be perfectly predictable from $X[0]$ as $X[1] = -(a_0/a_1)X[0] + (c/a_1)$. Therefore, we could have $\text{var}(Y) = \mathbf{a}^T\mathbf{C}_X\mathbf{a} = 0$, and $\mathbf{C}_X$ would only be positive *semi*definite.) Now if $X[n]$ is a zero mean WSS random process

$$\text{cov}(X[n_1], X[n_2]) = E(X[n_1]X[n_2]) = r_X[n_2 - n_1]$$

and the covariance matrix becomes

$$\mathbf{C}_X = \begin{bmatrix} r_X[0] & r_X[1] \\ r_X[-1] & r_X[0] \end{bmatrix} = \underbrace{\begin{bmatrix} r_X[0] & r_X[1] \\ r_X[1] & r_X[0] \end{bmatrix}}_{\mathbf{R}_X}.$$

Therefore, the covariance matrix, which we now denote by $\mathbf{R}_X$ and which is called the *autocorrelation matrix*, must be positive definite. This implies that all the *principal minors* (see Appendix C) are positive. For the 2×2 case this means that

$$\begin{aligned} r_X[0] &> 0 \\ r_X^2[0] - r_X^2[1] &> 0 \end{aligned} \tag{17.22}$$

with the first condition being consistent with Property 17.1 and the second condition producing $r_X[0] > |r_X[1]|$. The latter condition is nearly consistent with Property 17.3 with the slight difference, that $|r_X[1]|$ may equal $r_X[0]$ being excluded. This is because we assumed that $X[1]$ was not perfectly predictable from knowledge of $X[0]$. If we allow perfect predictability, then the autocorrelation matrix is only positive

semidefinite and the $>$ sign in the second equation of (17.22) would be replaced with $\geq$. In general the $N \times N$ autocorrelation matrix $\mathbf{R}_X$ is given as the covariance matrix of the *zero mean* random vector $\mathbf{X} = [X[0]\, X[1] \ldots X[N-1]]^T$ as

$$\mathbf{R}_X = \begin{bmatrix} r_X[0] & r_X[1] & r_X[2] & \ldots & r_X[N-1] \\ r_X[1] & r_X[0] & r_X[1] & \ldots & r_X[N-2] \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_X[N-1] & r_X[N-2] & r_X[N-3] & \ldots & r_X[0] \end{bmatrix}. \tag{17.23}$$

For a sequence to be a valid ACS the $N \times N$ autocorrelation matrix must be positive semidefinite for all $N = 1, 2, \ldots$ and positive definite if we exclude the possibility of perfect predictability [Brockwell and Davis 1987]. This imposes a large number of constraints on $r_X[k]$ and hence not all sequences satisfying Properties 17.1–17.3 are valid ACSs (see also Problem 17.19). In summary, for our last property of the ACS we have the following.

Property 17.6 – ACS is a positive semidefinite sequence.
Mathematically, this means that $r_X[k]$ must satisfy

$$\mathbf{a}^T \mathbf{R}_X \mathbf{a} \geq 0$$

for all $\mathbf{a} = [a_0\, a_1 \ldots a_{N-1}]^T$ and where $\mathbf{R}_X$ is the $N \times N$ autocorrelation matrix given by (17.23). This must hold for all $N \geq 1$.

□

17.5 Ergodicity and Temporal Averages

When a random process is WSS, its mean does not depend on time. Hence, the random variables $\ldots, X[-1], X[0], X[1], \ldots$ all have the same mean. Then, at least as far as the mean is concerned, when we observe a realization of a random process, it is as if we are observing multiple realizations of the same random variable. This suggests that we may be able to determine the value of the mean from a single infinite length realization. To pursue this idea further we plot three realizations of an IID random process whose marginal PDF is Gaussian with mean $\mu_X[n] = \mu = 1$ and a variance $\sigma_X^2[n] = \sigma^2 = 1$ in Figure 17.7. If we let $x_i[18]$ denote the ith realization at time $n = 18$, then by definition of $E[X[18]]$

$$\lim_{M \to \infty} \frac{1}{M} \sum_{m=1}^{M} x_m[18] = E[X[18]] = \mu_X[18] = \mu = 1. \tag{17.24}$$

This is because as we observe all realizations of the random variable $X[18]$ they will conform to the Gaussian PDF (recall that $X[n] \sim \mathcal{N}(1,1)$). In fact, the original definition of expected value was based on the relationship given in (17.24). This

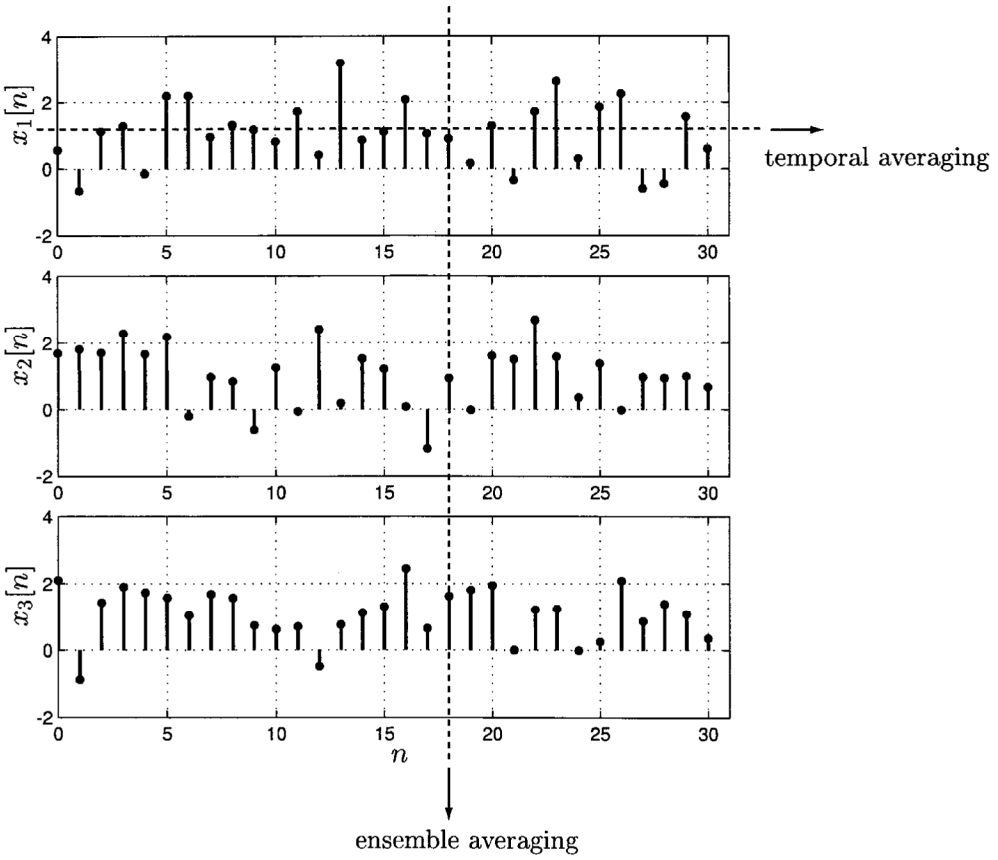


Figure 17.7: Several realizations of WSS random process with $\mu_X[n] = \mu = 1$. Vertical dashed line indicates "ensemble averaging" while horizontal dashed line indicates "temporal averaging."

type of averaging is called "averaging down the ensemble" and consequently is just a restatement of our usual notion of the expected value of a random variable. However, if we are given only a single realization such as $x_1[n]$, then it seems reasonable that

$$\hat{\mu}_N = \frac{1}{N} \sum_{n=0}^{N-1} x_1[n]$$

should also converge to μ as $N \to \infty$. This type of averaging is called "temporal averaging" since we are averaging the samples in time. If it is true that the temporal average converges to μ, then we can state that

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=0}^{N-1} x_1[n] = \mu = E[X[18]] = \lim_{M \to \infty} \frac{1}{M} \sum_{m=1}^{M} x_m[18]$$

and it is said that *temporal averaging is equivalent to ensemble averaging* or that the random process is *ergodic in the mean.* This property is of great practical importance since it assures us that by averaging enough samples of the realization, we can determine the mean of the random process. For the case of an IID random process ergodicity holds due to the law of large numbers (see Chapter 15). Recall that if $X_1, X_2, \ldots, X_N$ are IID random variables with mean μ and variance σ^2, then the sample mean random variable has the property that

$$\frac{1}{N}\sum_{i=1}^{N} X_i \to E[X] = \mu \qquad \text{as } N \to \infty.$$

Hence, if $X[n]$ is an IID random process, the conditions required for the law of large numbers to hold are satisfied, and we can immediately conclude that

$$\hat{\mu}_N = \frac{1}{N}\sum_{n=0}^{N-1} X[n] \to \mu. \tag{17.25}$$

Now the assumptions required for a random process to be IID are overly restrictive for (17.25) to hold. More generally, if $X[n]$ is a WSS random process, then since $E[X[n]] = \mu$, it follows that $E[\hat{\mu}_N] = (1/N)\sum_{n=0}^{N-1} E[X[n]] = \mu$. Therefore, the only further condition required for ergodicity in the mean is that

$$\lim_{N\to\infty} \text{var}(\hat{\mu}_N) = 0.$$

In the case of the IID random process it is easily shown that $\text{var}(\hat{\mu}_N) = \sigma^2/N \to 0$ as $N \to \infty$ and the condition is satisfied. More generally, however, the random process samples are correlated so that evaluation of this variance is slightly more complicated. We illustrate this computation next.

Example 17.6 – General MA random process

Consider the general MA random process given as $X[n] = (U[n] + U[n-1])/2$, where $E[U[n]] = \mu$ and $\text{var}(U[n]) = \sigma_U^2$ for $-\infty < n < \infty$ and the $U[n]$'s are all uncorrelated. This is similar to the MA process of Example 16.10 but is more general in that the mean of $U[n]$ is not necessarily zero, the samples of $U[n]$ are only uncorrelated, and hence, not necessarily independent, and the PDF of each sample need not be Gaussian. The general MA process $X[n]$ is easily shown to be WSS and to have a mean sequence $\mu_X[n] = \mu$ (see Problem 17.20). To determine if it is ergodic in the mean we must compute the $\text{var}(\hat{\mu}_N)$ and show that it converges to zero as $N \to \infty$. Now

$$\text{var}(\hat{\mu}_N) = \text{var}\left(\frac{1}{N}\sum_{n=0}^{N-1} X[n]\right).$$

Since the $X[n]$'s are now correlated, we use (9.26), where $\mathbf{a} = [a_0\, a_1 \ldots a_{N-1}]^T$ with $a_n = 1/N$, to yield

$$\text{var}(\hat{\mu}_N) = \text{var}\left(\sum_{n=0}^{N-1} a_n X[n]\right) = \mathbf{a}^T \mathbf{C}_X \mathbf{a}. \tag{17.26}$$

The covariance matrix has (i,j) element

$$[\mathbf{C}_X]_{ij} = E[(X[i]-E[X[i]])(X[j]-E[X[j]])] \qquad i = 0,1,\ldots,N-1; j = 0,1,\ldots,N-1.$$

But

$$\begin{aligned} X[n] - E[X[n]] &= \frac{1}{2}(U[n] + U[n-1]) - \frac{1}{2}(\mu + \mu) \\ &= \frac{1}{2}[(U[n] - \mu) + (U[n-1] - \mu)] \\ &= \frac{1}{2}[\bar{U}[n] + \bar{U}[n-1]] \end{aligned}$$

where $\bar{U}[n]$ is a *zero mean* random variable for each value of n. Thus,

$$\begin{aligned} [\mathbf{C}_X]_{ij} &= \frac{1}{4}E[(\bar{U}[i] + \bar{U}[i-1])(\bar{U}[j] + \bar{U}[j-1])] \\ &= \frac{1}{4}\left(E[\bar{U}[i]\bar{U}[j]] + E[\bar{U}[i]\bar{U}[j-1]] + E[\bar{U}[i-1]\bar{U}[j]] + E[\bar{U}[i-1]\bar{U}[j-1]]\right) \end{aligned}$$

and since $E[\bar{U}[n_1]\bar{U}[n_2]] = \text{cov}(U[n_1], U[n_2]) = \sigma_U^2 \delta[n_2 - n_1]$ (all the $U[n]$'s are uncorrelated), we have

$$[\mathbf{C}_X]_{ij} = \frac{1}{4}\left(\sigma_U^2\delta[j-i] + \sigma_U^2\delta[j-1-i] + \sigma_U^2\delta[j-i+1] + \sigma_U^2\delta[j-i]\right).$$

Finally, we have the required covariance matrix

$$[\mathbf{C}_X]_{ij} = \begin{cases} \frac{1}{2}\sigma_U^2 & i = j \\ \frac{1}{4}\sigma_U^2 & |i-j| = 1 \\ 0 & \text{otherwise.} \end{cases} \tag{17.27}$$

Using this in (17.26) produces

$$\text{var}(\hat{\mu}_N)$$

$$
\begin{aligned}
&= \mathbf{a}^T \mathbf{C}_X \mathbf{a} \\
&= \begin{bmatrix} \frac{1}{N} & \frac{1}{N} & \cdots & \frac{1}{N} \end{bmatrix}
\begin{bmatrix}
\frac{\sigma_U^2}{2} & \frac{\sigma_U^2}{4} & 0 & 0 & \cdots & 0 & 0 & 0 \\
\frac{\sigma_U^2}{4} & \frac{\sigma_U^2}{2} & \frac{\sigma_U^2}{4} & 0 & \cdots & 0 & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
0 & 0 & 0 & 0 & \cdots & \frac{\sigma_U^2}{4} & \frac{\sigma_U^2}{2} & \frac{\sigma_U^2}{4} \\
0 & 0 & 0 & 0 & \cdots & 0 & \frac{\sigma_U^2}{4} & \frac{\sigma_U^2}{2}
\end{bmatrix}
\begin{bmatrix} \frac{1}{N} \\ \frac{1}{N} \\ \vdots \\ \frac{1}{N} \end{bmatrix} \\
&= \frac{1}{N^2} \sum_{i=0}^{N-1} \frac{\sigma_U^2}{2} + \frac{1}{N^2} \sum_{i=0}^{N-2} \frac{\sigma_U^2}{4} + \frac{1}{N^2} \sum_{i=1}^{N-1} \frac{\sigma_U^2}{4} \\
&= \frac{\sigma_U^2}{2N} + \frac{\sigma_U^2}{4} \frac{N-1}{N^2} + \frac{\sigma_U^2}{4} \frac{N-1}{N^2} \to 0 \qquad \text{as } N \to \infty.
\end{aligned}
$$

Finally, we see that the general MA random process is ergodic in the mean.

♢

In general, it can be shown that for a WSS random process to be ergodic in the mean, the variance of the sample mean

$$\text{var}(\hat{\mu}_N) = \frac{1}{N} \sum_{k=-(N-1)}^{N-1} \left(1 - \frac{|k|}{N}\right) (r_X[k] - \mu^2) \tag{17.28}$$

must converge to zero as $N \to \infty$ (see Problem 17.23 for the derivation of (17.28)). For this to occur, the covariance sequence $r_X[k] - \mu^2$ must decay to zero at a fast enough rate as $k \to \infty$, which is to say that as the samples are spaced further and further apart, they must eventually become uncorrelated. A little reflection on the part of the reader will reveal that ergodicity requires a single realization of the random process to display the behavior of the entire ensemble of realizations. If not, ergodicity will not hold. Consider the following simple nonergodic random process.

Example 17.7 – Random DC level

Define a random process as $X[n] = A$ for $-\infty < n < \infty$, where $A \sim \mathcal{N}(0, 1)$. Some realizations are shown in Figure 17.8. This random process is WSS since

$$
\begin{aligned}
\mu_X[n] &= E[X[n]] = E[A] = 0 = \mu \qquad -\infty < n < \infty \quad \text{(not dependent on } n\text{)} \\
r_X[k] &= E[X[n]X[n+k]] = E[A^2] = 1 \qquad\qquad\qquad\quad \text{(not dependent on } n\text{)}.
\end{aligned}
$$

However, it should be clear that $\hat{\mu}_N$ will not converge to $\mu = 0$. Referring to the realization $x_1[n]$ in Figure 17.8, the sample mean will produce -0.43 no matter how large N becomes. In addition, it can be shown that $\text{var}(\hat{\mu}_N) = 1$ (see Problem 17.24). Each realization is *not* representative of the *ensemble of realizations.*

♢

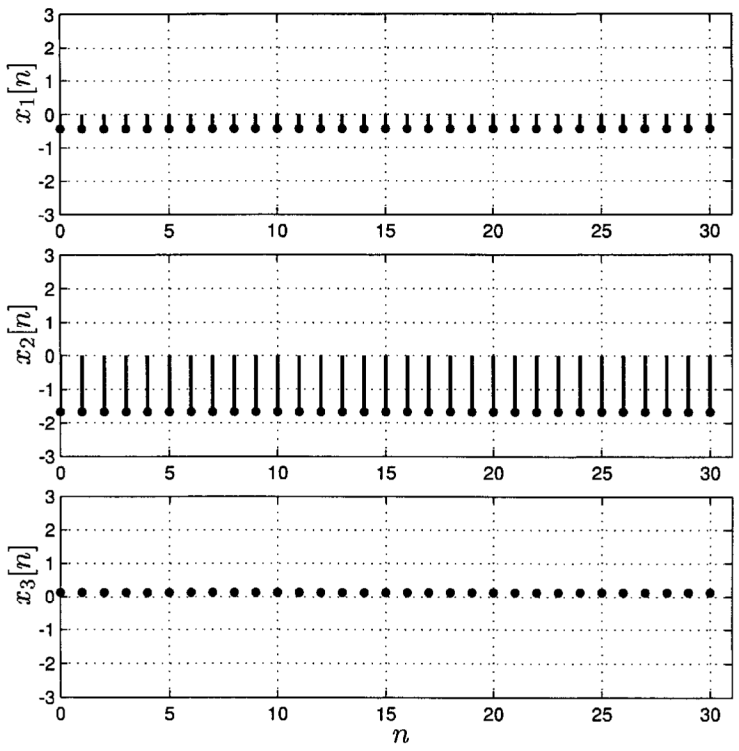


Figure 17.8: Several realizations of the random DC level process.

17.6 The Power Spectral Density

The ACS measures the correlation between samples of a WSS random process. For example, the AR random process was shown to have the ACS

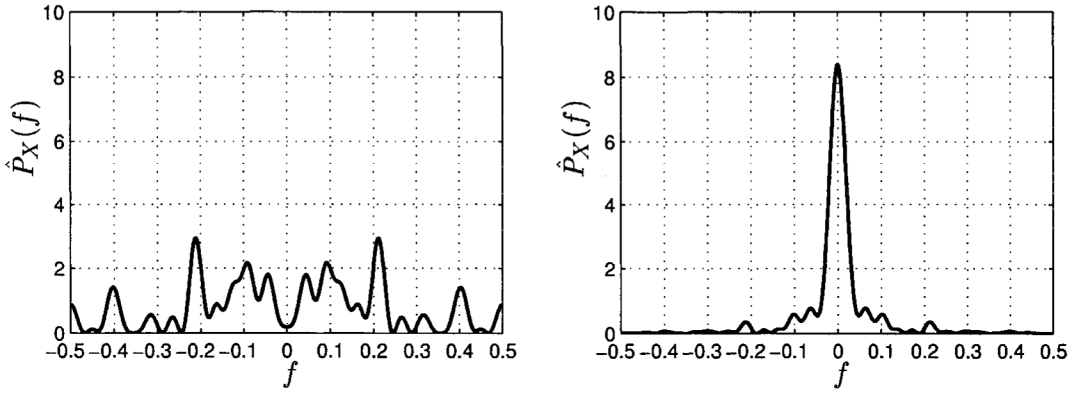
$$r_X[k] = \frac{\sigma_U^2}{1-a^2}a^{|k|}$$

which for $a = 0.25$ and $a = 0.98$ is shown in Figure 17.6, along with some typical realizations in Figure 17.5. Note that when the ACS dies out rapidly (see Figure 17.6a), the realization is more rapidly varying in time (see Figure 17.5a). In contrast, when the ACS decays slowly (see Figure 17.6b), the realization varies slowly (see Figure 17.5b). It would seem that the ACS is related to the rate of change of the random process. For deterministic signals the rate of change is usually measured by examining a discrete-time Fourier transform [Jackson 1991]. Signals with high frequency content exhibit rapid fluctutations in time while signals with only low frequency content exhibit slow variations in time. For WSS random processes we will be interested in the *power* at the various frequencies. In particular, we will introduce the measure known as the *power spectral density* (PSD) and show that it

quantifies the distribution of power with frequency. Before doing so, however, we consider the following deterministically motivated measure of power with frequency based on the discrete-time Fourier transform

$$\hat{P}_X(f) = \frac{1}{N}\left|\sum_{n=0}^{N-1} X[n]\exp(-j2\pi fn)\right|^2 . \tag{17.29}$$

This is a normalized version of the magnitude-squared discrete-time Fourier transform of the random process over the time interval $0 \leq n \leq N-1$. It is called the *periodogram* since its original purpose was to find periodicities in random data sets [Schuster 1898]. In (17.29) f denotes the discrete-time frequency, which is assumed to be in the range $-1/2 \leq f \leq 1/2$ for reasons that will be elucidated later. The $1/N$ factor is required to normalize $\hat{P}_X(f)$ to be interpretable as a power spectral *density* or power per unit frequency. The use of a "hat" is meant to convey the notion that this quantity is an estimator. As we now show, the periodogram is not a suitable measure of the distribution of power with frequency, although it would be for some deterministic signals (such as periodic discrete-time signals with period N). As an example, we plot $\hat{P}_X(f)$ in Figure 17.9 for the realizations given in Figure 17.5. We



(a) $a = 0.25$, $\sigma_U^2 = 1 - a^2$ (b) $a = 0.98$, $\sigma_U^2 = 1 - a^2$

Figure 17.9: Periodogram for autoregressive random process with different parameters. The realizations shown in Figure 17.5 were used to generate these estimates.

see that the periodogram in Figure 17.9a exhibits many random fluctuations. Other realizations will also produce similar seemingly random curves. However, it does seem to produce a reasonable result—for the periodogram in Figure 17.9a there is more high frequency power than for the periodogram in Figure 17.9b. The reason for the random nature of the plot is that (17.29) is a function of N random variables and hence is a random variable itself for each frequency. As such, it exhibits the

variability of a random process for which the usual dependence on time is replaced by frequency. What we would actually like is an *average* measure of the power distribution with frequency, suggesting the need for an expected value. Also, to ensure that we capture the entire random process behavior, an infinite length realization is required. We are therefore led to the following more suitable definition of the PSD

$$P_X(f) = \lim_{M\to\infty} \frac{1}{2M+1} E\left[\left|\sum_{n=-M}^{M} X[n]\exp(-j2\pi fn)\right|^2\right]. \tag{17.30}$$

The function $P_X(f)$ is called the *power spectral density* (PSD) and when integrated provides a measure of the average power within a band of frequencies. It is completely analogous to the PDF in that to find the *average power* of the random process in the frequency band $f_1 \leq f \leq f_2$ we should find the area under the PSD curve.

Fourier analysis of a random process yields no phase information.

In our definition of the PSD we are using the magnitude-squared of the Fourier transform. It is obvious then, that the PSD does not tell us anything about the phases of the Fourier transform of the random process. This is in contrast to a Fourier transform of a deterministic signal. There the inverse Fourier transform can be viewed as a decomposition of the signal into sinusoids of different frequencies with deterministic amplitudes and phases. For a random process a similar decomposition called the *spectral representation theorem* [Brockwell and Davis 1987] yields sinusoids of different frequencies with *random amplitudes and random phases.* The PSD is essentially the *expected value of the power of the random sinusoidal amplitudes* per unit of frequency. No phase information is retained and therefore no phase information can be extracted from knowledge of the PSD.

We next give an example of the computation of a PSD.

Example 17.8 – White noise

Assume that $X[n]$ is white noise (see Example 17.2) and therefore, has a zero mean

and ACS $r_X[k] = \sigma^2\delta[k]$. Then,

$$\begin{aligned} P_X(f) &= \lim_{M\to\infty} \frac{1}{2M+1} E\left[\sum_{n=-M}^{M} X[n]\exp(j2\pi fn) \sum_{m=-M}^{M} X[m]\exp(-j2\pi fm)\right] \\ &= \lim_{M\to\infty} \frac{1}{2M+1} \sum_{n=-M}^{M}\sum_{m=-M}^{M} \underbrace{E[X[n]X[m]]}_{r_X[m-n]} \exp[-j2\pi f(m-n)] \qquad (17.31) \\ &= \lim_{M\to\infty} \frac{1}{2M+1} \sum_{n=-M}^{M}\sum_{m=-M}^{M} \sigma^2\delta[m-n]\exp[-j2\pi f(m-n)] \\ &= \lim_{M\to\infty} \frac{1}{2M+1} \sum_{n=-M}^{M} \sigma^2 \\ &= \lim_{M\to\infty} \sigma^2 = \sigma^2. \qquad (17.32) \end{aligned}$$

Hence, for white noise the PSD is

$$P_X(f) = \sigma^2 \qquad -1/2 \le f \le 1/2.$$

As first mentioned in Chapter 16 white noise contains equal contributions of average power at all frequencies.

$\Diamond$

A more straightforward approach to obtaining the PSD is based on knowledge of the ACS. From (17.31) we see that

$$P_X(f) = \lim_{M\to\infty} \frac{1}{2M+1} \sum_{n=-M}^{M}\sum_{m=-M}^{M} r_X[m-n]\exp[-j2\pi f(m-n)]. \qquad (17.33)$$

This can be simplified using the formula (see Problem 17.26)

$$\sum_{n=-M}^{M}\sum_{m=-M}^{M} g[m-n] = \sum_{k=-2M}^{2M} (2M+1-|k|)g[k]$$

which results from considering $g[m-n]$ as an element of the $(2M+1)\times(2M+1)$ matrix $\mathbf{G}$ with elements $[\mathbf{G}]_{mn} = g[m-n]$ for $m = -M,\ldots,M$ and $n = -M,\ldots,M$ and then summing all the elements. Using this relationship in (17.33) produces

$$\begin{aligned} P_X(f) &= \lim_{M\to\infty} \frac{1}{2M+1} \sum_{k=-2M}^{2M} (2M+1-|k|)r_X[k]\exp(-j2\pi fk) \\ &= \lim_{M\to\infty} \sum_{k=-2M}^{2M} \left(1 - \frac{|k|}{2M+1}\right) r_X[k]\exp(-j2\pi fk). \end{aligned}$$

Assuming that $\sum_{k=-\infty}^{\infty} |r_X[k]| < \infty$, the limit can be shown to produce the final result (see Problem 17.27)

$$P_X(f) = \sum_{k=-\infty}^{\infty} r_X[k] \exp(-j2\pi fk) \tag{17.34}$$

which says that *the power spectral density is the discrete-time Fourier transform of the ACS.* This relationship is known as the *Wiener-Khinchine* theorem. Some examples follow.

Example 17.9 – White noise

From Example 17.2 $r_X[k] = \sigma^2\delta[k]$ and so

$$\begin{aligned} P_X(f) &= \sum_{k=-\infty}^{\infty} r_X[k] \exp(-j2\pi fk) \\ &= \sum_{k=-\infty}^{\infty} \sigma^2\delta[k] \exp(-j2\pi fk) \\ &= \sigma^2. \end{aligned}$$

This is shown in Figure 17.10. Note that the total average power in $X[n]$, which is $r_X[0] = \sigma^2$, is given by the area under the PSD curve.

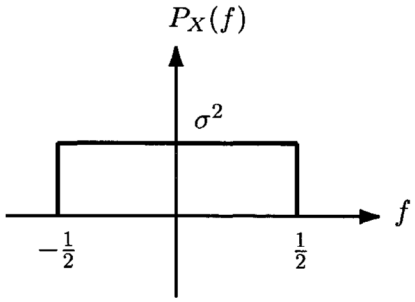


Figure 17.10: PSD of white noise.

♢

Example 17.10 – AR random process

From (17.21) we have that

$$r_X[k] = \frac{\sigma_U^2}{1-a^2} a^{|k|} \qquad -\infty < k < \infty$$

and from (17.34)

$$\begin{aligned} P_X(f) &= \sum_{k=-\infty}^{\infty} r_X[k]\exp(-j2\pi fk) \\ &= \frac{\sigma_U^2}{1-a^2}\sum_{k=-\infty}^{\infty} a^{|k|}\exp(-j2\pi fk) \\ &= \frac{\sigma_U^2}{1-a^2}\left[\sum_{k=-\infty}^{-1} a^{-k}\exp(-j2\pi fk) + \sum_{k=0}^{\infty} a^k\exp(-j2\pi fk)\right] \\ &= \frac{\sigma_U^2}{1-a^2}\left[\sum_{k=1}^{\infty}[a\exp(j2\pi f)]^k + \sum_{k=0}^{\infty}[a\exp(-j2\pi f)]^k\right]. \end{aligned}$$

Since $|a\exp(\pm j2\pi f)| = |a| < 1$, we can use the formula $\sum_{k=k_0}^{\infty} z^k = z^{k_0}/(1-z)$ for z a complex number with $|z| < 1$ to evaluate the sums. This produces

$$\begin{aligned} P_X(f) &= \frac{\sigma_U^2}{1-a^2}\left(\frac{a\exp(j2\pi f)}{1-a\exp(j2\pi f)} + \frac{1}{1-a\exp(-j2\pi f)}\right) \\ &= \frac{\sigma_U^2}{1-a^2}\frac{a\exp(j2\pi f)(1-a\exp(-j2\pi f)) + (1-a\exp(j2\pi f))}{(1-a\exp(j2\pi f))(1-a\exp(-j2\pi f))} \\ &= \frac{\sigma_U^2}{1-a^2}\frac{1-a^2}{|1-a\exp(-j2\pi f)|^2} \\ &= \frac{\sigma_U^2}{|1-a\exp(-j2\pi f)|^2.} \end{aligned} \tag{17.35}$$

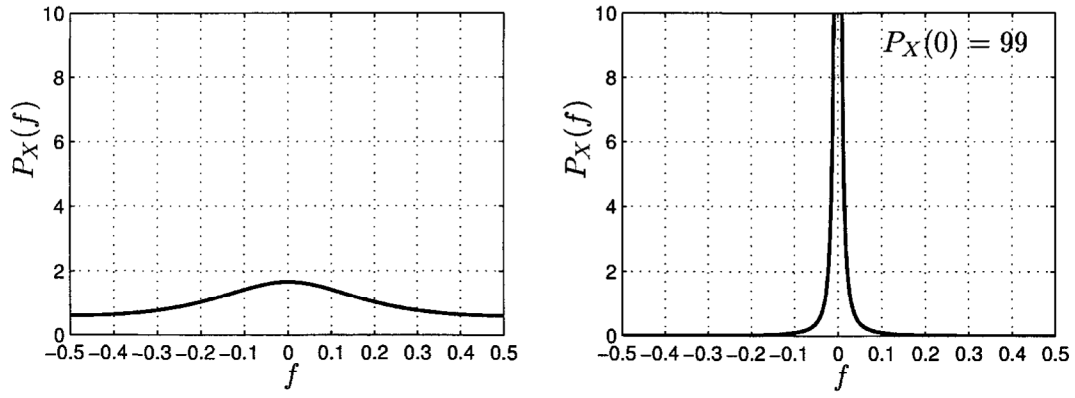
This can also be written in real form as

$$P_X(f) = \frac{\sigma_U^2}{1+a^2-2a\cos(2\pi f)} \qquad -1/2 \le f \le 1/2. \tag{17.36}$$

For $a = 0.25$ and $a = 0.98$ and $\sigma_U^2 = 1-a^2$, the PSDs are plotted in Figure 17.11. Note that the total average power in each PSD is the same, being $r_X[0] = \sigma_U^2/(1-a^2) = 1$. As expected the more noise-like random process has a PSD (see Figure 17.11a) with more high frequency average power than the slowly varying random process (see Figure 17.11b) which has all its average power near $f = 0$ (or at DC).

$\Diamond$

From the previous example, we observe that the PSD exhibits the properties of being a real nonnegative function of frequency, consistent with our notion of power as a nonnegative physical quantity, of being symmetric about $f = 0$, and of being periodic with period one (see (17.36)). We next prove that these properties are true in general.



(a) $a = 0.25$, $\sigma_U^2 = 1 - a^2$

(b) $a = 0.98$, $\sigma_U^2 = 1 - a^2$

Figure 17.11: Power spectral densities for autoregressive random process with different parameters. The periodograms, which are estimated PSDs, were given in Figure 17.9.

Property 17.7 – PSD is a real function.

The PSD is also given by the real function

$$P_X(f) = \sum_{k=-\infty}^{\infty} r_X[k] \cos(2\pi f k). \tag{17.37}$$

Proof:

$$\begin{aligned} P_X(f) &= \sum_{k=-\infty}^{\infty} r_X[k] \exp(-j2\pi f k) \\ &= \sum_{k=-\infty}^{\infty} r_X[k](\cos(2\pi f k) - j \sin(2\pi f k)) \\ &= \sum_{k=-\infty}^{\infty} r_X[k] \cos(2\pi f k) - j \sum_{k=-\infty}^{\infty} r_X[k] \sin(2\pi f k). \end{aligned}$$

But

$$\sum_{k=-\infty}^{\infty} r_X[k] \sin(2\pi f k) = \sum_{k=-\infty}^{-1} r_X[k] \sin(2\pi f k) + \sum_{k=1}^{\infty} r_X[k] \sin(2\pi f k)$$

since the $k = 0$ term is zero, and letting $l = -k$ in the first sum we have

$$\begin{aligned}\sum_{k=-\infty}^{\infty} r_X[k]\sin(2\pi fk) &= \sum_{l=1}^{\infty} r_X[-l]\sin(2\pi f(-l)) + \sum_{k=1}^{\infty} r_X[k]\sin(2\pi fk) \\ &= \sum_{k=1}^{\infty} r_X[k](-\sin(2\pi fk) + \sin(2\pi fk)) = 0 \quad (r_X[-l] = r_X[l])\end{aligned}$$

from which (17.37) follows.

□

Property 17.8 – PSD is nonnegative.

$$P_X(f) \geq 0$$

Proof: Follows from (17.30) but can also be shown to follow from the positive semidefinite property of the ACS [Brockwell and Davis 1987]. (See also Problem 17.19.)

□

Property 17.9 – PSD is symmetric about $f = 0$.

$$P_X(-f) = P_X(f)$$

Proof: Follows from (17.37).

□

Property 17.10 – PSD is periodic with period one.

$$P_X(f+1) = P_X(f)$$

Proof: From (17.37) we have

$$\begin{aligned}P_X(f+1) &= \sum_{k=-\infty}^{\infty} r_X[k]\cos(2\pi(f+1)k) \\ &= \sum_{k=-\infty}^{\infty} r_X[k]\cos(2\pi fk + 2\pi k) \\ &= \sum_{k=-\infty}^{\infty} r_X[k]\cos(2\pi fk) \qquad (\cos(2\pi k) = 1,\ \sin(2\pi k) = 0) \\ &= P_X(f)\end{aligned}$$

□

Property 17.11 – ACS recovered from PSD using inverse Fourier transform

$$r_X[k] = \int_{-\frac{1}{2}}^{\frac{1}{2}} P_X(f)\exp(j2\pi fk)df \qquad -\infty < k < \infty \tag{17.38}$$

$$= \int_{-\frac{1}{2}}^{\frac{1}{2}} P_X(f)\cos(2\pi fk)df \qquad -\infty < k < \infty \tag{17.39}$$

Proof: (17.38) follows from properties of discrete-time Fourier transform [Jackson 1991]. (17.39) follows from Property 17.9 (see Appendix B.5 and also Problem 17.49).

□

Property 17.12 – PSD yields average power over band of frequencies. To obtain the average power in the frequency band $f_1 \leq f \leq f_2$ we need only find the area under the PSD curve for this band. The average *physical* power is obtained as twice this area since the negative frequencies account for half of the average power (recall Property 17.9). Hence,

$$\text{Average physical power in } [f_1, f_2] = 2\int_{f_1}^{f_2} P_X(f)df. \tag{17.40}$$

The proof of this property requires some concepts to be described in the next chapter, and thus, we defer the proof until Section 18.4. Note, however, that if $f_1 = 0$ and $f_2 = 1/2$, then the average power in this band is

$$\begin{aligned}
\text{Average physical power in } [0, 1/2] &= 2\int_0^{1/2} P_X(f)df \\
&= \int_{-\frac{1}{2}}^{\frac{1}{2}} P_X(f)df \quad \text{(due to symmetry of PSD)} \\
&= \int_{-\frac{1}{2}}^{\frac{1}{2}} P_X(f)\exp(j2\pi f(0))df \\
&= r_X[0] \qquad \text{(from (17.38))}
\end{aligned}$$

which we have already seen yields the total average power since $r_X[0] = E[X^2[n]]$. Hence, we see that the *total average power* is obtained by integrating the PSD over all frequencies to yield

$$r_X[0] = \int_{-\frac{1}{2}}^{\frac{1}{2}} P_X(f)df. \tag{17.41}$$

□

Definitions of PSD are not consistent.

In some texts, especially ones describing the use of the PSD for physical measurements, the definition of the PSD is slightly different. The alternative definition relies on the relationship of (17.40) to define the PSD as $G_X(f) = 2P_X(f)$. It is called the *one-sided PSD* and its advantage is that it yields directly the average power over a band when integrated over the band. As can be seen from (17.40)

$$\text{Average physical power in } [f_1, f_2] = \int_{f_1}^{f_2} G_X(f)df.$$

A final comment concerns the periodicity of the PSD. We have chosen the frequency interval $[-1/2, 1/2]$ over which to display the PSD. The rationale for this choice arises from the practical situation in which a *continuous-time* WSS random process (see Section 17.8) is sampled to produce a discrete-time WSS random process. Then, if the continuous-time random process $X(t)$ has a PSD that is bandlimited to W Hz and is sampled at F_s samples/sec, the discrete-time PSD $P_X(f)$ will have discrete-time frequency units of W/F_s. For Nyquist rate sampling of $F_s = 2W$, the maximum discrete-time frequency will be $f = W/F_s = 1/2$. Hence, our choice of the frequency interval $[-1/2, 1/2]$ corresponds to the continuous-time frequency interval of $[-W, W]$ Hz. The discrete-time frequency is also referred to as the *normalized frequency*, the normalizing factor being F_s.

17.7 Estimation of the ACS and PSD

Recall from our discussion of ergodicity that in the problem of mean estimation for a WSS random process, we were restricted to observing only a finite number of samples of *one realization* of the random process. If the random process is ergodic in the mean, then we saw that as the number of samples increases to infinity, the temporal average $\hat{\mu}_N$ will converge to the ensemble average μ. To apply this result to estimation of the ACS consider the problem of estimating the ACS for lag $k = k_0$ which is

$$r_X[k_0] = E[X[n]X[n + k_0]].$$

Then by defining the product random process $Y[n] = X[n]X[n + k_0]$ we see that

$$r_X[k_0] = E[Y[n]] \qquad -\infty < n < \infty$$

or the desired quantity to be estimated is just the mean of the random process $Y[n]$. The mean of $Y[n]$ does not depend on n. This suggests that we replace the observed values of $X[n]$ with those of $Y[n]$ by using $y[n] = x[n]x[n+k_0]$, and then use a temporal average to estimate the ensemble average. Hence, we have the temporal average estimate

$$\begin{aligned}\hat{r}_X[k_0] &= \frac{1}{N}\sum_{n=0}^{N-1} y[n] \\ &= \frac{1}{N}\sum_{n=0}^{N-1} x[n]x[n+k_0]. \end{aligned} \tag{17.42}$$

Also, since $r_X[-k] = r_X[k]$, we need only estimate the ACS for $k \geq 0$. There is one slight modification that we need to make to the estimate. Assuming that $\{x[0], x[1], \ldots, x[N-1]\}$ are observed, we must choose the upper limit on the summation in (17.42) to satisfy the constraint $n + k_0 \leq N-1$. This is because $x[n+k_0]$ is unobserved for $n + k_0 > N-1$. With this modification we have as our estimate of the ACS (and now replacing the specific lag of k_0 by the more general lag k)

$$\hat{r}_X[k] = \frac{1}{N-k}\sum_{n=0}^{N-1-k} x[n]x[n+k] \qquad k = 0, 1, \ldots, N-1. \tag{17.43}$$

We have also changed the $1/N$ averaging factor to $1/(N-k)$. This is because the number of terms in the sum is only $N-k$. For example, if $N = 4$ so that we observe $\{x[0], x[1], x[2], x[3]\}$, then (17.43) yields the estimates

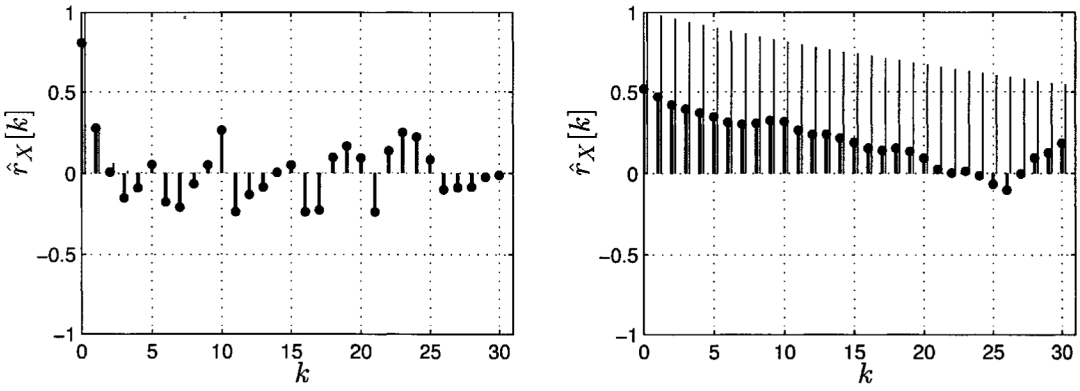
$$\begin{aligned}\hat{r}_X[0] &= \frac{1}{4}(x^2[0] + x^2[1] + x^2[2] + x^2[3]) \\ \hat{r}_X[1] &= \frac{1}{3}(x[0]x[1] + x[1]x[2] + x[2]x[3]) \\ \hat{r}_X[2] &= \frac{1}{2}(x[0]x[2] + x[1]x[3]) \\ \hat{r}_X[3] &= x[0]x[3]. \end{aligned}$$

As k increases, the distance between the samples increases and so there are less products available for averaging. In fact, for $k > N-1$, we cannot estimate the value of the ACS at all. With the estimate given in (17.43) we see that $E[\hat{r}_X[k]] = r_X[k]$ for $k = 0, 1, \ldots, N-1$. In order for the estimate to converge to the true value as $N \to \infty$, i.e, for the random process to be *ergodic in the autocorrelation* or

$$\lim_{N\to\infty} \hat{r}_X[k] = \lim_{N\to\infty} \frac{1}{N-k}\sum_{n=0}^{N-1-k} x[n]x[n+k] = r_X[k] \qquad k = 0, 1, \ldots$$

we require that $\mathrm{var}(\hat{r}_X[k]) \to 0$ as $N \to \infty$. This will generally be true if $r_X[k] \to 0$ as $k \to \infty$ for a zero mean random process but see Problem 17.25 for a case where

this is not required. To illustrate the estimation performance consider the AR random process described in Example 17.5. The true ACS and the estimated one using (17.43) and based on the realizations shown in Figure 17.5 are shown in Figure 17.12. The estimated ACS is shown as the dark lines while the true ACS as given by (17.21) is shown as light lines, which are slightly displaced to the right for easier viewing. Note that in Figure 17.12 the estimated values for k large exhibit



(a) $a = 0.25$, $\sigma_U^2 = 1 - a^2$ (b) $a = 0.98$, $\sigma_U^2 = 1 - a^2$

Figure 17.12: Estimated ACSs (dark lines) and the true ACSs given in Figure 17.6 (light lines) for the AR random process realizations shown in Figure 17.5.

a large error. This is due to the fewer number of products, i.e., $N - k = 31 - k$, that are available for averaging in (17.43). In the case of $k = 30$ the estimate is $\hat{r}_X[30] = x[0]x[30]$, which as you might expect is very poor since there is no averaging at all! Clearly, for accurate estimates of the ACS we require that $k_{\max} \ll N$. The MATLAB code used to estimate the ACS for Figure 17.12 is given below.

```
n=[0:30]';N=length(n);
a1=0.25;a2=0.98;
varu1=1-a1^2;varu2=1-a2^2;
r1true=(varu1/(1-a1^2))*a1.^n; % see (17.21)
r2true=(varu2/(1-a2^2))*a2.^n;
for k=0:N-1
   r1est(k+1,1)=(1/(N-k))*sum(x1(1:N-k).*x1(1+k:N));
   r2est(k+1,1)=(1/(N-k))*sum(x2(1:N-k).*x2(1+k:N));
end
```

To estimate the PSD requires somewhat more care than the ACS. We have already seen that the periodogram estimate of (17.29) is not suitable. There are many ways to estimate the PSD based on either (17.30) or (17.34). We illustrate

one approach based on (17.30). Others may be found in [Jenkins and Watts 1968, Kay 1988]. Since we only have a segment of a single realization of the random process, we cannot implement the expectation operation required in (17.30). Note that the operation of $E[\cdot]$ represents an average down the ensemble or equivalently an average over multiple realizations. To obtain some averaging, however, we can break up the data $\{x[0], x[1], \ldots, x[N-1]\}$ into I nonoverlapping blocks, with each block having a total of L samples. We assume for simplicity that there is an integer number of blocks so that $N = IL$. The implicit assumption in doing so is that each block exhibits the statistical characteristics of a single realization and so we can mimic the averaging down the ensemble by averaging temporally across successive blocks of data. Once again, the assumption of ergodicity is being employed. Thus, we first break up the data set into the I nonoverlapping data blocks

$$y_i[n] = x[n + iL] \qquad n = 0, 1, \ldots, L-1; i = 0, 1, \ldots, I-1$$

where each data block has a length of L samples. Then, for each data block we compute a periodogram as

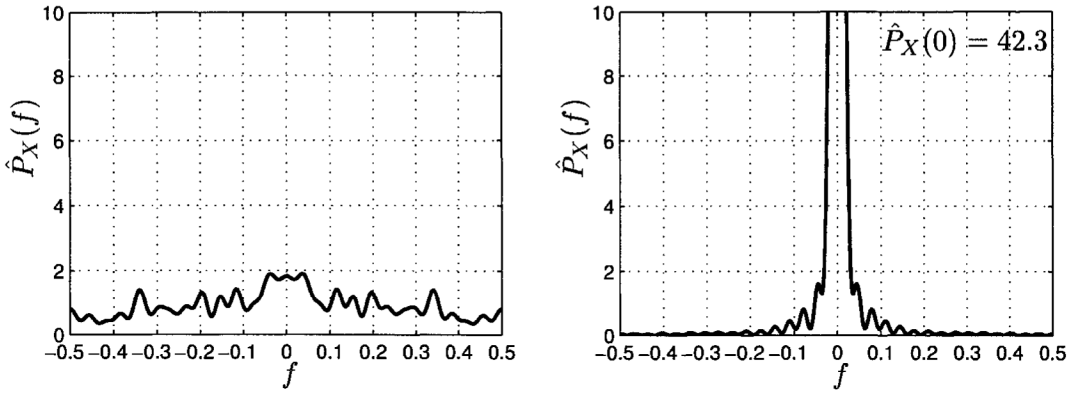
$$\hat{P}_X^{(i)}(f) = \frac{1}{L} \left| \sum_{n=0}^{L-1} y_i[n] \exp(-j2\pi f n) \right|^2 \tag{17.44}$$

and then average all the periodograms together to yield the final PSD estimate as

$$\hat{P}_{\text{av}}(f) = \frac{1}{I} \sum_{i=0}^{I-1} \hat{P}_X^{(i)}(f). \tag{17.45}$$

This estimate is called the *averaged periodogram*. It can be shown that under some conditions, $\lim_{N\to\infty} \hat{P}_{\text{av}}(f) = P_X(f)$. Once again we are calling upon an ergodicity type of property in that we are averaging the periodograms obtained in time instead of the theoretical ensemble averaging. Of course, for convergence to hold as $N \to \infty$, we must have $L \to \infty$ and $I \to \infty$ as well.

As an example, we examine the averaged periodogram estimates for the two AR processes whose PSDs are shown in Figure 17.11. The number of data samples was $N = 310$, which was broken up into $I = 10$ nonoverlapping blocks of data with $L = 31$ samples in each one. By comparing the spectral estimates in Figure 17.13 with those of Figure 17.9, it is seen that the averaging has yielded a better estimate. Of course, the price paid is that the data set needs to be $I = 10$ times as long! The MATLAB code used to implement the averaged periodogram estimate is given next. A fast Fourier transform (FFT) is used to compute the Fourier transform of the $y_i[n]$ sequences at the frequencies $f = -0.5 + k\Delta_f$, where $k = 0, 1, \ldots, 1023$ and $\Delta_f = 1/1024$ (see [Kay 1988] for a more detailed description).



(a) $a = 0.25$, $\sigma_U^2 = 1 - a^2$

(b) $a = 0.98$, $\sigma_U^2 = 1 - a^2$

Figure 17.13: Power spectral density estimates using the averaged periodogram method for autoregressive processes with different parameters. The true PSDs are shown in Figure 17.11.

```
Nfft=1024; % set FFT size
Pav1=zeros(Nfft,1);Pav2=Pav1; % set up arrays with desired dimension
f=[0:Nfft-1]'/Nfft-0.5; % set frequencies for later plotting
                        % of PSD estimate
for i=0:I-1
   nstart=1+i*L;nend=L+i*L; % set up beginning and end points
                            % of ith block of data
   y1=x1(nstart:nend);
   y2=x2(nstart:nend);
% take FFT of block, since FFT outputs samples of Fourier
% transform over frequency range [0,1), must shift FFT outputs
% for [1/2,1) to [-1/2, 0), then take complex magnitude-squared,
% normalize by L and average
Pav1=Pav1+(1/(I*L))*abs(fftshift(fft(y1,Nfft))).^2;
Pav2=Pav2+(1/(I*L))*abs(fftshift(fft(y2,Nfft))).^2;
end
```

17.8 Continuous-Time WSS Random Processes

In this section we give the corresponding definitions and formulas for continuous-time WSS random processes. A more detailed description can be found in [Papoulis 1965]. Also, an important example is described to illustrate the use of these formulas.

A continuous-time random process $X(t)$ for $-\infty < t < \infty$ is defined to be WSS

if the *mean function* $\mu_X(t)$ satisfies

$$\mu_X(t) = E[X(t)] = \mu \qquad -\infty < t < \infty \tag{17.46}$$

which is to say it is constant in time and an *autocorrelation function* (ACF) can be defined as

$$r_X(\tau) = E[X(t)X(t+\tau)] \qquad -\infty < \tau < \infty \tag{17.47}$$

which is not dependent on the value of t. Thus, $E[X(t_1)X(t_2)]$ depends only on $|t_2 - t_1|$. Note the use of the "parentheses" indicates that the argument of the ACF is continuous and serves to distinguish $r_X[k]$ from $r_X(\tau)$. The ACF has the following properties.

Property 17.13 – ACF is positive for the zero lag or $r_X(0) > 0$.
The total average power is $r_X(0) = E[X^2(t)]$.

□

Property 17.14 – ACF is an even function or $r_X(-\tau) = r_X(\tau)$.

□

Property 17.15 – Maximum value of ACF is at $\tau = 0$ or $|r_X(\tau)| \leq r_X(0)$.

□

Property 17.16 – ACF measures the predictability of a random process.
The correlation coefficient for two samples of a zero mean WSS random process is

$$\rho_{X(t),X(t+\tau)} = \frac{r_X(\tau)}{r_X(0)}.$$

□

Property 17.17 – ACF approaches μ^2 as $\tau \to \infty$.
This assumes that the samples become uncorrelated for large lags, which is usually the case.

□

Property 17.18 – $r_X(\tau)$ is a positive semidefinite function.
See [Papoulis 1965] for the definition of a positive semidefinite function. This property assumes that the some samples of $X(t)$ may be perfectly predictable. If it is not, then the ACF is positive definite.

□

The PSD is defined as

$$P_X(F) = \lim_{T\to\infty} \frac{1}{T} E\left[\left|\int_{-T/2}^{T/2} X(t)\exp(-j2\pi Ft)dt\right|^2\right] \quad -\infty < F < \infty \tag{17.48}$$

where F is the frequency in Hz. We use a capital F to denote continuous-time or analog frequency. By the Wiener-Khinchine theorem this is equivalent to the continuous-time Fourier transform of the ACF

$$P_X(F) = \int_{-\infty}^{\infty} r_X(\tau)\exp(-j2\pi F\tau)d\tau \tag{17.49}$$

$$= \int_{-\infty}^{\infty} r_X(\tau)\cos(2\pi F\tau)d\tau. \tag{17.50}$$

(See also Problem 17.49.) The PSD has the usual interpretation as the average power distribution with frequency. In particular, it is the average power per Hz. The average physical power in a frequency band $[F_1, F_2]$ is given by

$$\text{Average physical power in } [F_1, F_2] = 2\int_{F_1}^{F_2} P_X(F)dF$$

where again the 2 factor reflects the additional contribution of the negative frequencies. The properties of the PSD are as follows:

Property 17.19 – PSD is a real function.
The PSD is given by the real function

$$P_X(F) = \int_{-\infty}^{\infty} r_X(\tau)\cos(2\pi F\tau)d\tau$$

□

Property 17.20 – PSD is nonnegative.

$$P_X(F) \geq 0$$

□

Property 17.21 – PSD is symmetric about $F = 0$.

$$P_X(-F) = P_X(F)$$

□

Property 17.22 – ACF recovered from PSD using inverse Fourier transform

$$r_X(\tau) = \int_{-\infty}^{\infty} P_X(F)\exp(j2\pi F\tau)dF \qquad -\infty < \tau < \infty \tag{17.51}$$

$$= \int_{-\infty}^{\infty} P_X(F)\cos(2\pi F\tau)dF \qquad -\infty < \tau < \infty. \tag{17.52}$$

(See also Problem 17.49.)

□

Unlike the PSD for a discrete-time WSS random process, the PSD for a continuous-time WSS random process is *not* periodic. We next illustrate these definitions and formulas with an example of practical importance.

Example 17.11 – Obtaining discrete-time WGN from continuous-time WGN

A common model for a continuous-time noise random process $X(t)$ in a physical system is a WSS random process with a zero mean. In addition, due to the origin of noise as microscopic fluctuations of a large number of electrons, or molecules, etc., a central limit theorem can be employed to assert that $X(t)$ is a Gaussian random variable for all t. The average power of the noise in a band of frequencies is observed to be the same for all bands up to some upper frequency limit, at which the average power begins to decrease. For instance, consider thermal noise in a conductor due to random fluctuations of the electrons about some mean velocity. The average power versus frequency is predicted by physics to be constant until a cutoff frequency of about $F_c = 1000$ GHz at room temperature [Bell Telephone Labs 1970]. Hence, we can assume that the PSD of the noise has a PSD shown in Figure 17.14 as the true PSD. To further simplify the mathematically modeling without sacrificing the realism of the model, we can observe that all physical systems will only pass frequency components that are much lower than F_c—typically the bandwidth of the system is W Hz as shown in Figure 17.14. Any frequencies above W Hz will be cut off by the system. Therefore, the noise output of the system will be the same whether we use the true PSD or the modeled one shown in Figure 17.10. The modeled PSD is given by

$$P_X(F) = \frac{N_0}{2} \qquad -\infty < F < \infty.$$

This is clearly a physically impossible PSD in that the total average power is $r_X(0) = \int_{-\infty}^{\infty} P_X(F)df = \infty$. However, its use simplifies much systems analysis (see Problem 17.50). The corresponding ACF is from (17.51) the inverse Fourier transform, which is

$$r_X(\tau) = \frac{N_0}{2}\delta(\tau) \tag{17.53}$$

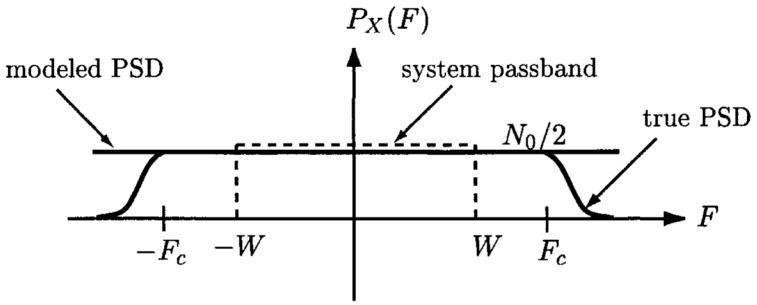


Figure 17.14: True and modeled PSDs for continuous-time white Gaussian noise.

and is seen to be an impulse at $\tau = 0$. Again the nonphysical nature of this model is manifest by the value $r_X(0) = \infty$. A continuous-time WSS Gaussian random process with zero mean and the ACF given by (17.53) is called continuous-time *white Gaussian noise* (WGN) (see also Example 20.6). It is a standard model in many disciplines.

Now as was previously mentioned, all physical systems are bandlimited to W Hz, which is typically chosen to ensure that a desired signal with a bandwidth of W Hz is not distorted. Modern signal processing hardware first bandlimits the continuous-time waveform to a maximum of W Hz using a lowpass filter and then samples the output of the filter at the Nyquist rate of $F_s = 2W$ samples/sec. The samples are then input into a digital computer. An important question to answer is: What are the statistical characteristics of the noise samples that are input to the computer? To answer this question we let Δ_t be the time interval between successive samples. Also, let $X(t)$ be the noise at the output of an ideal lowpass filter ($H(F) = 1$ for $|F| \leq W$ and $H(F) = 0$ for $|F| > W$) over the system passband shown in Figure 17.14. Then, the noise samples can be represented as

$$X(t)|_{t=n\Delta_t} = X[n] \qquad \text{for} -\infty < n < \infty.$$

Since $X(t)$ is bandlimited to W Hz and prior to filtering had the modeled PSD shown in Figure 17.14, its PSD is

$$P_X(F) = \begin{cases} \frac{N_0}{2} & |F| \leq W \\ 0 & |F| > W. \end{cases}$$

The noise samples $X[n]$ comprise a discrete-time random process. Its characteristics follow those of $X(t)$. Since $X(t)$ is Gaussian, then so is $X[n]$ (being just a sample). Also, since $X(t)$ is zero mean, so is $X[n]$ for all n. Finally, we inquire as to whether $X[n]$ is WSS, i.e., can we define an ACS? To answer this we first note that $X[n] = X(n\Delta_t)$ and recall that $X(t)$ is WSS. Then from the definition of the ACS

$$\begin{aligned} E[X[n]X[n+k]] &= E[X(n\Delta_t)X((n+k)\Delta_t)] \\ &= r_X(k\Delta_t) \quad \text{(definition of continuous-time ACF)} \end{aligned}$$

which does not depend on n, and so $X[n]$ is a zero mean discrete-time WSS random process with ACS

$$r_X[k] = r_X(k\Delta_t). \tag{17.54}$$

It is seen to be a sampled version of the continuous-time ACF. To explicitly evaluate the ACS we have from (17.51)

$$\begin{aligned} r_X(\tau) &= \int_{-\infty}^{\infty} P_X(F)\exp(j2\pi F\tau)dF \\ &= \int_{-W}^{W} \frac{N_0}{2}\exp(j2\pi F\tau)dF \\ &= \frac{N_0}{2}\int_{-W}^{W}\cos(2\pi F\tau)dF \qquad \text{(sine component is odd function)} \\ &= \frac{N_0}{2}\frac{\sin(2\pi F\tau)}{2\pi\tau}\Big|_{-W}^{W} \\ &= N_0 W\frac{\sin(2\pi W\tau)}{2\pi W\tau} \end{aligned} \tag{17.55}$$

which is shown in Figure 17.15. Now since $r_X[k] = r_X(k\Delta_t) = r_X(k/(2W))$, we

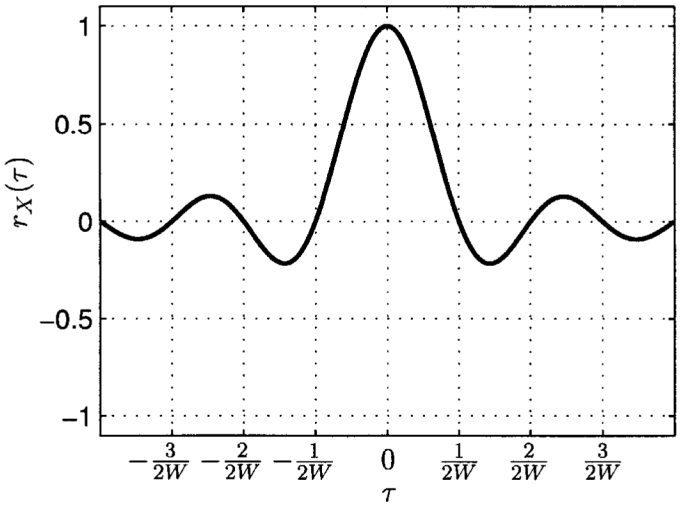


Figure 17.15: ACF for bandlimited continuous-time WGN with $N_0W = 1$.

see from Figure 17.15 that for $k = \pm 1, \pm 2, \ldots$ the ACS is zero, being the result of sampling the continuous-time ACF at its zeros. The only nonzero value is for $k = 0$, which is $r_X[0] = r_X(0) = N_0W$ from (17.55). Therefore, we finally observe that the ACS of the noise samples is

$$r_X[k] = N_0W\delta[k]. \tag{17.56}$$

The discrete-time noise random process is indeed WSS and has the ACS of (17.56). The PSD corresponding to this ACS has already been found and is shown in Figure 17.10, where $\sigma^2 = N_0 W$. Therefore, $X[n]$ *is a discrete-time white Gaussian noise random process.* This example justifies the use of the WGN model for discrete-time systems analysis.

$\diamondsuit$

Sampling faster gives only marginally better performance.

It is sometimes argued that by sampling the output of a system lowpass filter whose cutoff frequency is W Hz at a rate greater than $2W$, we can improve the performance of a signal processing system. For example, consider the estimation of the mean μ based on samples $Y[n] = \mu + X[n]$ for $n = 0, 1, \ldots, N-1$ where $E[X[n]] = 0$, $\text{var}(X[n]) = \sigma^2$, and the $X[n]$ samples are uncorrelated. The obvious estimate is the sample mean or $(1/N)\sum_{n=0}^{N-1} Y[n]$, whose expectation is μ and whose variance is σ^2/N. Clearly, if we could increase N, then the variance could be reduced and a better estimate would result. This suggests sampling the continuous-time random process at a rate faster than $2W$ samples/sec. The fallacy, however, is that as the sampling rate increases, *the noise samples become correlated* as can be seen by considering a sampling rate of $4W$ for which the time interval between samples becomes $\tau = \Delta_t/2 = 1/(4W)$. Then, as observed from Figure 17.15, the correlation between successive samples is $r_X(1/(4W)) = 0.6$. In effect, by sampling faster we are not obtaining any new realizations of the noise samples but nearly repetitions of the same noise samples. As a result, the variance will *not* decrease as $1/N$ but at a slower rate (see also Problem 17.51).

17.9 Real-World Example – Random Vibration Testing

Anyone who has ever traveled in a jet knows that upon landing, the cabin can vibrate greatly. This is due to the air currents outside the cabin which interact with the metallic aircraft surface. These pressure variations give rise to vibrations which are referred to as *turbulent boundary layer noise.* A manufacturer that intends to attach an antenna or other device to an aircraft must be cognizant of this vibration and plan for it. It is customary then to subject the antenna to a random vibration test in the lab to make sure it is not adversely affected in flight [McConnell 1995]. To do so the antenna would be mounted on a shaker table and the table shaken in a manner to simulate the turbulent boundary layer (TBL) noise. The problem the manufacturer faces is how to provide the proper vibration signal to the table, which

presumably will then be transmitted to the antenna. We now outline a possible solution to this problem.

The National Aeronautics and Space Administration (NASA) has determined PSD models for the TBL noise through physical modeling and experimentation. A reasonable model for the *one-sided* PSD of TBL noise upon reentry of a space vehicle, such as the space shuttle, into the earth's atmosphere is given by [NASA 2001]

$$G_X(F) = \begin{cases} G_X(500) & 0 \leq F < 500 \text{ Hz} \\ \frac{9 \times 10^{14} r^2}{F + 11364} & 500 \leq F \leq 50000 \text{ Hz} \end{cases}$$

where r represents a reference value which is 20μPa. A μPa is a unit of pressure equal to 10^{-6} nt/m^2. This PSD is shown in Figure 17.16 referenced to the standard unit so that $r = 1$. Note that it has a lowpass type of characteristic. In order

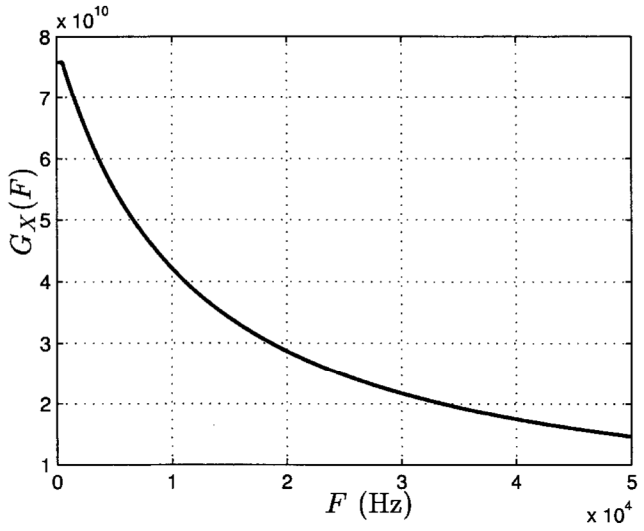


Figure 17.16: Continuous-time one-sided PSD for TBL noise.

to provide a signal to the shaker table that is random and has the PSD shown in Figure 17.16, we will assume that the signal is produced in a digital computer and then converted via a digital-to-analog convertor to a continuous-time signal. Hence, we need to produce a discrete-time WSS random process within the computer that has the proper PSD. Recalling our discussion in Section 17.8 we know that $r_X[k] = r_X(k\Delta_t)$ and since the highest frequency in the PSD is $W = 50,000$ Hz, we choose $\Delta_t = 1/(2W) = 1/100,000$. This produces the discrete-time PSD shown in Figure 17.17 and is given by $P_X(f) = (1/(2\Delta_t))G_X(f/\Delta_t)$. (We have divided by two to obtain the usual *two-sided* PSD. Also, the sampling operation introduces a factor of $1/\Delta_t$ [Jackson 1991].) To generate a realization of a discrete-time WSS random process with PSD given in Figure 17.17 we will use the AR model introduced in

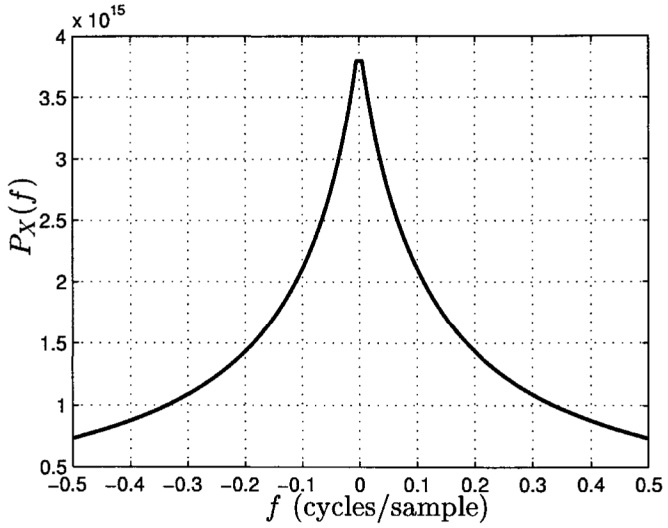


Figure 17.17: Discrete-time PSD for TBL noise.

Example 17.5. From the ACS we can determine values of a and σ_U^2 if we know $r_X[0]$ and $r_X[1]$ since

$$a = \frac{r_X[1]}{r_X[0]} \tag{17.57}$$

$$\sigma_U^2 = r_X[0](1-a^2) = r_X[0]\left[1-\left(\frac{r_X[1]}{r_X[0]}\right)^2\right]. \tag{17.58}$$

Knowing a and σ_U^2 will allow us to use the defining recursive difference equation, $X[n] = aX[n-1] + U[n]$, of an AR random process to generate the realization. To obtain the first two lags of the ACS we use (17.39)

$$r_X[0] = \int_{-\frac{1}{2}}^{\frac{1}{2}} P_X(f)df$$

$$r_X[1] = \int_{-\frac{1}{2}}^{\frac{1}{2}} P_X(f)\cos(2\pi f)df$$

where $P_X(f)$ is given in Figure 17.17. These can be evaluated numerically by replacing the integrals with approximating sums to yield $r_X[0] = 1.5169 \times 10^{15}$ and $r_X[1] = 4.8483 \times 10^{14}$. Then, using (17.57) and (17.58), we have the AR parameters $a = 0.3196$ and $\sigma_U^2 = 1.362 \times 10^{15}$. With these parameters the AR PSD (see (17.36)) and the true PSD (shown in Figure 17.17) are plotted in Figure 17.18. The agreement between them is fairly good except near $f = 0$. Hence, with these values of the parameters a random process realization could be synthesized within a digital computer and then converted to analog form to drive the shaker table.

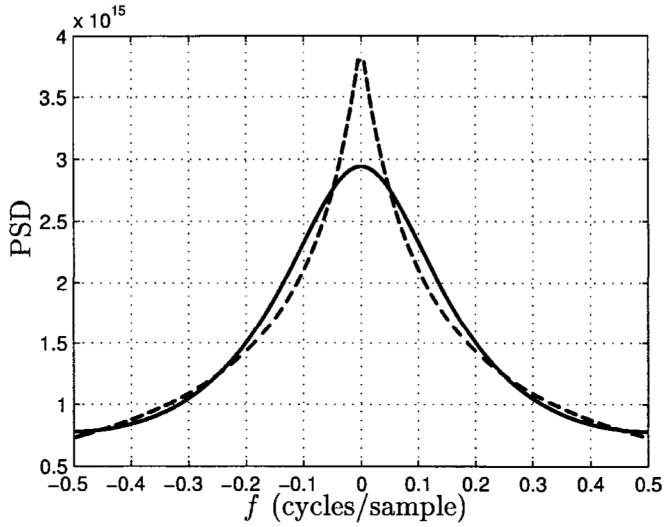


Figure 17.18: Discrete-time PSD and its AR PSD model for TBL noise. The true PSD is shown as the dashed line and the AR PSD model as the solid line.

References

Bell Telephone Laboratories, *Transmission Systems for Communications*, Western Electric Co, Winston-Salem, NC, 1970.

Box, E.P.G., G.M. Jenkins, *Time Series Analysis, Forecasting and Control*, Holden-Day, San Francisco, 1970.

Brockwell, P.J., R.A. Davis, *Time Series: Theory and Methods*, Springer-Verlag, New York, 1987.

Harvey, A.C., *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press, New York, 1989.

Jackson, L.B., *Signals, Systems, and Transforms*, Addison-Wesley, Reading, MA 1991.

Jenkins, G.M., D.G. Watts, *Spectral Analysis and Its Applications*, Holden-Day, San Francisco, 1968.

Kay, S., *Modern Spectral Estimation, Theory and Applications*, Prentice-Hall, Englewood Cliffs, NJ, 1988.

McConnell, K.G., *Vibration Testing*, John Wiley & Sons, New York, 1995.

NASA, *Dynamic Environmental Criteria, NASA Technical Handbook*, pp. 50–51, NASA-HDBK-7005, March 13, 2001.

Papoulis, A., *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, New York, 1965.

Robinson, E.A., *Multichannel Time Series Analysis with Digital Computer Programs*, Holden-Day, San Francisco, 1967.

Schuster, A., "On the Investigation of Hidden Periodicities with Applications to a Supposed 26 Day Period of Meterological Phenomena," *Terr. Magn.*, Vol. 3, pp. 13–41, March 1898.

Schwartz, M., L. Shaw, *Signal Processing, Discrete Spectral Analysis, Detection, and Estimation*, McGraw-Hill, New York, 1975.

Taylor, S., *Modelling Financial Time Series*, John Wiley & Sons, New York, 1986.

Tong, H., *Non-linear Time Series*, Oxford University Press, New York, 1990.

Problems

17.1 (☺) (w) A Bernoulli random process $X[n]$ for $-\infty < n < \infty$ consists of independent random variables with each random variable taking on the values $+1$ and -1 with probabilities p and $1-p$, respectively. Is this random process WSS? If it is WSS, find its mean sequence and autocorrelation sequence.

17.2 (w) Consider the random process defined as $X[n] = a_0 U[n] + a_1 U[n-1]$ for $-\infty < n < \infty$, where a_0 and a_1 are constants, and $U[n]$ is an IID random process with each $U[n]$ having a mean of zero and a variance of one. Is this random process WSS? If it is WSS, find its mean sequence and autocorrelation sequence.

17.3 (w) A sinusoidal random process is defined as $X[n] = A\cos(2\pi f_0 n)$ for $-\infty < n < \infty$, where $0 < f_0 < 0.5$ is a discrete-time frequency, and $A \sim \mathcal{N}(0,1)$. Is this random process WSS? If it is WSS, find its mean sequence and autocorrelation sequence.

17.4 (f) A WSS random process has $E[X[0]] = 1$ and a covariance sequence $c_X[n_1, n_2] = 2\delta[n_2 - n_1]$. Find the ACS and plot it.

17.5 (☺) (w) A random process $X[n]$ for $-\infty < n < \infty$ consists of independent random variables with

$$X[n] \sim \begin{cases} \mathcal{N}(0,1) & \text{for } n \text{ even} \\ \mathcal{U}(-\sqrt{3}, \sqrt{3}) & \text{for } n \text{ odd.} \end{cases}$$

Is this random process WSS? Is it stationary?

17.6 (w) The random processes $X[n]$ and $Y[n]$ are both WSS. Every sample of $X[n]$ is independent of every sample of $Y[n]$. Is $Z[n] = X[n] + Y[n]$ WSS? If it is WSS, find its mean sequence and autocorrelation sequence.

17.7 (w) The random processes $X[n]$ and $Y[n]$ are both WSS. Every sample of $X[n]$ is independent of every sample of $Y[n]$. Is $Z[n] = X[n]Y[n]$ WSS? If it is WSS, find its mean sequence and autocorrelation sequence.

17.8 (f) For the ACS $r_X[k] = (1/2)^k$ for $k \geq 0$ and $r_X[k] = (1/2)^{-k}$ for $k < 0$, verify that Properties 17.1–17.3 are satisfied.

17.9 (☺) (w) For the sequence $r_X[k] = ab^{|k|}$ for $-\infty < k < \infty$, determine the values of a and b that will result in a valid ACS.

17.10 (w) A periodic WSS random process with period P is defined to be a random process $X[n]$ whose ACS satisfies $r_X[k + P] = r_X[k]$ for all k. An example is the randomly phased sinusoid of Example 17.10 for which $P = 10$. Show that the correlation coefficient for two samples of a *zero mean* periodic random process that are separated by P samples is one. Comment on the predictability of $X[n + P]$ based on $X[n] = x[n]$.

17.11 (w) A WSS random process has an ACS $r_X[k]$ and mean μ. Find the correlation coefficient for two samples of the random process that are separated by k samples.

17.12 (☺) (w) Which of the sequences in Figure 17.19 cannot be valid ACSs? If the sequence cannot be an ACS, explain why not.

17.13 (w) For the randomly phased sinusoid described in Example 17.4 find the optimal linear prediction of $X[1]$ based on observing $X[0] = x[0]$, and also of $X[10]$ based on observing $X[0] = x[0]$. Can either of these samples be perfectly predicted? Explain why or why not.

17.14 (w) For the AR random process described in Example 17.10 find the optimal linear prediction of $X[n_0+k_0]$ based on observing $X[n_0] = x[n_0]$. How accurate is your prediction in terms of MSE as k_0 increases?

17.15 (t) In this problem we derive $r_X[0]$ for the AR random process described in Example 17.5. To do so assume that $X[n]$ can be written as

$$X[n] = \sum_{k=0}^{\infty} a^k U[n-k]. \tag{17.59}$$

This was shown to be true in Example 17.5. Then verify that $r_X[0]$ can be written as

$$r_X[0] = \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} a^k a^l E[U[n-k]U[n-l]]$$

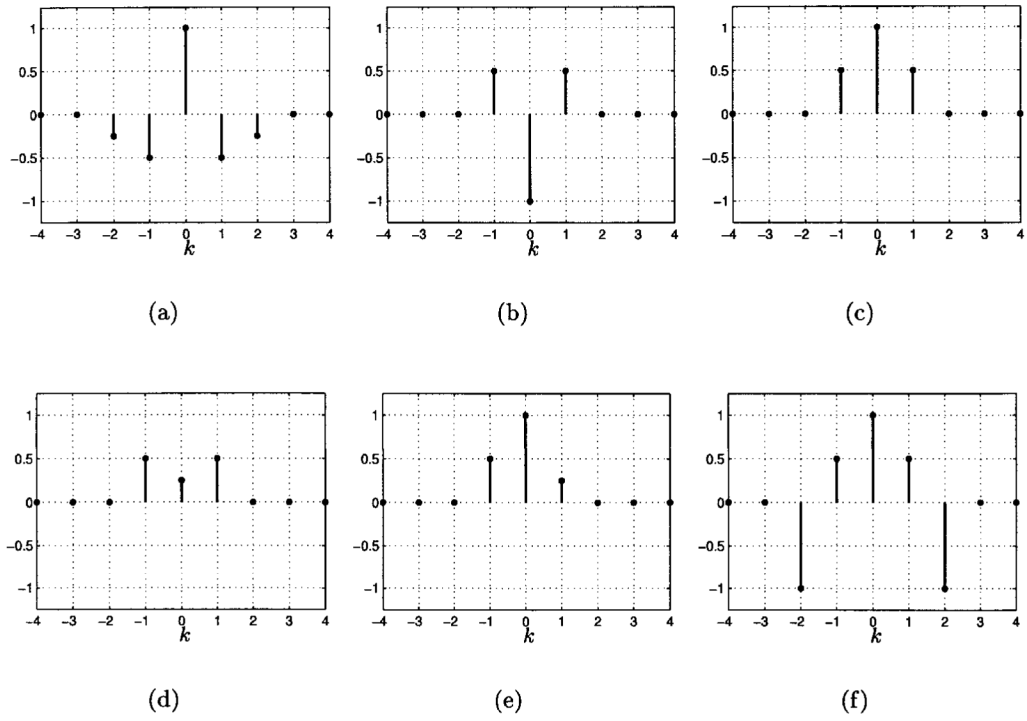


Figure 17.19: Possible ACSs for Problem 17.12.

and use the properties of the $U[n]$ random process to finish the derivation.

17.16 (t) Using a similar approach to the one used in Problem 17.15 derive the ACS for the AR random process described in Example 17.5. Hint: Start with the definition of the ACS and use (17.59).

17.17 (☺) (w) To generate a realization of an AR process on the computer we can use the recursive difference equation $X[n] = aX[n-1] + U[n]$ for $n \geq 0$. However, in doing so, we soon realize that the initial condition $X[-1]$ is required. Assume that we set $X[-1] = 0$ and use the recursion $X[0] = U[0], X[1] = aX[0] + U[1], \ldots$. Determine the mean and variance of $X[n]$ for $n \geq 0$, where as usual $U[n]$ consists of uncorrelated random variables with zero mean and variance σ_U^2. Does the mean depend on n? Does the variance depend on n? What happens as $n \to \infty$? Hint: First show that $X[n]$ can be written as $X[n] = \sum_{k=0}^{n} a^k U[n-k]$ for $n \geq 0$.

17.18 (w) This problem continues Problem 17.17. Instead of letting $X[-1] = 0$, set $X[-1]$ equal to a random variable with mean 0 and a variance of $\sigma_U^2/(1-a^2)$ and that is uncorrelated with $U[n]$ for $n \geq 0$. Find the mean and variance of

$X[0]$. Explain your results and why this makes sense.

17.19 (⌣) **(w)** An example of a sequence that is not positive semidefinite is $r[0] = 1$, $r[-1] = r[1] = -7/8$ and equals zero otherwise. Compute the determinant of the 1×1 principal minor, the 2×2 principal minor, and the 3×3 principal minor of the 3×3 autocorrelation matrix $\mathbf{R}_X$ using these values. Also, plot the discrete-time Fourier transform of $r[k]$. Why do you think the positive semidefinite property is important?

17.20 (⌣) **(w)** For the general MA random process of Example 17.6 show that the process is WSS.

17.21 (f) Use (17.28) to show that the MA random process defined in Example 17.6 is ergodic in the mean.

17.22 (t,f) Show that a WSS random process whose ACS satisfies $r_X[k] = \mu^2$ for $k > k_0 \geq 0$ must be ergodic in the mean.

17.23 (t) Prove (17.28) by using the relationship

$$\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} g[i-j] = \sum_{k=-(N-1)}^{N-1} (N - |k|) g[k].$$

Try verifying this relationship for $N = 3$.

17.24 (f) For the random DC level defined in Example 17.7 prove that $\text{var}(\hat{\mu}_N) = 1$.

17.25 (f) Explain why the randomly phased sinusoid defined in Example 17.4 is ergodic in the mean. Next show that it is ergodic in the ACS in that

$$\lim_{N \to \infty} \hat{r}_X[k] = \lim_{N \to \infty} \frac{1}{N-k} \sum_{n=0}^{N-1-k} X[n]X[n+k] = \frac{1}{2} \cos(2\pi(0.1)k) = r_X[k] \quad k \geq 0$$

by computing $\hat{r}_X[k]$ directly. Hint: Use the fact that $\lim_{N \to \infty} (1/(N-k)) \sum_{n=0}^{N-1-k} \cos(2\pi f n + \phi) = 0$ for any $0 < f < 1$ and any phase angle ϕ. This is because the temporal average of an infinite duration sinusoid is zero.

17.26 (t) Show that the formula

$$\sum_{m=-M}^{M} \sum_{n=-M}^{M} g[m-n] = \sum_{k=-2M}^{2M} (2M + 1 - |k|) g[k]$$

is true for $M = 1$.

17.27 (t) Argue that

$$\lim_{M\to\infty} \sum_{k=-2M}^{2M} \underbrace{\left(1 - \frac{|k|}{2M+1}\right)}_{w[k]} r_X[k] \exp(-j2\pi fk) = \sum_{k=-\infty}^{\infty} r_X[k] \exp(-j2\pi fk)$$

by drawing pictures of $r_X[k]$, which decays to zero, and overlay it with $w[k]$ as M increases.

17.28 (:-) (w) For the differenced random process defined in Example 17.1 determine the PSD. Explain your results.

17.29 (f) Determine the PSD for the randomly phased sinusoid described in Example 17.4. Is this result reasonable? Hint: The discrete-time Fourier transform of $\exp(j2\pi f_0 n)$ for $-1/2 < f_0 < 1/2$ is $\delta(f - f_0)$ over the frequency interval $-1/2 \leq f \leq 1/2$.

17.30 (:-) (w) A random process is defined as $X[n] = AU[n]$, where $A \sim \mathcal{N}(0, \sigma_A^2)$ and $U[n]$ is white noise with variance σ_U^2. The random variable A is independent of all the samples of $U[n]$. Determine the PSD of $X[n]$.

17.31 (w) Find the PSD of the random process $X[n] = (1/2)^{|n|}U[n]$ for $-\infty < n < \infty$, where $U[n]$ is white noise with variance σ_U^2.

17.32 (w) Find the PSD of the random process $X[n] = a_0 U[n] + a_1 U[n-1]$, where a_0, a_1 are constants and $U[n]$ is white noise with variance $\sigma_U^2 = 1$.

17.33 (w) A Bernoulli random process consists of IID Bernoulli random variables taking on values $+1$ and -1 with equal probabilities. Determine the PSD and explain your results.

17.34 (:-) (w) A random process is defined as $X[n] = U[n] + \mu$ for $-\infty < n < \infty$, where $U[n]$ is white noise with variance σ_U^2. Find the ACS and PSD and plot your results.

17.35 (w,c) Consider the AR random process defined in Example 17.5 and described further in Example 17.10 with $-1 < a < 0$ and for some $\sigma_U^2 > 0$. Plot the PSD for several values of a and explain your results.

17.36 (f,c) Plot the corresponding PSD for the ACS

$$r_X[k] = \begin{cases} 1 & k = 0 \\ 1/2 & k = \pm 1 \\ 1/4 & k = \pm 2 \\ 0 & \text{otherwise.} \end{cases}$$

17.37 (w) If a random process has the PSD $P_X(f) = 1 + \cos(2\pi f)$, are the samples of the random process uncorrelated?

17.38 (☺) (f) If a random process has the PSD $P_X(f) = |1 + \exp(-j2\pi f) + (1/2)\exp(-j4\pi f)|^2$, determine the ACS.

17.39 (c) For the AR random processes whose ACSs are shown in Figure 17.6 generate a realization of $N = 2000$ samples for each process. Use the MATLAB code segment given in Section 17.4 to do this. Then, estimate the ACS for $k = 0, 1, \ldots, 30$ and plot the results. Compare your results to those shown in Figure 17.12 and explain.

17.40 (☺) (w) A PSD is given as $P_X(f) = a + b\cos(2\pi f)$ for some constants a and b. What values of a and b will result in a valid PSD?

17.41 (f) A PSD is given as

$$P_X(f) = \begin{cases} 2 - 8x & 0 \leq f \leq 1/4 \\ 0 & 1/4 < f \leq 1/2. \end{cases}$$

Plot the PSD and find the total average power in the random process.

17.42 (☺) (c) Plot 50 realizations of the randomly phased sinusoid described in Example 17.4 with $N = 50$, and overlay the samples in a scatter diagram plot such as shown in Figure 16.15. Explain the results by referring to the PDF of Figure 16.12. . Next estimate the following quantities: $E[X[10]], E[X[12]]$, $E[X[10]X[12]], E[X[12]X[14]]$ by averaging down the ensemble, and compare your simulated results to the theoretical values.

17.43 (c) In this problem we support the results of Problem 17.18 by using a computer simulation. Specifically, generate $M = 10,000$ realizations of the AR random process $X[n] = 0.95X[n-1] + U[n]$ for $n = 0, 1, \ldots, 49$, where $U[n]$ is WGN with $\sigma_U^2 = 1$. Do so two ways: for the first set of realizations let $X[-1] = 0$ and for the second set of realizations let $X[-1] \sim \mathcal{N}(0, \sigma_U^2/(1-a^2))$, using a different random variable for each realization. Now estimate the variance for each sample time n, which is $r_X[0]$, by averaging $X^2[n]$ down the ensemble of realizations. Do you obtain the theoretical result of $r_X[0] = \sigma_U^2/(1-a^2)$?

17.44 (☺) (c) Generate a realization of discrete-time white Gaussian noise with variance $\sigma_X^2 = 1$. For $N = 64$, $N = 128$, and $N = 256$, plot the periodogram. What is the true PSD? Does your estimated PSD get closer to the true PSD as N increases? If not, how could you improve your estimate?

17.45 (c) Generate a realization of an AR random process of length $N = 31,000$ with $a = 0.25$ and $\sigma_U^2 = 1 - a^2$. Break up the data set into 1000 nonoverlapping blocks of data and compute the periodogram for each block. Finally, average

the periodograms together for each point in frequency to determine the final averaged periodogram estimate. Compare your results to the theoretical PSD shown in Figure 17.11a.

17.46 (f) A continuous-time randomly phased sinusoid is defined by $X(t) = \cos(2\pi F_0 t + \Theta)$, where $\Theta \sim \mathcal{U}(0, 2\pi)$. Determine the mean function and ACF for this random process.

17.47 (:)) (f) For the PSD $P_X(F) = \exp(-|F|)$, determine the average power in the band $[10, 100]$ Hz.

17.48 (w) If a PSD is given as $P_X(F) = \exp(-|F/F_0|)$, what happens to the ACF as F_0 increases and also as $F_0 \to \infty$?

17.49 (t) Based on (17.49) derive (17.50), and also based on (17.51) derive (17.52).

17.50 (:)) (w) A continuous-time white noise random process $U(t)$ whose PSD is given as $P_U(F) = N_0/2$ is integrated to form the continuous-time MA random process

$$X(t) = \frac{1}{T} \int_{t-T}^{t} U(\xi) d\xi.$$

Determine the mean function and the variance of $X(t)$. Does $X(t)$ have infinite total average power?

17.51 (:)) (w,c) Consider a continuous-time random process $X(t) = \mu + U(t)$, where $U(t)$ is zero mean and has the ACF given in Figure 17.15. If $X(t)$ is sampled at twice the Nyquist rate, which is $F_s = 4W$, determine the ACS of $X[n]$. Next using (17.28) find the variance of the sample mean estimator $\hat{\mu}_N$ for $N = 20$. Is it half of the variance of the sample mean estimator if we had sampled at the Nyquist rate and used $N = 10$ samples in our estimate? Note that in either case the total length of the data interval in seconds is the same, which is $20/(4W) = 10/(2W)$.

17.52 (f) A PSD is given as

$$P_X(f) = \left| 1 + \frac{1}{2} \exp(-j2\pi f) \right|^2.$$

Model this PSD by using an AR PSD as was done in Section 17.9. Plot the true PSD and the AR model PSD.

Chapter 18

Linear Systems and Wide Sense Stationary Random Processes

18.1 Introduction

Most physical systems are conveniently modeled by a *linear system.* These include electrical circuits, mechanical machines, human biological functions, and chemical reactions, just to name a few. When the system is capable of responding to a continuous-time input, its effect can be described using a linear differential equation. For a system that responds to a discrete-time input a linear difference equation can be used to characterize the effect of the system. Furthermore, for systems whose characteristics do not change with time, the coefficients of the differential or difference equation are constants. Such a system is termed a *linear time invariant* (LTI) system for continuous-time inputs/outputs and a *linear shift invariant* (LSI) system for discrete-time inputs/outputs. In this chapter we explore the effect of these systems on wide sense stationary (WSS) random process inputs. The reader who is unfamiliar with the basic concepts of linear systems should first read Appendix D for a brief introduction. Many excellent books are available to supplement this material [Jackson 1991, Oppenheim, Willsky, and Nawab 1997, Poularikas and Seely 1985]. We will now consider only discrete-time systems and discrete-time WSS random processes. A summary of the analogous concepts for the continuous-time case is given in Section 18.6.

The importance of LSI systems is that they maintain the wide sense stationarity of the random process. That is to say, *if the input to an LSI system is a WSS random process, then the output is also a WSS random process.* The mean and ACS, or equivalently the PSD, however, are modified by the action of the system. We will be able to obtain simple formulas yielding these quantities at the system output. In effect, the linear system modifies the first two moments of the random process but in an easily determined and intuitively pleasing way. This allows us to assess the effect of a linear system on a WSS random process and therefore provides a means

to produce a WSS random process at the output with some desired characteristics. Furthermore, the theory is easily extended to the case of multiple random processes and multiple linear systems as we will see in the next chapter.

18.2 Summary

For the linear shift invariant system shown in Figure 18.1 the output random process is given by (18.2). If the input random process is WSS, then the output random process is also WSS. The output random process has a mean given by (18.9), an ACS given by (18.10), and a PSD given by (18.11). If the input WSS random process is white noise, then the output random process has the ACS of (18.15). In Section 18.4 the PSD is interpreted, using the results of Theorem 18.3.1, as the average power in a narrow frequency band divided by the width of the frequency band. The application of discrete-time linear systems to estimation of samples of a random process is explored in Section 18.5. Generically known as Wiener filtering, there are four separate problems defined, of which the smoothing and prediction problems are solved. For smoothing of a random process signal in noise the estimate is given by (18.20) and the optimal filter has the frequency response of (18.25). A specific application is given in Example 18.4 to estimation of an AR signal that has been corrupted by white noise. The minimum MSE of the optimal Wiener smoother is given by (18.27). One-step linear prediction of a random process sample based on the current and all past samples as given by (18.21) leads to the optimal filter impulse response satisfying the infinite set of linear equations of (18.28). The general solution is summarized in Section 18.5.2 and then illustrated in Example 18.6. For linear prediction based on the current sample and a finite number of past samples the optimal impulse response is given by the solution of the Wiener-Hopf equations of (18.36). The corresponding minimum MSE is given by (18.37). In particular, if the random process is an AR random process of order p, the Wiener-Hopf equations are the same as the Yule-Walker equations of (18.38) and the minimum mean square error equation of (18.37) is the same as for the white noise variance of (18.39). In Section 18.6 the corresponding formulas for a continuous-time random process that is input to a linear time invariant system are summarized. The mean at the output is given by (18.40), the ACF is given by (18.41), and the PSD is given by (18.42). Example 18.7 illustrates the use of these formulas. In Section 18.7 the application of AR random process modeling to speech synthesis is described. In particular, it is shown how a segment of speech can first be modeled, and then how for an actual segment of speech, the parameters of the model can be extracted. The model with its estimated parameters can then be used for speech synthesis.

18.3 Random Process at Output of Linear System

We wish to consider the effect of an LSI system on a discrete-time WSS random process. We will from time to time refer to the linear system as a *filter*, a term that

is synonomous. In Section 18.6 we summarize the results for a continuous-time WSS random process that is input to an LTI system. To proceed, let $U[n]$ be the WSS random process input and $X[n]$ be the random process output of the system. We generally represent an LSI system schematically with its input and output as shown in Figure 18.1. Previously, in Chapters 16 and 17 we have seen several examples

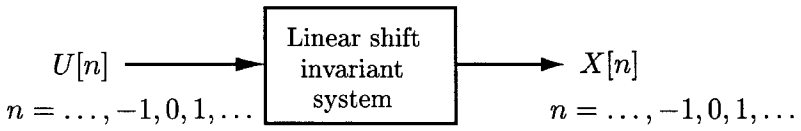


Figure 18.1: Linear shift invariant system with random process input and output.

of LSI systems with WSS random process inputs. One example is the MA random process (see Example 16.7) for which $X[n] = (U[n] + U[n-1])/2$, with $U[n]$ a white Gaussian noise process with variance σ_U^2. (Recall that discrete-time white noise is a zero mean WSS random process with ACS $r_U[k] = \sigma_U^2 \delta[k]$.) We may view the MA random process as the output $X[n]$ of an LSI filter excited at the input by the white Gaussian noise random process $U[n]$. (In this chapter we will be considering only the *first two moments* of $X[n]$. That $U[n]$ is a random process consisting of *Gaussian* random variables is of no consequence to these discussions. The same results are obtained for any white noise random process $U[n]$ irregardless of the marginal PDFs. In Chapter 20, however, we will consider the joint PDF of samples of $X[n]$, and in that case, the fact that $U[n]$ is white *Gaussian* noise will be very important.) The averaging operation can be thought of as a filtering by the LSI filter having an *impulse response*

$$h[k] = \begin{cases} \frac{1}{2} & k = 0 \\ \frac{1}{2} & k = 1 \\ 0 & \text{otherwise.} \end{cases} \tag{18.1}$$

(Recall that the impulse response $h[n]$ is the output of the LSI system when the input $u[n]$ is a unit impulse $\delta[n]$.) This is because the output of an LSI filter is obtained using the convolution sum formula

$$X[n] = \sum_{k=-\infty}^{\infty} h[k]U[n-k] \tag{18.2}$$

so that upon using (18.1) in (18.2) we have

$$\begin{aligned} X[n] &= h[0]U[n] + h[1]U[n-1] \\ &= \frac{1}{2}U[n] + \frac{1}{2}U[n-1] \\ &= \frac{1}{2}(U[n] + U[n-1]). \end{aligned}$$

In general, the LSI system will be specified by giving its impulse response $h[k]$ for $-\infty < k < \infty$ or equivalently by giving its *system function*, which is defined as the z-transform of the impulse response. The system function is thus given by

$$\mathcal{H}(z) = \sum_{k=-\infty}^{\infty} h[k]z^{-k}. \tag{18.3}$$

In addition, we will have need for the *frequency response* of the LSI system, which is defined as the discrete-time Fourier transform of the impulse response. It is therefore given by

$$H(f) = \sum_{k=-\infty}^{\infty} h[k]\exp(-j2\pi fk). \tag{18.4}$$

This function assesses the effect of the system on a complex sinusoidal input sequence $u[n] = \exp(j2\pi f_0 n)$ for $-\infty < n < \infty$. It can be shown that the response of the system to this input is $x[n] = H(f_0)\exp(j2\pi f_0 n) = H(f_0)u[n]$ (use (18.2) with the deterministic input $u[n] = \exp(j2\pi f_0 n)$). Hence, its name derives from the fact that the system action is to modify the amplitude of the complex sinusoid by $|H(f_0)|$ and the phase of the complex sinusoid by $\angle H(f_0)$, but otherwise retains the complex sinusoidal sequence. It should also be noted that the frequency response is easily obtained from the system function as $H(f) = \mathcal{H}(\exp(j2\pi f))$. For the MA random process we have upon using (18.1) in (18.3) that the system function is

$$\mathcal{H}(z) = \frac{1}{2} + \frac{1}{2}z^{-1}$$

and the frequency response is the system function when z is replaced by $\exp(j2\pi f)$, yielding

$$H(f) = \frac{1}{2} + \frac{1}{2}\exp(-j2\pi f).$$

It is said that the system function has been evaluated "on the unit circle in the z-plane".

We next give an example to determine the characteristics of the output random process of an LSI system with a WSS input random process. The previous example is generalized slightly to prepare for the theorem to follow.

Example 18.1 – Output random process characteristics

Let $U[n]$ be a WSS random process with mean μ_U and ACS $r_U[k]$. This random process is input to an LSI system with impulse response

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

This linear system is called a finite impulse response (FIR) filter since its impulse response has only a finite number of nonzero samples. We wish to determine if

a. the output random process is WSS and if so

b. its mean sequence and ACS.

The output of the linear system is from (18.2)

$$X[n] = h[0]U[n] + h[1]U[n-1].$$

The mean sequence is found as

$$\begin{aligned} E[X[n]] &= h[0]E[U[n]] + h[1]E[U[n-1]] \\ &= h[0]\mu_U + h[1]\mu_U \\ &= (h[0] + h[1])\mu_U \end{aligned}$$

so that the mean is constant with time and is given by

$$\mu_X = (h[0] + h[1])\mu_U.$$

It can also be written from (18.4) as

$$\mu_X = \sum_{k=-\infty}^{\infty} h[k] \exp(-j2\pi fk)\Big|_{f=0} \mu_U = H(0)\mu_U.$$

The mean at the output of the LSI system is seen to be modified by the frequency response evaluated at $f = 0$. Does this seem reasonable? Next, if $E[X[n]X[n+k]]$ is found not to depend on n, we will be able to conclude that $X[n]$ is WSS. Continuing we have

$$\begin{aligned} E[X[n]X[n+k]] &= E[(h[0]U[n] + h[1]U[n-1])(h[0]U[n+k] + h[1]U[n+k-1])] \\ &= h^2[0]E[U[n]U[n+k]] + h[0]h[1]E[U[n]U[n+k-1]] \\ &\quad + h[1]h[0]E[U[n-1]U[n+k]] + h^2[1]E[U[n-1]U[n+k-1]] \\ &= (h^2[0] + h^2[1])r_U[k] + h[0]h[1]r_U[k-1] + h[1]h[0]r_U[k+1] \end{aligned}$$

and is seen not to depend on n. Hence, $X[n]$ is WSS and its ACS is

$$r_X[k] = (h^2[0] + h^2[1])r_U[k] + h[0]h[1]r_U[k-1] + h[1]h[0]r_U[k+1]. \qquad (18.5)$$

$\diamondsuit$

Using the previous example for sake of illustration, we next show that the ACS of the output random process of an LSI system can be written as a multiple convolution of sequences. To do so consider (18.5) and let

$$\begin{aligned} g[0] &= h^2[0] + h^2[1] \\ g[1] &= h[0]h[1] \\ g[-1] &= h[1]h[0] \end{aligned}$$

and zero otherwise. Then

$$\begin{aligned} r_X[k] &= g[0]r_U[k] + g[1]r_U[k-1] + g[-1]r_U[k+1] \\ &= \sum_{j=-1}^{1} g[j]r_U[k-j] \\ &= g[k] \star r_U[k] \quad \text{(definition of convolution sum)} \end{aligned} \tag{18.6}$$

where $\star$ denotes convolution. Also, it is easily shown by direct computation that

$$\begin{aligned} g[k] &= \sum_{j=-1}^{0} h[-j]h[k-j] \\ &= h[-k] \star h[k] \end{aligned} \tag{18.7}$$

and therefore from (18.6) and (18.7) we have the final result

$$\begin{aligned} r_X[k] &= (h[-k] \star h[k]) \star r_U[k] \\ &= h[-k] \star h[k] \star r_U[k]. \end{aligned} \tag{18.8}$$

The parentheses can be omitted in (18.8) since the order in which the convolutions are carried out is immaterial (due to associative and commutative property of convolution).

To find the PSD of $X[n]$ we note from (18.4) that the Fourier transform of the impulse response is the frequency response and therefore

$$\begin{aligned} \mathcal{F}\{h[k]\} &= H(f) \\ \mathcal{F}\{h[-k]\} &= H^*(f) \end{aligned}$$

where $\mathcal{F}$ indicates the discrete-time Fourier transform. Fourier transforming (18.8) produces

$$P_X(f) = H^*(f)H(f)P_U(f)$$

or finally we have

$$P_X(f) = |H(f)|^2 P_U(f).$$

This is the fundamental relationship for the PSD at the output of an LSI system—*the output PSD is the input PSD multiplied by the magnitude-squared of the frequency response.* We summarize the foregoing results in a theorem.

Theorem 18.3.1 (Random Process Characteristics at LSI System Output) *If a WSS random process $U[n]$ with mean μ_U and ACS $r_U[k]$ is input to an LSI system which has an impulse response $h[k]$ and frequency response $H(f)$, then the output random process $X[n] = \sum_{k=-\infty}^{\infty} h[k]U[n-k]$ is also WSS and*

$$\mu_X = \sum_{k=-\infty}^{\infty} h[k]\mu_U = H(0)\mu_U \tag{18.9}$$

$$r_X[k] = h[-k] \star h[k] \star r_U[k] \tag{18.10}$$

$$P_X(f) = |H(f)|^2 P_U(f). \tag{18.11}$$

Proof: The mean sequence at the output is

$$\begin{aligned}\mu_X[n] &= E[X[n]] = E\left[\sum_{k=-\infty}^{\infty} h[k]U[n-k]\right] \\ &= \sum_{k=-\infty}^{\infty} h[k]E[U[n-k]] \\ &= \sum_{k=-\infty}^{\infty} h[k]\mu_U = H(0)\mu_U \qquad (U[n] \text{ is WSS})\end{aligned}$$

and is not dependent on n. To determine if an ACS can be defined, we consider $E[X[n]X[n+k]]$. This becomes

$$\begin{aligned}E[X[n]X[n+k]] &= E\left[\sum_{i=-\infty}^{\infty} h[i]U[n-i]\sum_{j=-\infty}^{\infty} h[j]U[n+k-j]\right] \\ &= \sum_{i=-\infty}^{\infty}\sum_{j=-\infty}^{\infty} h[i]h[j]\underbrace{E[U[n-i]U[n+k-j]]}_{r_U[k-j+i]}\end{aligned}$$

since $U[n]$ was assumed to be WSS. It is seen that there is no dependence on n and hence $X[n]$ is WSS. The ACS is

$$\begin{aligned}r_X[k] &= \sum_{i=-\infty}^{\infty}\sum_{j=-\infty}^{\infty} h[i]h[j]r_U[(k+i)-j] \\ &= \sum_{i=-\infty}^{\infty} h[i]\underbrace{\sum_{j=-\infty}^{\infty} h[j]r_U[(k+i)-j]}_{g[k+i]}\end{aligned}$$

where

$$g[m] = h[m] \star r_U[m]. \tag{18.12}$$

Now we have

$$\begin{aligned}r_X[k] &= \sum_{i=-\infty}^{\infty} h[i]g[k+i] \\ &= \sum_{l=-\infty}^{\infty} h[-l]g[k-l] \qquad (\text{let } l=-i) \\ &= h[-k] \star g[k].\end{aligned}$$

But from (18.12) $g[k] = h[k] \star r_U[k]$ and therefore

$$\begin{aligned}r_X[k] &= h[-k] \star (h[k] \star r_U[k]) \\ &= h[-k] \star h[k] \star r_U[k]\end{aligned} \tag{18.13}$$

due to the associate and commutative properties of convolution. The last result of (18.11) follows by taking the Fourier transform of (18.13) and noting that $\mathcal{F}\{h[-k]\} = H^*(f)$.

△

A special case of particular interest occurs when the input to the system is white noise. Then using $P_U(f) = \sigma_U^2$ in (18.11), the output PSD becomes

$$P_X(f) = |H(f)|^2 \sigma_U^2. \tag{18.14}$$

Using $r_U[k] = \sigma_U^2 \delta[k]$ in (18.10), the output ACS becomes

$$r_X[k] = h[-k] \star h[k] \star \sigma_U^2 \delta[k]$$

and noting that $h[k] \star \delta[k] = h[k]$

$$\begin{aligned} r_X[k] &= \sigma_U^2 h[-k] \star h[k] \\ &= \sigma_U^2 \sum_{i=-\infty}^{\infty} h[-i]h[k-i]. \end{aligned}$$

Finally, letting $m = -i$ we have the result

$$r_X[k] = \sigma_U^2 \sum_{m=-\infty}^{\infty} h[m]h[m+k] \qquad -\infty < k < \infty. \tag{18.15}$$

This formula is useful for determining the output ACS, as is illustrated next.

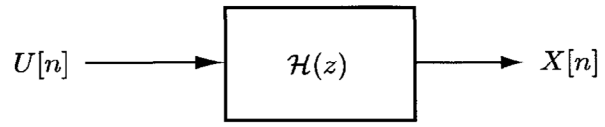
Example 18.2 – AR random process

In Examples 17.5 and 17.10 we derived the ACS and PSD for an AR random process. We now rederive these quantities using the linear systems concepts just described. Recall that an AR random process is defined as $X[n] = aX[n-1] + U[n]$ and can be viewed as the output of an LSI filter with system function

$$\mathcal{H}(z) = \frac{1}{1 - az^{-1}}$$

with white Gaussian noise $U[n]$ at the input. This is shown in Figure 18.2 and follows from the definition of the system function $\mathcal{H}(z)$ as the z-transform of the output sequence divided by the z-transform of the input sequence. To see this let $u[n]$ be a deterministic input sequence with z-transform $\mathcal{U}(z)$ and $x[n]$ be the corresponding deterministic output sequence with z-transform $\mathcal{X}(z)$. Then we have by the definition of the system function

$$\mathcal{H}(z) = \frac{\mathcal{X}(z)}{\mathcal{U}(z)}$$



$$\mathcal{H}(z) = \frac{1}{1-az^{-1}}$$

Figure 18.2: Linear system model for AR random process. The input random process $U[n]$ is white Gaussian noise with variance σ_U^2.

and therefore for the given system function

$$\begin{aligned} \mathcal{X}(z) &= \mathcal{H}(z)\mathcal{U}(z) \\ &= \frac{1}{1-az^{-1}}\mathcal{U}(z). \end{aligned}$$

Thus,

$$\mathcal{X}(z) - az^{-1}\mathcal{X}(z) = \mathcal{U}(z)$$

and taking the inverse z-transform yields the recursive difference equation

$$x[n] - ax[n-1] = u[n] \tag{18.16}$$

which is equivalent to our AR random process definition when the input and output sequences are replaced by random processes.

The output PSD is now found by using (18.14) to yield

$$\begin{aligned} P_X(f) &= |\mathcal{H}(\exp(j2\pi f))|^2\sigma_U^2 \\ &= \frac{\sigma_U^2}{|1-a\exp(-j2\pi f)|^2} \end{aligned} \tag{18.17}$$

which agrees with our previous results. To determine the ACS we can either take the inverse Fourier transform of (18.17) or use (18.15). The latter approach is generally easier. To find the impulse response we can use (18.16) with the input set to $\delta[n]$ so that the output is by definition $h[n]$. Since the LSI system is assumed to be causal, we need to determine the solution of the difference equation $h[n] = ah[n-1] + \delta[n]$ for $n \geq 0$ with initial condition $h[-1] = 0$. The reason that the initial condition is set equal to zero is *our assumption that the LSI system is causal.* A causal system cannot produce an output which is nonzero, in this case $h[-1]$, before the input is applied, in this case at $n = 0$ since the input is $\delta[n]$. This produces $h[n] = a^n u_s[n]$, where we now use $u_s[n]$ to denote the unit step in order to avoid confusion with the random process realization $u[n]$ (see Appendix D.3). Thus, (18.15) becomes for

$k \geq 0$

$$\begin{aligned} r_X[k] &= \sigma_U^2 \sum_{m=-\infty}^{\infty} a^m u_s[m] a^{m+k} u_s[m+k] \\ &= \sigma_U^2 a^k \sum_{m=0}^{\infty} a^{2m} \qquad (m \geq 0 \text{ and } m+k \geq 0 \text{ for nonzero term in sum}) \\ &= \sigma_U^2 \frac{a^k}{1-a^2} \qquad (\text{since } |a| < 1) \end{aligned}$$

and therefore for all k

$$r_X[k] = \sigma_U^2 \frac{a^{|k|}}{1-a^2}.$$

Again the ACS is in agreement with our previous results. Note that the linear system shown in Figure 18.2 is called an *infinite impulse response* (IIR) filter. This is because the impulse response $h[n] = a^n u_s[n]$ is infinite in length.

Fourier and z-transforms of WSS random process don't exist.

To determine the system function in the previous example we assumed the input to the linear system was a deterministic sequence $u[n]$. The corresponding output $x[n]$, therefore, was also a deterministic sequence. This is because formally the z-transform (and also the Fourier transform) cannot exist for a WSS random process. Existence requires the sequence to decay to zero as time becomes large. But of course if the random process is WSS, then we know that $E[X^2[n]]$ is constant as $n \to \pm\infty$ and so we cannot have $|X[n]| \to 0$ as $n \to \pm\infty$.

Example 18.3 – MA random process

In Example 17.3 we derived the ACS for an MA random process. We now show how to accomplish this more easily using (18.15). Recall the definition of the MA random process in Example 17.3 as $X[n] = (U[n] + U[n-1])/2$, with $U[n]$ being white Gaussian noise. This may be interpreted as the output of an LSI filter with white Gaussian noise at the input. In fact, it should now be obvious that the system function is $\mathcal{H}(z) = 1/2 + (1/2)z^{-1}$ and therefore the impulse response is $h[m] = 1/2$

for $m = 0, 1$ and zero otherwise. Using (18.15) we have

$$\begin{aligned} r_X[k] &= \sigma_U^2 \sum_{m=-\infty}^{\infty} h[m]h[m+k] \\ &= \sigma_U^2 \sum_{m=0}^{1} h[m]h[m+k] \end{aligned}$$

and so for $k \geq 0$

$$r_X[k] = \begin{cases} \sigma_U^2 \sum_{m=0}^{1} h^2[m] & k = 0 \\ \sigma_U^2 \sum_{m=0}^{1} h[m]h[m+1] & k = 1 \\ 0 & k \geq 2. \end{cases}$$

Finally, we have

$$r_X[k] = \begin{cases} \sigma_U^2[(\frac{1}{2})^2 + (\frac{1}{2})^2] = \sigma_U^2/2 & k = 0 \\ \sigma_U^2(\frac{1}{2})(\frac{1}{2}) = \sigma_U^2/4 & k = 1 \\ 0 & k \geq 2 \end{cases}$$

which is the same as previously obtained.

♢

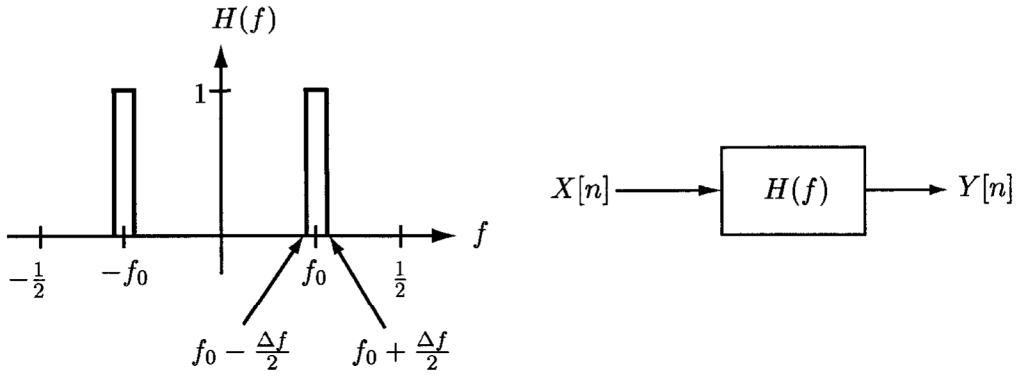
18.4 Interpretation of the PSD

We are now in a position to prove that *the PSD, when integrated over a band of frequencies yields the average power within that band.* In doing so, the PSD may then be interpreted as the average power per unit frequency. We next consider a method to measure the average power of a WSS random process within a very narrow band of frequencies. To do so we filter the random process with an ideal narrowband filter whose frequency response is

$$H(f) = \begin{cases} 1 & -f_0 - \frac{\Delta f}{2} \leq f \leq -f_0 + \frac{\Delta f}{2},\ f_0 - \frac{\Delta f}{2} \leq f \leq f_0 + \frac{\Delta f}{2} \\ 0 & \text{otherwise} \end{cases}$$

and which is shown in Figure 18.3a. The width of the passband of the filter Δf is assumed to be very small. If a WSS random process $X[n]$ is input to this filter, then the output WSS random process $Y[n]$ will be composed of frequency components within the Δf frequency band, the remaining ones having been "filtered out". The total average power in the output random process $Y[n]$ (which is WSS by Theorem 18.3.1) is $r_Y[0]$ and represents the sum of the average powers in $X[n]$ within the bands $[-f_0 - \Delta f/2, -f_0 + \Delta f/2]$ and $[f_0 - \Delta f/2, f_0 + \Delta f/2]$. It can be found from

$$r_Y[0] = \int_{-\frac{1}{2}}^{\frac{1}{2}} P_Y(f)df \qquad \text{(from (17.38))}.$$



(a) (b)

Figure 18.3: Narrowband filtering of random process to measure power within a band of frequencies.

Now using (18.11) and the definition of the narrowband filter frequency response we have

$$\begin{aligned} r_Y[0] &= \int_{-\frac{1}{2}}^{\frac{1}{2}} P_Y(f)df \\ &= \int_{-\frac{1}{2}}^{\frac{1}{2}} |H(f)|^2 P_X(f)df \qquad \text{(from (18.11))} \\ &= \int_{-f_0-\Delta f/2}^{-f_0+\Delta f/2} 1 \cdot P_X(f)df + \int_{f_0-\Delta f/2}^{f_0+\Delta f/2} 1 \cdot P_X(f)df \\ &= 2\int_{f_0-\Delta f/2}^{f_0+\Delta f/2} 1 \cdot P_X(f)df \qquad \text{(since } P_X(-f) = P_X(f)\text{)}. \end{aligned}$$

If we let $\Delta f \to 0$, so that $P_X(f) \to P_X(f_0)$ within the integration interval, this becomes approximately

$$r_Y[0] = 2P_X(f_0)\Delta f$$

or

$$P_X(f_0) = \frac{1}{2}\frac{r_Y[0]}{\Delta f}.$$

Since $r_Y[0]$ is the total average power due to the frequency components within the bands shown in Figure 18.3a, which is twice the total average power in the positive frequency band, we have that

$$P_X(f_0) = \frac{\text{Total average power in band } [f_0 - \Delta f/2, f_0 + \Delta f/2]}{\Delta f}. \qquad (18.18)$$

This says that *the PSD $P_X(f_0)$ is the average power of $X[n]$ in a small band of frequencies about $f = f_0$ divided by the width of the band.* It justifies the name of power spectral *density.* Furthermore, to obtain the average power within a frequency band from knowledge of the PSD, we can reverse (18.18) to obtain

$$\text{Total average power in band } [f_0 - \Delta f/2, f_0 + \Delta f/2] = P_X(f_0)\Delta f$$

which is the *area* under the PSD curve. More generally, we have for an arbitrary frequency band

$$\text{Total average power in band } [f_1, f_2] = \int_{f_1}^{f_2} P_X(f)df$$

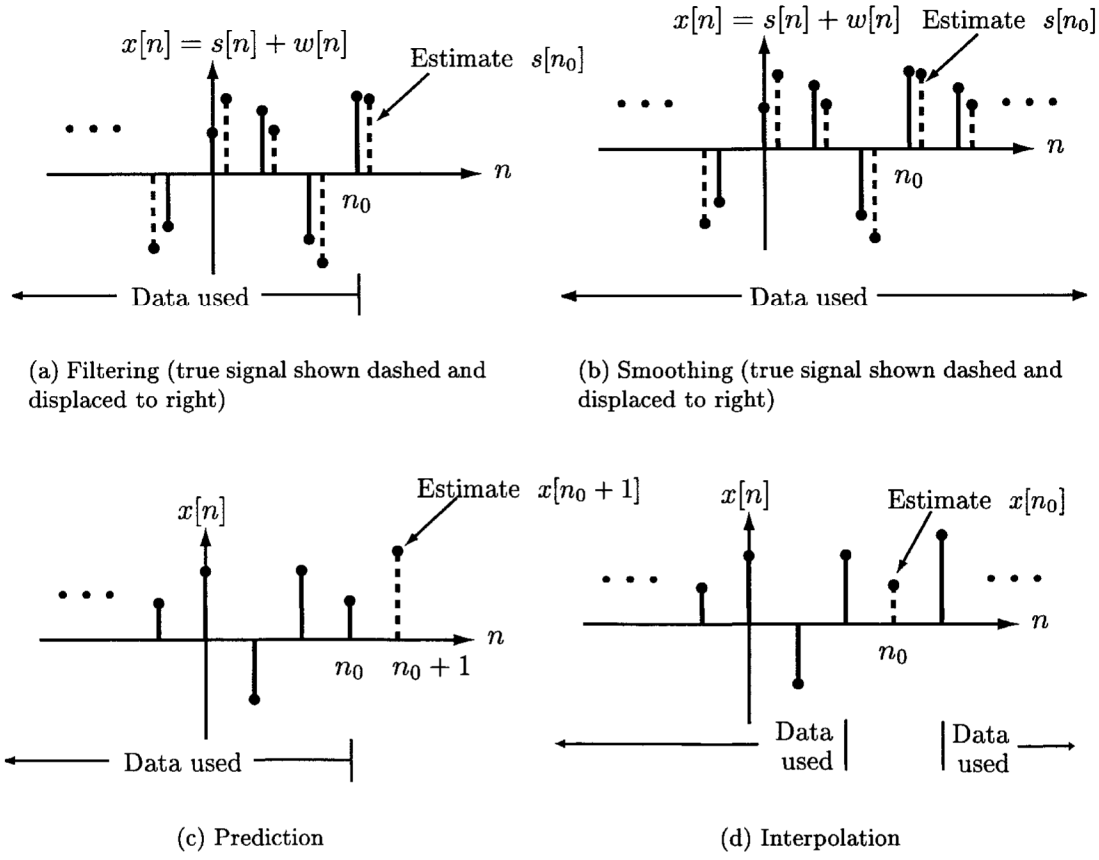
which was previously asserted.

18.5 Wiener Filtering

Armed with the knowledge of the mean and ACS or equivalently the mean and PSD of a WSS random process, there are several important problems that can be solved. Because the required knowledge consists of only the first two moments of the random process (which in practice can be estimated), the solutions to these problems have found widespread application. The generic approach that results is termed *Wiener filtering*, although there are actually four slightly different problems and corresponding solutions. These problems are illustrated in Figure 18.4 and are referred to as *filtering*, *smoothing*, *prediction*, and *interpolation* [Wiener 1949]. In the filtering problem (see Figure 18.4a) it is assumed that a signal $S[n]$ has been corrupted by additive noise $W[n]$ so that the observed random process is $X[n] = S[n] + W[n]$. It is desired to estimate $S[n]$ by filtering $X[n]$ with an LSI filter having an impulse response $h[k]$. The filter will hopefully reduce the noise but pass the signal. The filter estimates a particular sample of the signal, say $S[n_0]$, by processing the current data sample $X[n_0]$ and the past data samples $\{X[n_0-1], X[n_0-2], \ldots\}$. Hence, the filter is assumed to be causal with an impulse response $h[k] = 0$ for $k < 0$. This produces the estimator

$$\hat{S}[n_0] = \sum_{k=0}^{\infty} h[k]X[n_0 - k] \tag{18.19}$$

which depends on the current sample, containing the signal sample of interest, and past observed data samples. Presumably, the past signal samples are correlated with the present signal sample and hence the use of past samples of $X[n]$ should enhance the estimation performance. This type of processing is called *filtering* and can be implemented in *real time.*



(a) Filtering (true signal shown dashed and displaced to right)

(b) Smoothing (true signal shown dashed and displaced to right)

(c) Prediction

(d) Interpolation

Figure 18.4: Definition of Wiener "filtering" problems.

What are we really estimating here?

In Section 7.9 we attempted to estimate the outcome of a random variable, which was unobserved, based on the outcome of another random variable, which was observed. The correlation between the two random variables allowed us to do this. Here we have essentially the same problem, except that the outcome of interest to us is of the random variable $S[n_0]$. The random variables that are observed are $\{X[n_0], X[n_0-1], \ldots\}$ or we have access to the *realization* (another name for outcome) $\{x[n_0], x[n_0-1], \ldots\}$. Thus, we are attempting to estimate the *realization* of $S[n_0]$ based on the realization $\{x[n_0], x[n_0-1], \ldots\}$. This should be kept in mind since our notation of $\hat{S}[n_0] = \sum_{k=0}^{\infty} h[k]X[n_0-k]$ seems to indicate that we are attempting to estimate a random variable $S[n_0]$ based on other random variables $\{X[n_0], X[n_0-1], \ldots\}$. What we are actually trying to accomplish is a *procedure* of

estimating a realization of a random variable based on realizations of other random variables that will work *for all realizations*. Hence, we employ the capital letter notation for random variables to indicate our interest in all realizations and to allow us to employ expectation operations on the random variables.

The second problem is called smoothing (see Figure 18.4b). It differs from filtering in that the filter is not constrained to be causal. Therefore, the estimator becomes

$$\hat{S}[n_0] = \sum_{k=-\infty}^{\infty} h[k]X[n_0 - k] \tag{18.20}$$

where $\hat{S}[n_0]$ now depends on present, past, and *future* samples of $X[n]$. Clearly, this is not realizable in real time but can be approximated if we allow a delay before determining the estimate. The delay is necessary to accumulate the samples $\{X[n_0 + 1], X[n_0 + 2], \ldots\}$ before computing $\hat{S}[n_0]$. Within a digital computer we would store these "future" samples.

For problems three and four we observe samples of the WSS random process $X[n]$ and wish to estimate an unobserved sample. For *prediction*, which is also called *extrapolation* and *forecasting*, we observe the current and past samples $\{X[n_0], X[n_0 - 1], \ldots\}$ and wish to estimate a future sample, $X[n_0 + L]$, for some positive integer L. The prediction is called an *L-step prediction*. We will only consider one-step prediction or $L = 1$ (see Figure 18.4c). The reader should see [Yaglom 1962] for the more general case and also Problem 18.26 for an example. The predictor then becomes

$$\hat{X}[n_0 + 1] = \sum_{k=0}^{\infty} h[k]X[n_0 - k] \tag{18.21}$$

which of course uses a causal filter. For *interpolation* (see Figure 18.4d) we observe samples $\{\ldots, X[n_0 - 1], X[n_0 + 1], \ldots\}$ and wish to estimate $X[n_0]$. The interpolator then becomes

$$\hat{X}[n_0] = \sum_{\substack{k=-\infty \\ k \neq 0}}^{\infty} h[k]X[n_0 - k] \tag{18.22}$$

which is a noncausal filter. For practical implementation of (18.19)–(18.22) we must truncate the impulse responses to some finite number of samples.

To determine the optimal filter impulse responses we adopt the mean square error (MSE) criterion. Estimators that consist of LSI filters whose impulses are chosen to minimize a MSE are generically referred to as *Wiener filters* [Wiener 1949]. Of the four problems mentioned, we will solve the smoothing and prediction problems. The solution for the filtering problem can be found in [Orfanidis 1985] while that for the interpolation problem is described in [Yaglom 1962] (see also Problem 18.27).

18.5.1 Wiener Smoothing

We observe $X[n] = S[n] + W[n]$ for $-\infty < n < \infty$ and wish to estimate $S[n_0]$ using (18.20). It is assumed that $S[n]$ and $W[n]$ are both zero mean WSS random processes with *known ACSs (PSDs)*. Also, since there is usually no reason to assume otherwise, we assume that the signal and noise random processes are uncorrelated. This means that any sample of $S[n]$ is uncorrelated with any sample of $W[n]$ or $E[S[n_1]W[n_2]] = 0$ for all n_1 and n_2. The MSE for this problem is defined as

$$\text{mse} = E[\epsilon^2[n_0]] = E[(S[n_0] - \hat{S}[n_0])^2]$$

where $\epsilon[n_0] = S[n_0] - \hat{S}[n_0]$ is the error. To minimize the MSE we utilize the orthogonality principle described in Section 14.7 which states that the error should be orthogonal, i.e., uncorrelated, with the data. Since the data consists of $X[n]$ for all n, the orthogonality principle produces the requirement

$$E[\epsilon[n_0]X[n_0 - l]] = 0 \qquad -\infty < l < \infty.$$

Thus, we have that

$$\begin{aligned} E[(S[n_0] - \hat{S}[n_0])X[n_0 - l]] &= 0 \\ E\left[\left(S[n_0] - \sum_{k=-\infty}^{\infty} h[k]X[n_0 - k]\right)X[n_0 - l]\right] &= 0 \qquad \text{(from (18.20))} \end{aligned}$$

which results in

$$E[S[n_0]X[n_0 - l]] = \sum_{k=-\infty}^{\infty} h[k]E[X[n_0 - k]X[n_0 - l]]. \tag{18.23}$$

But

$$\begin{aligned} E[S[n_0]X[n_0 - l]] &= E[S[n_0](S[n_0 - l] + W[n_0 - l])] \\ &= E[S[n_0]S[n_0 - l]] \qquad (S[n] \text{ and } W[n] \text{ are} \\ &\qquad\qquad\qquad\qquad \text{uncorrelated and zero mean}) \\ &= r_S[l] \end{aligned}$$

and

$$\begin{aligned} E[X[n_0 - k]X[n_0 - l]] &= E[(S[n_0 - k] + W[n_0 - k])(S[n_0 - l] + W[n_0 - l])] \\ &= E[S[n_0 - k]S[n_0 - l]] + E[W[n_0 - k]W[n_0 - l]] \\ &= r_S[l - k] + r_W[l - k]. \end{aligned}$$

The infinite set of simultaneous linear equations becomes from (18.23)

$$r_S[l] = \sum_{k=-\infty}^{\infty} h[k](r_S[l - k] + r_W[l - k]) \qquad -\infty < l < \infty. \tag{18.24}$$

Note that the equations *do not depend on* n_0 and therefore the solution for the optimal impulse response is the same for any n_0. This is due to the WSS assumption coupled with the LSI assumption for the estimator, which together imply that a shift in the sample to be estimated results in the same filtering operation but shifted. To solve this set of equations we can use transform techniques since the right-hand side of (18.24) is seen to be a discrete-time convolution. It follows then that

$$r_{SX}[l] = h[l] \star (r_S[l] + r_W[l])$$

and taking Fourier transforms of both sides yields

$$P_{SX}(f) = H(f)(P_S(f) + P_W(f))$$

or finally *the frequency response of the optimal Wiener smoothing filter* is

$$H_{\text{opt}}(f) = \frac{P_{SX}(f)}{P_S(f) + P_W(f)}. \tag{18.25}$$

The optimal impulse response can be found by taking the inverse Fourier transform of (18.25). We next give an example.

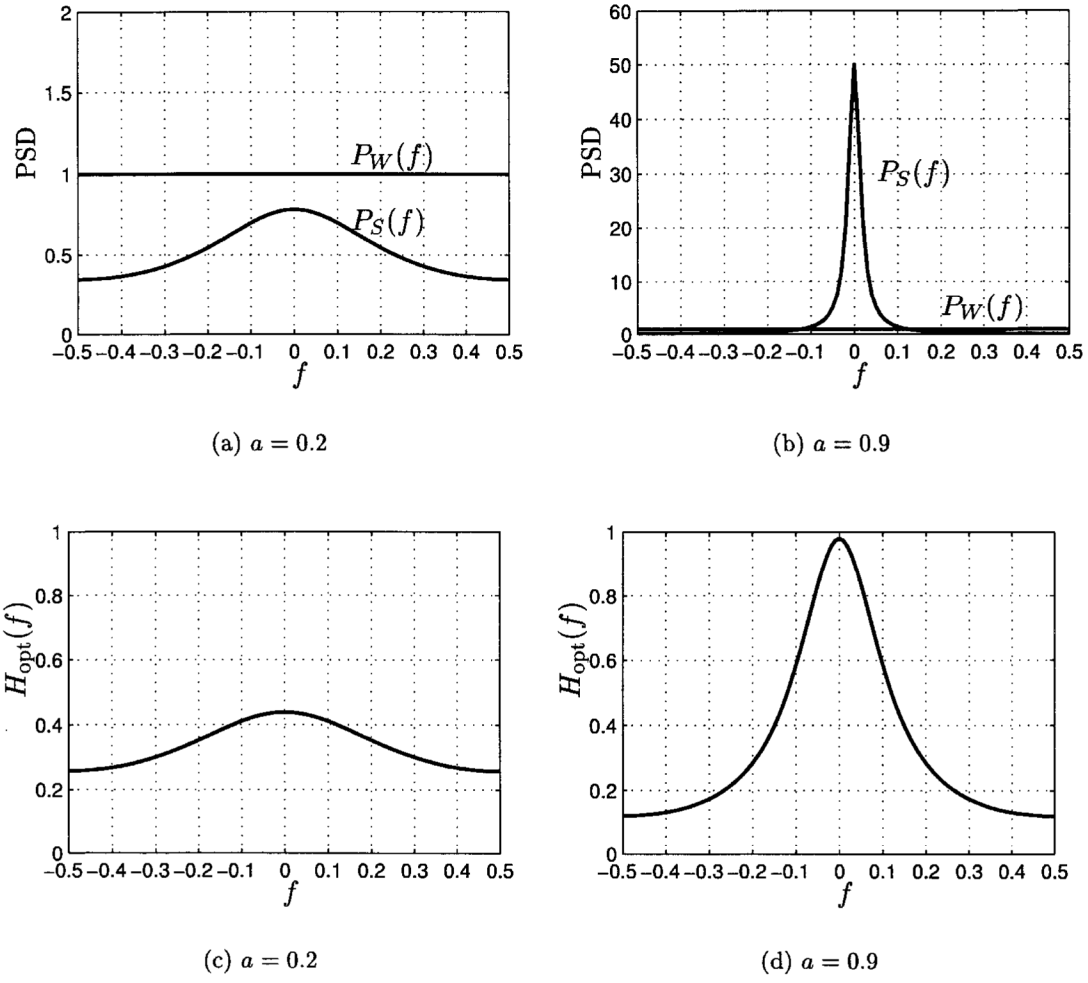
Example 18.4 – Wiener smoother for AR signal in white noise

Consider a signal that is an AR random process corrupted by additive white noise with variance σ_W^2. Then, the PSDs are

$$\begin{aligned} P_S(f) &= \frac{\sigma_U^2}{|1 - a\exp(-j2\pi f)|^2} \\ P_W(f) &= \sigma_W^2. \end{aligned}$$

The PSDs and corresponding Wiener smoother frequency responses are shown in Figure 18.5. In both cases the white noise variance is the same, $\sigma_W^2 = 1$, and the AR input noise variance is the same, $\sigma_U^2 = 0.5$, but the AR filter parameter a has been chosen to yield a wide PSD and a very narrow PSD. As an example, consider the case of $a = 0.9$, which results in a lowpass signal random process as shown in Figure 18.5b. Then, the results of a computer simulation are shown in Figure 18.6. In Figure 18.6a the signal realization $s[n]$ is shown as the dashed curve and the noise corrupted signal realization $x[n]$ is shown as the solid curve. The points have been connected by straight lines for easier viewing. Applying the Wiener smoother results in the estimated signal shown in Figure 18.6b as the solid curve. Once again the true signal realization is shown as dashed. Note that the estimated signal shown in Figure 18.6b exhibits less noise fluctuations but having been smoothed, also exhibits a reduced ability to follow the signal when the signal changes rapidly (see the estimated signal from $n = 25$ to $n = 35$). This is a standard tradeoff in that noise smoothing is obtained at the price of poorer signal following dynamics.

$\diamondsuit$



(a) $a = 0.2$

(b) $a = 0.9$

(c) $a = 0.2$

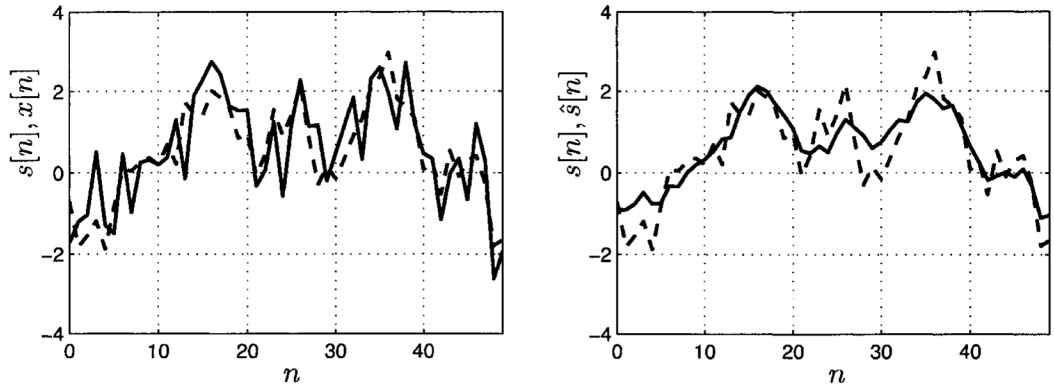
(d) $a = 0.9$

Figure 18.5: Power spectral densities of the signal and noise and corresponding frequency responses of Wiener smoother.

In order to implement the Wiener smoother for the previous example the data was filtered in the frequency domain and converted back into the time domain. This was done using the inverse discrete-time Fourier transform

$$\hat{s}[n] = \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{P_S(f)}{P_S(f) + \sigma_W^2} X_N(f) \exp(j2\pi f n) df \qquad n = 0, 1, \ldots, N-1$$

where $X_N(f)$ is the Fourier transform of the available data $\{x[0], x[1], \ldots, x[N-1]\}$,



(a) True (dashed) and noisy (solid) signal

(b) True (dashed) and estimated (solid) signal

Figure 18.6: Example of Wiener smoother for additive noise corrupted AR signal. The true PSDs are shown in Figure 18.5b. In a) the true signal is shown as the dashed curve and the noisy signal as the solid curve and in b) the true signal is shown as the dashed curve and the Wiener smoothed signal estimate (using the Wiener smoother shown in Figure 18.5d) as the solid curve.

which is

$$X_N(f) = \sum_{n=0}^{N-1} x[n] \exp(-j2\pi fn)$$

($N = 50$ for the previous example). The actual implementation used an inverse FFT to approximate the integral as is shown in the MATLAB code given next. Note that in using the FFT and inverse FFT to calculate the Fourier transform and inverse Fourier transform, respectively, the frequency interval has been changed to $[0, 1]$. Because the Fourier transform is periodic with period one, however, this will not affect the result.

```
clear all
randn('state',0)
a=0.9;varu=0.5;vars=varu/(1-a^2);varw=1;N=50; % set up parameters
for n=0:N-1 % generate signal realization
   nn=n+1;
   if n==0 % use Gaussian random processes
      s(nn,1)=sqrt(vars)*randn(1,1); % initialize first sample
                                     % to avoid transient
   else
      s(nn,1)=a*s(nn-1)+sqrt(varu)*randn(1,1);
```

```
   end
end
x=s+sqrt(varw)*randn(N,1); % add white Gaussian noise
Nfft=1024; % set up FFT length
% compute PSD of signal, frequency interval is [0,1]
Ps=varu./(abs(1-a*exp(-j*2*pi*[0:Nfft-1]'/Nfft)).^2);
Hf=Ps./(Ps+varw); % form Wiener smoother
sestf=Hf.*fft(x,Nfft);  % signal estimate in frequency domain,
                        % frequency interval is [0,1]
sest=real(ifft(sestf,Nfft)); % inverse Fourier transform
```

One can also determine the minimum MSE to assess how well the smoother performs. This is

$$\begin{aligned} \text{mse}_{\min} &= E[(S[n_0] - \hat{S}[n_0])^2] \\ &= E[(S[n_0] - \hat{S}[n_0])S[n_0]] - E[(S[n_0] - \hat{S}[n_0])\hat{S}[n_0]]. \end{aligned}$$

But the second term is zero since by the orthogonality principle

$$\begin{aligned} E[(S[n_0] - \hat{S}[n_0])\hat{S}[n_0]] &= E\left[\epsilon[n_0] \sum_{k=-\infty}^{\infty} h_{\text{opt}}[k]X[n_0 - k]\right] \\ &= \sum_{k=-\infty}^{\infty} h_{\text{opt}}[k] \underbrace{E[\epsilon[n_0]X[n_0 - k]]}_{=0} = 0. \end{aligned}$$

Thus, we have

$$\begin{aligned} \text{mse}_{\min} &= E[(S[n_0] - \hat{S}[n_0])S[n_0]] \\ &= r_S[0] - E\left[\sum_{k=-\infty}^{\infty} h_{\text{opt}}[k]X[n_0 - k]S[n_0]\right] \\ &= r_S[0] - \sum_{k=-\infty}^{\infty} h_{\text{opt}}[k] \underbrace{E[(S[n_0 - k] + W[n_0 - k])S[n_0]]}_{=E[S[n_0-k]S[n_0]]=r_S[k]} \end{aligned}$$

since $S[n_1]$ and $W[n_2]$ are uncorrelated for all n_1 and n_2 and also are zero mean. The minimum MSE becomes

$$\text{mse}_{\min} = r_S[0] - \sum_{k=-\infty}^{\infty} h_{\text{opt}}[k]r_S[k]. \tag{18.26}$$

This can also be written in the frequency domain by using Parseval's theorem to

yield

$$\begin{aligned}
\text{mse}_{\min} &= \int_{-\frac{1}{2}}^{\frac{1}{2}} P_S(f)df - \int_{-\frac{1}{2}}^{\frac{1}{2}} H_{\text{opt}}(f)P_S(f)df \qquad \text{((17.38) and Parseval)} \\
&= \int_{-\frac{1}{2}}^{\frac{1}{2}} (1 - H_{\text{opt}}(f))P_S(f)df \\
&= \int_{-\frac{1}{2}}^{\frac{1}{2}} \left(1 - \frac{P_S(f)}{P_S(f) + P_W(f)}\right) P_S(f)df \\
&= \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{P_W(f)}{P_S(f) + P_W(f)} P_S(f)df
\end{aligned}$$

and finally letting $\rho(f) = P_S(f)/P_W(f)$ be the signal-to-noise ratio in the frequency domain we have

$$\text{mse}_{\min} = \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{P_S(f)}{1 + \rho(f)} df. \tag{18.27}$$

It is seen that the frequency bands for which the contribution to the minimum MSE is largest, are the bands for which the signal-to-noise ratio is smallest or for which $\rho(f) \ll 1$.

18.5.2 Prediction

We consider only the case of $L = 1$ or *one-step prediction.* The more general case can be found in [Yaglom 1962] (see also Problem 18.26). As before, the criterion of MSE is used to design the predictor so that from (18.21)

$$\begin{aligned}
\text{mse} &= E[(X[n_0 + 1] - \hat{X}[n_0 + 1])^2] \\
&= E\left[\left(X[n_0 + 1] - \sum_{k=0}^{\infty} h[k]X[n_0 - k]\right)^2\right]
\end{aligned}$$

is to be minimized over $h[k]$ for $k \geq 0$. Invoking the orthogonality principle leads to the infinite set of simultaneous linear equations

$$E\left[\left(X[n_0 + 1] - \sum_{k=0}^{\infty} h[k]X[n_0 - k]\right) X[n_0 - l]\right] = 0 \qquad l = 0, 1, \ldots .$$

These equations become

$$E[X[n_0 + 1]X[n_0 - l]] = \sum_{k=0}^{\infty} h[k]E[X[n_0 - k]X[n_0 - l]]$$

or finally

$$r_X[l+1] = \sum_{k=0}^{\infty} h[k] r_X[l-k] \qquad l = 0, 1, \ldots . \tag{18.28}$$

Note that once again the optimal impulse response does not depend upon n_0 so that we obtain the same predictor for any sample. Although it appears that we should be able to solve these simultaneous linear equations using the previous Fourier transform approach, this is not so. Because the equations are only valid for $l \geq 0$ and not for $l < 0$, a z-transform cannot be used. Consider forming the z-transform of the left-hand-side as $\sum_{l=0}^{\infty} r_X[l+1] z^{-l}$ and note that it is *not* equal to $z\mathcal{P}(z)$. (See also Problem 18.15 to see what would happen if we blindly went ahead with this approach.)

The minimum MSE is evaluated by using a similar argument as for the Wiener smoother

$$\begin{aligned} \text{mse}_{\min} &= E\left[\left(X[n_0+1] - \sum_{k=0}^{\infty} h_{\text{opt}}[k] X[n_0-k]\right) X[n_0+1]\right] \\ &= r_X[0] - \sum_{k=0}^{\infty} h_{\text{opt}}[k] r_X[k+1] \end{aligned} \tag{18.29}$$

where $h_{\text{opt}}[k]$ is the impulse response solution from (18.28). A simple example for which the equations of (18.28) can be solved is given next.

Example 18.5 – Prediction of AR random process

Consider an AR random process for which the ACS is given by $r_X[k] = (\sigma_U^2/(1-a^2))a^{|k|} = r_X[0]a^{|k|}$. Then from (18.28)

$$r_X[0]a^{|l+1|} = \sum_{k=0}^{\infty} h[k] r_X[0] a^{|l-k|} \qquad l = 0, 1, \ldots$$

and if we let $h[k] = 0$ for $k \geq 1$, we have

$$a^{|l+1|} = h[0]a^{|l|} \qquad l = 0, 1, \ldots .$$

Since $l \geq 0$, the solution is easily seen to be

$$h_{\text{opt}}[0] = \frac{a^{l+1}}{a^l} = a$$

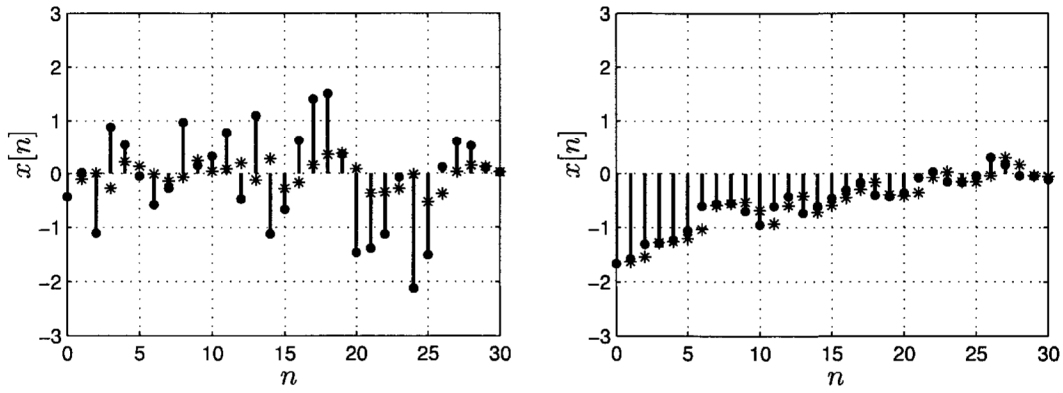
or finally

$$\hat{X}[n_0+1] = aX[n_0].$$

Also, since this is true for any n_0, we can replace the specific sample by a more general sample by replacing n_0 by $n-1$. This results in

$$\hat{X}[n] = aX[n-1]. \tag{18.30}$$

Recalling that the AR random process is defined as $X[n] = aX[n-1] + U[n]$, it is now seen that the optimal one-step linear predictor is obtained from the definition by ignoring the term $U[n]$. This is because $U[n]$ cannot be predicted from the past samples $\{X[n-1], X[n-2], \ldots\}$, which are uncorrelated with $U[n]$ (see also Example 17.5). Furthermore, the prediction error is $\epsilon[n] = X[n] - \hat{X}[n] = X[n] - aX[n-1] = U[n]$. Finally, note that the prediction only depends on the most recent sample and not on the past samples of $X[n]$. In effect, to predict $X[n_0+1]$ all the past information of the random process is embodied in the sample $X[n_0]$. To illustrate the prediction solution consider the AR random process whose parameters and realizations were shown in Figure 17.5. The realizations, along with the one-step predictions, shown as the "*"s, are given in Figure 18.7. Note the good predictions



(a) $a = 0.25$, $\sigma_U^2 = 1 - a^2$ (b) $a = 0.98$, $\sigma_U^2 = 1 - a^2$

Figure 18.7: Typical realizations of autoregressive random process with different parameters and their one-step linear predictions indicated by the "*"s as $\hat{X}[n+1] = ax[n]$.

for the AR random process with $a = 0.98$ but the relatively poor ones for the AR random process with $a = 0.25$. Can you justify these results by comparing the minimum MSEs? (See Problem 18.17.)

$\diamondsuit$

The general solution of (18.28) is fairly complicated. The details are given in Appendix 18A. We now summarize the solution and then present an example.

1. Assume that the z-transform of the ACS, which is

$$\mathcal{P}_X(z) = \sum_{k=-\infty}^{\infty} r_X[k] z^{-k}$$

can be written as

$$\mathcal{P}_X(z) = \frac{\sigma_U^2}{\mathcal{A}(z)\mathcal{A}(z^{-1})} \tag{18.31}$$

where

$$\mathcal{A}(z) = 1 - \sum_{k=1}^{\infty} a[k]z^{-k}.$$

It is required that $\mathcal{A}(z)$ have all its zeros inside the unit circle of the z-plane, i.e., the filter with z-transform $1/\mathcal{A}(z)$ is a stable and causal filter [Jackson 1991].

2. The solution of (18.28) for the impulse response is

$$h_{\text{opt}}[k] = a[k+1] \qquad k = 0, 1, \ldots$$

and the minimum MSE is

$$\text{mse}_{\text{min}} = E[(X[n_0+1] - \hat{X}[n_0+1])^2] = \sigma_U^2.$$

3. The optimal linear predictor becomes from (18.21)

$$\hat{X}[n_0+1] = \sum_{k=0}^{\infty} a[k+1]X[n_0-k] \tag{18.32}$$

and has the minimum MSE, $\text{mse}_{\text{min}} = \sigma_U^2$.

Clearly, the most difficult part of the solution is putting $\mathcal{P}_X(z)$ into the required form of (18.31). In terms of the PSD the requirement is

$$\begin{aligned} P_X(f) = \mathcal{P}_X(\exp(j2\pi f)) &= \frac{\sigma_U^2}{\mathcal{A}(\exp(j2\pi f))\mathcal{A}(\exp(-j2\pi f))} \\ &= \frac{\sigma_U^2}{\mathcal{A}(\exp(j2\pi f))\mathcal{A}^*(\exp(j2\pi f))} \\ &= \frac{\sigma_U^2}{|\mathcal{A}(\exp(j2\pi f))|^2} \\ &= \frac{\sigma_U^2}{|1 - \sum_{k=1}^{\infty} a[k]\exp(-j2\pi fk)|^2}. \end{aligned}$$

But the form of the PSD is seen to be a generalization of the PSD for the AR random process. In fact, if we truncate the sum so that the required PSD becomes

$$P_X(f) = \frac{\sigma_U^2}{\left|1 - \sum_{k=1}^{p} a[k]\exp(-j2\pi fk)\right|^2}$$

then we have the PSD of what is referred to as an AR random process of *order* p, which is also denoted by the symbolism AR(p). In this case, the random process is defined as

$$X[n] = \sum_{k=1}^{p} a[k]X[n-k] + U[n] \tag{18.33}$$

where as usual $U[n]$ is white Gaussian noise with variance σ_U^2. Of course, for $p = 1$ we have our previous definition of the AR random process, which is an AR(1) random process with $a[1] = a$. Assuming an AR(p) random process so that $a[l] = 0$ for $l > p$, the solution for the optimal one-step linear predictor is from (18.32)

$$\hat{X}[n_0 + 1] = \sum_{l=0}^{p-1} a[l+1]X[n_0 - l]$$

and letting $k = l + 1$ produces

$$\hat{X}[n_0 + 1] = \sum_{k=1}^{p} a[k]X[n_0 + 1 - k] \tag{18.34}$$

and the minimum MSE is σ_U^2. Another example follows.

Example 18.6 – One-step linear prediction of MA random process

Consider the zero mean WSS random process given by $X[n] = U[n] - bU[n-1]$, where $|b| < 1$ and $U[n]$ is white Gaussian noise with variance σ_U^2 (also called an MA random process). This random process is a special case of that used in Example 18.1 for which $h[0] = 1$ and $h[1] = -b$ and $U[n]$ is white Gaussian noise. To find the optimal linear predictor we need to put the z-transform of the ACS into the required form. First we determine the PSD. Since the system function is easily shown to be $\mathcal{H}(z) = 1 - bz^{-1}$, the frequency response follows as $H(f) = 1 - b\exp(-j2\pi f)$. From (18.14) the PSD becomes

$$P_X(f) = H(f)H^*(f)\sigma_U^2 = (1 - b\exp(-j2\pi f))(1 - b\exp(j2\pi f))\sigma_U^2$$

and hence replacing $\exp(j2\pi f)$ by z, we have

$$\mathcal{P}_X(z) = (1 - bz^{-1})(1 - bz)\sigma_U^2. \tag{18.35}$$

By equating (18.35) to the required form for $\mathcal{P}_X(z)$ given in (18.31) we have

$$\mathcal{A}(z) = \frac{1}{1 - bz^{-1}}.$$

To convert this to $1 - \sum_{k=1}^{\infty} a[k]z^{-k}$, we take the inverse z-transform, assuming a stable and causal sequence, to yield

$$\mathcal{Z}^{-1}\{\mathcal{A}(z)\} = \begin{cases} b^k & k \geq 0 \\ 0 & k < 0 \end{cases}$$

and so $a[k] = -b^k$ for $k \geq 1$. (Note why $|b| < 1$ is required or else $a[n]$ would not be stable.) The optimal predictor is from (18.32)

$$\begin{aligned}\hat{X}[n_0+1] &= \sum_{k=0}^{\infty} a[k+1]X[n_0-k] \\ &= \sum_{k=0}^{\infty}(-b^{k+1})X[n_0-k] \\ &= -bX[n_0] - b^2X[n_0-1] - b^3X[n_0-2] - \cdots\end{aligned}$$

and the minimum MSE is

$$\text{mse}_{\text{min}} = \sigma_U^2.$$

$\diamondsuit$

As a special case of practical interest, we next consider a *finite length* one-step linear predictor. By finite length we mean that the prediction can only depend on the present sample and past $M-1$ samples. In a derivation similar to the infinite length predictor it is easy to show (see the discussion in Section 14.8 and also Problem 18.20) that if the predictor is given by

$$\hat{X}[n_0+1] = \sum_{k=0}^{M-1} h[k]X[n_0-k]$$

which is just (18.21) with $h[k] = 0$ for $k \geq M$, then the optimal impulse response satisfies the M simultaneous linear equations

$$r_X[l+1] = \sum_{k=0}^{M-1} h[k]r_X[l-k] \qquad l = 0, 1, \ldots, M-1.$$

(If $M \to \infty$, these equations are identical to (18.28)). The equations can be written in vector/matrix form as

$$\underbrace{\begin{bmatrix} r_X[0] & r_X[1] & \cdots & r_X[M-1] \\ r_X[1] & r_X[0] & \cdots & r_X[M-2] \\ \vdots & \vdots & \ddots & \vdots \\ r_X[M-1] & r_X[M-2] & \cdots & r_X[0] \end{bmatrix}}_{\mathbf{R}_X} \begin{bmatrix} h[0] \\ h[1] \\ \vdots \\ h[M-1] \end{bmatrix} = \begin{bmatrix} r_X[1] \\ r_X[2] \\ \vdots \\ r_X[M] \end{bmatrix}. \tag{18.36}$$

The corresponding minimum MSE is given by

$$\text{mse}_{\text{min}} = r_X[0] - \sum_{k=0}^{M-1} h_{\text{opt}}[k]r_X[k+1]. \tag{18.37}$$

These equations are called the *Wiener-Hopf equations*. In general, they must be solved numerically but there are many efficient algorithms to do so [Kay 1988]. The algorithms take advantage of the structure of the matrix which is seen to be an autocorrelation matrix $\mathbf{R}_X$ as first described in Section 17.4. As such, it is symmetric, positive definite, and has the *Toeplitz* property. The Toeplitz property asserts that the elements along each northwest-southeast diagonal are identical. Another important connection between the linear prediction equations and an AR(p) random process is made by letting $M = p$ in (18.36). Then, since for an AR(p) process, we have that $h[n] = a[n+1]$ for $n = 0, 1, \ldots, p-1$ (recall from (18.34) that $\hat{X}[n_0+1] = \sum_{k=1}^{p} a[k]X[n_0+1-k]$) the Wiener-Hopf equations become

$$\begin{bmatrix} r_X[0] & r_X[1] & \cdots & r_X[p-1] \\ r_X[1] & r_X[0] & \cdots & r_X[p-2] \\ \vdots & \vdots & \ddots & \vdots \\ r_X[p-1] & r_X[p-2] & \cdots & r_X[0] \end{bmatrix} \begin{bmatrix} a[1] \\ a[2] \\ \vdots \\ a[p] \end{bmatrix} = \begin{bmatrix} r_X[1] \\ r_X[2] \\ \vdots \\ r_X[p] \end{bmatrix}. \tag{18.38}$$

It is important to note that for *an AR(p) random process*, the optimal one-step linear predictor based on the infinite number of samples $\{X[n_0], X[n_0-1], \ldots\}$ is the same as that based on only the finite number of samples $\{X[n_0], X[n_0-1], \ldots, X[n_0-(p-1)]\}$ [Kay 1988]. The equations of (18.38) are now referred to as the *Yule-Walker equations*. In this form they relate the ACS samples $\{r_X[0], r_X[1], \ldots r_X[p]\}$ to the AR filter parameters $\{a[1], a[2], \ldots, a[p]\}$. If the ACS samples are known, then the AR filter parameters can be obtained by solving the equations. Furthermore, once the filter parameters have been found from (18.38), the variance of the white noise random process $U[n]$ is found from

$$\sigma_U^2 = \text{mse}_{\min} = r_X[0] - \sum_{k=1}^{p} a[k]r_X[k] \tag{18.39}$$

which follows by letting $h_{\text{opt}}[k] = a[k+1]$ with $M = p$ in (18.37). In the real-world example of Section 18.7 we will see how these equations can provide a method to synthesize speech.

18.6 Continuous-Time Definitions and Formulas

For a continuous-time WSS random process as defined in Section 17.8 the linear system of interest is a linear time invariant (LTI) system. It is characterized by its impulse response $h(\tau)$. If a random process $U(t)$ is input to an LTI system with impulse response $h(\tau)$, the output random process $X(t)$ is

$$X(t) = \int_{-\infty}^{\infty} h(\tau)U(t-\tau)d\tau.$$

The integral is referred to as a *convolution integral* and in shorthand notation the output is given by $X(t) = h(t) \star U(t)$. If $U(t)$ is WSS with constant mean μ_U and ACF $r_U(\tau)$, then the output random process $X(t)$ is also WSS. It has a mean function

$$\mu_X = \left(\int_{-\infty}^{\infty} h(\tau) d\tau \right) \mu_U = H(0)\mu_U \tag{18.40}$$

where

$$H(F) = \int_{-\infty}^{\infty} h(\tau) \exp(-j2\pi F\tau) d\tau$$

is the frequency response of the LTI system. The ACF of the output random process $X(t)$ is

$$r_X(\tau) = h(-\tau) \star h(\tau) \star r_U(\tau) \tag{18.41}$$

and therefore the PSD becomes

$$P_X(F) = |H(F)|^2 P_U(F). \tag{18.42}$$

An example follows.

Example 18.7 – Inteference rejection filter

A signal, which is modeled as a WSS random process $S(t)$, is corrupted by an additive interference $I(t)$, which can be modeled as a randomly phased sinusoid with a frequency of $F_0 = 60$ Hz. The corrupted signal is $X(t) = S(t) + I(t)$. It is desired to filter out the interference but if possible, to avoid altering the PSD of the signal due to the filtering. Since the sinusoidal interference has a period of $T = 1/F_0 = 1/60$ seconds, it is proposed to filter $X(t)$ with the differencing filter

$$Y(t) = X(t) - X(t - T). \tag{18.43}$$

The motivation for choosing this type of filter is that a periodic signal with period T will have the same value at any two time instants separated by T seconds. Hence, the difference should be zero for all t. We wish to determine the PSD at the filter output. We will assume that the interference is uncorrelated with the signal. This assumption means that the ACF of $X(t)$ is the sum of the ACFs of $S(t)$ and $I(t)$ and consequently the PSDs sum as well (see Problem 18.33). The differencing filter is an LTI system and so its output can be written as

$$Y(t) = \int_{-\infty}^{\infty} h(\tau) X(t - \tau) d\tau \tag{18.44}$$

for the appropriate choice of the impulse response. The impulse response is obtained by equating (18.44) to (18.43) from which it follows that

$$h(\tau) = \delta(\tau) - \delta(\tau - T) \tag{18.45}$$

as can easily be verified. By taking the Fourier transform, the frequency response becomes

$$\begin{aligned} H(F) &= \int_{-\infty}^{\infty} (\delta(\tau) - \delta(\tau - T)) \exp(-j2\pi F\tau) d\tau \\ &= 1 - \exp(-j2\pi FT). \end{aligned} \tag{18.46}$$

To determine the PSD at the filter output we use (18.42) and note that for the randomly phased sinusoid with amplitude A and frequency F_0, the ACF is (see Problem 17.46)

$$r_I(\tau) = \frac{A^2}{2} \cos(2\pi F_0 \tau)$$

and therefore its PSD, which is the Fourier transform, is given by

$$P_I(F) = \frac{A^2}{4}\delta(F + F_0) + \frac{A^2}{4}\delta(F - F_0).$$

The PSD at the filter input is $P_X(F) = P_S(F) + P_I(F)$ (the PSDs add due to the uncorrelated assumption) and therefore the PSD at the filter output is

$$\begin{aligned} P_Y(F) &= |H(F)|^2 P_X(F) = |H(F)|^2 (P_S(F) + P_I(F)) \\ &= |1 - \exp(-j2\pi FT)|^2 (P_S(F) + P_I(F)). \end{aligned}$$

The magnitude-squared of the frequency response of (18.46) can also be written in real form as

$$|H(F)|^2 = 2 - 2\cos(2\pi FT)$$

and is shown in Figure 18.8. Note that it exhibits zeros at multiples of $F = 1/T =$

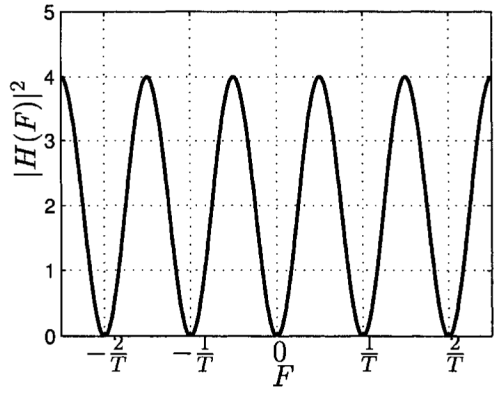


Figure 18.8: Magnitude-squared frequency response of interference canceling filter with $F_0 = 1/T$.

F_0. Hence, $|H(F_0)|^2 = 0$ and so the interfering sinusoid is filtered out. The PSD at the filter output then becomes

$$\begin{aligned} P_Y(F) &= |H(F)|^2 P_S(F) \\ &= 2(1 - \cos(2\pi FT))P_S(F). \end{aligned}$$

Unfortunately, the signal PSD has also been modified. What do you think would happen if the signal were periodic with period $1/(2F_0)$?

$\diamondsuit$

18.7 Real-World Example – Speech Synthesis

It is commonplace to hear computer generated speech when asking for directory assistance in obtaining telephone numbers, in using text to speech conversion programs in computers, and in playing with a multitude of children's toys. One of the earliest applications of computer speech synthesis was the Texas Instruments Speak and Spell[1]. The approach to producing intelligible, if not exactly human sounding, speech, is to mimic the human speech production process. A speech production model is shown in Figure 18.9 [Rabiner and Schafer 1978]. It is well known that speech sounds can be delineated into two classes—*voiced speech* such as a vowel sound and *unvoiced speech* such as a consonant sound. A voiced sound such as "ahhh" (the o in "lot" for example) is produced by the vibration of the vocal cords, while an unvoiced sound such as "sss" (the s in "runs" for example) is produced by passing air over a constriction in the mouth. In either case, the sound is the output of the vocal tract with the difference being the excitation sound and the subsequent filtering of that sound. For voiced sounds the excitation is modeled as a train of impulses to produce a periodic sound while for an unvoiced sound it is modeled as white noise to produce a noise-like sound (see Figure 18.9). The excita-

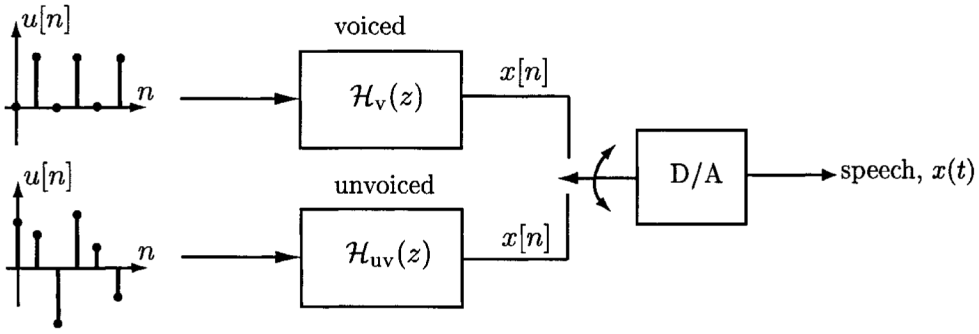


Figure 18.9: Speech production model.

tion is modified by the vocal tract, which can be modeled by an LSI filter. Knowing

[1]Registered trademark of Texas Instruments

the excitation waveform and the vocal tract system function allows us to synthesize speech. For the unvoiced sound we pass discrete white Gaussian noise through an LSI filter with system function $\mathcal{H}_{\text{uv}}(z)$. We next concentrate on the synthesis of unvoiced sounds with the synthesis of voiced sounds being similar.

It has been found that a good model for the vocal tract is the LSI filter with system function

$$\mathcal{H}_{\text{uv}}(z) = \frac{1}{1 - \sum_{k=1}^{p} a[k] z^{-k}}$$

which is an *all-pole filter*. Typically, the order of the filter p, which is the number of poles, is chosen to be $p = 12$. The output of the filter $X[n]$ for a white Gaussian noise random process input $U[n]$ with variance σ_U^2 is given as the WSS random process

$$X[n] = \sum_{k=1}^{p} a[k] X[n-k] + U[n]$$

which is recognized as the defining difference equation for an AR(p) random process. Hence, unvoiced speech sounds can be synthesized using this difference equation for an appropriate choice of the parameters $\{a[1], a[2], \ldots, a[p], \sigma_U^2\}$. The parameters will be different for each unvoiced sound to be synthesized. To determine the parameters for a given sound, a segment of the target speech sound is used to estimate the ACS. Estimation of the ACS was previously described in Section 17.7. Then, the parameters $a[k]$ for $k = 1, 2, \ldots, p$ can be obtained by solving the Yule-Walker equations (same as Wiener-Hopf equations). The theoretical ACS samples required are replaced by estimated ones to yield the set of simultaneous linear equations from (18.38) as

$$\begin{bmatrix} \hat{r}_X[0] & \hat{r}_X[1] & \cdots & \hat{r}_X[p-1] \\ \hat{r}_X[1] & \hat{r}_X[0] & \cdots & \hat{r}_X[p-2] \\ \vdots & \vdots & \ddots & \vdots \\ \hat{r}_X[p-1] & \hat{r}_X[p-2] & \cdots & \hat{r}_X[0] \end{bmatrix} \begin{bmatrix} a[1] \\ a[2] \\ \vdots \\ a[p] \end{bmatrix} = \begin{bmatrix} \hat{r}_X[1] \\ \hat{r}_X[2] \\ \vdots \\ \hat{r}_X[p] \end{bmatrix} \tag{18.47}$$

which are solved to yield the $\hat{a}[k]$'s. Then, the white noise variance estimate is found from (18.39) as

$$\hat{\sigma}_U^2 = \hat{r}_X[0] - \sum_{k=1}^{p} \hat{a}[k] \hat{r}_X[k] \tag{18.48}$$

where $\hat{a}[k]$ is given by the solution of the Yule-Walker equations of (18.47). Hence, we estimate the ACS for lags $k = 0, 1, \ldots, p$ based on an actual speech sound and then solve the equations of (18.47) to obtain $\{\hat{a}[1], \hat{a}[2], \ldots, \hat{a}[p]\}$ and finally, determine $\hat{\sigma}_U^2$ using (18.48). The only modification that is commonly made is to the ACS estimate, which is chosen to be

$$\hat{r}_X[k] = \frac{1}{N} \sum_{n=0}^{N-1-k} x[n] x[n+k] \qquad k = 0, 1, \ldots, p \tag{18.49}$$

and which differs from the one given in Section 17.7 in that the normalizing factor is N instead of $N-k$. For $N \gg p$ this will have minimal effect on the parameter estimates but has the benefit of ensuring a stable filter estimate, i.e., the poles of $\hat{\mathcal{H}}_{\text{uv}}(z)$ will lie inside the unit circle [Kay 1988]. This method of estimating the AR parameters is called the *autocorrelation method of linear prediction.* The entire procedure of modeling speech by an AR(p) model is referred to as *linear predictive coding* (LPC). The name originated with the connection of (18.47) as a set of linear prediction equations, although the ultimate goal here is not linear prediction but speech modeling [Makhoul 1975].

To demonstrate the modeling of an unvoiced sound consider the spoken word "seven" shown in Figure 18.10. A portion of the "sss" utterance is shown in Figure

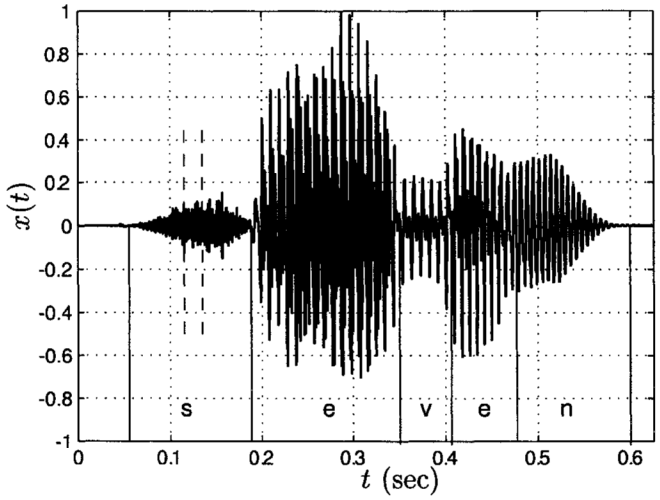


Figure 18.10: Waveform for the utterance "seven" [Allu 2005].

18.11 and as expected is noise-like. It is composed of the samples indicated between the dashed vertical lines in Figure 18.10. Typically, in analyzing speech sounds to estimate its AR parameters, we sample at 8 KHz and use a block of data 20 msec (about 160 samples) in length. The samples of $x(t)$ in Figure 18.10 from $t = 115$ msec to $t = 135$ msec are shown in Figure 18.11. With a model order of $p = 12$ we use (18.49) to estimate the ACS lags and then solve the Yule-Walker equations of (18.47) and also use (18.48) to yield the estimated parameters $\{\hat{a}[1], \hat{a}[2], \ldots, \hat{a}[p], \hat{\sigma}_U^2\}$. If the model is reasonably accurate, then the synthesized sound should be perceived as being similar to the original sound. It has been found through experimentation that if the PSDs are similar, then this will be the case. Hence, the estimated PSD

$$\hat{P}_X(f) = \frac{\hat{\sigma}_U^2}{\left|1 - \sum_{k=1}^{p} \hat{a}[k] \exp(-j2\pi fk)\right|^2} \tag{18.50}$$

should be a good match to the normalized and squared-magnitude of the Fourier

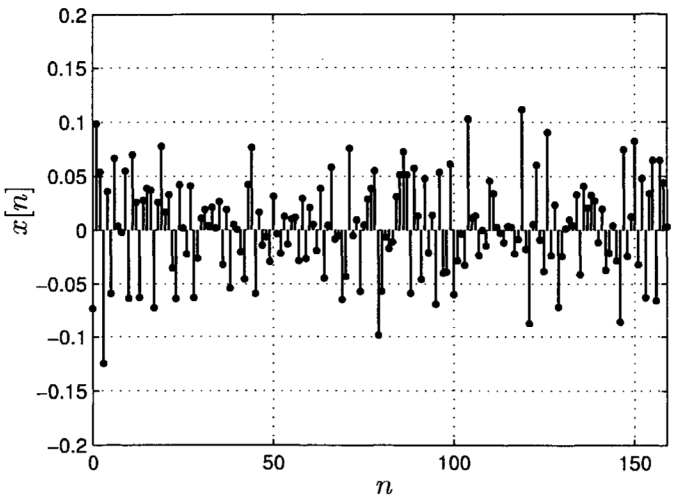


Figure 18.11: A 20 msec segment of the waveform for "sss". See Figure 18.10 for segment extracted as indicated by the vertical dashed lines.

transform of the speech sound. The latter is of course the periodogram. We need only consider the match in power since it is well known that the ear is relatively insensitive to the phase of the speech waveform [Rabiner and Schafer 1978].

As an example, for the portion of the "sss" sound shown in Figure 18.11 a periodogram as well as the AR PSD model of (18.50), is compared in Figure 18.12. Both PSDs are plotted in dB quantities, which is obtained by taking $10\log_{10}$ of the PSD. Note that the resonances, i.e., the portions of the PSD that are large and which are most important for intelligibility, are well matched by the model. This verifies the validity of the AR model. Finally, to synthesize the "sss" sound we compute

$$x[n] = \sum_{k=1}^{p} \hat{a}[k]x[n-k] + u[n]$$

where $u[n]$ is a *pseudorandom* Gaussian noise sequence [Knuth 1981] with variance $\hat{\sigma}_U^2$, for a total of about 20 msec. Then, the samples are converted to an analog sound using a digital-to-analog (D/A) convertor (see Figure 18.9). The TI Speak and Spell used $p = 10$ and stored the AR parameters in memory for each sound. The MATLAB code used to generate Figure 18.12 is given below.

```
N=length(xseg); % xseg is the data shown in Figure 18.11
Nfft=1024; % set up FFT length for Fourier transforms
freq=[0:Nfft-1]'/Nfft-0.5; % PSD frequency points to be plotted
P_per=(1/N)*abs(fftshift(fft(xseg,Nfft))).^2; % compute periodogram
p=12; % dimension of autocorrelation matrix
for k=1:p+1 % estimate ACS for k=0,1,...,p (MATLAB indexes
```

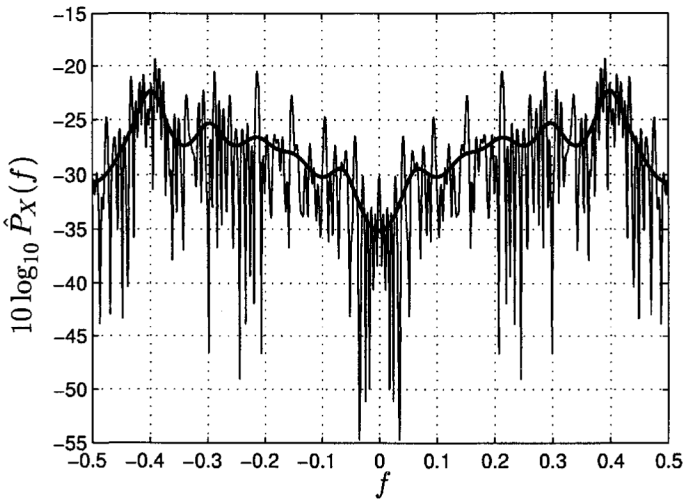


Figure 18.12: Periodogram, shown as the light line, and AR PSD model, shown as the darker line for speech segment of Figure 18.11.

```
            % must start at 1)
   rX(k,1)=(1/N)*sum(xseg(1:N-k+1).*xseg(k:N));
end
r=rX(2:p+1); % fill in right-hand-side vector
for i=1:p % fill in autocorrelation matrix
   for j=1:p
      R(i,j)=rX(abs(i-j)+1);
   end
end
a=inv(R)*r; % solve linear equations to find AR filter parameters
varu=rX(1)-a'*r; % find excitation noise variance
den=abs(fftshift(fft([1;-a],Nfft))).^2; % compute denominator of AR PSD
P_AR=varu./den; % compute AR PSD
```

See also Problem 18.34 for an application of AR modeling to spectral estimation [Kay 1988].

References

Allu, G., personal communication of speech data, 2005.

Jackson, L.B., *Signals, Systems, and Transforms*, Addison-Wesley, Reading, MA, 1991.

Kay, S., *Modern Spectral Estimation: Theory and Application*, Prentice-Hall, Englewood Cliffs, NJ, 1988.

Knuth, D.E., *The Art of Computer Programming, Vol. 2*, Addison-Wesley, Reading, MA, 1981.

Makhoul, J., "Linear Prediction: A Tutorial Review," *IEEE Proceedings*, Vol. 63, pp. 561–580, 1975.

Oppenheim, A.V., A.S. Willsky, S.H. Nawab, *Signal and Systems*, Prentice-Hall, Upper Saddle River, NJ, 1997.

Orfanidis, S.J., *Optimum Signal Processing, An Introduction*, Macmillan, New York, 1985.

Poularikas, A.D., S. Seely, *Signals and Systems*, PWS, Boston, 1985.

Rabiner, L.R., R.W. Schafer, *Digital Processing of Speech Signals*, Prentice-Hall, Englewood Cliffs, NJ, 1978.

Wiener, N., *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*, MIT Press, Cambridge, MA, 1949.

Yaglom, A.M., *An Introduction to the Theory of Stationary Random Functions*, Dover, New York, 1962.

Problems

18.1 (‿) **(f)** An LSI system with system function $\mathcal{H}(z) = 1 - z^{-1} - z^{-2}$ is used to filter a discrete-time white noise random process with variance $\sigma_U^2 = 1$. Determine the ACS and PSD of the output random process.

18.2 (f) A discrete-time WSS random process with mean $\mu_U = 2$ is input to an LSI system with impulse response $h[n] = (1/2)^n$ for $n \geq 0$ and $h[n] = 0$ for $n < 0$. Find the mean sequence at the system output.

18.3 (w) A discrete-time white noise random process $U[n]$ is input to a system to produce the output random process $X[n] = a^{|n|}U[n]$ for $|a| < 1$. Determine the output PSD.

18.4 (‿) **(w)** A randomly phased sinusoid $X[n] = \cos(2\pi(0.25)n + \Theta)$ with $\Theta \sim \mathcal{U}(0, 2\pi)$ is input to an LSI system with system function $\mathcal{H}(z) = 1 - b_1 z^{-1} - b_2 z^{-2}$. Determine the filter coefficients b_1, b_2 so that the sinusoid will have zero power at the filter output.