C H A P T E R   11

Wiener Filtering

INTRODUCTION

In this chapter we will consider the use of LTI systems in order to perform minimum mean-square-error (MMSE) estimation of a WSS random process of interest, given measurements of another related process. The measurements are applied to the input of the LTI system, and the system is designed to produce as its output the MMSE estimate of the process of interest.

We first develop the results in discrete time, and for convenience assume (unless otherwise stated) that the processes we deal with are *zero-mean*. We will then show that exactly analogous results apply in continuous time, although their derivation is slightly different in certain parts.

Our problem in the DT case may be stated in terms of Figure 11.1.

Here $x[n]$ is a WSS random process that we have measurements of. We want to determine the unit sample response or frequency response of the above LTI system such that the filter output $\widehat{y}[n]$ is the minimum-mean-square-error (MMSE) estimate of some "target" process $y[n]$ that is jointly WSS with $x[n]$. Defining the error $e[n]$ as

$$e[n] \triangleq \widehat{y}[n] - y[n] \ , \tag{11.1}$$

we wish to carry out the following minimization:

$$\min_{h[\,\cdot\,]} \epsilon = E\{e^2[n]\} \ . \tag{11.2}$$

The resulting filter $h[n]$ is called the **Wiener filter** for estimation of $y[n]$ from $x[n]$.

In some contexts it is appropriate or convenient to restrict the filter to be an FIR (finite-duration impulse response) filter of length N, e.g. $h[n] = 0$ except in the interval $0 \leq n \leq N-1$. In other contexts the filter impulse response can be of infinite duration and may either be restricted to be causal or allowed to be noncausal. In the next section we discuss the FIR and general noncausal IIR

$x[n] \longrightarrow \boxed{\text{LTI } h[n]} \longrightarrow \begin{aligned} &\widehat{y}[n] = \text{ estimate} \\ &y[n] = \text{ target process} \end{aligned}$
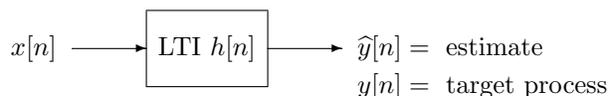
FIGURE 11.1 DT LTI filter for linear MMSE estimation.

(infinite-duration impulse response) cases. A later section deals with the more involved case where the filter is IIR but restricted to be causal.

If $x[n] = y[n] + v[n]$ where $y[n]$ is a signal and $v[n]$ is noise (both random processes), then the above estimation problem is called a *filtering* problem. If $y[n] = x[n + n_0]$ with n_0 positive, and if $h[n]$ is restricted to be causal, then we have a *prediction* problem. Both fit within the same general framework, but the solution under the restriction that $h[n]$ be causal is more subtle.

11.1 NONCAUSAL DT WIENER FILTER

To determine the optimal choice for $h[n]$ in (11.2), we first expand the error criterion in (11.2):

$$\epsilon = E\left\{ \left(\sum_{k=-\infty}^{+\infty} h[k]x[n-k] - y[n] \right)^2 \right\}. \tag{11.3}$$

The impulse response values that minimize ϵ can then be obtained by setting $\dfrac{\partial \epsilon}{\partial h[m]} = 0$ for all values of m for which $h[m]$ is not restricted to be zero (or otherwise pre-specified):

$$\frac{\partial \epsilon}{\partial h[m]} = E\left\{ 2 \underbrace{\left(\sum_{k} h[k]x[n-k] - y[n] \right)}_{e[n]} x[n-m] \right\} = 0 . \tag{11.4}$$

The above equation implies that

$$E\{e[n]x[n-m]\} = 0, \text{ or}$$
$$R_{ex}[m] = 0, \text{ for all } m \text{ for which } h[m] \text{ can be freely chosen.} \tag{11.5}$$

You may recognize the above equation (or constraint) on the relation between the input and the error as the familiar *orthogonality principle*: for the optimal filter, *the error is orthogonal to all the data that is used to form the estimate.* Under our assumption of zero-mean $x[n]$, orthogonality is equivalent to uncorrelatedness. As we will show shortly, the orthogonality principle also applies in continuous time.

Note that

$$\begin{aligned} R_{ex}[m] &= E\{e[n]x[n-m]\} \\ &= E\{(\widehat{y}[n] - y[n])x[n-m]\} \\ &= R_{\widehat{yx}}[m] - R_{yx}[m] . \end{aligned} \tag{11.6}$$

Therefore, an alternative way of stating the orthogonality principle (11.5) is that

$$R_{\widehat{yx}}[m] = R_{yx}[m] \text{ for all appropriate } m . \tag{11.7}$$

In other words, for the optimal system, the cross-correlation between the input and output of the estimator equals the cross-correlation between the input and target output.

To actually find the impulse response values, observe that since $\widehat{y}[n]$ is obtained by filtering $x[n]$ through an LTI system with impulse response $h[n]$, the following relationship applies:

$$R_{\widehat{y}x}[m] = h[m] * R_{xx}[m] \ . \tag{11.8}$$

Combining this with the alternative statement of the orthogonality condition, we can write

$$h[m] * R_{xx}[m] = R_{yx}[m] \ , \tag{11.9}$$

or equivalently,

$$\sum_k h[k] R_{xx}[m - k] = R_{yx}[m] \tag{11.10}$$

Equation (11.10) represents a set of linear equations to be solved for the impulse response values. If the filter is FIR of length N, then there are N equations in the N unrestricted values of $h[n]$. For instance, suppose that $h[n]$ is restricted to be zero except for $n \in [0, N-1]$. The condition (11.10) then yields as many equations as unknowns, which can be arranged in the following matrix form, which you may recognize as the appropriate form of the *normal equations* for LMMSE estimation, which we introduced in Chapter 8:

$$\begin{bmatrix} R_{xx}[0] & R_{xx}[-1] & \cdots & R_{xx}[1-N] \\ R_{xx}[1] & R_{xx}[0] & \cdots & R_{xx}[2-N] \\ \vdots & \vdots & \ddots & \vdots \\ R_{xx}[N-1] & R_{xx}[N-2] & \cdots & R_{xx}[0] \end{bmatrix} \begin{bmatrix} h[0] \\ h[1] \\ \vdots \\ h[N-1] \end{bmatrix} = \begin{bmatrix} R_{yx}[0] \\ R_{yx}[1] \\ \vdots \\ R_{yx}[N-1] \end{bmatrix} \ . \tag{11.11}$$

These equations can now be solved for the impulse response values. Because of the particular structure of these equations, there are efficient methods for solving for the unknown parameters, but further discussion of these methods is beyond the scope of our course.

In the case of an IIR filter, equation (11.10) must hold for an infinite number of values of m and, therefore, cannot simply be solved by the methods used for a finite number of linear equations. However, if $h[n]$ is not restricted to be causal or FIR, the equation (11.10) must hold for all values of m from $-\infty$ to $+\infty$, so the z-transform can be applied to equation (11.10) to obtain

$$H(z)S_{xx}(z) = S_{yx}(z) \tag{11.12}$$

The optimal transfer function, i.e. the transfer function of the resulting (Wiener) filter, is then

$$\boxed{H(z) = S_{yx}(z)/S_{xx}(z)} \tag{11.13}$$

If either of the correlation functions involved in this calculation does not possess a z-transform but if both possess Fourier transforms, then the calculation can be carried out in the Fourier transform domain.

Note the similarity between the above expression for the optimal filter and the expression we obtained in Chapters 5 and 7 for the gain σ_{YX}/σ_{XX} that multiplies a zero-mean random variable X to produce the LMMSE estimator for a zero-mean random variables Y. In effect, by going to the transform domain or frequency domain, we have decoupled the design into a problem that — at each frequency — is as simple as the one we solved in the earlier chapters.

As we will see shortly, in continuous time the results are exactly the same:

$$R_{\widehat{yx}}(\tau) = R_{yx}(\tau), \tag{11.14}$$

$$h(\tau) * R_{xx}(\tau) = R_{yx}(\tau), \tag{11.15}$$

$$H(s)S_{xx}(s) = S_{yx}(s), \text{ and} \tag{11.16}$$

$$\boxed{H(s) = S_{yx}(s)/S_{xx}(s)} \tag{11.17}$$

The mean-square-error corresponding to the optimum filter, i.e. the *minimum* MSE, can be determined by straightforward computation. We leave you to show that

$$R_{ee}[m] = R_{yy}[m] - R_{\widehat{yy}}[m] = R_{yy}[m] - h[m] * R_{xy}[m] \tag{11.18}$$

where $h[m]$ is the impulse response of the optimal filter. The MMSE is then just $R_{ee}[0]$. It is illuminating to rewrite this in the frequency domain, but dropping the argument $e^{j\Omega}$ on the power spectra $S_{**}(e^{j\Omega})$ and frequency response $H(e^{j\Omega})$ below to avoid notational clutter:

$$\begin{aligned}
\text{MMSE} = R_{ee}[0] &= \frac{1}{2\pi} \int_{-\pi}^{\pi} S_{ee} \, d\Omega \\
&= \frac{1}{2\pi} \int_{-\pi}^{\pi} (S_{yy} - HS_{xy}) \, d\Omega \\
&= \frac{1}{2\pi} \int_{-\pi}^{\pi} S_{yy} \left(1 - \frac{S_{yx}S_{xy}}{S_{yy}S_{xx}}\right) d\Omega \\
&= \frac{1}{2\pi} \int_{-\pi}^{\pi} S_{yy} \left(1 - \rho_{yx}\rho_{yx}^*\right) d\Omega \, . \tag{11.19}
\end{aligned}$$

The function $\rho_{yx}(e^{j\Omega})$ defined by

$$\rho_{yx}(e^{j\Omega}) = \frac{S_{yx}(e^{j\Omega})}{\sqrt{S_{yy}(e^{j\Omega})S_{xx}(e^{j\Omega})}} \tag{11.20}$$

evidently plays the role of a frequency-domain correlation coefficient (compare with our earlier definition of the correlation coefficient between two random variables). This function is sometimes referred to as the *coherence function* of the two processes. Again, note the similarity of this expression to the expression $\sigma_{YY}(1-\rho_{YX}^2)$ that we obtained in a previous lecture for the (minimum) mean-square-error after LMMSE

estimation of a random variable Y using measurements of a random variable X.

EXAMPLE 11.1   Signal Estimation in Noise (Filtering)

Consider a situation in which $x[n]$, the sum of a target process $y[n]$ and noise $v[n]$, is observed:

$$x[n] = y[n] + v[n] \ . \tag{11.21}$$

We would like to estimate $y[n]$ from our observations of $x[n]$. Assume that the signal and noise are uncorrelated, i.e. $R_{vy}[m] = 0$. Then

$$R_{xx}[m] = R_{yy}[m] + R_{vv}[m], \tag{11.22}$$

$$R_{yx}[m] = R_{yy}[m], \tag{11.23}$$

$$H(e^{j\Omega}) = \frac{S_{yy}(e^{j\Omega})}{S_{yy}(e^{j\Omega}) + S_{vv}(e^{j\Omega})} \ . \tag{11.24}$$

At values of Ω for which the signal power is much greater than the noise power, $H(e^{j\Omega}) \approx 1$. Where the noise power is much greater than the signal power, $H(e^{j\Omega}) \approx 0$. For example, when

$$S_{yy}(e^{j\Omega}) = (1 + e^{-j\Omega})(1 + e^{j\Omega}) = 2(1 + \cos\Omega) \tag{11.25}$$

and the noise is white, the optimal filter will be a low-pass filter with a frequency response that is appropriately shaped, shown in Figure 11.2. Note that the filter in
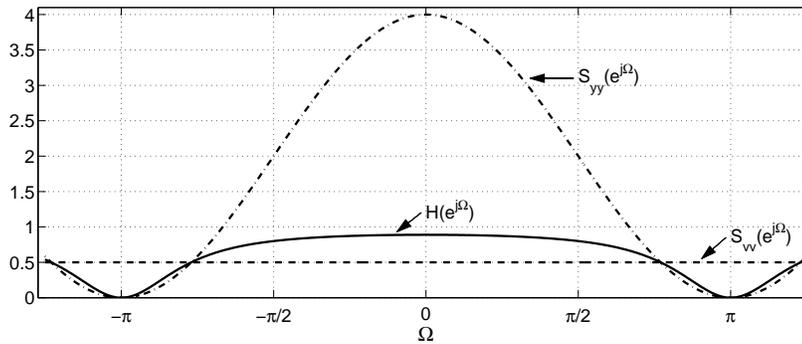

FIGURE 11.2   Optimal filter frequency response, $H(e^{j\Omega})$, input signal PSD signal, $S_{yy}(e^{j\Omega})$, and PSD of white noise, $S_{vv}(e^{j\Omega})$.

this case must have an impulse response that is an even function of time, since its frequency response is a real – and hence even – function of frequency.

Figure 11.3 shows a simulation example of such a filter in action (though for a different $S_{yy}(e^{j\Omega})$). The top plot is the PSD of the signal of interest; the middle plot shows both the signal $s[n]$ and the measured signal $x[n]$; and the bottom plot compares the estimate of $s[n]$ with $s[n]$ itself.

Power spectral density of AR(1) process

(a) Signal and Data

(b) Signal and Signal Estimate
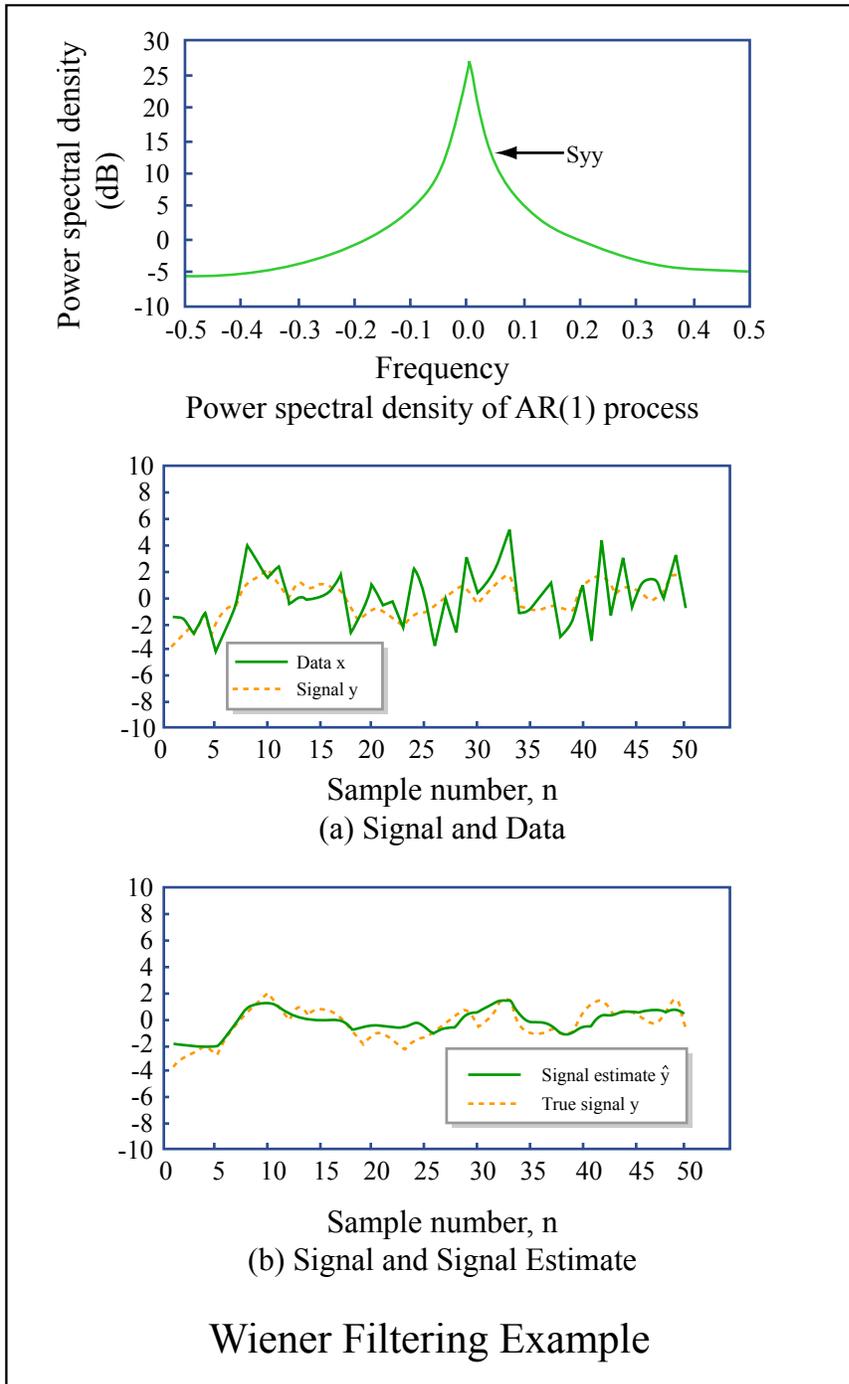
Wiener Filtering Example

Image by MIT OpenCourseWare, adapted from *Fundamentals of Statistical Signal Processing: Estimation Theory*, Steven Kay. Prentice Hall, 1993.

FIGURE 11.3 Wiener filtering example. (From S.M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, Prentice Hall, 1993. Figures 11.9 and 11.10.)

© *Alan V. Oppenheim and George C. Verghese, 2010*

EXAMPLE 11.2   Prediction

Suppose we wish to predict the measured process n_0 steps ahead, so

$$y[n] = x[n + n_0] \ . \tag{11.26}$$

Then

$$R_{yx}[m] = R_{xx}[m + n_0] \tag{11.27}$$

so the optimum filter has system function

$$H(z) = z^{n_0} \ . \tag{11.28}$$

This is of course not surprising: since we're allowing the filter to be noncausal, prediction is not a difficult problem! Causal prediction is much more challenging and interesting, and we will examine it later in this chapter.
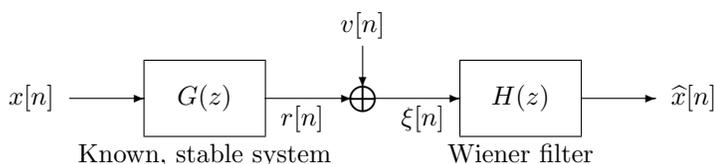

EXAMPLE 11.3   Deblurring (or Deconvolution)

FIGURE 11.4 Wiener filtering of a blurred and noisy signal.

In the Figure 11.4, $r[n]$ is a filtered or "blurred" version of the signal of interest, $x[n]$, while $v[n]$ is additive noise that is uncorrelated with $x[n]$. We wish to design a filter that will deblur the noisy measured signal $\xi[n]$ and produce an estimate of the input signal $x[n]$. Note that in the absence of the additive noise, the inverse filter $1/G(z)$ will recover the input exactly. However, this is not a good solution when noise is present, because the inverse filter accentuates precisely those frequencies where the measurement power is small relative to that of the noise. We shall therefore design a Wiener filter to produce an estimate of the signal $x[n]$.

We have shown that the cross-correlation between the measured signal, which is the input to the Wiener filter, and the estimate produced at its output is equal to the cross-correlation between the measurement process and the target process. In the transform domain, the statement of this condition is

$$S_{\widehat{x}\xi}(z) = S_{x\xi}(z) \tag{11.29}$$

or

$$S_{\xi\xi}(z)H(z) = S_{\widehat{x}\xi}(z) = S_{x\xi}(z) \ . \tag{11.30}$$

©*Alan V. Oppenheim and George C. Verghese, 2010*

We also know that

$$S_{\xi\xi}(z) = S_{vv}(z) + S_{xx}(z)G(z)G(1/z) \tag{11.31}$$
$$S_{x\xi}(z) = S_{xr}(z) \tag{11.32}$$
$$= S_{xx}(z)G(1/z), \tag{11.33}$$

where we have (in the first equality above) used the fact that $S_{vr}(z) = G(1/z)S_{vx}(z) = 0$. We can now write

$$H(z) = \frac{S_{xx}(z)G(1/z)}{S_{vv}(z) + S_{xx}(z)G(z)G(1/z)} \ . \tag{11.34}$$

We leave you to check that this system function assumes reasonable values in the limiting cases where the noise power is very small, or very large. It is also interesting to verify that the same overall filter is obtained if we first find an MMSE estimate $\widehat{r}[n]$ from $\xi[n]$ (as in Example 11.1), and then pass $\widehat{r}[n]$ through the inverse filter $1/G(z)$.

EXAMPLE 11.4   "De-Multiplication"

A message $s[n]$ is transmitted over a multiplicative channel (e.g. a fading channel) so that the received signal $r[n]$ is

$$r[n] = s[n]f[n] \ . \tag{11.35}$$

Suppose $s[n]$ and $f[n]$ are zero mean and independent. We wish to estimate $s[n]$ from $r[n]$ using a Wiener filter.

Again, we have

$$R_{sr}[m] = R_{\widehat{sr}}[m]$$
$$= h[m] * \underbrace{R_{rr}[m]}_{R_{ss}[m]R_{ff}[m]} \ . \tag{11.36}$$

But we also know that $R_{sr}[m] = 0$. Therefore $h[m] = 0$. This example emphasizes that the optimality of a filter satisfying certain constraints and minimizing some criterion does not necessarily make the filter a good one. The constraints on the filter and the criterion have to be relevant and appropriate for the intended task. For instance, if $f[n]$ was known to be i.i.d. and $+1$ or -1 at each time, then simply squaring the received signal $r[n]$ at any time would have at least given us the value of $s^2[n]$, which would seem to be more valuable information than what the Wiener filter produces in this case.

11.2 NONCAUSAL CT WIENER FILTER

In the previous discussion we derived and illustrated the discrete-time Wiener filter for the FIR and noncausal IIR cases. In this section we derive the continuous-time counterpart of the result for the noncausal IIR Wiener filter. The DT derivation involved taking derivatives with respect to a (countable) set of parameters $h[m]$, but in the CT case the impulse response that we seek to compute is a *CT function* $h(t)$, so the DT derivation cannot be directly copied. However, you will see that the results take the same form as in the DT case; furthermore, the derivation below has a natural DT counterpart, which provides an alternate route to the results in the preceding section.

Our problem is again stated in terms of Figure 11.5.

Estimator

$$x(t) \longrightarrow \boxed{h(t),\ H(j\omega)} \longrightarrow \widehat{y}(t) = \text{ estimate}$$
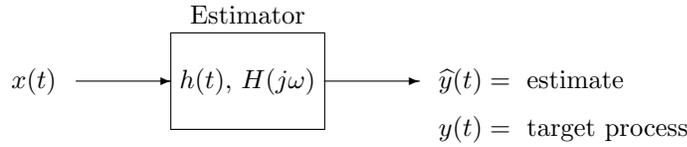$$y(t) = \text{ target process}$$

FIGURE 11.5 CT LTI filter for linear MMSE estimation.

Let $x(t)$ be a (zero-mean) WSS random process that we have measurements of. We want to determine the impulse response or frequency response of the above LTI system such that the filter output $\widehat{y}(t)$ is the LMMSE estimate of some (zero-mean) "target" process $y(t)$ that is jointly WSS with $x(t)$. We can again write

$$e(t) \overset{\Delta}{=} y(t) - \widehat{y}(t)$$
$$\min_{h(\,\cdot\,)} \epsilon = E\{e^2(t)\} \ . \tag{11.37}$$

Assuming the filter is stable (or at least has a well-defined frequency response), the process $\widehat{y}(t)$ is jointly WSS with $x(t)$. Furthermore,

$$E[\widehat{y}(t+\tau)y(t)] = h(\tau) * R_{xy}(\tau) = R_{\widehat{y}y}(\tau) \ , \tag{11.38}$$

The quantity we want to minimize can again be written as

$$\epsilon = E\{e^2(t)\} = R_{ee}(0) \ , \tag{11.39}$$

where the error autocorrelation function $R_{ee}(\tau)$ is — using the definition in (11.37) — evidently given by

$$R_{ee}(\tau) = R_{yy}(\tau) + R_{\widehat{y}\widehat{y}}(\tau) - R_{y\widehat{y}}(\tau) - R_{\widehat{y}y}(\tau) \ . \tag{11.40}$$

Thus

$$\epsilon = E\{e^2(t)\} = R_{ee}(0) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_{ee}(j\omega)\,d\omega$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} \left(S_{yy}(j\omega) + S_{\widehat{yy}}(j\omega) - S_{y\widehat{y}}(j\omega) - S_{\widehat{y}y}(j\omega) \right) d\omega$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} (S_{yy} + HH^*S_{xx} - H^*S_{yx} - HS_{xy})\,d\omega \ , \quad (11.41)$$

where we have dropped the argument $j\omega$ from the PSDs in the last line above for notational simplicity, and have used H^* to denote the complex conjugate of $H(j\omega)$, namely $H(-j\omega)$. The expression in this last line is obtained by using the fact that $x(t)$ and $\widehat{y}(t)$ are the WSS input and output, respectively, of a filter whose frequency response is $H(j\omega)$. Note also that because $R_{yx}(\tau) = R_{xy}(-\tau)$ we have

$$S_{yx} = S_{yx}(j\omega) = S_{xy}(-j\omega) = S_{xy}^* \ . \quad (11.42)$$

Our task is now to choose $H(j\omega)$ to minimize the integral in (11.41). We can do this by minimizing the integrand for each ω. The first term in the integrand does not involve or depend on H, so in effect we need to minimize

$$HH^*S_{xx} - H^*S_{yx} - HS_{xy} = HH^*S_{xx} - H^*S_{yx} - HS_{yx}^* \ . \quad (11.43)$$

If all the quantities in this equation were real, this minimization would be straightforward. Even with a complex H and S_{yx}, however, the minimization is not hard.

The key to the minimization is an elementary technique referred to as *completing the square*. For this, we write the quantity in (11.43) in terms of the squared magnitude of a term that is linear in H. This leads to the following rewriting of (11.43):

$$\left(H\sqrt{S_{xx}} - \frac{S_{yx}}{\sqrt{S_{xx}}} \right) \left(H^*\sqrt{S_{xx}} - \frac{S_{yx}^*}{\sqrt{S_{xx}}} \right) - \frac{S_{yx}S_{yx}^*}{S_{xx}} \ . \quad (11.44)$$

In writing $\sqrt{S_{xx}}$, we have made use of the fact that $S_{xx}(j\omega)$ is real and nonnegative. We have also felt free to divide by $\sqrt{S_{xx}(j\omega)}$ because for any ω where this quantity is 0 it can be shown that $S_{yx}(j\omega) = 0$ also. The optimal choice of $H(j\omega)$ is therefore arbitrary at such ω, as evident from (11.43). We thus only need to compute the optimal H at frequencies where $\sqrt{S_{xx}(j\omega)} > 0$.

Notice that the second term in parentheses in (11.44) is the complex conjugate of the first term, so the product of these two terms in parentheses is real and *nonnegative*. Also, the last term does not involve H at all. To cause the terms in parentheses to vanish and their product to thereby become 0, which is the best we can do, we evidently must choose as follows (assuming there are no additional constraints such as causality on the estimator):

$$H(j\omega) = \frac{S_{yx}(j\omega)}{S_{xx}(j\omega)} \quad (11.45)$$

This expression has the same form as in the DT case. The formula for $H(j\omega)$ causes it to inherit the symmetry properties of $S_{yx}(j\omega)$, so $H(j\omega)$ has a real part that is

even in ω, and an imaginary part that is odd in ω. Its inverse transform is thus a *real* impulse response $h(t)$, and the expression in (11.45) is the frequency response of the optimum (Wiener) filter.

With the choice of optimum filter frequency response in (11.45), the mean-square-error expression in (11.41) reduces (just as in the DT case) to:

$$
\begin{aligned}
\mathrm{MMSE} = R_{ee}(0) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} S_{ee}\, d\omega \\
&= \frac{1}{2\pi} \int_{-\infty}^{\infty} (S_{yy} - H S_{xy})\, d\omega \\
&= \frac{1}{2\pi} \int_{-\infty}^{\infty} S_{yy}\left(1 - \frac{S_{yx} S_{xy}}{S_{yy} S_{xx}}\right) d\omega \\
&= \frac{1}{2\pi} \int_{-\infty}^{\infty} S_{yy}(1 - \rho\rho^*)\, d\omega
\end{aligned}
\tag{11.46}
$$

where the function $\rho(j\omega)$ is defined by

$$
\rho(j\omega) = \frac{S_{yx}(j\omega)}{\sqrt{S_{yy}(j\omega) S_{xx}(j\omega)}}
\tag{11.47}
$$

and evidently plays the role of a (complex) frequency-by-frequency correlation coefficient, analogous to that played by the correlation coefficient of random variables Y and X.

11.2.1 Orthogonality Property

Rearranging the equation for the optimal Wiener filter, we find

$$
H\, S_{xx} = S_{yx}
\tag{11.48}
$$

or

$$
S_{\widehat{yx}} = S_{yx}\ ,
\tag{11.49}
$$

or equivalently

$$
R_{\widehat{yx}}(\tau) = R_{yx}(\tau) \text{ for all } \tau\ .
\tag{11.50}
$$

Again, for the optimal system, the cross-correlation between the input and output of the estimator equals the cross-correlation between the input and target output.

Yet another way to state the above result is via the following orthogonality property:

$$
R_{ex}(\tau) = R_{\widehat{yx}}(\tau) - R_{yx}(\tau) = 0 \text{ for all } \tau\ .
\tag{11.51}
$$

In other words, for the optimal system, the error is orthogonal to the data.

11.3 CAUSAL WIENER FILTERING

In the preceding discussion we developed the Wiener filter with no restrictions on the filter frequency response $H(j\omega)$. This allowed us to minimize a frequency-domain integral by choosing $H(j\omega)$ at each ω to minimize the integrand. However,

if we constrain the filter to be *causal*, then the frequency response cannot be chosen arbitrarily at each frequency, so the previous approach needs to be modified. It can be shown that for a causal system the real part of $H(j\omega)$ can be determined from the imaginary part, and vice versa, using what is known as a Hilbert transform. This shows that $H(j\omega)$ is constrained in the causal case. (We shall not need to deal explicitly with the particular constraint relating the real and imaginary parts of $H(j\omega)$, so we will not pursue the Hilbert transform connection here.) The development of the Wiener filter in the causal case is therefore subtler than the unrestricted case, but you know enough now to be able to follow the argument.

Recall our problem, described in terms of Figure 11.6.

$$\text{Estimator}$$

$$x(t) \longrightarrow \boxed{h(t),\ H(j\omega)} \longrightarrow \widehat{y}(t) = \text{ estimate}$$
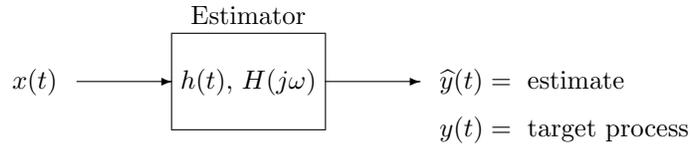
$$y(t) = \text{ target process}$$

FIGURE 11.6   Representation of LMMSE estimation using an LTI system.

The input $x(t)$ is a (zero-mean) WSS random process that we have measurements of, and we want to determine the impulse response or frequency response of the above LTI system such that the filter output $\widehat{y}(t)$ is the LMMSE estimate of some (zero-mean) "target" process $y(t)$ that is jointly WSS with $x(t)$:

$$e(t) \overset{\Delta}{=} y(t) - \widehat{y}(t)$$

$$\min_{h(\,\cdot\,)} \epsilon = E\{e^2(t)\} \ . \tag{11.52}$$

We shall now require, however, that the filter be *causal*. This is essential in, for example, the problem of prediction, where $y(t) = x(t + t_0)$ with $t_0 > 0$.

We have already seen that the quantity we want to minimize can be written as

$$\epsilon = E\{e^2(t)\} = R_{ee}(0) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_{ee}(j\omega)\, d\omega$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} \left(S_{yy}(j\omega) + S_{\widehat{y}\widehat{y}}(j\omega) - S_{y\widehat{y}}(j\omega) - S_{\widehat{y}y}(j\omega) \right) d\omega$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} \left(S_{yy} + HH^* S_{xx} - H^* S_{yx} - H S_{xy} \right) d\omega \quad (11.53)$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} \left| H\sqrt{S_{xx}} - \frac{S_{yx}}{\sqrt{S_{xx}}} \right|^2 d\omega + \frac{1}{2\pi} \int_{-\infty}^{\infty} \left(S_{yy} - \frac{S_{yx}S_{yx}^*}{S_{xx}} \right) d\omega \ .$$

$$\tag{11.54}$$

The last equality was the result of "completing the square" on the integrand in the preceding integral. In the case where H is unrestricted, we can set the first integral of the last equation to 0 by choosing

$$H(j\omega) = \frac{S_{yx}(j\omega)}{S_{xx}(j\omega)} \tag{11.55}$$

at each frequency. The second integral of the last equation is unaffected by our choice of H, and determines the MMSE.

If the Wiener filter is required to be *causal*, then we have to deal with the integral

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} \left| H \sqrt{S_{xx}} - \frac{S_{yx}}{\sqrt{S_{xx}}} \right|^2 d\omega \tag{11.56}$$

as a whole when we minimize it, because causality imposes constraints on $H(j\omega)$ that prevent it being chosen freely at each ω. (Because of the Hilbert transform relationship mentioned earlier, we could for instance choose the real part of $H(j\omega)$ freely, but then the imaginary part would be totally determined.) We therefore have to proceed more carefully.

Note first that the expression we obtained for the integrand in (11.56) by completing the square is actually not quite as general as we might have made it. Since we may need to use all the flexibility available to us when we tackle the constrained problem, we should explore how generally we can complete the square. Specifically, instead of using the real square root $\sqrt{S_{xx}}$ of the PSD S_{xx}, we could choose a *complex* square root M_{xx}, defined by the requirement that

$$S_{xx} = M_{xx} M_{xx}^* \qquad \text{or} \qquad S_{xx}(j\omega) = M_{xx}(j\omega) M_{xx}(-j\omega) \ , \tag{11.57}$$

and correspondingly rewrite the criterion in (11.56) as

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} \left| H M_{xx} - \frac{S_{yx}}{M_{xx}^*} \right|^2 d\omega \ , \tag{11.58}$$

which is easily verified to be the same criterion, although written differently. The quantity $M_{xx}(j\omega)$ is termed a spectral factor of $S_{xx}(j\omega)$ or a modeling filter for the process x. The reason for the latter name is that passing (zero-mean) unit-variance white noise through a filter with frequency response $M_{xx}(j\omega)$ will produce a process with the PSD $S_{xx}(j\omega)$, so we can model the process x as being the result of such a filtering operation. Note that the real square root $\sqrt{S_{xx}}(j\omega)$ we used earlier is a special case of a spectral factor, but others exist. In fact, multiplying $\sqrt{S_{xx}}(j\omega)$ by an all-pass frequency response $A(j\omega)$ will yield a modeling filter:

$$A(j\omega) \sqrt{S_{xx}}(j\omega) = M_{xx}(j\omega) \ , \qquad A(j\omega)A(-j\omega) = 1 \ . \tag{11.59}$$

Conversely, it is easy to show that the frequency response of any modeling filter can be written as the product of an all-pass frequency response and $\sqrt{S_{xx}}(j\omega)$.

It turns out that under fairly mild conditions (which we shall not go into here) a PSD is guaranteed to have a spectral factor that is the frequency response of a *stable and causal* system, and whose *inverse* is also the frequency response of a *stable and causal* system. (To simplify how we talk about such factors, we shall adopt an abuse of terminology that is common when talking about Fourier transforms, referring to the factor itself — rather than the system whose frequency response is this factor — as being stable and causal, with a stable and causal inverse.) For instance, if

$$S_{xx}(j\omega) = \frac{\omega^2 + 9}{\omega^2 + 4} \ , \tag{11.60}$$

then the required factor is

$$M_{xx}(j\omega) = \frac{j\omega + 3}{j\omega + 2} \; . \tag{11.61}$$

We shall limit ourselves entirely to S_{xx} that have such a spectral factor, and assume for the rest of the derivation that the M_{xx} introduced in the criterion (11.58) is such a factor. (Keep in mind that wherever we ask for a stable system here, we can actually make do with a system with a well-defined frequency response, even if it's not BIBO stable, except that our results may then need to be interpreted more carefully.)

With these understandings, it is evident that the term HM_{xx} in the integrand in (11.58) is causal, as it is the cascade of two causal terms. The other term, S_{yx}/M_{xx}^*, is generally not causal, but we may separate its causal part out, denoting the transform of its causal part by $[S_{yx}/M_{xx}^*]_+$, and the transform of its anti-causal part by $[S_{yx}/M_{xx}^*]_-$. (In the DT case, the latter would actually denote the transform of the *strictly* anti-causal part, i.e., at times -1 and earlier; the value at time 0 would be retained with the causal part.)

Now consider rewriting (11.58) in the time domain, using Parseval's theorem. If we denote the inverse transform operation by $\mathcal{I}\{\cdot\}$, then the result is the following rewriting of our criterion:

$$\int_{-\infty}^{\infty} \left| \mathcal{I}\{HM_{xx}\} - \mathcal{I}\{[S_{yx}/M_{xx}^*]_+\} - \mathcal{I}\{[S_{yx}/M_{xx}^*]_-\} \right|^2 dt \tag{11.62}$$

Since the term $\mathcal{I}\{HM_{xx}\}$ is causal (i.e., zero for negative time), the best we can do with it, as far as minimizing this integral is concerned, is to cancel out all of $\mathcal{I}\{[S_{yx}/M_{xx}^*]_+\}$. In other words, our best choice is

$$HM_{xx} = [S_{yx}/M_{xx}^*]_+ \; , \tag{11.63}$$

or

$$H(j\omega) = \frac{1}{M_{xx}(j\omega)} \left[\frac{S_{yx}(j\omega)}{M_{xx}(-j\omega)} \right]_+ \; . \tag{11.64}$$

Note that the stability and causality of the *inverse* of M_{xx} guarantee that this last step preserves stability and causality, respectively, of the solution.

The expression in (11.64) is the solution of the Wiener filtering problem under the causality constraint. It is also evident now that the MMSE is larger than in the unconstrained (noncausal) case by the amount

$$\Delta\text{MMSE} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \left| \left[\frac{S_{yx}}{M_{xx}^*} \right]_- \right|^2 d\omega \; . \tag{11.65}$$

EXAMPLE 11.5   DT Prediction

Although the preceding results were developed for the CT case, exactly analogous expressions with obvious modifications (namely, using the DTFT instead of the

CTFT, with integrals from $-\pi$ to π rather than $-\infty$ to ∞, etc.) apply to the DT case.

Consider a process $x[n]$ that is the result of passing (zero-mean) white noise of unit variance through a (modeling) filter with frequency response

$$M_{xx}(e^{j\Omega}) = \alpha_0 + \alpha_1 e^{-j\Omega} \,, \tag{11.66}$$

where both α_0 and α_1 are assumed nonzero. This filter is stable and causal, and if $|\alpha_1| < |\alpha_0|$ then the inverse is stable and causal too. We assume this condition holds. (If it doesn't, we can always find another modeling filter for which it does, by multiplying the present filter by an appropriate allpass filter.)

Suppose we want to do causal one-step prediction for this process, so $y[n] = x[n+1]$. Then $R_{yx}[m] = R_{xx}[m+1]$, so

$$S_{yx} = e^{j\Omega} S_{xx} = e^{j\Omega} M_{xx} M_{xx}^* \,. \tag{11.67}$$

Thus

$$\left[\frac{S_{yx}}{M_{xx}^*} \right]_+ = [e^{j\Omega} M_{xx}]_+ = \alpha_1 \,, \tag{11.68}$$

and so the optimum filter, according to (11.64), has frequency response

$$H(e^{j\Omega}) = \frac{\alpha_1}{\alpha_0 + \alpha_1 e^{-j\Omega}} \,. \tag{11.69}$$

The associated MMSE is evaluated by the expression in (11.65), and turns out to be simply α_0^2 (which can be compared with the value of $\alpha_0^2 + \alpha_1^2$ that would have been obtained if we estimated $x[n+1]$ by just its mean value, namely zero).

11.3.1   Dealing with Nonzero Means

We have so far considered the case where both x and y have zero means (and the practical consequence has been that we haven't had to worry about their PSDs having impulses at the origin). If their means are nonzero, then we can do a better job of estimating $y(t)$ if we allow ourselves to adjust the estimates produced by the LTI system, by adding appropriate constants (to make an *affine* estimator). For this, we can first consider the problem of estimating $y - \mu_y$ from $x - \mu_x$, illustrated in Figure 11.7
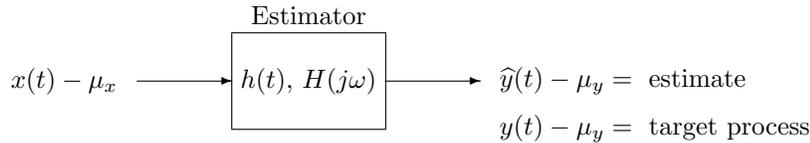

FIGURE 11.7 Wiener filtering with non-zero means.

Denoting the transforms of the *covariances* $C_{xx}(\tau)$ and $C_{yx}(\tau)$ by $D_{xx}(j\omega)$ and $D_{yx}(j\omega)$ respectively (these transforms are sometimes referred to as *covariance*

PSDs), the optimal unconstrained Wiener filter for our task will evidently have a frequency response given by

$$H(j\omega) = \frac{D_{yx}(j\omega)}{D_{xx}(j\omega)} \ . \tag{11.70}$$

We can then add μ_y to the output of this filter to get our LMMSE estimate of $y(t)$.

6.011 Introduction to Communication, Control, and Signal Processing
Spring 2010