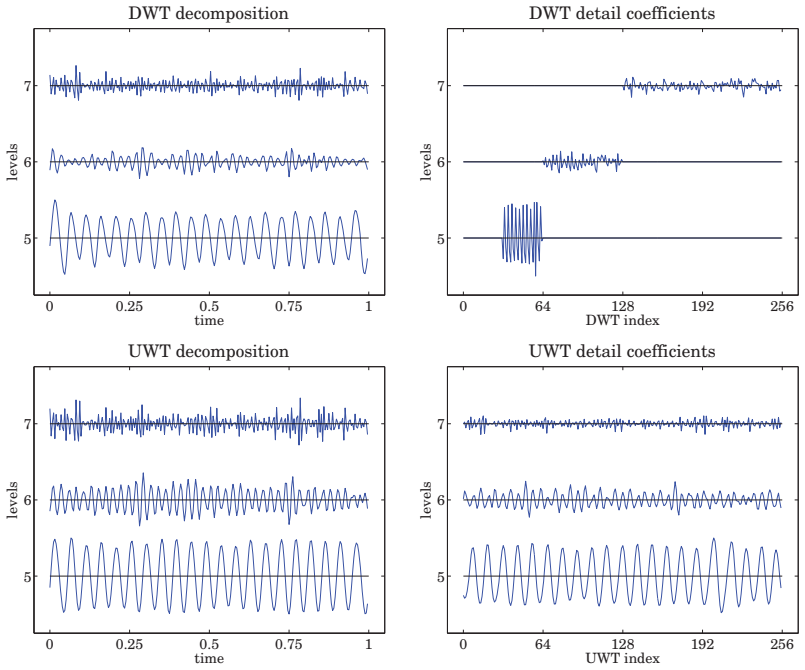


Fig. 10.9.1 UWT/DWT decompositions and wavelet coefficients of housing data.

10.3 Prove the downsampling replication property (10.4.11) by working backwards, that is, start from the Fourier transform expression and show that

$$\frac{1}{L}\sum_{m=0}^{L-1} X(f - mf_s^{\text{down}}) = \sum_k s(k)x(k)e^{-2\pi jfk/f_s} = \sum_n x(nL)e^{-2\pi jfnL/f_s} = Y_{\text{down}}(f)$$

where $s(k)$ is the periodic "sampling function" with the following representations:

$$s(k) = \frac{1}{L}\sum_{m=0}^{L-1} e^{-2\pi jkm/L} = \frac{1}{L}\frac{1-e^{-2\pi jk}}{1-e^{-2\pi jk/L}} = \sum_n \delta(k-nL)$$

Moreover, show that the above representations are nothing but the inverse L-point DFT of the DFT of one period of the periodic pulse train:

$$s(k) = [\ldots, 1, \underbrace{0,0,\ldots,0}_{L-1 \text{ zeros}}, 1, \underbrace{0,0,\ldots,0}_{L-1 \text{ zeros}}, 1, \underbrace{0,0,\ldots,0}_{L-1 \text{ zeros}}, \ldots\ ] = \sum_n \delta(k-nL)$$

10.4 Show that the solution to the optimization problem (10.7.7) is the soft-thresholding rule of Eq. (10.7.8).

10.5 Study the "Tikhonov regularizer" wavelet thresholding function:

$$d_{\text{thr}} = f(d,\lambda,a) = d\,\frac{|d|^a}{|d|^a + \lambda^a}, \quad a > 0,\ \lambda > 0$$

11

Wiener Filtering

The problem of estimating one signal from another is one of the most important in signal processing. In many applications, the desired signal is not available or observable directly. Instead, the observable signal is a degraded or distorted version of the original signal. The signal estimation problem is to recover, in the best way possible, the desired signal from its degraded replica.

We mention some typical examples: (1) The desired signal may be corrupted by strong additive noise, such as weak evoked brain potentials measured against the strong background of ongoing EEGs; or weak radar returns from a target in the presence of strong clutter. (2) An antenna array designed to be sensitive towards a particular "look" direction may be vulnerable to strong jammers from other directions due to sidelobe leakage; the signal processing task here is to null the jammers while at the same time maintaining the sensitivity of the array towards the desired look direction. (3) A signal transmitted over a communications channel can suffer phase and amplitude distortions and can be subject to additive channel noise; the problem is to recover the transmitted signal from the distorted received signal. (4) A Doppler radar processor tracking a moving target must take into account dynamical noise—such as small purely random accelerations—affecting the dynamics of the target, as well as measurement errors. (5) An image recorded by an imaging system is subject to distortions such as blurring due to motion or to the finite aperture of the system, or other geometric distortions; the problem here is to undo the distortions introduced by the imaging system and restore the original image. A related problem, of interest in medical image processing, is that of reconstructing an image from its projections. (6) In remote sensing and inverse scattering applications, the basic problem is, again, to infer one signal from another; for example, to infer the temperature profile of the atmosphere from measurements of the spectral distribution of infrared energy; or to deduce the structure of a dielectric medium, such as the ionosphere, by studying its response to electromagnetic wave scattering; or, in oil exploration to infer the layered structure of the earth by measuring its response to an impulsive input near its surface.

In this chapter, we pose the signal estimation problem and discuss some of the criteria used in the design of signal estimation algorithms.

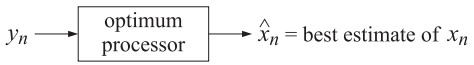
We do not present a complete discussion of all methods of signal recovery and estimation that have been invented for applications as diverse as those mentioned above.

Our emphasis is on traditional linear least-squares estimation methods, not only because they are widely used, but also because they have served as the motivating force for the development of other estimation techniques and as the yardstick for evaluating them.

We develop the theoretical solution of the Wiener filter both in the stationary and nonstationary cases, and discuss its connection to the orthogonal projection, Gram-Schmidt constructions, and correlation canceling ideas of Chap. 1. By means of an example, we introduce Kalman filtering concepts and discuss their connection to Wiener filtering and to signal modeling. Practical implementations of the Wiener filter are discussed in Chapters 12 and 16. Other signal recovery methods for deconvolution applications that are based on alternative design criteria are briefly discussed in Chap. 12, where we also discuss some interesting connections between Wiener filtering/linear prediction methods and inverse scattering methods.

11.1 Linear and Nonlinear Estimation of Signals

The signal estimation problem can be stated as follows: We wish to estimate a random signal x_n on the basis of available observations of a related signal y_n. The available signal y_n is to be processed by an optimal processor that produces the best possible estimate of x_n:



The resulting estimate $\hat{x}_n$ will be a function of the observations y_n. If the optimal processor is linear, such as a linear filter, then the estimate $\hat{x}_n$ will be a linear function of the observations. We are going to concentrate mainly on linear processors. However, we would like to point out that, depending on the estimation criterion, there are cases where the estimate $\hat{x}_n$ may turn out to be a nonlinear function of the y_ns.

We discuss briefly four major estimation criteria for designing such optimal processors. They are:

(1) The maximum a posteriori (MAP) criterion.
(2) The maximum likelihood (ML) criterion.
(3) The mean square (MS) criterion.
(4) The linear mean-square (LMS) criterion.

The LMS criterion is a special case of the MS criterion. It requires, *a priori*, that the estimate $\hat{x}_n$ be a *linear* function of the y_ns.[†] The main advantage of the LMS processor is that it requires only knowledge of second order statistics for its design, whereas the other, nonlinear, processors require more detailed knowledge of probability densities.

To explain the various estimation criteria, let us assume that the desired signal x_n is to be estimated over a finite time interval $n_a \le n \le n_b$ Without loss of generality, we may assume that the observed signal y_n is also available over the same interval. Define

[†]Note that the acronym LMS is also used in the context of adaptive filtering, for *least mean-square*.

the vectors

$$\mathbf{x} = \begin{bmatrix} x_{n_a} \\ x_{n_a+1} \\ \vdots \\ x_{n_b} \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} y_{n_a} \\ y_{n_a+1} \\ \vdots \\ y_{n_b} \end{bmatrix}$$

For each value of n, we seek the functional dependence

$$\hat{x}_n = \hat{x}_n(\mathbf{y})$$

of $\hat{x}_n$ on the given observation vector $\mathbf{y}$ that provides the *best estimate* of x_n.

1. The criterion for the MAP estimate is to maximize the a posteriori conditional density of x_n given that $\mathbf{y}$ already occurred; namely,

$$p(x_n|\mathbf{y}) = \text{maximum} \tag{11.1.1}$$

in other words, the optimal estimate $\hat{x}_n$ is that x_n that maximizes this quantity for the given vector $\mathbf{y}$; $\hat{x}_n$ is therefore the most probable choice resulting from the given observations $\mathbf{y}$.

2. The ML criterion, on the other hand, selects $\hat{x}_n$ to maximize the conditional density of $\mathbf{y}$ given x_n, that is,

$$p(\mathbf{y}|x_n) = \text{maximum} \tag{11.1.2}$$

This criterion selects $\hat{x}_n$ as though the already collected observations $\mathbf{y}$ were the most likely to occur.

3. The MS criterion minimizes the mean-square estimation error

$$\mathcal{E} = E[e_n^2] = \min, \quad \text{where} \quad e_n = x_n - \hat{x}_n \tag{11.1.3}$$

that is, the best choice of the functional dependence $\hat{x}_n = \hat{x}_n(\mathbf{y})$ is sought that minimizes this expression. We know from our results of Sec. 1.4 that the required solution is the corresponding *conditional mean*

$$\hat{x}_n = E[x_n|\mathbf{y}] = \text{MS estimate} \tag{11.1.4}$$

computed with respect to the conditional density $p(x_n|\mathbf{y})$.

4. Finally, the LMS criterion requires the estimate to be a linear function of the observations

$$\hat{x}_n = \sum_{i=n_a}^{n_b} h(n,i)\, y_i \tag{11.1.5}$$

For each n, the weights $h(n,i)$, $n_a \le i \le n_b$ are selected to *minimize* the mean-square estimation error

$$\mathcal{E} = E[e_n^2] = E\big[(x_n - \hat{x}_n)^2\big] = \text{minimum} \tag{11.1.6}$$

With the exception of the LMS estimate, all other estimates $\hat{x}_n(\mathbf{y})$ are, in general, nonlinear functions of $\mathbf{y}$.

Example 11.1.1: If both x_n and $\mathbf{y}$ are zero-mean and jointly gaussian, then Examples 1.4.1 and 1.4.2 imply that the MS and LMS estimates of x_n are the *same*. Furthermore, since $p(x_n|\mathbf{y})$ is gaussian it will be symmetric about its maximum, which occurs at its mean, that is, at $E[x_n|\mathbf{y}]$. Therefore, the MAP estimate of x_n is equal to the MS estimate. In conclusion, for *zero-mean jointly gaussian* x_n and $\mathbf{y}$, the three estimates MAP, MS, and LMS coincide. □

Example 11.1.2: To see the nonlinear character and the differences among the various estimates, consider the following example: A discrete-amplitude, constant-in-time signal x can take on the three values

$$x = -1, \quad x = 0, \quad x = 1$$

each with probability of $1/3$. This signal is placed on a known carrier waveform c_n and transmitted over a noisy channel. The received samples are of the form

$$y_n = c_n x + v_n\,, \quad n = 1, 2, \ldots, M$$

where v_n are zero-mean white gaussian noise samples of variance σ_v^2, assumed to be independent of x. The above set of measurements can be written in an obvious vector notation

$$\mathbf{y} = \mathbf{c}\,x + \mathbf{v}$$

(a) Determine the conditional densities $p(\mathbf{y}|x)$ and $p(x|\mathbf{y})$.

(b) Determine and compare the four alternative estimates MAP, ML, MS, and LMS.

Solution: To compute $p(\mathbf{y}|x)$, note that if x is given, then the only randomness left in $\mathbf{y}$ arises from the noise term $\mathbf{v}$. Since v_n are uncorrelated and gaussian, they will be independent; therefore,

$$p(\mathbf{y}|x) = p(\mathbf{v}) = \prod_{n=1}^{M} p(v_n) = (2\pi\sigma_v^2)^{-M/2} \exp\left[-\frac{1}{2\sigma_v^2}\sum_{n=1}^{M} v_n^2\right]$$

$$= (2\pi\sigma_v^2)^{-M/2} \exp\left[-\frac{1}{2\sigma_v^2}\mathbf{v}^2\right] = (2\pi\sigma_v^2)^{-M/2} \exp\left[-\frac{1}{2\sigma_v^2}(\mathbf{y}-\mathbf{c}\,x)^2\right]$$

Using Bayes' rule we find $p(x|\mathbf{y}) = p(\mathbf{y}|x)\,p(x)/p(\mathbf{y})$. Since

$$p(x) = \frac{1}{3}\left[\delta(x-1)+\delta(x)+\delta(x+1)\right]$$

we find

$$p(x|\mathbf{y}) = \frac{1}{A}\left[p(\mathbf{y}|1)\,\delta(x-1)+p(\mathbf{y}|0)\,\delta(x)+p(\mathbf{y}|-1)\,\delta(x+1)\right]$$

where the constant A is

$$A = 3p(\mathbf{y}) = 3\int p(\mathbf{y}|x)\,p(x)\,dx = p(\mathbf{y}|1)+p(\mathbf{y}|0)+p(\mathbf{y}|-1)$$

To find the MAP estimate of x, the quantity $p(x|\mathbf{y})$ must be maximized with respect to x. Since the expression for $p(x|\mathbf{y})$ forces x to be one of the three values $+1, 0, -1$, it follows

that the maximum among the three coefficients $p(\mathbf{y}|1)$, $p(\mathbf{y}|0)$, $p(\mathbf{y}|-1)$ will determine the value of x. Thus, for a given $\mathbf{y}$ we select that x that

$$p(\mathbf{y}|x) = \text{maximum of}\{p(\mathbf{y}|1),\, p(\mathbf{y}|0),\, p(\mathbf{y}|-1)\}$$

Using the gaussian nature of $p(\mathbf{y}|x)$, we find equivalently

$$(\mathbf{y}-\mathbf{c}\,x)^2 = \text{minimum of}\{(\mathbf{y}-\mathbf{c})^2,\, \mathbf{y}^2,\, (\mathbf{y}+\mathbf{c})^2\}$$

Subtracting $\mathbf{y}^2$ from both sides, dividing by $\mathbf{c}^T\mathbf{c}$, and denoting

$$\bar{y} = \frac{\mathbf{c}^T\mathbf{y}}{\mathbf{c}^T\mathbf{c}}$$

we find the equivalent equation

$$x^2 - 2x\bar{y} = \min\{1-2\bar{y},\, 0,\, 1+2\bar{y}\}$$

and in particular, applying these for $+1, 0, -1$, we find

$$\hat{x}_{\text{MAP}} = \begin{cases} 1, & \text{if } \bar{y} > \dfrac{1}{2} \\ 0, & \text{if } -\dfrac{1}{2} < \bar{y} < \dfrac{1}{2} \\ -1, & \text{if } \bar{y} < -\dfrac{1}{2} \end{cases}$$

To determine the ML estimate, we must maximize $p(\mathbf{y}|x)$ with respect to x. The ML estimate does not require knowledge of the a priori probability density $p(x)$ of x. Therefore, differentiating $p(\mathbf{y}|x)$ with respect to x and setting the derivative to zero gives

$$\frac{\partial}{\partial x}p(\mathbf{y}|x) = 0 \quad \text{or} \quad \frac{\partial}{\partial x}\ln p(\mathbf{y}|x) = 0 \quad \text{or} \quad \frac{\partial}{\partial x}(\mathbf{y}-\mathbf{c}\,x)^2 = 0$$

which gives

$$\hat{x}_{\text{ML}} = \frac{\mathbf{c}^T\mathbf{y}}{\mathbf{c}^T\mathbf{c}} = \bar{y}$$

The MS estimate is obtained by computing the conditional mean

$$E[x|\mathbf{y}] = \int x\,p(x|\mathbf{y})\,dx = \int x\frac{1}{A}\left[p(\mathbf{y}|1)\,\delta(x-1)+p(\mathbf{y}|0)\,\delta(x)+p(\mathbf{y}|-1)\,\delta(x+1)\right]dx$$

$$= \frac{1}{A}\left[p(\mathbf{y}|1)-p(\mathbf{y}|-1)\right], \quad \text{or,}$$

$$\hat{x}_{\text{MS}} = \frac{p(\mathbf{y}|1)-p(\mathbf{y}|-1)}{p(\mathbf{y}|1)+p(\mathbf{y}|0)+p(\mathbf{y}|-1)}$$

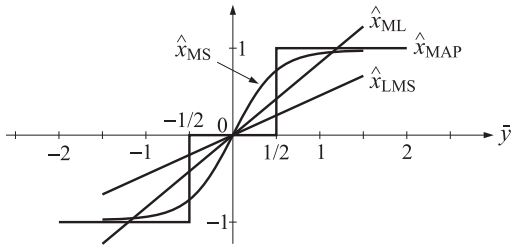
Canceling some common factors from the numerator and denominator, we find the simpler expression

$$\hat{x}_{\text{MS}} = \frac{2\sinh(2a\bar{y})}{e^a + 2\cosh(2a\bar{y})}, \quad \text{where} \quad a = \frac{\mathbf{c}^T\mathbf{c}}{2\sigma_v^2}$$

Finally, the LMS estimate can be computed as in Example 1.4.3. We find

$$\hat{x}_{\text{LMS}} = \frac{\mathbf{c}^T\mathbf{y}}{\dfrac{\sigma_v^2}{\sigma_x^2} + \mathbf{c}^T\mathbf{c}} = \frac{\mathbf{c}^T\mathbf{c}}{\dfrac{\sigma_v^2}{\sigma_x^2} + \mathbf{c}^T\mathbf{c}}\,\bar{y}$$

All four estimates have been expressed in terms of $\bar{y}$. Note that the ML estimate is linear but has a different slope than the LMS estimate. The nonlinearity of the various estimates is best seen in the following figure:



11.2 Orthogonality and Normal Equations

From now on, we will concentrate on the optimal linear estimate defined by Eqs. (11.1.5) and (11.1.6). For each time instant n at which an estimate $\hat{x}_n$ is sought, the optimal weights $h(n,i)$, $n_a \le i \le n_b$ must be determined that minimize the error criterion (11.1.6). In general, a new set of optimal weights must be computed for each time instant n. In the special case when the processes x_n and y_n are stationary and the observations are available for a long time, that is, $n_a = -\infty$, the weights become time-invariant in the sense that $h(n,i) = h(n-i)$, and the linear processor becomes an ordinary *time-invariant linear filter*. We will discuss the solution for $h(n,i)$ both for the time-invariant and the more general cases.

The problem of determining the optimal weights $h(n,i)$ according to the mean-square error minimization criterion (11.1.6) is in general referred to as the *Wiener filtering problem* [849–866]. An interesting historical account of the development of this problem and its ramifications is given in the review article by Kailath [866]. Wiener filtering problems are conventionally divided into three types:

1. The optimal *smoothing* problem,
2. The optimal *filtering* problem, and
3. The optimal *prediction* problem.

In all cases, the optimal estimate of x_n at a given time instant n is given by an expression of the form (11.1.5), as a linear combination of the available observations

y_n in the interval $n_a \le n \le n_b$. The division into three types of problems depends on which of the available observations in that interval are taken into account in making up the linear combination (11.1.5).

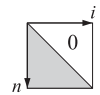
In the smoothing problem, *all* the observations in the interval $[n_a, n_b]$ are taken into account. The shaded part in the following figure denotes the range of observations that are used in the summation of Eq. (11.1.5):

$$\hat{x}_n = \sum_{i=n_a}^{n_b} h(n,i)\,y_i$$

Since some of the observations are to the future of x_n, the linear operation is not causal. This does not present a problem if the sequence y_n is already available and stored in memory.

The optimal filtering problem, on the other hand, requires the linear operation (11.1.5) to be *causal*, that is, only those observations that are in the present and past of the current sample x_n must be used in making up the estimate $\hat{x}_n$. This requires that the matrix of optimal weights $h(n,i)$ be lower triangular, that is,

$$h(n,i) = 0, \quad \text{for } n < i$$



Thus, in reference to the figure below, only the shaded portion of the observation interval is used at the current time instant:

$$\hat{x}_n = \sum_{i=n_a}^{n} h(n,i)\,y_i$$

The estimate $\hat{x}_n$ depends on the present and all the past observations, from the fixed starting point n_a to the current time instant n. As n increases, more and more observations are taken into account in making up the estimate, and the actual computation of $\hat{x}_n$ becomes less and less efficient. It is desirable, then, to be able to recast the expression for $\hat{x}_n$ a time-recursive form. This is what is done in Kalman filtering. But, there is another way to make the Wiener filter computationally manageable. Instead of allowing a growing number of observations, only the current and the past M observations y_i, $i = n, n-1, \ldots, n-M$ are taken into account. In this case, only $(M+1)$ filter weights are to be computed at each time instant n. This is depicted below:

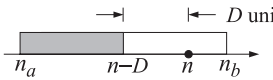
$$\hat{x}_n = \sum_{i=n-M}^{n} h(n,i)\,y_i = \sum_{m=0}^{M} h(n,n-m)\,y_{n-m}$$

This is referred to as the *finite impulse response* (FIR) Wiener filter. Because of its simple implementation, the FIR Wiener filter has enjoyed widespread popularity. Depending on the particular application, the practical implementation of the filter may vary. In Sec. 11.3 we present the theoretical formulation that applies to the stationary case; in Chap. 12 we reconsider it as a waveshaping and spiking filter and discuss a number of deconvolution applications. In Chap. 16, we consider its adaptive implementation using the Widrow-Hoff LMS algorithm and discuss a number of applications such

as channel equalization and echo cancellation; we also discuss two alternative adaptive implementations—the so-called "gradient lattice," and the "recursive least-squares."

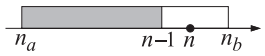
Finally, the linear prediction problem is a special case of the optimal filtering problem with the additional stipulation that observations only up to time instant $n-D$ must be used in obtaining the current estimate $\hat{x}_n$; this is equivalent to the problem of predicting D units of time into the future. The range of observations used in this case is shown below:

$$\hat{x}_n = \sum_{i=n_a}^{n-D} h(n,i)\,y_i$$



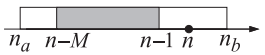
Of special interest to us will be the case of one-step prediction, corresponding to the choice $D=1$. This is depicted below:

$$\hat{x}_n = \sum_{i=n_a}^{n-1} h(n,i)\,y_i$$



If we demand that the prediction be based only on the past M samples (from the current sample), we obtain the FIR version of the prediction problem, referred to as *linear prediction based on the past M samples*, which is depicted below:

$$\hat{x}_n = \sum_{i=n-M}^{n-1} h(n,i)\,y_i = \sum_{m=1}^{M} h(n,n-m)\,y_{n-m}$$



Next, we set up the orthogonality and normal equations for the optimal weights. We begin with the smoothing problem. The estimation error is in this case

$$e_n = x_n - \hat{x}_n = x_n - \sum_{i=n_a}^{n_b} h(n,i)\,y_i \tag{11.2.1}$$

Differentiating the mean-square estimation error (11.1.6) with respect to each weight $h(n,i)$, $n_a \le i \le n_b$, and setting the derivative to zero, we obtain the orthogonality equations that are enough to determine the weights:

$$\frac{\partial \mathcal{E}}{\partial h(n,i)} = 2E\left[e_n \frac{\partial e_n}{\partial h(n,i)}\right] = -2E[e_n y_i] = 0\,, \quad \text{for } n_a \le i \le n_b\,, \quad \text{or,}$$

$$R_{ey}(n,i) = E[e_n y_i] = 0 \quad \text{(orthogonality equations)} \tag{11.2.2}$$

for $n_a \le i \le n_b$. Thus, the estimation error e_n is orthogonal (uncorrelated) to each observation y_i used in *making up* the estimate $\hat{x}_n$. The orthogonality equations provide exactly as many equations as there are unknown weights.

Inserting Eq. (11.2.1) for e_n, the orthogonality equations may be written in an equivalent form, known as the *normal equations*

$$E\big[\big(x_n - \sum_{k=n_a}^{n_b} h(n,k)\,y_k\big)\,y_i\big] = 0\,, \quad \text{or,}$$

$$E[x_n y_i] = \sum_{k=n_a}^{n_b} h(n,k)\,E[y_k y_i] \quad \text{(normal equations)} \tag{11.2.3}$$

These determine the optimal weights at the current time instant n. In the vector notation of Sec. 11.1, we write Eq. (11.2.3) as

$$E[\mathbf{x}\mathbf{y}^T] = HE[\mathbf{y}\mathbf{y}^T]$$

where H is the matrix of weights $h(n,i)$. The optimal H and the estimate are then

$$\hat{\mathbf{x}} = H\mathbf{y} = E[\mathbf{x}\mathbf{y}^T]E[\mathbf{y}\mathbf{y}^T]^{-1}\mathbf{y}$$

This is identical to the correlation canceler of Sec. 1.4. The orthogonality equations (11.2.2) are precisely the correlation cancellation conditions. Extracting the nth row of this matrix equation, we find an explicit expression for the nth estimate $\hat{x}_n$

$$\hat{x}_n = E[x_n\mathbf{y}^T]E[\mathbf{y}\mathbf{y}^T]^{-1}\mathbf{y}$$

which is recognized as the projection of the random variable x_n onto the subspace spanned by the available observations; namely, $Y = \{y_{n_a}, y_{n_a+1}, \ldots, y_{n_b}\}$. This is a general result: The minimum mean-square linear estimate $\hat{x}_n$ is the projection of x_n onto the subspace spanned by all the observations that are used to make up that estimate. This result is a direct consequence of the quadratic minimization criterion (11.1.6) and the orthogonal projection theorem discussed in Sec. 1.6.

Using the methods of Sec. 1.4, the minimized estimation error at time instant n is easily computed by

$$\mathcal{E}_n = E[e_n e_n] = E[e_n x_n] = E\big[\big(x_n - \sum_{i=n_a}^{n_b} h(n,i)\,y_i\big)x_n\big]$$

$$= E[x_n^2] - \sum_{i=n_a}^{n_b} h(n,i)\,E[y_i x_n] = E[x_n^2] - E[x_n\mathbf{y}^T]E[\mathbf{y}\mathbf{y}^T]^{-1}E[\mathbf{y}x_n]$$

which corresponds to the diagonal entries of the covariance matrix of the estimation error $\mathbf{e}$:

$$R_{ee} = E[\mathbf{e}\mathbf{e}^T] = E[\mathbf{x}\mathbf{x}^T] - E[\mathbf{x}\mathbf{y}^T]E[\mathbf{y}\mathbf{y}^T]^{-1}E[\mathbf{y}\mathbf{x}^T]$$

The *optimum filtering* problem is somewhat more complicated because of the causality condition. In this case, the estimate at time n is given by

$$\hat{x}_n = \sum_{i=n_a}^{n} h(n,i)\,y_i \tag{11.2.4}$$

Inserting this into the minimization criterion (11.1.6) and differentiating with respect to $h(n,i)$ for $n_a \le i \le n$, we find again the orthogonality conditions

$$R_{ey}(n,i) = E[e_n y_i] = 0 \quad \text{for } n_a \le i \le n \tag{11.2.5}$$

where the most important difference from Eq. (11.2.2) is the restriction on the range of i, that is, e_n is decorrelated only from the present and past values of y_i. Again, the estimation error e_n is orthogonal to each observation y_i that is being used to make up

the estimate. The orthogonality equations can be converted into the normal equations as follows:

$$E[e_n y_i] = E\big[(x_n - \sum_{k=n_a}^{n} h(n,k)y_k)y_i\big] = 0\,, \quad \text{or,}$$

$$E[x_n y_i] = \sum_{k=n_a}^{n} h(n,k)E[y_k y_i] \quad \text{for } n_a \le i \le n\,, \quad \text{or,} \tag{11.2.6}$$

$$R_{xy}(n,i) = \sum_{k=n_a}^{n} h(n,k)R_{yy}(k,i) \quad \text{for } n_a \le i \le n \tag{11.2.7}$$

Such equations are generally known as *Wiener-Hopf* equations. Introducing the vector of observations *up to* the current time n, namely,

$$\mathbf{y}_n = [y_{n_a}, y_{n_a+1}, \ldots, y_n]^T$$

we may write Eq. (11.2.6) in vector form as

$$E[x_n \mathbf{y}_n^T] = [h(n,n_a), h(n,n_a+1), \ldots, h(n,n)]E[\mathbf{y}_n \mathbf{y}_n^T]$$

which can be solved for the vector of weights

$$[h(n,n_a), h(n,n_a+1), \ldots, h(n,n)] = E[x_n \mathbf{y}_n^T]E[\mathbf{y}_n \mathbf{y}_n^T]^{-1}$$

and for the estimate $\hat{x}_n$:

$$\hat{x}_n = E[x_n \mathbf{y}_n^T]E[\mathbf{y}_n \mathbf{y}_n^T]^{-1}\mathbf{y}_n \tag{11.2.8}$$

Again, $\hat{x}_n$ is recognized as the *projection* of x_n onto the space spanned by the observations that are used in making up the estimate; namely, $Y_n = \{y_{n_a}, y_{n_a+1}, \ldots, y_n\}$. This solution of Eqs. (11.2.5) and (11.2.7) will be discussed in more detail in Sec. 11.8, using covariance factorization methods.

11.3 Stationary Wiener Filter

In this section, we make two assumptions that simplify the structure of Eqs. (11.2.6) and (11.2.7). The first is to assume *stationarity* for all signals so that the cross-correlation and autocorrelation appearing in Eq. (11.2.7) become functions of the *differences* of their arguments. The second assumption is to take the initial time n_a to be the *infinite past*, $n_a = -\infty$, that is, the observation interval is $Y_n = \{y_i,\ -\infty < i \le n\}$.

The assumption of stationarity can be used as follows: Suppose we have the solution of $h(n,i)$ of Eq. (11.2.7) for the best weights to estimate x_n, and wish to determine the best weights $h(n+d,i)$, $n_a \le i \le n+d$ for estimating the sample x_{n+d} at the future time $n+d$. Then, the new weights will satisfy the same equations as (11.2.7) with the changes

$$R_{xy}(n+d,i) = \sum_{k=n_a}^{n+d} h(n+d,k)R_{yy}(k,i), \quad \text{for } n_a \le i \le n+d \tag{11.3.1}$$

Making a change of variables $i \to i+d$ and $k \to k+d$, we rewrite Eq. (11.3.1) as

$$R_{xy}(n+d,i+d) = \sum_{k=n_a-d}^{n} h(n+d,k+d)R_{yy}(k+d,i+d), \quad \text{for } n_a - d \le i \le n \tag{11.3.2}$$

Now, if we assume stationarity, Eqs. (11.2.7) and (11.3.2) become

$$\begin{aligned} R_{xy}(n-i) &= \sum_{k=n_a}^{n} h(n,k)R_{yy}(k-i)\,, \quad \text{for } n_a \le i \le n \\ R_{xy}(n-i) &= \sum_{k=n_a-d}^{n} h(n+d,k+d)R_{yy}(k-i)\,, \quad \text{for } n_a - d \le i \le n \end{aligned} \tag{11.3.3}$$

If it were not for the differences in the ranges of i and k, these two equations would be the same. But this is exactly what happens when we make the second assumption that $n_a = -\infty$. Therefore, by uniqueness of the solution, we find in this case

$$h(n+d,k+d) = h(n,k)$$

and since d is arbitrary, it follows that $h(n,k)$ must be a function of the difference of its arguments, that is,

$$h(n,k) = h(n-k) \tag{11.3.4}$$

Thus, the optimal linear processor becomes a *shift-invariant causal linear filter* and the estimate is given by

$$\hat{x}_n = \sum_{i=-\infty}^{n} h(n-i)y_i = \sum_{i=0}^{\infty} h(i)y_{n-i} \tag{11.3.5}$$

and Eq. (11.3.3) becomes in this case

$$R_{xy}(n-i) = \sum_{k=-\infty}^{n} h(n,k)R_{yy}(k-i)\,, \quad \text{for } -\infty < i \le n$$

With the change of variables $n-i \to n$ and $n-k \to k$, we find

$$R_{xy}(n) = \sum_{k=0}^{\infty} R_{yy}(n-k)h(k)\,, \quad \text{for } n \ge 0 \tag{11.3.6}$$

and written in matrix form

$$\begin{bmatrix} R_{yy}(0) & R_{yy}(1) & R_{yy}(2) & R_{yy}(3) & \cdots \\ R_{yy}(1) & R_{yy}(0) & R_{yy}(1) & R_{yy}(2) & \cdots \\ R_{yy}(2) & R_{yy}(1) & R_{yy}(0) & R_{yy}(1) & \cdots \\ R_{yy}(3) & R_{yy}(2) & R_{yy}(1) & R_{yy}(0) & \cdots \\ \vdots & \vdots & \vdots & \vdots & \end{bmatrix} \begin{bmatrix} h(0) \\ h(1) \\ h(2) \\ h(3) \\ \vdots \end{bmatrix} = \begin{bmatrix} R_{xy}(0) \\ R_{xy}(1) \\ R_{xy}(2) \\ R_{xy}(3) \\ \vdots \end{bmatrix} \tag{11.3.7}$$

These are the *discrete-time Wiener-Hopf equations*. Were it not for the restriction $n \ge 0$ (which reflects the requirement of causality), they could be solved easily by z-transform methods. As written above, they require methods of *spectral factorization* for their solution.

Before we discuss such methods, we mention in passing the continuous-time version of the Wiener-Hopf equation:

$$R_{xy}(t) = \int_0^\infty R_{yy}(t-t')h(t')\,dt', \quad t \geq 0$$

We also consider the FIR Wiener filtering problem in the stationary case. The observation interval in this case is $Y_n = \{y_i,\ n-M \leq i \leq n\}$. Using the same arguments as above we have $h(n,i) = h(n-i)$, and the estimate $\hat{x}_n$ is obtained by an ordinary FIR linear filter

$$\hat{x}_n = \sum_{i=n-M}^{n} h(n-i)y_i = h(0)y_n + h(1)y_{n-1} + \cdots + h(M)y_{n-M} \tag{11.3.8}$$

where the $(M+1)$ filter weights $h(0), h(1), \ldots, h(M)$ are obtained by the $(M+1)\times(M+1)$ matrix version of the Wiener-Hopf normal equations:

$$\begin{bmatrix} R_{yy}(0) & R_{yy}(1) & R_{yy}(2) & \cdots & R_{yy}(M) \\ R_{yy}(1) & R_{yy}(0) & R_{yy}(1) & \cdots & R_{yy}(M-1) \\ R_{yy}(2) & R_{yy}(1) & R_{yy}(0) & \cdots & R_{yy}(M-2) \\ \vdots & \vdots & \vdots & & \vdots \\ R_{yy}(M) & R_{yy}(M-1) & R_{yy}(M-2) & \cdots & R_{yy}(0) \end{bmatrix} \begin{bmatrix} h(0) \\ h(1) \\ h(2) \\ \vdots \\ h(M) \end{bmatrix} = \begin{bmatrix} R_{xy}(0) \\ R_{xy}(1) \\ R_{xy}(2) \\ \vdots \\ R_{xy}(M) \end{bmatrix} \tag{11.3.9}$$

Exploiting the Toeplitz property of the matrix R_{yy}, the above matrix equation can be solved efficiently using Levinson's algorithm. This will be discussed in Chap. 12. In Chap. 16, we will consider adaptive implementations of the FIR Wiener filter which produce the optimal filter weights adaptively without requiring prior knowledge of the autocorrelation and cross-correlation matrices R_{yy} and R_{xy} and without requiring any matrix inversion.

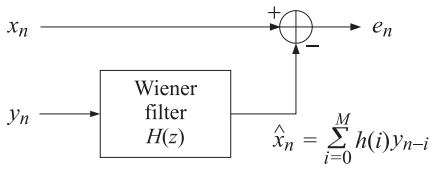


Fig. 11.3.1 Time-Invariant Wiener Filter.

We summarize our results on the stationary Wiener filter in Fig. 11.3.1. The optimal filter weights $h(n)$, $n = 0, 1, 2, \ldots$ are computed from Eq. (11.3.7) or Eq. (11.3.9). The action of the filter is precisely that of the correlation canceler: The filter processes the observation signal y_n *causally* to produce the best possible estimate $\hat{x}_n$ of x_n, and then it proceeds to cancel it from the output e_n. As a result, the output e_n is no longer correlated with any of the present and past values of y_n, that is, $E[e_n y_{n-i}] = 0$, for $i = 0, 1, 2, \ldots$. As we remarked in Sec. 1.4, it is better to think of $\hat{x}_n$ as the optimal estimate of *that part* of the primary signal x_n which happens to be correlated with the secondary signal y_n. This follows from the property that if $x_n = x_1(n) + x_2(n)$ with

$R_{x_2 y} = 0$, then $R_{xy} = R_{x_1 y}$. Therefore, the solution of Eq. (11.3.7) for the best weights to estimate x_n is also the solution for the best weights to estimate $x_1(n)$. The filter may also be thought of as the *optimal signal separator* of the two signal components $x_1(n)$ and $x_2(n)$.

11.4 Construction of the Wiener Filter by Prewhitening

The normal equations (11.3.6) would have a trivial solution if the sequence y_n were a white-noise sequence with delta-function autocorrelation. Thus, the solution procedure is first to whiten the sequence y_n and then solve the normal equations. To this end, let y_n have a signal model, as guaranteed by the spectral factorization theorem

$$S_{yy}(z) = \sigma_\epsilon^2 B(z)B(z^{-1}) \qquad \epsilon_n \longrightarrow \boxed{B(z)} \longrightarrow y_n \tag{11.4.1}$$

where ϵ_n is the driving white noise, and $B(z)$ a minimal-phase filter. The problem of estimating x_n in terms of the sequence y_n becomes equivalent to the problem of estimating x_n in terms of the white-noise sequence ϵ_n:

$$\epsilon_n \longrightarrow \boxed{B(z)} \longrightarrow y_n \longrightarrow \boxed{H(z)} \longrightarrow \hat{x}_n$$

If we could determine the combined filter

$$F(z) = B(z)H(z) \qquad \epsilon_n \longrightarrow \boxed{F(z)} \longrightarrow \hat{x}_n$$

we would then solve for the desired Wiener filter $H(z)$

$$H(z) = \frac{F(z)}{B(z)} \tag{11.4.2}$$

Since $B(z)$ is minimum-phase, the indicated inverse $1/B(z)$ is guaranteed to be *stable and causal*. Let f_n be the causal impulse response of $F(z)$. Then, it satisfies the normal equations of the type of Eq. (11.3.6):

$$R_{x\epsilon}(n) = \sum_{i=0}^{\infty} f_i R_{\epsilon\epsilon}(n-i), \quad n \geq 0 \tag{11.4.3}$$

Since $R_{\epsilon\epsilon}(n-i) = \sigma_\epsilon^2 \delta(n-i)$, Eq. (11.4.3) collapses to

$$R_{x\epsilon}(n) = \sigma_\epsilon^2 f_n, \quad n \geq 0, \quad \text{or,}$$

$$f_n = \frac{R_{x\epsilon}(n)}{\sigma_\epsilon^2}, \quad \text{for } n \geq 0 \tag{11.4.4}$$

Next, we compute the corresponding z-transform $F(z)$

$$F(z) = \sum_{n=0}^{\infty} f_n z^{-n} = \frac{1}{\sigma_\epsilon^2} \sum_{n=0}^{\infty} R_{x\epsilon}(n) z^{-n} = \frac{1}{\sigma_\epsilon^2} [S_{x\epsilon}(z)]_+ \tag{11.4.5}$$

where $[S_{x\epsilon}(z)]_+$ denotes the *causal part* of the double-sided z-transform $S_{x\epsilon}(z)$. Generally, the causal part of a z-transform

$$G(z) = \sum_{n=-\infty}^{\infty} g_n z^{-n} = \sum_{n=-\infty}^{-1} g_n z^{-n} + \sum_{n=0}^{\infty} g_n z^{-n}$$

is defined as

$$[G(z)]_+ = \sum_{n=0}^{\infty} g_n z^{-n}$$

The causal instruction in Eq. (11.4.5) was necessary since the above solution for f_n was valid only for $n \geq 0$. Since y_n is the output of the filter $B(z)$ driven by ϵ_n, it follows that

$$S_{xy}(z) = S_{x\epsilon}(z)B(z^{-1}) \quad \text{or} \quad S_{x\epsilon}(z) = \frac{S_{xy}(z)}{B(z^{-1})}$$

Combining Eqs. (11.4.2) and (11.4.5), we finally find

$$H(z) = \frac{1}{\sigma_\epsilon^2 B(z)} \left[\frac{S_{xy}(z)}{B(z^{-1})} \right]_+ \quad \text{(Wiener filter)} \tag{11.4.6}$$

Thus, the construction of the optimal filter first requires the spectral factorization of $S_{yy}(z)$ to obtain $B(z)$, and then use of the above formula. This is the optimal *realizable* Wiener filter based on the *infinite past*. If the causal instruction is ignored, one obtains the optimal *unrealizable* Wiener filter

$$H_{\text{unreal}}(z) = \frac{S_{xy}(z)}{\sigma_\epsilon^2 B(z)B(z^{-1})} = \frac{S_{xy}(z)}{S_{yy}(z)} \tag{11.4.7}$$

The *minimum* value of the mean-square estimation error can be conveniently expressed by a contour integral, as follows

$$\begin{aligned} \mathcal{E} &= E[e_n^2] = E[e_n(x_n - \hat{x}_n)] = E[e_n x_n] - E[e_n \hat{x}_n] = E[e_n x_n] = R_{ex}(0) \\ &= \oint_{\text{u.c.}} S_{ex}(z) \frac{dz}{2\pi j z} = \oint_{\text{u.c.}} [S_{xx}(z) - S_{\hat{x}x}(z)] \frac{dz}{2\pi j z}, \quad \text{or,} \\ \mathcal{E} &= \oint_{\text{u.c.}} [S_{xx}(z) - H(z)S_{yx}(z)] \frac{dz}{2\pi j z} \end{aligned} \tag{11.4.8}$$

11.5 Wiener Filter Example

This example, in addition to illustrating the above ideas, will also serve as a short introduction to *Kalman filtering*. It is desired to estimate the signal x_n on the basis of noisy observations

$$y_n = x_n + v_n$$

where v_n is white noise of unit variance, $\sigma_v^2 = 1$, uncorrelated with x_n. The signal x_n is a first order Markov process, having a signal model

$$x_{n+1} = 0.6x_n + w_n$$

where w_n is white noise of variance $\sigma_w^2 = 0.82$. Enough information is given above to determine the required power spectral densities $S_{xy}(z)$ and $S_{yy}(z)$. First, we note that the signal generator transfer function for x_n is

$$w_n \longrightarrow \boxed{M(z)} \longrightarrow x_n \qquad M(z) = \frac{1}{z - 0.6}$$

so that

$$S_{xx}(z) = \sigma_w^2 M(z)M(z^{-1}) = \frac{0.82}{(z - 0.6)(z^{-1} - 0.6)} = \frac{0.82}{(1 - 0.6z^{-1})(1 - 0.6z)}$$

Then, we find

$$\begin{aligned} S_{xy}(z) &= S_{x(x+v)}(z) = S_{xx}(z) + S_{xv}(z) = S_{xx}(z) = \frac{0.82}{(1 - 0.6z^{-1})(1 - 0.6z)} \\ S_{yy}(z) &= S_{(x+v)(x+v)}(z) = S_{xx}(z) + S_{xv}(z) + S_{vx}(z) + S_{vv}(z) = S_{xx}(z) + S_{vv}(z) \\ &= \frac{0.82}{(1 - 0.6z^{-1})(1 - 0.6z)} + 1 = \frac{0.82 + (1 - 0.6z^{-1})(1 - 0.6z)}{(1 - 0.6z^{-1})(1 - 0.6z)} \\ &= \frac{2(1 - 0.3z^{-1})(1 - 0.3z)}{(1 - 0.6z^{-1})(1 - 0.6z)} = 2 \cdot \frac{1 - 0.3z^{-1}}{1 - 0.6z^{-1}} \cdot \frac{1 - 0.3z}{1 - 0.6z} \\ &= \sigma_\epsilon^2 B(z)B(z^{-1}) \end{aligned}$$

Then according to Eq. (11.4.6), we must compute the causal part of

$$G(z) = \frac{S_{xy}(z)}{B(z^{-1})} = \frac{\dfrac{0.82}{(1 - 0.6z^{-1})(1 - 0.6z)}}{\dfrac{1 - 0.3z}{1 - 0.6z}} = \frac{0.82}{(1 - 0.6z^{-1})(1 - 0.3z)}$$

This may be done by partial fraction expansion, but the fastest way is to use the contour inversion formula to compute g_k for $k \geq 0$, and then resum the z-transform:

$$\begin{aligned} g_k &= \oint_{\text{u.c.}} G(z) z^k \frac{dz}{2\pi j z} = \oint_{\text{u.c.}} \frac{0.82 z^k}{(1 - 0.3z)(z - 0.6)} \frac{dz}{2\pi j} \\ &= (\text{residue at } z = 0.6) = \frac{0.82(0.6)^k}{1 - (0.3)(0.6)} = (0.6)^k, \quad k \geq 0 \end{aligned}$$

Resumming, we find the causal part

$$[G(z)]_+ = \sum_{k=0}^{\infty} g_k z^{-k} = \frac{1}{1 - 0.6z^{-1}}$$

Finally, the optimum Wiener estimation filter is

$$H(z) = \frac{1}{\sigma_\epsilon^2 B(z)} \left[\frac{S_{xy}(z)}{B(z^{-1})} \right]_+ = \frac{[G(z)]_+}{\sigma_\epsilon^2 B(z)} = \frac{0.5}{1 - 0.3z^{-1}} \tag{11.5.1}$$

which can be realized as the difference equation

$$\hat{x}_n = 0.3\hat{x}_{n-1} + 0.5y_n \qquad y_n \longrightarrow \boxed{H(z)} \longrightarrow \hat{x}_n \tag{11.5.2}$$

The estimation error is also easily computed using the contour formula of Eq. (11.4.8):

$$\mathcal{E} = E[e_n^2] = \sigma_e^2 = \oint_{\text{u.c.}} [S_{xx}(z) - H(z)S_{yx}(z)]\,\frac{dz}{2\pi jz} = 0.5$$

To appreciate the improvement afforded by filtering, this error must be compared with the error in case no processing is made and y_n is itself taken to represent a noisy estimate of x_n. The estimation error in the latter case is $y_n - x_n = v_n$, so that $\sigma_v^2 = 1$. Thus, the gain afforded by processing is

$$\frac{\sigma_e^2}{\sigma_v^2} = 0.5 \quad \text{or} \quad 3\ \text{dB}$$

11.6 Wiener Filter as Kalman Filter

We would like to cast this example in a Kalman filter form. The difference equation Eq. (11.5.2) for the Wiener filter seems to have the "wrong" state transition matrix; namely, 0.3 instead of 0.6, which is the state matrix for the state model of x_n. However, it is not accidental that the Wiener filter difference equation may be rewritten in the alternative form

$$\hat{x}_n = 0.6\hat{x}_{n-1} + 0.5(y_n - 0.6\hat{x}_{n-1})$$

The quantity $\hat{x}_n$ is the best estimate of x_n, at time n, based on all the observations up to that time, that is, $Y_n = \{y_i,\ -\infty < i \le n\}$. To simplify the subsequent notation, we denote it by $\hat{x}_{n/n}$. It is the projection of x_n on the space Y_n. Similarly, $\hat{x}_{n-1}$ denotes the best estimate of x_{n-1}, based on the observations up to time $n-1$, that is, $Y_{n-1} = \{y_i,\ -\infty < i \le n-1\}$. The above filtering equation is written in this notation as

$$\hat{x}_{n/n} = 0.6\hat{x}_{n-1/n-1} + 0.5(y_n - 0.6\hat{x}_{n-1/n-1}) \tag{11.6.1}$$

It allows the computation of the *current best* estimate $\hat{x}_{n/n}$, in terms of the previous best estimate $\hat{x}_{n-1/n-1}$ and the new observation y_n that becomes available at the current time instant n.

The various terms of Eq. (11.6.1) have nice interpretations: Suppose that the best estimate $\hat{x}_{n-1/n-1}$ of the previous sample x_{n-1} is available. Even before the next observation y_n comes in, we may use this estimate to make a reasonable prediction as to what the next best estimate ought to be. Since we know the system dynamics of x_n, we may try to "boost" $\hat{x}_{n-1/n-1}$ to the next time instant n according to the system dynamics, that is, we take

$$\hat{x}_{n/n-1} = 0.6\hat{x}_{n-1/n-1} = \text{prediction of } x_n \text{ on the basis of } Y_{n-1} \tag{11.6.2}$$

Since $y_n = x_n + v_n$, we may use this prediction of x_n to make a prediction of the next measurement y_n, that is, we take

$$\hat{y}_{n/n-1} = \hat{x}_{n/n-1} = \text{prediction of } y_n \text{ on the basis of } Y_{n-1} \tag{11.6.3}$$

If this prediction were perfect, and if the next observation y_n were noise free, then this would be the value that we would observe. Since we actually observe y_n, the observation or innovations residual will be

$$\alpha_n = y_n - \hat{y}_{n/n-1} \tag{11.6.4}$$

This quantity represents that part of y_n that *cannot* be predicted on the basis of the previous observations Y_{n-1}. It represents the truly new information contained in the observation y_n. Actually, if we are making the best prediction possible, then the most we can expect of our prediction is to make the innovations residual a white-noise (uncorrelated) signal, that is, what remains after we make the best possible prediction should be unpredictable. According to the general discussion of the relationship between signal models and linear prediction given in Sec. 1.17, it follows that if $\hat{y}_{n/n-1}$ is the best predictor of y_n then α_n must be the whitening sequence that drives the signal model of y_n. We shall verify this fact shortly. This establishes an intimate connection between the *Wiener/Kalman filtering* problem and the *signal modeling* problem. If we overestimate the observation y_n the innovation residual will be negative; and if we underestimate it, the residual will be positive. In either case, we would like to correct our tentative estimate in the right direction. This may be accomplished by

$$\hat{x}_{n/n} = \hat{x}_{n/n-1} + G(y_n - \hat{y}_{n/n-1}) = 0.6\hat{x}_{n-1/n-1} + G(y_n - 0.6\hat{x}_{n-1/n-1}) \tag{11.6.5}$$

where the gain G, known as the *Kalman gain*, should be a positive quantity. The *prediction/correction* procedure defined by Eqs. (11.6.2) through (11.6.5) is known as the *Kalman filter*. It should be clear that any value for the gain G will provide an estimate, even if suboptimal, of x_n. Our solution for the Wiener filter has precisely the above structure with a gain $G = 0.5$. This value is optimal for the given example. It is a very instructive exercise to show this in two ways: First, with G arbitrary, the estimation filter of Eq. (11.6.5) has transfer function

$$H(z) = \frac{G}{1 - 0.6(1-G)z^{-1}} \qquad y_n \longrightarrow \boxed{H(z)} \longrightarrow \hat{x}_{n/n}$$

Insert this expression into the mean-square estimation error $\mathcal{E} = E[e_n^2]$, where $e_n = x_n - \hat{x}_{n/n}$, and minimize it with respect to the parameter G. This should give $G = 0.5$.

Alternatively, G should be such that to render the innovations residual (11.6.4) a white noise signal. In requiring this, it is useful to use the spectral factorization model for y_n, that is, the fact that y_n is the output of $B(z)$ when driven by the white noise signal ϵ_n. Working with z-transforms, we have:

$$\begin{aligned}
\alpha(z) &= Y(z) - 0.6z^{-1}\hat{X}(z) = Y(z) - 0.6z^{-1}H(z)Y(z) \\
&= \left[1 - 0.6z^{-1}\,\frac{G}{1-0.6(1-G)z^{-1}}\right]Y(z) = \left[\frac{1-0.6z^{-1}}{1-0.6(1-G)z^{-1}}\right]Y(z) \\
&= \left[\frac{1-0.6z^{-1}}{1-0.6(1-G)z^{-1}}\right]\left[\frac{1-0.3z^{-1}}{1-0.6z^{-1}}\right]\epsilon(z) = \left[\frac{1-0.3z^{-1}}{1-0.6(1-G)z^{-1}}\right]\epsilon(z)
\end{aligned}$$

Since ϵ_n is white, it follows that the transfer function relationship between α_n and ϵ_n must be trivial; otherwise, there will be sequential correlations present in α_n. Thus,

we must have $0.6(1-G)=0.3$, or $G=0.5$; and in this case, $\alpha_n=\epsilon_n$. It is also possible to set $0.6(1-G)=1/0.3$, but this would correspond to an unstable filter.

We have obtained a most interesting result; namely, that when the Wiener filtering problem is recast into its Kalman filter form given by Eq. (11.6.1), then the innovations residual α_n, which is computable on line with the estimate $\hat{x}_{n/n}$, is identical to the whitening sequence ϵ_n of the signal model of y_n. In other words, the Kalman filter can be thought of as the *whitening filter* for the observation signal y_n.

To appreciate further the connection between Wiener and Kalman filters and between Kalman filters and the whitening filters of signal models, we consider a generalized version of the above example and cast it in standard Kalman filter notation.

It is desired to estimate x_n from y_n. The signal model for x_n is taken to be the first-order autoregressive model

$$x_{n+1}=ax_n+w_n \qquad \text{(state model)} \tag{11.6.6}$$

with $|a|<1$. The observation signal y_n is related to x_n by

$$y_n=cx_n+v_n \qquad \text{(measurement model)} \tag{11.6.7}$$

It is further assumed that the state and measurement noises, w_n and v_n, are zero-mean, mutually uncorrelated, white noises of variances Q and R, respectively, that is,

$$E[w_nw_i]=Q\delta_{ni}\,,\quad E[v_nv_i]=R\delta_{ni}\,,\quad E[w_nv_i]=0 \tag{11.6.8}$$

We also assume that v_n is uncorrelated with the initial value of x_n so that v_n and x_n will be uncorrelated for all n. The parameters a,c,Q,R are assumed to be known. Let $x_1(n)$ be the time-advanced version of x_n:

$$x_1(n)=x_{n+1}$$

and consider the two related Wiener filtering problems of estimating x_n and $x_1(n)$ on the basis of $Y_n=\{y_i,\ -\infty<i\le n\}$, depicted below

$$y_n \longrightarrow \boxed{H(z)} \longrightarrow \hat{x}_{n/n} \qquad\qquad y_n \longrightarrow \boxed{H_1(z)} \longrightarrow \hat{x}_1(n)=\hat{x}_{n+1/n}$$

The problem of estimating $x_1(n)=x_{n+1}$ is equivalent to the problem of one-step prediction into the future on the basis of the past and present. Therefore, we will denote this estimate by $\hat{x}_1(n)=\hat{x}_{n+1/n}$. The state equation (11.6.6) determines the spectral density of x_n:

$$S_{xx}(z)=\frac{1}{(z-a)(z^{-1}-a)}\,S_{ww}(z)=\frac{Q}{(1-az^{-1})(1-az)}$$

The observation equation (11.6.7) determines the cross-densities

$$\begin{aligned}S_{xy}(z)&=cS_{xx}(z)+S_{xv}(z)=cS_{xx}(z)\\ S_{x_1y}(z)&=zS_{xy}(z)=zcS_{xx}(z)\end{aligned}$$

where we used the filtering equation $X_1(z)=zX(z)$. The spectral density of y_n can be factored as follows:

$$\begin{aligned}S_{yy}(z)&=c^2S_{xx}(z)+S_{vv}(z)=\frac{c^2Q}{(1-az^{-1})(1-az)}+R\\ &=\frac{c^2Q+R(1-az^{-1})(1-az)}{(1-az^{-1})(1-az)}\equiv\sigma_\epsilon^2\left(\frac{1-fz^{-1}}{1-az^{-1}}\right)\left(\frac{1-fz}{1-az}\right)\end{aligned}$$

where f and σ_ϵ^2 satisfy the equations

$$f\sigma_\epsilon^2=aR \tag{11.6.9}$$

$$(1+f^2)\sigma_\epsilon^2=c^2Q+(1+a^2)R \tag{11.6.10}$$

and f has magnitude less than one. Thus, the corresponding signal model for y_n is

$$B(z)=\frac{1-fz^{-1}}{1-az^{-1}} \tag{11.6.11}$$

Next, we compute the causal parts as required by Eq. (11.4.6):

$$\left[\frac{S_{xy}(z)}{B(z^{-1})}\right]_+=\left[\frac{cQ}{(1-az^{-1})(1-fz)}\right]_+=\frac{cQ}{1-fa}\,\frac{1}{1-az^{-1}}$$

$$\left[\frac{S_{x_1y}(z)}{B(z^{-1})}\right]_+=\left[\frac{cQz}{(1-az^{-1})(1-fz)}\right]_+=\frac{cQa}{1-fa}\,\frac{1}{1-az^{-1}}$$

Using Eq. (11.4.6), we determine the Wiener filters $H(z)$ and $H_1(z)$ as follows:

$$H(z)=\frac{1}{\sigma_\epsilon^2B(z)}\left[\frac{S_{xy}(z)}{B(z^{-1})}\right]_+=\frac{\dfrac{cQ/(1-fa)}{(1-az^{-1})}}{\sigma_\epsilon^2\left(\dfrac{1-fz^{-1}}{1-az^{-1}}\right)}=\frac{\left(\dfrac{cQ}{\sigma_\epsilon^2(1-fa)}\right)}{1-fz^{-1}}$$

or, defining the gain G by

$$G=\frac{cQ}{\sigma_\epsilon^2(1-fa)} \tag{11.6.12}$$

we finally find

$$H(z)=\frac{G}{1-fz^{-1}} \tag{11.6.13}$$

$$H_1(z)=aH(z)=\frac{K}{1-fz^{-1}} \tag{11.6.14}$$

where in Eq. (11.6.14) we defined a related gain, also called the Kalman gain, as follows:

$$K=aG=\frac{cQa}{\sigma_\epsilon^2(1-fa)} \tag{11.6.15}$$

Eq. (11.6.14) immediately implies that

$$\hat{x}_{n+1/n}=a\hat{x}_{n/n} \tag{11.6.16}$$

which is the precise justification of Eq. (11.6.2). The difference equations of the two filters are

$$\begin{aligned}\hat{x}_{n+1/n} &= f\hat{x}_{n/n-1} + Ky_n \\ \hat{x}_{n/n} &= f\hat{x}_{n-1/n-1} + Gy_n\end{aligned} \tag{11.6.17}$$

Using the results of Problem 1.50, we may express all the quantities f, σ_ϵ^2, K, and G in terms of a single positive quantity P which satisfies the *algebraic Riccati equation*:

$$Q = P - \frac{PRa^2}{R + c^2P} \tag{11.6.18}$$

Then, we find the interrelationships

$$K = aG = \frac{acP}{R + c^2P}\,, \quad \sigma_\epsilon^2 = R + c^2P\,, \quad f = a - cK = \frac{Ra}{R + c^2P} \tag{11.6.19}$$

It is left as an exercise to show that the minimized mean-square estimation errors are given in terms of P by

$$E[e_{n/n-1}^2] = P\,, \quad E[e_{n/n}^2] = \frac{RP}{R + c^2P}$$

where

$$e_{n/n-1} = x_n - \hat{x}_{n/n-1}\,, \quad e_{n/n} = x_n - \hat{x}_{n/n}$$

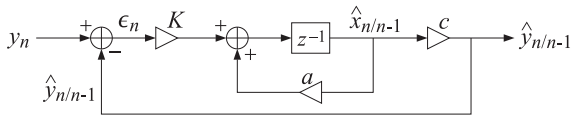
are the corresponding estimation errors for the optimally predicted and filtered estimates, respectively. Using Eq. (11.6.19)), we may rewrite the filtering equation (11.6.17) in the following forms:

$$\begin{aligned}\hat{x}_{n+1/n} &= (a - cK)\hat{x}_{n/n-1} + Ky_n\,, \quad \text{or,} \\ \hat{x}_{n+1/n} &= a\hat{x}_{n/n-1} + K(y_n - c\hat{x}_{n/n-1})\,, \quad \text{or,} \\ \hat{x}_{n+1/n} &= a\hat{x}_{n/n-1} + K(y_n - \hat{y}_{n/n-1})\end{aligned} \tag{11.6.20}$$

where we set

$$\hat{y}_{n/n-1} = c\hat{x}_{n/n-1} \tag{11.6.21}$$

A realization of the estimation filter based on (11.6.20) is shown below:



Replacing $K = aG$ and using Eq. (11.6.16) in (11.6.20), we also find

$$\hat{x}_{n/n} = \hat{x}_{n/n-1} + G(y_n - \hat{y}_{n/n-1}) \tag{11.6.22}$$

The quantity $\hat{y}_{n/n-1}$ defined in Eq. (11.6.21) is the best estimate of y_n based on its past Y_{n-1}. This can be seen in two ways: First, using the results of Problem 1.7 on the linearity of the estimates, we find

$$\hat{y}_{n/n-1} = \widehat{cx_n + v_n} = c\hat{x}_{n/n-1} + \hat{v}_{n/n-1} = c\hat{x}_{n/n-1}$$

where the term $\hat{v}_{n/n-1}$ was dropped. This term represents the estimate of v_n on the basis of the past ys; that is, Y_{n-1}. Since v_n is white and also uncorrelated with x_n, it follows that it will be uncorrelated with all past ys; therefore, $\hat{v}_{n/n-1} = 0$. The second way to show that $\hat{y}_{n/n-1}$ is the best prediction of y_n is to show that the innovations residual

$$\alpha_n = y_n - \hat{y}_{n/n-1} = y_n - c\hat{x}_{n/n-1} \tag{11.6.23}$$

is a white-noise sequence and coincides with the whitening sequence ϵ_n of y_n. Indeed, working in the z-domain and using Eq. (11.6.17) and the signal model of y_n we find

$$\begin{aligned}\alpha(z) &= Y(z) - cz^{-1}\hat{X}_1(z) = Y(z) - cz^{-1}H_1(z)Y(z) \\ &= \left[1 - cz^{-1}\frac{K}{1 - fz^{-1}}\right]Y(z) = \left[\frac{1 - (f + cK)z^{-1}}{1 - fz^{-1}}\right]Y(z) \\ &= \left[\frac{1 - az^{-1}}{1 - fz^{-1}}\right]Y(z) = \frac{1}{B(z)}\,Y(z) = \epsilon(z)\end{aligned}$$

which implies that

$$\alpha_n = \epsilon_n$$

Finally, we note that the recursive updating of the estimate of x_n given by Eq. (11.6.22) is identical to the result of Problem 1.11.

Our purpose in presenting this example was to tie together a number of ideas from Chapter 1 (correlation canceling, estimation, Gram-Schmidt orthogonalization, linear prediction, and signal modeling) to ideas from this chapter on Wiener filtering and its recursive reformulation as a Kalman filter.

We conclude this section by presenting a simulation of this example defined by the following choice of parameters:

$$a = 0.95\,, \quad c = 1\,, \quad Q = 1 - a^2\,, \quad R = 1$$

The above choice for Q normalizes the variance of x_n to unity. Solving the Riccati equation (11.6.18) and using Eq. (11.6.19), we find

$$P = 0.3122\,, \quad K = 0.2261\,, \quad G = 0.2380\,, \quad f = a - cK = 0.7239$$

Fig. 11.6.1 shows 100 samples of the observed signal y_n together with the desired signal x_n. The signal y_n processed through the Wiener filter $H(z)$ defined by the above parameters is shown in Fig. 11.6.2 together with x_n. The tracking properties of the filter are evident from the graph. It should be emphasized that this is the best one can do by means of ordinary causal linear filtering.

11.7 Construction of the Wiener Filter by the Gapped Function

Next, we would like to give an alternative construction of the optimal Wiener filter based on the concept of the gapped function. This approach is especially useful in linear prediction. The gapped function is defined as the cross-correlation between the estimation error e_n and the observation sequence y_n, as follows:

$$g(k) = R_{ey}(k) = E[e_n y_{n-k}]\,, \quad \text{for} \quad -\infty < k < \infty \tag{11.7.1}$$

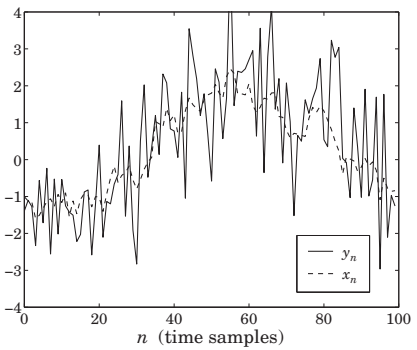


Fig. 11.6.1 Desired signal and its noisy observation.

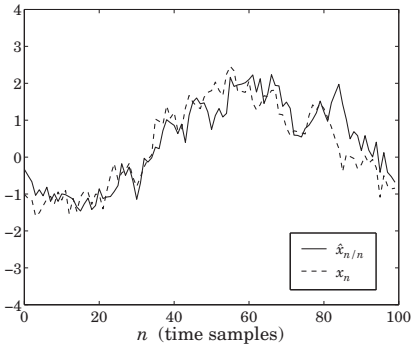
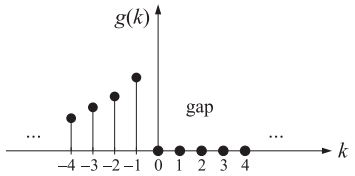


Fig. 11.6.2 Best estimate of desired signal.

This definition is motivated by the orthogonality equations which state that the prediction error e_n must be orthogonal to all of the available observations; namely, $Y_n = \{y_i\,,\ -\infty < i \le n\} = \{y_{n-k}\,,\ k \ge 0\}$. That is, for the optimal set of filter weights we must have

$$g(k) = R_{ey}(k) = E[e_n y_{n-k}] = 0\,, \quad \text{for } k \ge 0 \tag{11.7.2}$$



and $g(k)$ develops a right-hand side gap. On the other hand, $g(k)$ may be written in

the alternative form

$$g(k) = E[e_n y_{n-k}] = E\big[\,(x_n - \sum_{i=0}^{\infty} h_i y_{n-i})\,y_{n-k}\,\big] = R_{xy}(k) - \sum_{i=0}^{\infty} h_i R_{yy}(k-i)\,, \quad \text{or,}$$

$$g(k) = R_{ey}(k) = R_{xy}(k) - \sum_{i=0}^{\infty} h_i R_{yy}(k-i) \tag{11.7.3}$$

Taking z-transforms of both sides we find

$$G(z) = S_{ey}(z) = S_{xy}(z) - H(z)S_{yy}(z)$$

Because of the gap conditions, the left-hand side contains only positive powers of z, whereas the right-hand side contains both positive and negative powers of z. Thus, the non-positive powers of z must drop out of the right side. This condition precisely determines $H(z)$. Introducing the spectral factorization of $S_{yy}(z)$ and dividing both sides by $B(z^{-1})$ we find

$$G(z) = S_{xy}(z) - H(z)S_{yy}(z) = S_{xy}(z) - H(z)\sigma_\epsilon^2 B(z)B(z^{-1})$$

$$\frac{G(z)}{B(z^{-1})} = \frac{S_{xy}(z)}{B(z^{-1})} - \sigma_\epsilon^2 H(z)B(z)$$

The z-transform $B(z^{-1})$ is anticausal and, because of the gap conditions, so is the ratio $G(z)/B(z^{-1})$. Therefore, taking causal parts of both sides and noting that the product $H(z)B(z)$ is already causal, we find

$$0 = \left[\frac{S_{xy}(z)}{B(z^{-1})}\right]_+ - \sigma_\epsilon^2 H(z)B(z)$$

which may be solved for $H(z)$ to give Eq. (11.4.6).

11.8 Construction of the Wiener Filter by Covariance Factorization

In this section, we present a generalization of the gapped-function method to the more general non-stationary and/or finite-past Wiener filter. This is defined by the Wiener-Hopf equations (11.2.7), which are equivalent to the orthogonality equations (11.2.5). The latter are the non-stationary versions of the gapped function of the previous section. The best way to proceed is to cast Eqs. (11.2.5) in matrix form as follows: Without loss of generality we may take the starting point $n_a = 0$. The final point n_b is left arbitrary. Introduce the vectors

$$\mathbf{x} = \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_{n_b} \end{bmatrix}\,, \quad \mathbf{y} = \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_{n_b} \end{bmatrix}$$

and the corresponding correlation matrices

$$R_{xy} = E[\mathbf{x}\mathbf{y}^T], \quad R_{yy} = E[\mathbf{y}\mathbf{y}^T]$$

The filtering equation (11.2.4) may be written in vector form as

$$\hat{\mathbf{x}} = H\mathbf{y} \tag{11.8.1}$$

where H is the matrix of optimal weights $\{h(n,i)\}$. The *causality* of the filtering operation (11.8.1), requires H to be *lower-triangular*. The minimization problem becomes equivalent to the problem of minimizing the mean-square estimation error subject to the constraint that H be lower-triangular. The minimization conditions are the normal equations (11.2.5) which, in this matrix notation, state that the matrix R_{ey} has no lower-triangular (causal) part; or, equivalently, that R_{ey} is *strictly* upper-triangular (i.e., even the main diagonal of R_{ey} is zero), therefore

$$R_{ey} = \text{strictly upper triangular} \qquad \boxed{0} \tag{11.8.2}$$

Inserting Eq. (11.8.1) into R_{ey} we find

$$R_{ey} = E[\mathbf{e}\mathbf{y}^T] = E[(\mathbf{x} - H\mathbf{y})\mathbf{y}^T], \quad \text{or,}$$

$$R_{ey} = R_{xy} - HR_{yy} \tag{11.8.3}$$

The minimization conditions (11.8.2) require H to be that lower-triangular matrix which renders the combination (11.8.3) upper-triangular. In other words, H should be such that the lower triangular part of the right-hand side must vanish. To solve Eqs. (11.8.2) and (11.8.3), we introduce the *LU Cholesky factorization* of the covariance matrix R_{yy} given by

$$R_{yy} = BR_{\epsilon\epsilon}B^T \tag{11.8.4}$$

where B is unit lower-triangular, and $R_{\epsilon\epsilon}$ is diagonal. This was discussed in Sec. 1.6. Inserting this into Eq. (11.8.3) we find

$$R_{ey} = R_{xy} - HR_{yy} = R_{xy} - HBR_{\epsilon\epsilon}B^T \tag{11.8.5}$$

Multiplying by the inverse transpose of B we obtain

$$R_{ey}B^{-T} = R_{xy}B^{-T} - HBR_{\epsilon\epsilon} \tag{11.8.6}$$

Now, the matrix B^{-T} is unit upper-triangular, but R_{ey} is strictly upper, therefore, the product $R_{xy}B^{-T}$ will be strictly upper. This can be verified easily for any two such matrices. Extracting the lower-triangular parts of both sides of Eq. (11.8.6) we find

$$0 = [R_{xy}B^{-T}]_+ - HBR_{\epsilon\epsilon}$$

where we used the fact that the left-hand side was strictly upper and that the term $HBR_{\epsilon\epsilon}$ was already lower-triangular. The notation $[\ ]_+$ denotes the lower triangular part of a matrix including the diagonal. We find finally

$$H = [R_{xy}B^{-T}]_+ R_{\epsilon\epsilon}^{-1}B^{-1} \tag{11.8.7}$$

This is the most general solution of the Wiener filtering problem [18, 19]. It includes the results of the stationary case, as a special case. Indeed, if all the signals are stationary, then the matrices R_{xy}, B, and B^T become Toeplitz and have a z-transform associated with them as discussed in Problem 1.51. Using the results of that problem, it is easily seen that Eq. (11.8.7) is the time-domain equivalent of Eq. (11.4.6).

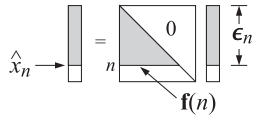
The prewhitening approach of Sec. 11.4 can also be understood in the present matrix framework. Making the change of variables

$$\mathbf{y} = B\boldsymbol{\epsilon}$$

we find that $R_{xy} = E[\mathbf{x}\mathbf{y}^T] = E[\mathbf{x}\boldsymbol{\epsilon}^T]B^T = R_{x\epsilon}B^T$, and therefore, $R_{xy}B^{-T} = R_{x\epsilon}$ and the filter H becomes $H = [R_{x\epsilon}]_+ R_{\epsilon\epsilon}^{-1}B^{-1}$. The corresponding estimate is then

$$\hat{\mathbf{x}} = H\mathbf{y} = HB\boldsymbol{\epsilon} = F\boldsymbol{\epsilon}, \quad \text{where } F = HB = [R_{x\epsilon}]_+ R_{\epsilon\epsilon}^{-1} \tag{11.8.8}$$

This is the matrix equivalent of Eq. (11.4.5). The matrix F is lower-triangular by construction. Therefore, to extract the nth component $\hat{x}_n$ of Eq. (11.8.8), it is enough to consider the $n\times n$ submatrices as shown below:



The nth row of F is $\mathbf{f}(n)^T = E[x_n\boldsymbol{\epsilon}_n^T]E[\boldsymbol{\epsilon}_n\boldsymbol{\epsilon}_n^T]^{-1}$. Therefore, the nth estimate becomes

$$\hat{x}_n = \mathbf{f}(n)^T\boldsymbol{\epsilon}_n = E[x_n\boldsymbol{\epsilon}_n^T]E[\boldsymbol{\epsilon}_n\boldsymbol{\epsilon}_n^T]^{-1}\boldsymbol{\epsilon}_n$$

which may also be written in the recursive form

$$\hat{x}_{n/n} = \sum_{i=0}^{n} E[x_n\epsilon_i]E[\epsilon_i\epsilon_i]^{-1}\epsilon_i = \sum_{i=0}^{n-1} E[x_n\epsilon_i]E[\epsilon_i\epsilon_i]^{-1}\epsilon_i + G_n\epsilon_n, \quad \text{or,}$$

$$\hat{x}_{n/n} = \hat{x}_{n/n-1} + G_n\epsilon_n \tag{11.8.9}$$

where we made an obvious change in notation, and $G_n = E[x_n\epsilon_n]E[\epsilon_n\epsilon_n]^{-1}$. This is identical to Eq. (11.6.22); in the stationary case, G_n is a constant, independent of n. We can also recast the nth estimate in "batch" form, expressed directly in terms of the observation vector $\mathbf{y}_n = [y_0, y_1, \ldots, y_n]^T$. By considering the $n\times n$ subblock part of the Gram-Schmidt construction, we may write $\mathbf{y}_n = B_n\boldsymbol{\epsilon}_n$, where B_n is unit lower-triangular. Then, $\hat{x}_n$ can be expressed as

$$\hat{x}_n = E[x_n\boldsymbol{\epsilon}_n^T]E[\boldsymbol{\epsilon}_n\boldsymbol{\epsilon}_n^T]^{-1}\boldsymbol{\epsilon}_n = E[x_n\mathbf{y}_n^T]E[\mathbf{y}_n\mathbf{y}_n^T]^{-1}\mathbf{y}_n$$

which is identical to Eq. (11.2.8).

11.9 The Kalman Filter

The Kalman filter discussion of Sec. 11.6 and its equivalence to the Wiener filter was based on the asymptotic Kalman filter for which the observations were available from the infinite past to the present, namely, $\{y_i\,,\ -\infty < i \le n\}$. In Sec. 11.7, we solved the most general Wiener filtering problem based on the finite past for which the observation space was

$$Y_n = \{y_0, y_1, \ldots, y_n\} \tag{11.9.1}$$

Here, we recast these results in a time-recursive form and obtain the time-varying Kalman filter for estimating x_n based on the finite observation subspace Y_n. We also discuss its asymptotic properties for large n and show that it converges to the steady-state Kalman filter of Sec. 11.6.

Our discussion is based on Eq. (11.8.9), which is essentially the starting point in Kalman's original derivation [852]. To make Eq. (11.8.9) truly recursive, we must have a means of recursively computing the required gain G_n from one time instant to the next. As in Sec. 11.8, we denote by $\hat{x}_{n/n}$ and $\hat{x}_{n/n-1}$ the optimal estimates of x_n based on the observation subspaces Y_n and Y_{n-1}, defined in Eq. (11.9.1), with the initial condition $\hat{x}_{0/-1} = 0$. Iterating the state and measurement models (11.6.6) and (11.6.7) starting at $n = 0$, we obtain the following two results, previously derived for the steady-state case

$$\hat{x}_{n+1/n} = a\hat{x}_{n/n}\,, \quad \hat{y}_{n/n-1} = c\hat{x}_{n/n-1} \tag{11.9.2}$$

The proof of both is based on the linearity property of estimates; for example,

$$\hat{x}_{n+1/n} = a\widehat{x_n + w_n} = a\hat{x}_{n/n} + \hat{w}_{n/n} = a\hat{x}_{n/n}$$

where $\hat{w}_{n/n}$ was set to zero because w_n does not depend on any of the observations Y_n. This is seen as follows. The iteration of the state equation (11.6.6) leads to the expression $x_n = a^n x_0 + a^{n-1}w_0 + a^{n-2}w_1 + \cdots + aw_{n-2} + w_{n-1}$. It follows from this and Eq. (11.6.7) that the observation subspace Y_n will depend only on

$$\{x_0, w_0, w_1, \ldots, w_{n-1}, v_0, v_1, \ldots, v_n\}$$

Making the additional assumption that x_0 is uncorrelated with w_n it follows that w_n will be uncorrelated with all random variables in the above set, and thus, with Y_n. The second part of Eq. (11.9.2) is shown by similar arguments. Next, we develop the recursions for the gain G_n. Using Eq. (11.8.9), the estimation and prediction errors may be related as follows

$$e_{n/n} = x_n - \hat{x}_{n/n} = x_n - \hat{x}_{n/n-1} - G_n\epsilon_n = e_{n/n-1} - G_n\epsilon_n$$

Taking the correlation of both sides with x_n we find

$$E[e_{n/n}x_n] = E[e_{n/n-1}x_n] - G_nE[\epsilon_n x_n] \tag{11.9.3}$$

Using the orthogonality properties $E[e_{n/n}\hat{x}_{n/n}] = 0$ and $E[e_{n/n-1}\hat{x}_{n/n-1}] = 0$, which follow from the optimality of the two estimates $\hat{x}_{n/n}$ and $\hat{x}_{n/n-1}$, we can write the mean-square estimation and prediction errors as

$$P_{n/n} = E[e_{n/n}^2] = E[e_{n/n}x_n]\,, \quad P_{n/n-1} = E[e_{n/n-1}^2] = E[e_{n/n-1}x_n] \tag{11.9.4}$$

We find also

$$\epsilon_n = y_n - \hat{y}_{n/n-1} = (cx_n + v_n) - c\hat{x}_{n/n-1} = ce_{n/n-1} + v_n$$

Using the fact that $e_{n/n-1}$ depends only on x_n and Y_{n-1}, it follows that the two terms in the right-hand side are uncorrelated with each other. Thus,

$$E[\epsilon_n^2] = c^2E[e_{n/n-1}^2] + E[v_n^2] = c^2P_{n/n-1} + R \tag{11.9.5}$$

also

$$E[\epsilon_n x_n] = cE[e_{n/n-1}x_n] + E[v_n x_n] = cP_{n/n-1} \tag{11.9.6}$$

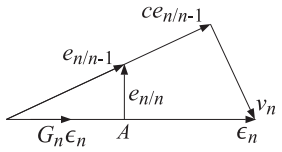
Therefore, the gain G_n is computable by

$$G_n = \frac{E[\epsilon_n x_n]}{E[\epsilon_n^2]} = \frac{cP_{n/n-1}}{R + c^2P_{n/n-1}} \tag{11.9.7}$$

Using Eqs. (11.9.4), (11.9.6), and (11.9.7) into Eq. (11.9.3), we obtain

$$P_{n/n} = P_{n/n-1} - G_ncP_{n/n-1} = P_{n/n-1} - \frac{c^2P_{n/n-1}}{R + c^2P_{n/n-1}} = \frac{RP_{n/n-1}}{R + c^2P_{n/n-1}} \tag{11.9.8}$$

The subtracted term in (11.9.8) represents the *improvement* in estimating x_n using $\hat{x}_{n/n}$ over using $\hat{x}_{n/n-1}$. Equations (11.9.3), (11.9.7), and (11.9.8) admit a nice geometrical interpretation [867]. The two right-hand side terms in $\epsilon_n = ce_{n/n-1} + v_n$ are orthogonal and can be represented by the orthogonal triangle



where the prediction error $e_{n/n-1}$ has been scaled up by the factor c. Thus, Eq. (11.9.5) is the statement of the Pythagorean theorem for this triangle. Next, write the equation $e_{n/n} = e_{n/n-1} - G_n\epsilon_n$ as

$$e_{n/n-1} = e_{n/n} + G_n\epsilon_n$$

Because $e_{n/n}$ is orthogonal to all the observations in Y_n and ϵ_n is a linear combination of the same observations, it follows that the two terms in the right-hand side will be orthogonal. Thus, $e_{n/n-1}$ may be resolved in two orthogonal parts, one being in the direction of ϵ_n. This is represented by the smaller orthogonal triangle in the previous diagram. Clearly, the length of the side $e_{n/n}$ is minimized at right angles at point A. It follows from the similarity of the two orthogonal triangles that

$$\frac{G_n\sqrt{E[\epsilon_n^2]}}{\sqrt{E[e_{n/n-1}^2]}} = \frac{c\sqrt{E[e_{n/n-1}^2]}}{\sqrt{E[\epsilon_n^2]}}$$

which is equivalent to Eq. (11.9.7). Finally, the Pythagorean theorem applied to the smaller triangle implies $E[e_{n/n-1}^2] = E[e_{n/n}^2] + G_n^2E[\epsilon_n^2]$, which is equivalent to Eq. (11.9.8).

To obtain a truly recursive scheme, we need next to find a relationship between $P_{n/n}$ and the next prediction error $P_{n+1/n}$. It is found as follows. From the state model (11.6.6) and (11.9.2), we have

$$e_{n+1/n} = x_{n+1} - \hat{x}_{n+1/n} = (ax_n + w_n) - a\hat{x}_{n/n} = ae_{n/n} + w_n$$

Because $e_{n/n}$ depends only on x_n and Y_n, it follows that the two terms in the right-hand side will be uncorrelated. Therefore, $E[e^2_{n+1/n}] = a^2 E[e^2_{n/n}] + E[w_n^2]$, or,

$$P_{n+1/n} = a^2 P_{n/n} + Q \tag{11.9.9}$$

The first term corresponds to the propagation of the estimate $\hat{x}_{n/n}$ forward in time according to the system dynamics; the second term represents the worsening of the estimate due to the presence of the dynamical noise w_n. The Kalman filter algorithm is now complete. It is summarized below:

0. Initialize by $\hat{x}_{0/-1} = 0$ and $P_{0/-1} = E[x_0^2]$.
1. At time n, $\hat{x}_{n/n-1}$, $P_{n/n-1}$, and the new measurement y_n are available.
2. Compute $\hat{y}_{n/n-1} = c\hat{x}_{n/n-1}$, $\epsilon_n = y_n - \hat{y}_{n/n-1}$, and the gain G_n using Eq. (11.9.7).
3. Correct the predicted estimate $\hat{x}_{n/n} = \hat{x}_{n/n-1} + G_n\epsilon_n$ and compute its mean-square error $P_{n/n}$, using Eq. (11.9.8).
4. Predict the next estimate $\hat{x}_{n+1/n} = a\hat{x}_{n/n}$, and compute the mean-square prediction error $P_{n+1/n}$, using Eq. (11.9.9).
5. Go to the next time instant, $n \to n+1$.

The optimal predictor $\hat{x}_{n/n-1}$ satisfies the Kalman filtering equation

$$\hat{x}_{n+1/n} = a\hat{x}_{n/n} = a(\hat{x}_{n/n-1} + G_n\epsilon_n) = a\hat{x}_{n/n-1} + aG_n(y_n - c\hat{x}_{n/n-1}) \,, \quad \text{or,}$$

$$\hat{x}_{n+1/n} = f_n\hat{x}_{n/n-1} + K_n y_n \tag{11.9.10}$$

where we defined

$$K_n = aG_n \,, \quad f_n = a - cK_n \tag{11.9.11}$$

These are the time-varying analogs of Eqs. (11.6.17) and (11.6.19). Equations (11.9.8) and (11.9.9) may be combined into one updating equation for $P_{n/n-1}$, known as the discrete Riccati *difference* equation

$$P_{n+1/n} = \frac{a^2 R P_{n/n-1}}{R + c^2 P_{n/n-1}} + Q \tag{11.9.12}$$

It is the time-varying version of Eq. (11.6.18). We note that in deriving all of the above results, we did not need to assume that the model parameters $\{a, c, Q, R\}$ were constants, independent of time. They can just as well be replaced by time-varying model parameters:

$$\{a_n, c_n, Q_n, R_n\}$$

The asymptotic properties of the Kalman filter depend, of course, on the particular time variations in the model parameters. In the time-invariant case, with $\{a, c, Q, R\}$ constant, we expect the solution of the Riccati equation (11.9.12) to converge, for large n, to some steady-state value $P_{n/n-1} \to P$. In this limit, the Riccati difference equation (11.9.12) tends to the steady-state algebraic Riccati equation (11.6.18), which determines the limiting value P. The Kalman filter parameters will converge to the limiting values $f_n \to f$, $K_n \to K$, and $G_n \to G$ given by Eq. (11.6.19).

It is possible to solve Eq. (11.9.12) in closed form and explicitly demonstrate these convergence properties. Using the techniques of [871,872], we obtain

$$P_{n/n-1} = P + \frac{f^{2n} E_0}{1 + S_n E_0} \,, \quad \text{for } n = 0, 1, 2, \ldots, \tag{11.9.13}$$

where $E_0 = P_{0/-1} - P$ and

$$S_n = B\,\frac{1 - f^{2n}}{1 - f^2} \,, \quad B = \frac{c^2}{R + c^2 P}$$

We have already mentioned (see Problem 1.50) that the stability of the signal model and the positivity of the asymptotic solution P imply the minimum phase condition $|f| < 1$. Thus, the second term of Eq. (11.9.13) converges to zero exponentially with a time constant determined by f.

Example 11.9.1: Determine the closed form solutions of the time-varying Kalman filter for the state and measurement models:

$$x_{n+1} = x_n + w_n \,, \quad y_n = x_n + v_n$$

with $Q = 0.5$ and $R = 1$. Thus, $a = 1$ and $c = 1$. The Riccati equations are

$$P_{n+1/n} = \frac{P_{n/n-1}}{1 + P_{n/n-1}} + 0.5 \,, \quad P = \frac{P}{1 + P} + 0.5$$

The solution of the algebraic Riccati equation is $P = 1$. This implies that $f = aR/(R + c^2 P) = 0.5$. To illustrate the solution (11.9.13), we take the initial condition to be zero $P_{0/-1} = 0$. We find $B = c^2/(R + c^2 P) = 0.5$ and

$$S_n = \frac{2}{3}\left[1 - (0.5)^{2n}\right]$$

Thus,

$$P_{n/n-1} = 1 - \frac{(0.5)^{2n}}{1 - \frac{2}{3}\left[1 - (0.5)^{2n}\right]} = \frac{1 - (0.5)^{2n}}{1 + 2(0.5)^{2n}}$$

The first few values calculated from this formula are

$$P_{1/0} = \frac{1}{2} \,, \quad P_{2/1} = \frac{5}{6} \,, \quad P_{3/2} = \frac{21}{22} \,, \ldots$$

and quickly converge to $P = 1$. They may also be obtained by iterating Eq. (11.9.12). □

11.10 Problems

11.1 Let $\mathbf{x} = [x_{n_a}, \ldots, x_{n_b}]^T$ and $\mathbf{y} = [y_{n_a}, \ldots, y_{n_b}]^T$ be the desired and available signal vectors. The relationship between $\mathbf{x}$ and $\mathbf{y}$ is assumed to be linear of the form

$$\mathbf{y} = C\mathbf{x} + \mathbf{v}$$

where C represents a linear degradation and $\mathbf{v}$ is a vector of zero-mean independent gaussian samples with a common variance σ_v^2. Show that the maximum likelihood (ME) estimation criterion is in this case equivalent to the following least-squares criterion, based on the quadratic vector norm:

$$\mathcal{E} = \|\mathbf{y} - C\mathbf{x}\|^2 = \text{minimum with respect to } \mathbf{x}$$

Show that the resulting estimate is given by

$$\hat{\mathbf{x}} = (C^T C)^{-1} C^T \mathbf{y}$$

11.2 Let $\hat{\mathbf{x}} = H\mathbf{y}$ be the optimal linear smoothing estimate of $\mathbf{x}$ given by Eq. (11.1.5). It is obtained by minimizing the mean-square estimation error $\mathcal{E}_n = E[e_n^2]$ for each n in the interval $[n_a, n_b]$.

(a) Show that the solution for H also minimizes the error covariance matrix

$$R_{ee} = E[\mathbf{e}\mathbf{e}^T]$$

where $\mathbf{e}$ is the vector of estimation errors $\mathbf{e} = [e_{n_a}, \ldots, e_{n_b}]^T$.

(b) Show that H also minimizes every quadratic index of the form, for any positive semi-definite matrix Q:

$$E[\mathbf{e}^T Q \mathbf{e}] = \min$$

(c) Explain how the minimization of each $E[e_n^2]$ can be understood in terms of part (b).

11.3 Consider the smoothing problem of estimating the signal vector $\mathbf{x}$ from the signal vector $\mathbf{y}$. Assume that $\mathbf{x}$ and $\mathbf{y}$ are linearly related by

$$\mathbf{y} = C\mathbf{x} + \mathbf{v}$$

and that $\mathbf{v}$ and $\mathbf{x}$ are uncorrelated from each other, and that the covariance matrices of $\mathbf{x}$ and $\mathbf{v}$, R_{xx} and R_{vv}, are known. Show that the smoothing estimate of $\mathbf{x}$ is in this case

$$\hat{\mathbf{x}} = R_{xx} C^T [C R_{xx} C^T + R_{vv}]^{-1} \mathbf{y}$$

11.4 A stationary random signal has autocorrelation function $R_{xx}(k) = \sigma_x^2 a^{|k|}$, for all k. The observation signal is $y_n = x_n + v_n$, where v_n is a zero-mean, white noise sequence of variance σ_v^2, uncorrelated from x_n.

(a) Determine the optimal FIR Wiener filter of order $M = 1$ for estimating x_n from y_n.

(b) Repeat for the optimal linear predictor of order $M = 2$ for predicting x_n on the basis of the past two samples y_{n-1} and y_{n-2}.

11.5 A stationary random signal $x(n)$ has autocorrelation function $R_{xx}(k) = \sigma_x^2 a^{|k|}$, for all k. Consider a time interval $[n_a, n_b]$. The random signal $x(n)$ is known only at the end-points of that interval; that is, the only available observations are

$$y(n_a) = x(n_a), \quad y(n_b) = x(n_b)$$

Determine the optimal estimate of $x(n)$ based on just these two samples in the form

$$\hat{x}(n) = h(n, n_a) y(n_a) + h(n, n_b) y(n_b)$$

for the following values of n: (a) $n_a \le n \le n_b$, (b) $n \le n_a$, (c) $n \ge n_b$.

11.6 A stationary random signal x_n is to be estimated on the basis of the noisy observations

$$y_n = x_n + v_n$$

It is given that

$$S_{xx}(z) = \frac{1}{(1 - 0.5z^{-1})(1 - 0.5z)}, \quad S_{vv}(z) = 5, \quad S_{xv}(z) = 0$$

(a) Determine the optimal realizable Wiener filter for estimating the signal x_n on the basis of the observations $Y_n = \{y_i,\ i \le n\}$. Write the difference equation of this filter. Compute the mean-square estimation error.

(b) Determine the optimal realizable Wiener filter for predicting one step into the future; that is, estimate x_{n+1} on the basis of Y_n.

(c) Cast the results of (a) and (b) in a predictor/corrector Kalman filter form, and show explicitly that the innovations residual of the observation signal y_n is identical to the corresponding whitening sequence ϵ_n driving the signal model of y_n.

11.7 Repeat the previous problem for the following choice of state and measurement models

$$x_{n+1} = x_n + w_n, \quad y_n = x_n + v_n$$

where w_n and v_n have variances $Q = 0.5$ and $R = 1$, respectively.

11.8 Consider the state and measurement equations

$$x_{n+1} = a x_n + w_n, \quad y_n = c x_n + v_n$$

as discussed in Sec. 11.6. For *any* value of the Kalman gain K, consider the Kalman predictor/corrector algorithm defined by the equation

$$\hat{x}_{n+1/n} = a\hat{x}_{n/n-1} + K(y_n - c\hat{x}_{n/n-1}) = f\hat{x}_{n/n-1} + K y_n \tag{P.1}$$

where $f = a - cK$. The stability requirement of this estimation filter requires further that K be such that $|f| < 1$.

(a) Let $e_{n/n-1} = x_n - \hat{x}_{n/n-1}$ be the corresponding estimation error. Assuming that all signals are stationary, and working with z-transforms, show that the power spectral density of $e_{n/n-1}$ is given by

$$S_{ee}(z) = \frac{Q + K^2 R}{(1 - fz^{-1})(1 - fz)}$$

(b) Integrating $S_{ee}(z)$ around the unit circle, show that the mean-square value of the estimation error is given by

$$\mathcal{E} = E[e_{n/n-1}^2] = \frac{Q + K^2 R}{1 - f^2} = \frac{Q + K^2 R}{1 - (a - cK)^2} \tag{P.2}$$

(c) To select the optimal value of the Kalman gain K, differentiate $\mathcal{E}$ with respect to K and set the derivative to zero. Show that the resulting equation for K can be expressed in the form

$$K = \frac{caP}{R + c^2 P}$$

where P stands for the minimized value of $\mathcal{E}$; that is, $P = \mathcal{E}_{\min}$.

(d) Inserting this expression for K back into the expression (P.2) for $\mathcal{E}$, show that the quantity P must satisfy the algebraic Riccati equation

$$Q = P - \frac{a^2 RP}{R + c^2 P}$$

Thus, the resulting estimator filter is identical to the optimal one-step prediction filter discussed in Sec. 11.6.

11.9 Show that Eq. (P.2) of Problem 11.8 can be derived without using z-transforms, by using only stationarity, as suggested below: Using the state and measurement model equations and Eq. (P. l), show that the estimation error $e_{n/n-1}$ satisfies the difference equation

$$e_{n+1/n} = f e_{n/n-1} + w_n - K v_n$$

Then, invoking stationarity, derive Eq. (P.2). Using similar methods, show that the mean-square estimation error is given by

$$E[e_{n/n}^2] = \frac{RP}{R + c^2 P}$$

where $e_{n/n} = x_n - \hat{x}_{n/n}$ is the estimation error of the optimal filter (11.6.13).

11.10 Consider the general example of Sec. 11.6. It was shown there that the innovations residual was the same as the whitening sequence ϵ_n driving the signal model of y_n

$$\epsilon_n = y_n - \hat{y}_{n/n-1} = y_n - c\hat{x}_{n/n-1}$$

Show that it can be written as

$$\epsilon_n = c e_{n/n-1} + v_n$$

where $e_{n/n-1} = x_n - \hat{x}_{n/n-1}$ is the prediction error. Then, show that

$$\sigma_\epsilon^2 = E[\epsilon_n^2] = R + c^2 P$$

11.11 *Computer Experiment.* Consider the signal and measurement model defined by Eqs. (11.6.6) through (11.6.8), with the choices $a = 0.9$, $c = 1$, $Q = 1 - a^2$, and $R = 1$. Generate 1500 samples of the random noises w_n and v_n. Generate the corresponding signals x_n and y_n according to the state and measurement equations. Determine the optimal Wiener filter of the form (11.6.13) for estimating x_n on the basis of y_n. Filter the sequence y_n through the Wiener filter to generate the sequence $\hat{x}_{n/n}$.

(a) On the same graph, plot the desired signal x_n and the available noisy version y_n for n ranging over the last 100 values (i.e., $n = 1400$–1500.)

(b) On the same graph, plot the recovered signal $\hat{x}_{n/n}$ together with the original signal x_n for n ranging over the last 100 values.

(c) Repeat (a) and (b) using a different realization of w_n and v_n.

(d) Repeat (a), (b), and (c) for the choice $a = -0.9$.

11.12 Consider the optimal Wiener filtering problem in its matrix formulation of Sec. 11.8. Let $\mathbf{e} = \mathbf{x} - \hat{\mathbf{x}} = \mathbf{x} - H\mathbf{y}$ be the estimation error corresponding to a particular choice of the lower-triangular matrix H. Minimize the error covariance matrix $R_{ee} = E[\mathbf{e}\mathbf{e}^T]$ with respect to H subject to the *constraint* that H be lower-triangular. These constraints are $H_{ni} = 0$ for $n < i$. To do this, introduce a set of Lagrange multipliers Λ_{ni} for $n < i$, one for each constraint equation, and incorporate them into an effective performance index

$$J = E[\mathbf{e}\mathbf{e}^T] + \Lambda H^T + H\Lambda^T = \min$$

where the matrix Λ is strictly upper-triangular. Show that this formulation of the minimization problem yields exactly the same solution as Eq. (11.8.7).

11.13 *Exponential Moving Average as Wiener Filter.* The single EMA filter for estimating the local level of a signal that we discussed in Chap. 6 admits a nice Wiener-Kalman filtering interpretation. Consider the noisy random walk signal model,

$$\begin{aligned} x_{n+1} &= x_n + w_n \\ y_n &= x_n + v_n \end{aligned} \tag{11.10.1}$$

where w_n, v_n are mutually uncorrelated, zero-mean, white noise signals of variances $Q = \sigma_w^2$ and $R = \sigma_v^2$. Based on the material in Section 12.6, show that the optimum Wiener/Kalman filter for predicting x_n from y_n is equivalent to the exponential smoother, that is, show that it is given by,

$$\hat{x}_{n+1/n} = f\hat{x}_{n/n-1} + (1 - f)y_n \tag{11.10.2}$$

so that the forgetting-factor parameter λ of EMA is identified as the closed-loop parameter f of the Kalman filter, and show further that f is given in terms of Q, R as follows,

$$1 - f = \frac{\sqrt{Q^2 + 4QR} - Q}{2R}$$

Show also the $\hat{x}_{n+1/n} = \hat{x}_{n/n}$.

a. For the following values $\sigma_w = 0.1$ and $\sigma_v = 1$, generate $N = 300$ samples of x_n, y_n from Eq. (11.10.1) and run y_n through the equivalent Kalman filter of Eq. (11.10.2) to compute $\hat{x}_{n/n-1}$. On the same graph, plot all three signals $y_n, x_n, \hat{x}_{n/n-1}$ versus $0 \le n \le N - 1$. An example graph is shown at the end.

b. A possible way to determine λ or f from the data y_n is as follows. Assume a tentative value for λ, compute $\hat{x}_{n/n-1}$, then the error $e_{n/n-1} = x_n - \hat{x}_{n/n-1}$, and the mean-square error:

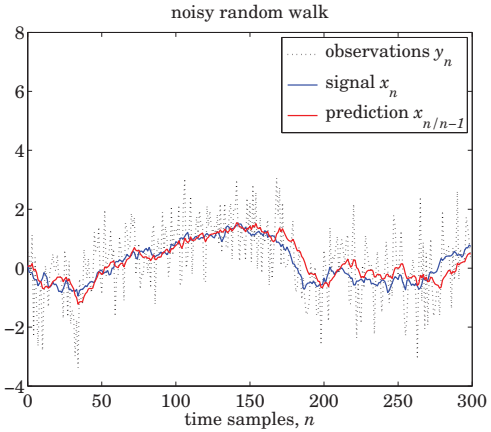
$$\text{MSE}(\lambda) = \sum_n e_{n/n-1}^2$$

Repeat the calculation of $\text{MSE}(\lambda)$ over a range of λs, for example, $0.80 \le \lambda \le 0.95$, chosen such that the interval $[0.80, 0.95]$ contain the true λ. Then find that λ that minimizes $\text{MSE}(\lambda)$, which should be close to the true value.

Because the estimated λ depends on the particular realization of the model (11.10.1), generate 20 different realizations of the pair x_n, y_n with the same Q, R, and for each realization carry out the estimate of λ as described above, and finally form the average of the 20 estimated λs. Discuss if this method generates an acceptable estimate of λ or f.

c. Repeat part (b), by replacing the MSE by the mean-absolute-error:

$$\text{MAE}(\lambda) = \sum_n |e_{n/n-1}|$$



12

Linear Prediction

12.1 Pure Prediction and Signal Modeling

In Sec. 1.17, we discussed the connection between linear prediction and signal modeling. Here, we rederive the same results by considering the linear prediction problem as a special case of the Wiener filtering problem, given by Eq. (11.4.6). Our aim is to cast the results in a form that will suggest a practical way to solve the prediction problem and hence also the modeling problem. Consider a stationary signal y_n having a signal model

$$S_{yy}(z) = \sigma_\epsilon^2 B(z) B(z^{-1}) \qquad \epsilon_n \longrightarrow \boxed{B(z)} \longrightarrow y_n \tag{12.1.1}$$

as guaranteed by the spectral factorization theorem. Let $R_{yy}(k)$ denote the autocorrelation of y_n:

$$R_{yy}(k) = E[y_{n+k} y_n]$$

The linear prediction problem is to predict the current value y_n on the basis of all the past values $Y_{n-1} = \{y_i\,,\ -\infty < i \le n-1\}$. If we define the delayed signal $y_1(n) = y_{n-1}$, then the linear prediction problem is equivalent to the optimal Wiener filtering problem of estimating y_n from the related signal $y_1(n)$. The optimal estimation filter $H(z)$ is given by Eq. (11.4.6), where we must identify x_n and y_n with y_n and $y_1(n)$ of the present notation. Using the filtering equation $Y_1(z) = z^{-1} Y(z)$, we find that y_n and $y_1(n)$ have the same spectral factor $B(z)$

$$S_{y_1 y_1}(z) = (z^{-1})(z) S_{yy}(z) = S_{yy}(z) = \sigma_\epsilon^2 B(z) B(z^{-1})$$

and also that

$$S_{y y_1}(z) = S_{yy}(z) z = z \sigma_\epsilon^2 B(z) B(z^{-1})$$

Inserting these into Eq. (11.4.6), we find for the optimal filter $H(z)$

$$H(z) = \frac{1}{\sigma_\epsilon^2 B(z)} \left[\frac{S_{y y_1}(z)}{B(z^{-1})} \right]_+ = \frac{1}{\sigma_\epsilon^2 B(z)} \left[\frac{z \sigma_\epsilon^2 B(z) B(z^{-1})}{B(z^{-1})} \right]_+ , \quad \text{or,}$$

$$H(z) = \frac{1}{B(z)} \left[z B(z) \right]_+ \tag{12.1.2}$$