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

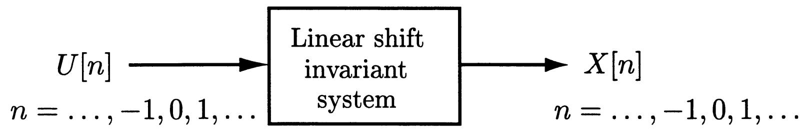


Figure 18.1: Linear shift invariant system with random process input and output.

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

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

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

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

$\diamondsuit$

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 $\mathcal{F}\{h[-k]\} = H^*(f)$.

$\triangle$

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.

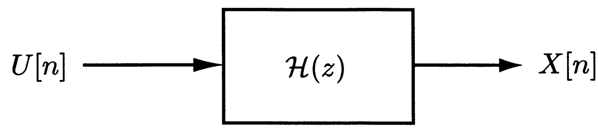
Example 18.2 – AR random process

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

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

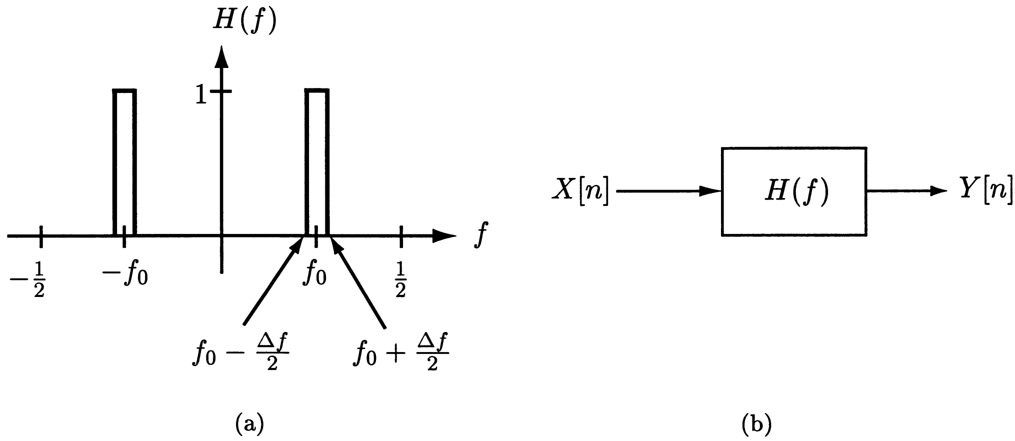


Figure 18.3: Narrowband filtering of random process to measure power within a band of frequencies.

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

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

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

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

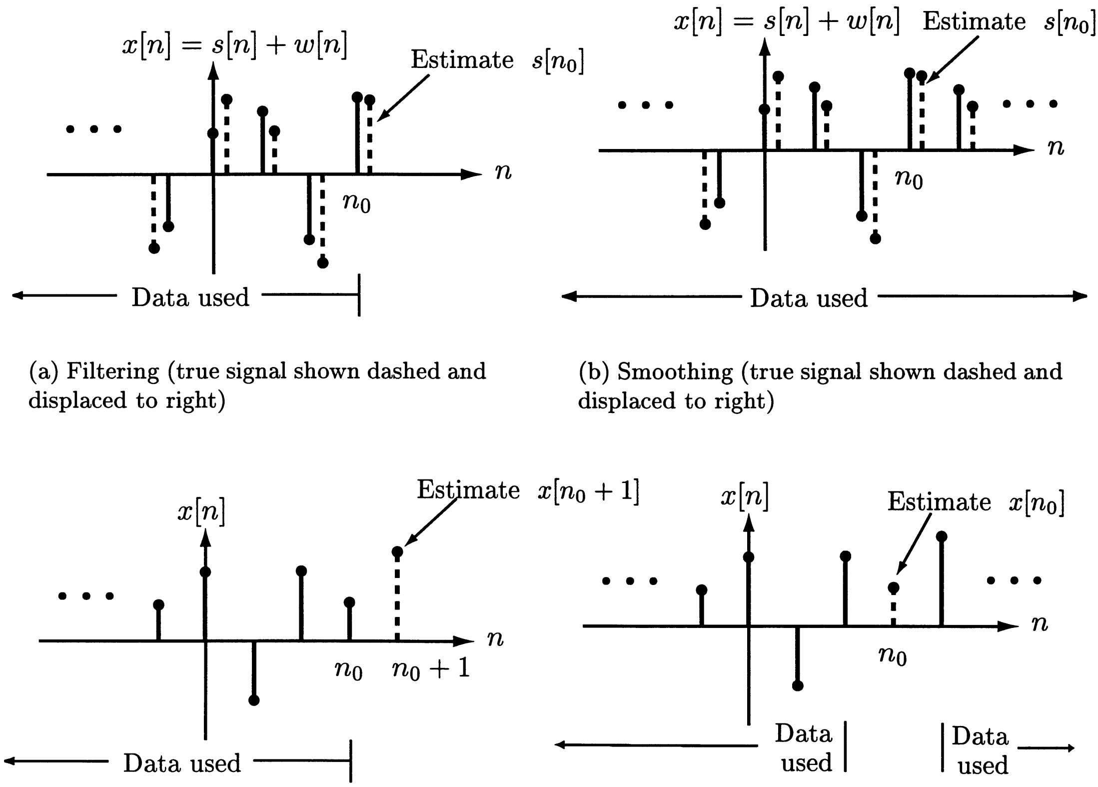
which was previously asserted.

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

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

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

$$r_S[l] = h[l] \star (r_S[l] + r_W[l])$$

and taking Fourier transforms of both sides yields

$$P_S(f) = H(f)(P_S(f) + P_W(f))$$

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

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

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

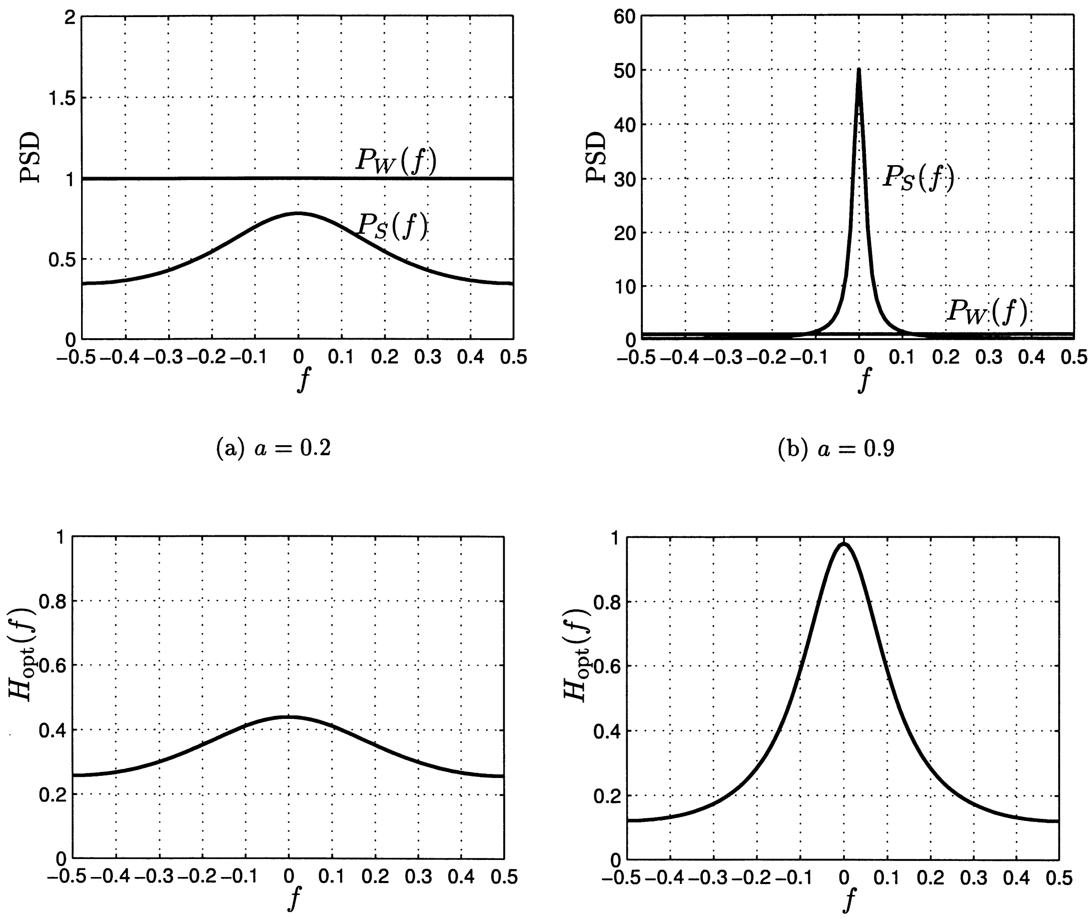
Example 18.4 – Wiener smoother for AR signal in white noise

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.

◇



(a) $a = 0.2$ (b) $a = 0.9$

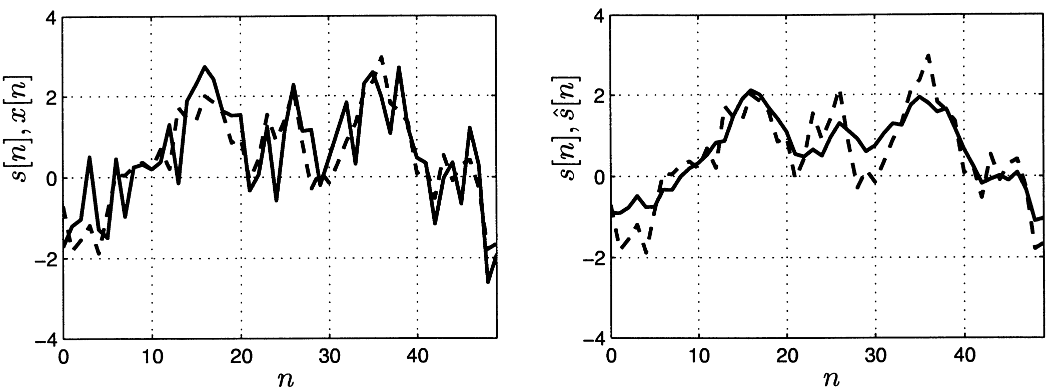
(c) $a = 0.2$ (d) $a = 0.9$

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

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

($N = 50$ 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

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

But the second term is zero since by the orthogonality principle

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

Thus, we have

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

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

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

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

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

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

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

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)

$$
\begin{aligned}
\mathrm{mse} &= E[(X[n_0 + 1] - \hat{X}[n_0 + 1])^2] \\
&= E\left[\left(X[n_0 + 1] - \sum_{k=0}^{\infty} h[k]X[n_0 - k]\right)^2\right]
\end{aligned}
$$

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

$$\begin{aligned} \mathrm{mse}_{\min} &= E\left[\left(X[n_0+1] - \sum_{k=0}^{\infty} h_{\mathrm{opt}}[k] X[n_0-k]\right) X[n_0+1]\right] \\ &= r_X[0] - \sum_{k=0}^{\infty} h_{\mathrm{opt}}[k] r_X[k+1] \end{aligned} \tag{18.29}$$

where $h_{\mathrm{opt}}[k]$ 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

$$h_{\mathrm{opt}}[0] = \frac{a^{l+1}}{a^l} = a$$

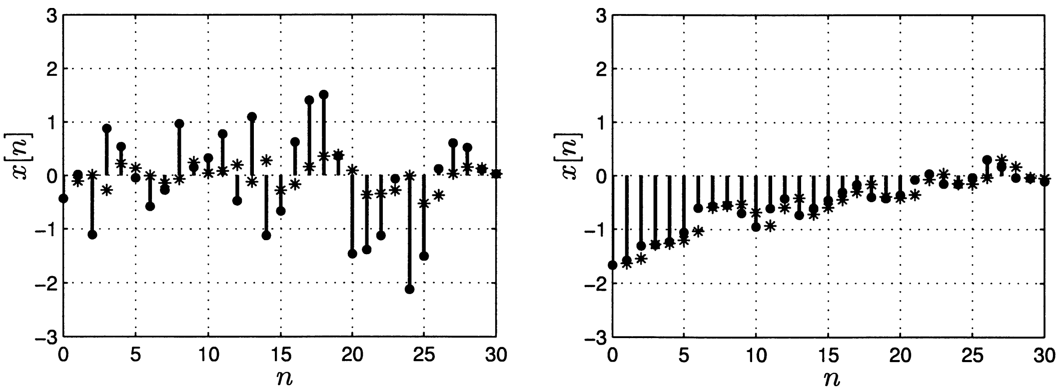
or finally

$$\hat{X}[n_0+1] = aX[n_0].$$

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.)

$\diamond$

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

$$h_{\mathrm{opt}}[k] = a[k+1] \qquad k = 0, 1, \ldots$$

and the minimum MSE is

$$\mathrm{mse_{min}} = E[(X[n_0+1] - \hat{X}[n_0+1])^2] = \sigma_U^2.$$

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, $\mathrm{mse_{min}} = \sigma_U^2$.

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

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

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

$$\mathrm{mse}_{\min} = \sigma_U^2.$$

$\diamond$

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

$$\underbrace{\begin{bmatrix} r_X[0] & r_X[1] & \ldots & r_X[M-1] \\ r_X[1] & r_X[0] & \ldots & r_X[M-2] \\ \vdots & \vdots & \ddots & \vdots \\ r_X[M-1] & r_X[M-2] & \ldots & r_X[0] \end{bmatrix}}_{\mathbf{R}_X} \begin{bmatrix} h[0] \\ h[1] \\ \vdots \\ h[M-1] \end{bmatrix} = \begin{bmatrix} r_X[1] \\ r_X[2] \\ \vdots \\ r_X[M] \end{bmatrix}. \tag{18.36}$$

The corresponding minimum MSE is given by

$$\mathrm{mse}_{\min} = r_X[0] - \sum_{k=0}^{M-1} h_{\mathrm{opt}}[k]r_X[k+1]. \tag{18.37}$$

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

$$\sigma_U^2 = \mathrm{mse}_{\min} = r_X[0] - \sum_{k=1}^{p} a[k]r_X[k] \tag{18.39}$$

which follows by letting $h_{\mathrm{opt}}[k] = a[k+1]$ 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 $F = 1/T =$

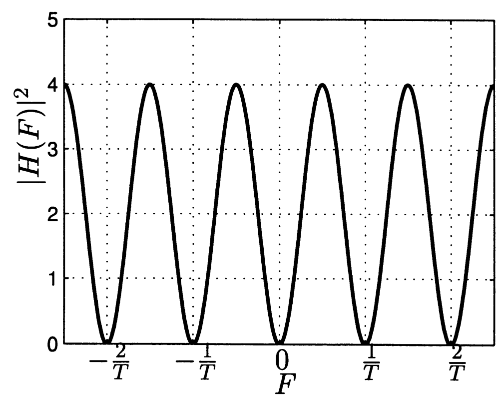


Figure 18.8: Magnitude-squared frequency response of interference canceling filter with $F_0 = 1/T$.

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 $1/(2F_0)$?

◇

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 excitation is modified by the vocal tract, which can be modeled by an LSI filter. Knowing

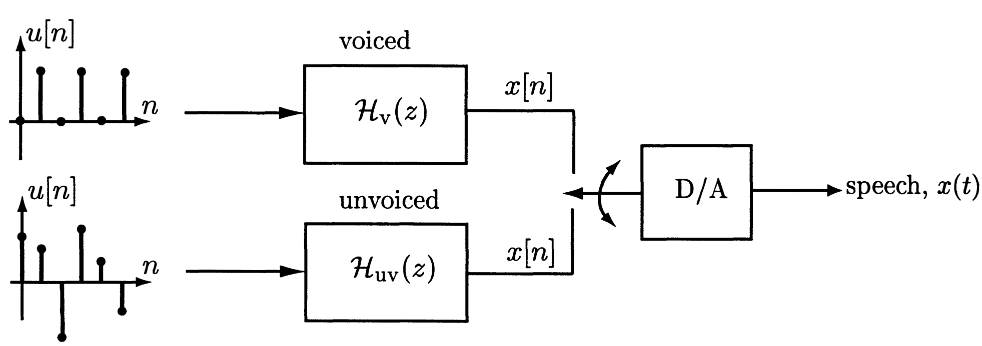


Figure 18.9: Speech production model.

[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 $\mathcal{H}_{\mathrm{uv}}(z)$. 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

$$\mathcal{H}_{\mathrm{uv}}(z) = \frac{1}{1 - \sum_{k=1}^{p} a[k] z^{-k}}$$

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

$$\begin{bmatrix} \hat{r}_X[0] & \hat{r}_X[1] & \ldots & \hat{r}_X[p-1] \\ \hat{r}_X[1] & \hat{r}_X[0] & \ldots & \hat{r}_X[p-2] \\ \vdots & \vdots & \ddots & \vdots \\ \hat{r}_X[p-1] & \hat{r}_X[p-2] & \ldots & \hat{r}_X[0] \end{bmatrix} \begin{bmatrix} a[1] \\ a[2] \\ \vdots \\ a[p] \end{bmatrix} = \begin{bmatrix} \hat{r}_X[1] \\ \hat{r}_X[2] \\ \vdots \\ \hat{r}_X[p] \end{bmatrix} \tag{18.47}$$

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 $\hat{\mathcal{H}}_{\mathrm{uv}}(z)$ 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

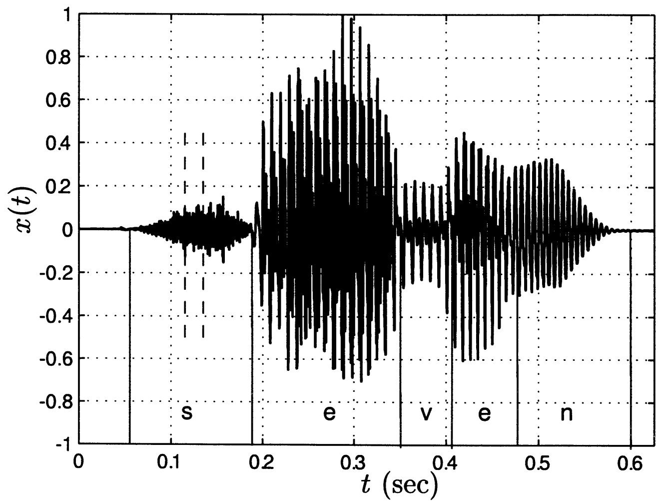


Figure 18.10: Waveform for the utterance "seven" [Allu 2005].

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

$$\hat{P}_X(f) = \frac{\hat{\sigma}_U^2}{\left|1 - \sum_{k=1}^{p} \hat{a}[k] \exp(-j2\pi f k)\right|^2} \tag{18.50}$$

should be a good match to the normalized and squared-magnitude of the Fourier

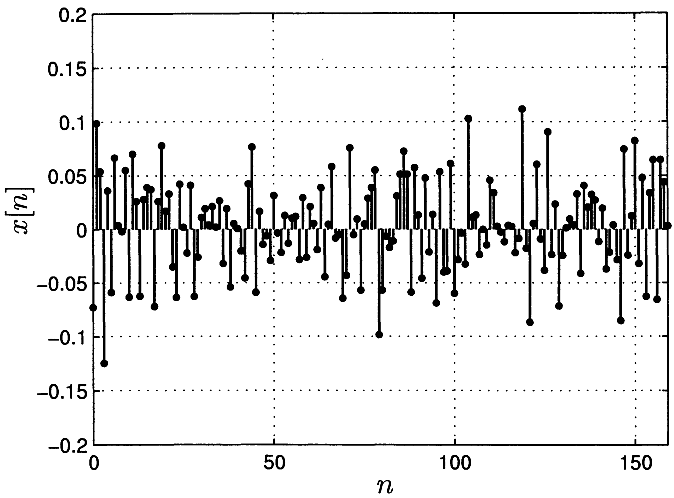


Figure 18.11: A 20 msec segment of the waveform for "sss". See Figure 18.10 for segment extracted as indicated by the vertical dashed lines.

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
```

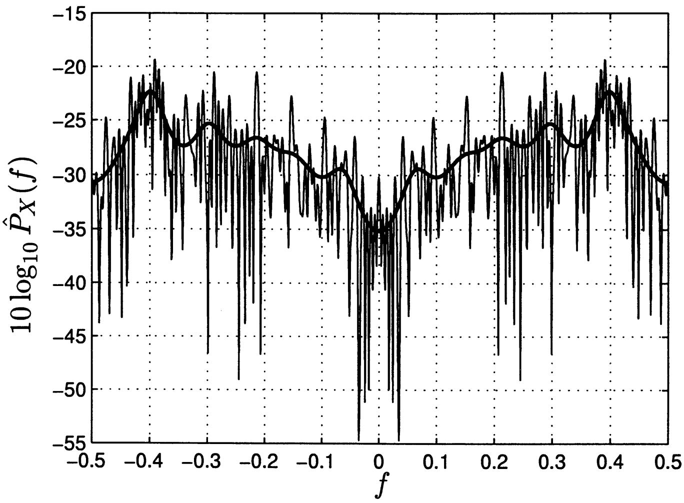


Figure 18.12: Periodogram, shown as the light line, and AR PSD model, shown as the darker line for speech segment of Figure 18.11.

```
          % 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.

18.5 (f,c) A discrete-time WSS random process $X[n]$ is defined by the difference equation $X[n] = aX[n-1] + U[n] - bU[n-1]$, where $U[n]$ is a discrete-time white noise random process with variance $\sigma_U^2 = 1$. Plot the PSD of $X[n]$ if $a = 0.9, b = 0.2$ and also if $a = 0.2, b = 0.9$ and explain your results.

18.6 (f) A discrete-time WSS random process $X[n]$ is defined by the difference equation $X[n] = 0.5X[n-1] + U[n] - 0.5U[n-1]$, where $U[n]$ is a discrete-time white noise random process with variance $\sigma_U^2 = 1$. Find the ACS and PSD of $X[n]$ and explain your results.

18.7 (⌣̈) (f) A differencer is given by $X[n] = U[n] - U[n-1]$. If the input random process $U[n]$ has the PSD $P_U(f) = 1 - \cos(2\pi f)$, determine the ACS and PSD at the output of the differencer.

18.8 (t) Verify that the discrete-time Fourier transform of $r_X[k]$ given in (18.15) is $\sigma_U^2 |H(f)|^2$.

18.9 (w) A discrete-time white noise random process is input to an LSI system which has $h[0] = 1$ with all the other impulse response samples nonzero. Can the output power of the filter ever be less than the input power?

18.10 (w) A random process with PSD

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

is to be filtered with an LSI system to produce a white noise random process $U[n]$ with variance $\sigma_U^2 = 4$ at the output. What should the difference equation of the LSI system be?

18.11 (w,c) An AR random process of order 2 is given by the recursive difference equation $X[n] = 2r\cos(2\pi f_0)X[n-1] - r^2X[n-2] + U[n]$, where $U[n]$ is white Gaussian noise with variance $\sigma_U^2 = 1$. For $r = 0.7, f_0 = 0.1$ and also for $r = 0.95, f_0 = 0.1$ plot the PSD of $X[n]$. Can you explain your results? Hint: Determine the pole locations of $\mathcal{H}(z)$.

18.12 (w) A signal, which is bandlimited to B cycles/sample with $B < 1/2$, is modeled as a WSS random process with zero mean and PSD $P_S(f)$. If white noise is added to the signal with $\sigma_W^2 = 1$, find the frequency response of the optimal Wiener smoother. Explain your results.

18.13 (⌣̈) (f,c) A zero mean signal with PSD $P_S(f) = 2 - 2\cos(2\pi f)$ is embedded in white noise with variance $\sigma_W^2 = 1$. Plot the frequency response of the optimal Wiener smoother. Also, compute the minimum MSE. Hint: For the MSE use a "sum" approximation to the integral (see Problem 1.14).

18.14 (c) In this problem we simulate the Wiener smoother. First generate $N = 50$ samples of a signal $S[n]$, which is an AR random process (assumes that $U[n]$ is white Gaussian noise) with $a = 0.25$ and $\sigma_U^2 = 0.5$. Remember to set the initial condition $S[-1] \sim \mathcal{N}(0, \sigma_U^2/(1-a^2)$. Next add white Gaussian noise $W[n]$ with $\sigma_W^2 = 1$ to the AR random process realization. Finally, use the MATLAB code in the chapter to smooth the noise-corrupted signal. Plot the true signal and the smoothed signal. How well does the smoother perform?

18.15 (w) To see that the linear prediction equations of (18.28) cannot be solved directly using z-transforms, take the z-transform of both sides of the equation. Next solve for $\mathcal{H}(z) = \mathcal{Z}\{h[k]\}$. Explain why the solution for the predictor cannot be correct.

18.16 (t) In this problem we rederive the optimal one-step linear predictor for the AR random process of Example 18.5. Assume that $X[n_0+1]$ is to be predicted based on observing the realization of $\{X[n_0], X[n_0-1], \ldots\}$. The random process $X[n]$ is assumed to be an AR random process described in Example 18.5. Prove that $\hat{X}[n_0+1] = aX[n_0]$ satisfies the orthogonality principle, making use of the result that $E[U[n_0+1]X[n_0-k]] = 0$ for $k = 0, 1, \ldots$. The latter result says that "future" samples of $U[n]$ must be uncorrelated with the present and past samples of $X[n]$. Explain why this is true. Hint: Recall that for an AR random process $X[n]$ can be rewritten as $X[n] = \sum_{l=0}^{\infty} a^l U[n-l]$.

18.17 (w) For the AR random process described in Example 18.5 show that the minimum MSE for the optimal predictor $\hat{X}[n_0+1] = aX[n_0]$ is given by $\mathrm{mse}_{\min} = r_X[0](1-a^2)$. Use this to explain why the results shown in Figure 18.7 are reasonable.

18.18 (⌣) (w) Express the minimum MSE given in the previous problem in terms of $r_X[0]$ and the correlation coefficient between $X[n_0]$ and $X[n_0+1]$. What happens to the minimum MSE if the correlation coefficient magnitude approaches one and also if it is zero?

18.19 (c) Consider an AR(2) random process given by $X[n] = -r^2 X[n-2] + U[n]$, where $U[n]$ is white Gaussian noise with variance σ_U^2 and $0 < r < 1$. This random process follows from (18.33) with $p = 2$ and $a[1] = 0$, $a[2] = -r^2$. The ACS for this random process can be shown to be $r_X[k] = (\sigma_U^2/(1-r^4))r^{|k|}\cos(k\pi/2)$ [Kay 1988]. Find the optimal one-step linear predictor based on the present and past samples of $X[n]$. Next perform a computer simulation to see how the predictor performs. Consider the two cases $r = 0.5, \sigma_U^2 = 1 - r^4$ and $r = 0.95, \sigma_U^2 = 1 - r^4$ so that the average power in each case is the same ($r_X[0] = 1$). Generate 150 samples of each process and discard the first 100 samples to make sure the generated samples are WSS. Then, plot the realization and its predicted values for each case. Which value of r results in a more predictable process?

18.20 (t) Derive the Wiener-Hopf equations given by (18.36) and the resulting minimum MSE given by (18.37) for the finite length predictor.

18.21 (f) For $M = 1$ solve the Wiener-Hopf equations given by (18.36) to find $h[0]$. Relate this to $\text{cov}(X,Y)/\text{var}(X)$ used in the prediction of Y given $X = x$.

18.22 (⌣) (f) The MA random process described in Example 18.6 and given by $X[n] = U[n] - bU[n-1]$ has as its ACS for $\sigma_U^2 = 1$

$$r_X[k] = \begin{cases} 1+b^2 & k=0 \\ -b & k=1 \\ 0 & k \geq 2. \end{cases}$$

For $M = 2$ solve the Wiener-Hopf equations to find this finite length predictor and then determine the minimum MSE. Compare this minimum MSE to that of the infinite length predictor given in Example 18.6.

18.23 (f) It is desired to predict white noise. Solve the Wiener-Hopf equations for $r_X[k] = \sigma_X^2 \delta[k]$ and explain your results.

18.24 (⌣) (f,c) For the MA random process $X[n] = U[n] - \frac{1}{2}U[n-1]$ where $U[n]$ is white Gaussian noise with $\sigma_U^2 = 1$ find the optimal finite length predictor $\hat{X}[n_0+1] = h[0]X[n_0] + h[1]X[n_0-1]$ and the corresponding minimum MSE. Next simulate the random process and compare the estimated minimum MSE with the theoretical one. Hint: Use your results from Problem 18.22.

18.25 (f) Consider the prediction of a randomly phased sinusoid whose ACS is $r_X[k] = \cos(2\pi f_0 k)$. For $M = 2$ solve the Wiener-Hopf equations to determine the optimal linear predictor and also the minimum MSE. Hint: You should be able to show that the minimum MSE is zero. Use the trigonometric identity $\cos(2\theta) = 2\cos^2(\theta) - 1$.

18.26 (t) In this problem we consider the L-step infinite length predictor of an AR random process. Let the predictor be given as

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

and show that the linear equations to be solved to determine the optimal $h[k]$'s are

$$r_X[l+L] = \sum_{k=0}^{\infty} h[k]r_X[l-k] \qquad l = 0,1,\ldots.$$

Next show that the minimum MSE is

$$\text{mse}_{\text{min}} = r_X[0] - \sum_{k=0}^{\infty} h_{\text{opt}}[k]r_X[k+L].$$

Finally, for an AR random process with ACS $r_X[k] = (\sigma_U^2/(1-a^2))a^{|k|}$ show that

$$\begin{aligned}\hat{X}[n_0+L] &= a^L X[n_0] \\ \text{mse}_{\min} &= r_X[0](1-a^{2L})\end{aligned}$$

for a predictor based on $\{X[n_0], X[n_0-1], \ldots\}$. To do so assume that $h[k]=0$ for $k \geq 1$ and show that the equations can be satisfied by choosing $h[0]$. Explain what happens to the quality of the prediction as L increases and why.

18.27 (⌣̈) **(t)** In this problem we consider the interpolation of a random process using a sample on either side of the sample to be interpolated. We wish to estimate or interpolate $X[n_0]$ using $\hat{X}[n_0] = h[-1]X[n_0+1] + h[1]X[n_0-1]$ for some impulse response values $h[-1], h[1]$. Find the optimal impulse response values by minimizing the MSE of the interpolated sample if $X[n]$ is the AR random process given by $X[n] = aX[n-1] + U[n]$. Does your interpolator average the samples on either side of $X[n_0]$? What happens as $a \to 1$ and as $a \to 0$?

18.28 (f) An LTI system has the impulse response $h(\tau) = \exp(-\tau)$ for $\tau \geq 0$ and is zero for $\tau < 0$. If continuous-time white noise with ACF $r_U(\tau) = (N_0/2)\delta(\tau)$ is input to the system, what is the PSD of the output random process? Sketch the PSD.

18.29 (⌣̈) **(f)** An LTI system has the impulse response $h(\tau) = 1$ for $0 \leq \tau \leq T$ and is zero otherwise. If continuous-time white noise with ACF $r_U(\tau) = (N_0/2)\delta(\tau)$ is input to the system, what is the PSD of the output random process? Sketch the PSD.

18.30 (f) A filter with frequency response $H(F) = \exp(-j2\pi F\tau_0)$ is used to filter a WSS random process with PSD $P_X(F)$. What is the PSD at the filter output and why?

18.31 (t) Prove that if a continuous-time white noise random process with ACF $r_U(\tau) = (N_0/2)\delta(\tau)$ is input to an LTI system with impulse response $h(\tau)$, then the ACF of the output random process is

$$r_X(\tau) = \frac{N_0}{2}\int_{-\infty}^{\infty} h(t)h(t+\tau)dt.$$

18.32 (⌣̈) **(w)** An RC electrical circuit with frequency response

$$H(F) = \frac{1/RC}{1/RC + j2\pi F}$$

is used to filter a white noise random process with ACF $r_U(\tau) = (N_0/2)\delta(\tau)$. Find the total average power at the filter output. Is it infinite? Hint: See previous problem.

18.33 (t) Two continuous-time WSS zero mean random processes $X(t)$ and $Y(t)$ are uncorrelated, which means that $E[X(t_1)Y(t_2)] = 0$ for all t_1 and t_2. Is the sum random process $Z(t) = X(t) + Y(t)$ also WSS, and if so, what is its ACF and PSD?

18.34 (c) In this problem we compare the periodogram spectral estimator to one based on an AR(2) model. This assumes, however, that the AR model is an accurate one for the random process. First generate $N = 50$ samples of a realization of the AR(2) random process described in Problem 18.19 with $r = 0.5$ and $\sigma_U^2 = 1 - r^4$. Next plot the periodogram of the realization (see Section 17.6). Using the estimate of the ACS given in (18.49) solve the Yule-Walker equations of (18.47) for $p = 2$ and then find $\hat{\sigma}_U^2$ from (18.48). Finally, plot the estimated PSD given by (18.50) and compare it to the periodogram as well as the true PSD. You may also wish to print out $\hat{a}[1]$ and $\hat{a}[2]$ and compare them to the theoretical values of $a[1] = 0$ and $a[2] = -r^2 = -0.25$. Hint: You can use the MATLAB code given in Section 18.7.

Appendix 18A

Solution for Infinite Length Predictor

The equations to be solved for the one-step predictor are from (18.28)

$$r_X[l+1] = \sum_{k=0}^{\infty} h[k] r_X[l-k] \qquad l = 0, 1, \ldots \tag{18A.1}$$

and the minimum MSE can be written from (18.29) as

$$\mathrm{mse}_{\min} = r_X[0] - \sum_{k=0}^{\infty} h_{\mathrm{opt}}[k] r_X[-1-k]. \tag{18A.2}$$

Now let $n = l + 1$ in (18A.1) so that

$$\begin{aligned} r_X[n] &= \sum_{k=0}^{\infty} h[k] r_X[n-1-k] \qquad n = 1, 2, \ldots \\ &= \sum_{j=1}^{\infty} h[j-1] r_X[n-j] \qquad (\text{let } j = k+1) \end{aligned} \tag{18A.3}$$

and also let $j = k + 1$ in (18A.2) to yield

$$r_X[0] = \sum_{j=1}^{\infty} h[j-1] r_X[-j] + \mathrm{mse}_{\min} \tag{18A.4}$$

where we drop the "opt" on $h_{\mathrm{opt}}[k]$ since $h[k]$ and $\mathrm{mse}_{\min}$ are unknowns that we wish to solve for. Then combining (18A.3) and (18A.4) we have

$$r_X[n] = \sum_{j=1}^{\infty} h[j-1] r_X[n-j] + \mathrm{mse}_{\min}\delta_{n0} \qquad n = 0, 1, \ldots$$

where $\delta_{n0} = 1$ for $n = 0$ and $\delta_{n0} = 0$ for $n \geq 1$. Next divide both sides by $\mathrm{mse}_{\min}$ to yield

$$\frac{r_X[n]}{\mathrm{mse}_{\min}} = \sum_{j=1}^{\infty} \frac{h[j-1]}{\mathrm{mse}_{\min}} r_X[n-j] + \delta_{n0} \qquad n = 0, 1, \ldots .$$

Let

$$g[j] = \begin{cases} 1/\mathrm{mse}_{\min} & j = 0 \\ -h[j-1]/\mathrm{mse}_{\min} & j = 1, 2, \ldots \end{cases} \tag{18A.5}$$

so that the equations become

$$r_X[n]g[0] = -\sum_{j=1}^{\infty} g[j] r_X[n-j] + \delta_{n0}$$

or

$$\sum_{j=0}^{\infty} g[j] r_X[n-j] = \delta_{n0} \quad n = 0, 1, \ldots . \tag{18A.6}$$

Now if (18A.6) can be solved for $g[j]$, then $h[j], \mathrm{mse}_{\min}$ can then be found from (18A.5). Note that (18A.6) is a discrete-time convolution that holds for $n \geq 0$. We therefore need to find a *causal sequence* $g[n]$ (since the sum in (18A.6) is only over $j \geq 0$), which when convolved with $r_X[n]$ yields 1 for $n = 0$ and 0 for $n > 0$. Note that the values of $g[n] \star r_X[n]$ for $n < 0$ are unspecified by the equations. Hence, $g[n] \star r_X[n]$ must be an *anticausal sequence* to be a solution of (18A.6). This can easily be solved if

$$\mathcal{P}_X(z) = \sum_{k=-\infty}^{\infty} r_X[n] z^{-n}$$

can be written as

$$\mathcal{P}_X(z) = \frac{\sigma_U^2}{\mathcal{A}(z)\mathcal{A}(z^{-1})} \tag{18A.7}$$

where

$$\mathcal{A}(z) = 1 - \sum_{k=1}^{\infty} a[k] z^{-k}$$

has all its zeros within the unit circle of the z plane. Now $1/\mathcal{A}(z)$ is the z-transform of a causal sequence. This is because if all the zeros of $\mathcal{A}(z)$ are within the unit circle, then all the poles of $1/\mathcal{A}(z)$ are within the unit circle. Thus, the z-transform $1/\mathcal{A}(z)$ must converge on and outside of the unit circle. Also, then $1/\mathcal{A}(z^{-1})$ is the z-transform of an anticausal sequence. Assuming this is possible (18A.6) becomes

$$\begin{aligned} \mathcal{Z}^{-1}\{\mathcal{G}(z)\mathcal{P}_X(z)\} &= \mathcal{Z}^{-1}\left\{\mathcal{G}(z)\frac{\sigma_U^2}{\mathcal{A}(z)\mathcal{A}(z^{-1})}\right\} \\ &= \begin{cases} 1 & n = 0 \\ 0 & n > 0 \end{cases} \end{aligned}$$

where $\mathcal{G}(z)$ is the z-transform of $g[n]$ and $\mathcal{Z}^{-1}$ denotes the inverse z-transform. Now if we choose

$$\mathcal{G}(z) = \frac{\mathcal{A}(z)}{\sigma_U^2} \tag{18A.8}$$

then

$$\begin{aligned} \mathcal{Z}^{-1}\left\{\mathcal{G}(z)\frac{\sigma_U^2}{\mathcal{A}(z)\mathcal{A}(z^{-1})}\right\} &= \mathcal{Z}^{-1}\left\{\frac{1}{\mathcal{A}(z^{-1})}\right\} \\ &= \begin{cases} 1 & n=0 \\ 0 & n>0 \end{cases} \end{aligned}$$

since $1/\mathcal{A}(z^{-1})$ is the z-transform of an anticausal sequence, and the equations are satisfied. The inverse z-transform for $n=0$ has been obtained by using the initial value theorem [Jackson 1991] which says that for an anticausal sequence $x[n]$

$$\mathcal{Z}^{-1}\left\{\sum_{n=-\infty}^{0} x[n]z^{-n}\right\}\Bigg|_{n=0} = \lim_{z\to 0}\sum_{n=-\infty}^{0} x[n]z^{-n} = x[0].$$

Therefore, we have that

$$\mathcal{Z}^{-1}\left\{\frac{1}{\mathcal{A}(z^{-1})}\right\}\Bigg|_{n=0} = \lim_{z\to 0}\frac{1}{\mathcal{A}(z^{-1})} = 1.$$

The solution for $g[n]$ is from (18A.8)

$$g[n] = \mathcal{Z}^{-1}\left\{\frac{\mathcal{A}(z)}{\sigma_U^2}\right\} = \begin{cases} 1/\sigma_U^2 & n=0 \\ -a[n]/\sigma_U^2 & n\ge 1 \end{cases}$$

and using (18A.5)

$$\begin{aligned} \frac{1}{\text{mse}_{\text{min}}} &= g[0] = \frac{1}{\sigma_U^2} \\ -\frac{h[j-1]}{\text{mse}_{\text{min}}} &= g[j] = -\frac{a[j]}{\sigma_U^2} \qquad j\ge 1. \end{aligned}$$

Finally, we have the result that

$$\begin{aligned} h[n] &= a[n+1] \qquad n=0,1,\ldots \\ \text{mse}_{\text{min}} &= \sigma_U^2. \end{aligned}$$