STOCHASTIC PROCESSES, DETECTION AND ESTIMATION

6.432 Course Notes

Alan S. Willsky, Gregory W. Wornell, and Jeffrey H. Shapiro

Department of Electrical Engineering and Computer Science
Massachusetts Institute of Technology
Cambridge, MA 02139

Fall 2003

7

Waveform Estimation, Wiener and Kalman Filtering: Addendum

7.1 LINEAR LEAST-SQUARES FILTERING: SUMMARY

We begin by summarizing our results for continuous-time problems. The canonical problem we have been considering is the following: given observations of a (zero-mean) random process $y(\cdot)$ of the form

$$\{y(\tau), \quad T_\mathrm{i} \le \tau \le T_\mathrm{f}\}$$

we seek estimates of a second (zero-mean) random process $x(\cdot)$ at some particular time t. We focus on a Bayesian formulation with a least-squares cost criterion, and constrain our estimates to be linear, i.e.,

$$\hat{x}(t) = \int_{T_\mathrm{i}}^{T_\mathrm{f}} h(t,\tau)\, y(\tau)\, d\tau, \tag{7.1}$$

where $h(\cdot,\cdot)$ is the linear (and generally time-varying) filter to be determined.

In Chapter 3, we saw that the corresponding problems involving random vectors required knowledge of only second-order statistics to determine the optimum estimator. This is also the case in this more general problem: the optimum filter $h(\cdot,\cdot)$ is a function of the cross-covariance $K_{xy}(\cdot,\cdot)$ and the autocovariance of the observations $K_{yy}(\cdot,\cdot)$. Indeed, the orthogonality condition

$$E\left[(\hat{x}(t) - x(t))\; y(\tau)\right] = 0, \qquad T_\mathrm{i} \le \tau \le T_\mathrm{f} \tag{7.2}$$

leads immediately to the *Wiener-Hopf* equation

$$\int_{T_\mathrm{i}}^{T_\mathrm{f}} h(t,\sigma)\, K_{yy}(\sigma,\tau)\, d\sigma = K_{xy}(t,\tau), \qquad T_\mathrm{i} \le \tau \le T_\mathrm{f}, \tag{7.3}$$

from which the optimum $h(\cdot, \cdot)$ can be determined.

To further determine the performance of the estimator requires, in addition, knowledge of $K_{xx}(\cdot, \cdot)$, the autocovariance of the process of interest. In particular, the estimation error

$$e(t) = x(t) - \hat{x}(t) \tag{7.4}$$

has variance

$$\begin{aligned}
\lambda_e(t) &= E\left[(x(t) - \hat{x}(t))\,(x(t) - \hat{x}(t))\right] \\
&= E\left[x(t)\,(x(t) - \hat{x}(t))\right] \\
&= K_{xx}(t,t) - \int_{T_\mathrm{i}}^{T_\mathrm{f}} h(t,\tau)\,K_{xy}(t,\tau)\,d\tau,
\end{aligned} \tag{7.5}$$

where the second equality follows from using the orthogonality condition (7.2), and where the last equality follows from substituting (7.1).

Analogous results apply in discrete-time problems. Summarizing briefly, in this case the observations take the form

$$\{y[k], \quad N_\mathrm{i} \le k \le N_\mathrm{f}\}$$

where $y[\cdot]$ is a (zero-mean) random process, and we seek estimates of a second (zero-mean) random process $x[\cdot]$ at some particular time n. Constraining our estimates to be linear, i.e.,

$$\hat{x}[n] = \sum_{k=N_\mathrm{i}}^{N_\mathrm{f}} h[n,k]\,y[k] \tag{7.6}$$

where $h[\cdot, \cdot]$ is the linear (and generally time-varying) filter to be determined, we obtain that the filter generating the minimum mean-square error estimate satisfies the discrete-time Wiener-Hopf equation

$$\sum_{l=N_\mathrm{i}}^{N_\mathrm{f}} h[n,l]\,K_{yy}[l,k] = K_{xy}[n,k], \qquad N_\mathrm{i} \le k \le N_\mathrm{f}, \tag{7.7}$$

as a result of the orthogonality condition

$$E\left[(\hat{x}[n] - x[n])\;y[k]\right] = 0, \qquad N_\mathrm{i} \le k \le N_\mathrm{f}. \tag{7.8}$$

In turn, the estimation error

$$e[n] = x[n] - \hat{x}[n] \tag{7.9}$$

has variance

$$\begin{aligned}
\lambda_e[n] &= E\left[(x[n] - \hat{x}[n])^2\right] \\
&= E\left[x[n]\,(x[n] - \hat{x}[n])\right] \\
&= K_{xx}[n,n] - \sum_{k=N_\mathrm{i}}^{N_\mathrm{f}} h[n,k]\,K_{xy}[n,k].
\end{aligned} \tag{7.10}$$

7.2 WIENER FILTERING

When the processes involved are jointly wide-sense stationary, we obtained more specialized results. In particular, when the observation interval was doubly-infinite ($-T_\mathrm{i} = T_\mathrm{f} \to \infty$ or $-N_\mathrm{i} = N_\mathrm{f} \to \infty$), we concluded that the optimum filter was linear and time-invariant (LTI), i.e.,

$$\hat{x}(t) = h(t) * y(t) = \int h(t-\tau)\, y(\tau)\, d\tau \tag{7.11}$$

in continuous-time, or

$$\hat{x}[n] = h[n] * y[n] = \sum_k h[n-k]\, y[k] \tag{7.12}$$

in discrete-time. Moreover, this optimum filter was generally noncausal.

In some real-time applications, causal filters are required, corresponding to a semi-infinite observation interval ($T_\mathrm{i} \to -\infty, T_\mathrm{f} = t$ or $N_\mathrm{i} \to -\infty, N_\mathrm{f} = n$). In this case, the optimum filter is still LTI, but has a causal impulse response:

$$h(t) = 0, \qquad t < 0 \tag{7.13}$$

in continuous-time, or

$$h[n] = 0, \qquad n \le -1 \tag{7.14}$$

in discrete-time.

When the spectra $S_{yy}(s)$ and $S_{xy}(s)$ (or $S_{yy}(z)$ and $S_{xy}(z)$) involved are rational, we were able to develop a convenient procedure for determining the optimum filters. This procedure exploited the concept of preprocessing the observations with a causal, stable whitening filter that is causally and stably invertible. The preprocessed, "causally equivalent" observations we referred to as the *innovations* process.

The filters that arise in both the noncausal and causal cases generally have infinite-impulse response (IIR) and may not be realizable. In the case of rational spectra, however, they are realizable; efficient implementations result from interpreting the required filtering as a linear constant coefficient differential (or difference) equation. Even when the spectra involved are not rational, however, realizable approximations can be obtained by appropriately truncating the impulse response.

When it is important that the filters involved have strictly finite impulse response (FIR), then it is natural to impose this constraint explicitly. In particular, this constraint corresponds to choosing an observation interval of the form $T_\mathrm{i} = t-T, T_\mathrm{f} = t$ for some T, or $N_\mathrm{i} = n-N+1, N_\mathrm{f} = n$ for some N, in which case the Wiener-Hopf equation (7.3) or (7.7) specializes further. In the discrete-time case, for example, this equation becomes

$$\sum_{k=0}^{N-1} h[k]\, K_{yy}[n-k] = K_{xy}[n], \qquad n = 0, 1, \dots, N-1, \tag{7.15}$$

which is a straightforward matrix equation for the taps of the FIR filter:

$$\begin{bmatrix} K_{yy}[0] & K_{yy}[1] & \cdots & K_{yy}[N-1] \\ K_{yy}[1] & K_{yy}[0] & \cdots & K_{yy}[N-2] \\ \vdots & \vdots & \ddots & \vdots \\ K_{yy}[N-1] & K_{yy}[N-2] & \cdots & K_{yy}[0] \end{bmatrix} \begin{bmatrix} h[0] \\ h[1] \\ \vdots \\ h[N-1] \end{bmatrix} = \begin{bmatrix} K_{xy}[0] \\ K_{xy}[1] \\ \vdots \\ K_{xy}[N-1] \end{bmatrix}. \quad (7.16)$$

As a historical note, much of the continuous-time theory summarized in this section is generally attributed to Wiener and Hopf, while much of the discrete-time theory was developed by Kolmogorov. Nevertheless, the term "Wiener filtering" is widely used to refer to both the continuous-time and discrete-time theory.

We also remark that Wiener filtering theory can be readily generalized to accommodate scenarios in which the constituent processes are vector-valued, i.e.,

$$\mathbf{x}[n] = \begin{bmatrix} x_1[n] \\ x_2[n] \\ \vdots \\ x_{L_\mathbf{x}}[n] \end{bmatrix}, \qquad \text{and} \qquad \mathbf{y}[n] = \begin{bmatrix} y_1[n] \\ y_2[n] \\ \vdots \\ y_{L_\mathbf{y}}[n] \end{bmatrix},$$

7.3 RECURSIVE ESTIMATION

When the processes involved are not wide-sense stationary, or when the observation intervals involved are such that LTI filtering cannot be used to generate the linear least-squares estimates, Wiener filtering results cannot be applied. Such is the case, for instance, when the observation interval takes the form $T_\mathrm{i} = 0, T_\mathrm{f} = t$ or $N_\mathrm{i} = 0, N_\mathrm{f} = n$.

From the perspective of the FIR Wiener filtering discussion of the last section, this would correspond to scenario in which estimating $x[n]$ for successive values of n would require progressively longer FIR filters—each having length $N = n+1$. In principle, the Wiener-Hopf equations (7.16) could be solved for each such n. However, this is obviously computationally unattractive, requiring $\mathcal{O}(n^3)$ computations per output sample of $\hat{x}[n]$ due to the matrix inversion!

Kalman filtering theory is the tool by which problems of this form can be solved in a computationally efficient manner. In this section we focus on the discrete-time case; analogous results can be developed for the continuous-time case.

7.3.1 Model

Kalman filtering exploits the fact that in some scenarios the statistical relationship between $x[n]$ and $y[n]$ is such that the estimates $\hat{x}[n]$ for successive values of n are closely related. By exploiting this relationship we can avoid re-solving the Wiener-Hopf equation independently each time, instead using preceding estimates in the construction of current estimates in a recursive manner.

One class of models for which this property holds has the relationship between $x[n]$ and $y[n]$ given in the following state-space form

$$x[n+1] = a\,x[n] + v[n] \tag{7.17a}$$
$$y[n] = c\,x[n] + w[n], \tag{7.17b}$$

where $v[n]$ and $w[n]$ are suitable zero-mean random processes. We refer to (7.17a) as the *state equation*; it describes the dynamics governing the statistical structure of $x[n]$. In this equation, the initial condition $x[0]$ is a zero-mean random variable with some variance $\lambda_x[0] = \sigma^2$. We refer to (7.17b) as the *observation equation*; it describes the manner by which the observations $y[\cdot]$ are correlated with the process $x[\cdot]$ to be estimated.

7.3.2 Kalman Filtering

The problem we consider in this section is the following. Given observations of the form

$$\{y[k], \quad 0 \leq k \leq n\} \tag{7.18}$$

determine an efficient recursive algorithm for the calculation of the linear least-squares estimate $\hat{x}[n]$ when $x[n]$ and $y[n]$ are related via the state-space model (7.17).

We adopt some convenient notation to facilitate our discussion. If $r[\cdot]$ is an arbitrary random process, then $\hat{r}[n|k]$ denotes the linear least-squares estimate of the process at time n based on observations through time k, i.e., the estimate of sample $r[n]$ based on observations of $\{y[0], y[1], \dots, y[k]\}$. Similarly, $\lambda_{\hat{r}}[n|k]$ will be used to denote the variance of this estimate, and $\lambda_e[n|k]$ to denote the variance in the error

$$e[n|k] \triangleq x[n] - \hat{x}[n|k]. \tag{7.19}$$

As in the case of our development of causal Wiener filtering, our derivation of the Kalman filter exploits the concept of preprocessing our observations with a whitening filter. For the moment, let us assume the existence of such a filter, producing equivalent whitened observations $z[0], z[1], \dots, z[n]$; we will develop how such "innovations" are produced later. Accordingly, we base our estimates on these innovations.

Step 1: Update Equations

In terms of our whitened observations, the Wiener-Hopf equations (i.e., normal equations) yield, simply,

$$\begin{aligned}\hat{x}[n|n] &= \sum_{k=0}^{n} \frac{K_{xz}[n,k]}{K_{zz}[k,k]} \, z[k] \\ &= \sum_{k=0}^{n-1} \frac{K_{xz}[n,k]}{K_{zz}[k,k]} \, z[k] + \frac{K_{xz}[n,n]}{K_{zz}[n,n]} \, z[n]. \end{aligned} \tag{7.20}$$

However, similar Wiener-Hopf equations yield the prediction

$$\hat{x}[n|n-1] = \sum_{k=0}^{n-1} \frac{K_{xz}[n,k]}{K_{zz}[k,k]} \, z[k], \tag{7.21}$$

so substituting (7.21) in (7.20) yields the recursion

$$\hat{x}[n|n] = \hat{x}[n|n-1] + g[n] \, z[n] \tag{7.22}$$

for combining the previous prediction with the new observation, where the weight (or "gain") $g[n]$ is

$$g[n] = \frac{K_{xz}[n,n]}{K_{zz}[n,n]}. \tag{7.23}$$

A similar recursion can be obtained for the corresponding mean-square estimation error. In particular,

$$\begin{aligned}\lambda_e[n|n] &= \lambda_x[n] - \lambda_{\hat{x}}[n|n] \\ &= \lambda_x[n] - \sum_{k=0}^{n} \frac{K_{xz}^2[n,k]}{K_{zz}[k,k]} \\ &= \left[\lambda_x[n] - \sum_{k=0}^{n-1} \frac{K_{xz}^2[n,k]}{K_{zz}[k,k]} \right] - \frac{K_{xz}^2[n,n]}{K_{zz}[n,n]} \end{aligned} \tag{7.24}$$

where we have used, in order, the orthogonality condition and the fact that the observations $z[\cdot]$ are white. Similarly, recognizing that

$$\lambda_e[n|n-1] = \lambda_x[n] - \lambda_{\hat{x}}[n|n-1] = \lambda_x[n] - \sum_{k=0}^{n-1} \frac{K_{xz}^2[n,k]}{K_{zz}[k,k]}, \tag{7.25}$$

we can substitute (7.25) into (7.24) to obtain the recursion

$$\lambda_e[n|n] = \lambda_e[n|n-1] - g[n] \, K_{xz}[n,n]. \tag{7.26}$$

Note that to obtain these results we have not yet exploited any of the particular structure of the signal model (7.17).

Step 2: Prediction Equations

In the previous section we developed recursions for obtaining $\hat{x}[n|n]$ from the previous predictions $\hat{x}[n|n-1]$. In this section, we exploit the fact that $x[n]$ are generated by the recursion (7.17a) to generate the predictions themselves. To begin, exploiting the superposition property of linear least-squares estimation we obtain

$$\hat{x}[n+1|n] = a\,\hat{x}[n|n] + \hat{v}[n|n]. \tag{7.27}$$

In turn, provided we constrain our model so that $v[n]$ is uncorrelated with $y[0]$, $y[1], \dots, y[n]$, then the last term in (7.27) is zero, so our prediction equation becomes

$$\hat{x}[n+1|n] = a\,\hat{x}[n|n]. \tag{7.28}$$

Using (7.28), we can compute the new prediction based on the current estimate.

We can similarly obtain a recursion for the prediction error variance. In particular, combining (7.17a) with (7.28) we obtain

$$\begin{aligned} e[n+1|n] &= x[n+1] - \hat{x}[n+1|n] \\ &= a\,(x[n] - \hat{x}[n|n]) + v[n] \\ &= a\,e[n|n] + v[n]. \end{aligned} \tag{7.29}$$

Finally, when, in addition, the process $v[n]$ is uncorrelated with $x[n]$, then (7.29) implies

$$\lambda_e[n+1|n] = a^2\,\lambda_e[n|n] + \sigma_v^2, \tag{7.30}$$

where σ_v^2 is the variance of the process $v[n]$. We ensure this additional requirement is met by constraining $v[n]$ to be white and uncorrelated with the initial condition $x[0]$.

Step 3: Gain Computation

It remains to complete our expression for the gain (7.23). This requires that we make use of two final pieces of information: the specific relationship between the original and whitened observations, and the observation equation (7.17b).

The former is addressed by the following result, derived in the context of Gram-Schmidt orthogonalization in Section 3.2.5.

Theorem 7.1 *The innovations process*

$$z[n] = y[n] - \hat{y}[n|n-1] \tag{7.31}$$

defined for $n \geq 0$ with $\hat{y}[0|-1] \triangleq 0$, which is generated causally from $y[n]$, is a white random process that is causally invertible.

Let us now turn our attention to the observation equation (7.17b). Again exploiting the superposition property of linear least-squares estimation we obtain

$$\hat{y}[n|n-1] = c\hat{x}[n|n-1] + \hat{w}[n|n-1]. \tag{7.32}$$

However, to ensure that $v[n]$ is uncorrelated with $y[0], y[1], \ldots, y[n]$ as required in the last section, it suffices to choose $w[\cdot]$ to be uncorrelated with both $v[\cdot]$ and $x[0]$. This property, when combined with an additional constraint that $w[n]$ also be white, ensures that $w[n]$ is uncorrelated with the $x[\cdot]$, so that $w[n]$ is also uncorrelated with $y[0], y[1], \ldots, y[n-1]$. The last term in (7.32) is therefore zero, and

$$\hat{y}[n|n-1] = c\hat{x}[n|n-1]. \tag{7.33}$$

Hence, using (7.33) with (7.31) we obtain

$$\begin{aligned} z[n] &= c\,(x[n] - \hat{x}[n|n-1]) + w[n] \\ &= c\, e[n|n-1] + w[n]. \end{aligned} \tag{7.34}$$

We can now calculate the quantities $K_{xz}[n,n]$ and $K_{zz}[n,n]$ we need. In particular, using (7.34) and the fact that $w[n]$ is uncorrelated with $x[n]$, we obtain

$$\begin{aligned} K_{xz}[n,n] &= E\,[x[n]\,z[n]] \\ &= cE\,[x[n]\,(x[n] - \hat{x}[n|n-1])] + E\,[x[n]\,w[n]] \\ &= c\,\lambda_e[n|n-1], \end{aligned} \tag{7.35}$$

where the last equality follows from (7.10), and

$$K_{zz}[n,n] = c^2\,\lambda_e[n|n-1] + \sigma_w^2, \tag{7.36}$$

where we have also used once more that $w[n]$ is uncorrelated with $y[0], y[1], \ldots, y[n-1]$. Using (7.35) and (7.36) in (7.23) we then obtain

$$g[n] = \frac{c\,\lambda_e[n|n-1]}{c^2\,\lambda_e[n|n-1] + \sigma_w^2} \tag{7.37}$$

and can rewrite (7.26) as

$$\lambda_e[n|n] = \lambda_e[n|n-1] - g[n]\,c\,\lambda_e[n|n-1]. \tag{7.38}$$

Summary

In this section, we collect together our constraints on the model and the resulting estimation equations developed above to obtain our complete recursive estimation algorithm.

In particular, when our model is of the form

$$\begin{aligned} x[n+1] &= a\,x[n] + v[n] \\ y[n] &= c\,x[n] + w[n], \end{aligned}$$

where $v[\cdot]$ and $w[\cdot]$ are uncorrelated, white, zero-mean random variables with variances σ_v^2 and σ_w^2, respectively, and uncorrelated with the zero-mean, variance σ^2 initial condition $x[0]$, then the following recursive algorithm computes the linear least-squares estimates of the process $x[n]$ based on the observations $y[n]$:

1. Initialize the prediction and its associated error variance according to

$$\hat{x}[0|-1] = 0 \tag{7.39a}$$
$$\lambda_e[0|-1] = \sigma^2, \tag{7.39b}$$

   and let $n = 0$.

2. Compute the Kalman gain

$$g[n] = \frac{c\,\lambda_e[n|n-1]}{c^2\,\lambda_e[n|n-1] + \sigma_w^2}. \tag{7.39c}$$

   and generate the filtered estimate and its associated error variance from the corresponding prediction quantities according to

$$\hat{x}[n|n] = \hat{x}[n|n-1] + g[n]\ (y[n] - c\,\hat{x}[n|n-1]) \tag{7.39d}$$
$$\lambda_e[n|n] = \lambda_e[n|n-1] - g[n]\,c\,\lambda_e[n|n-1]. \tag{7.39e}$$

3. Generate the next prediction and its associated error variance from the corresponding filtered quantities according to

$$\hat{x}[n+1|n] = a\,\hat{x}[n|n] \tag{7.39f}$$
$$\lambda_e[n+1|n] = a^2\,\lambda_e[n|n] + \sigma_v^2. \tag{7.39g}$$

4. Increment n and go to step 2.

These equations are referred to as the Kalman filter equations. Related recursions can be obtained to implement Kalman predictors and Kalman smoothers. Furthermore, it is important to keep in mind that although these are equations for the least-squares estimates of $x[n]$ when the estimators are specifically constrained to be linear, for the case in which $x[0]$, $v[\cdot]$, and $w[\cdot]$ are jointly Gaussian, the resulting estimates are also Bayes least-squares (and, for that matter, MAP) estimates without the linear constraint.

Several additional points should be emphasized. Note that the prediction and filtering error variances $\lambda_e[n+1|n]$ and $\lambda_e[n|n]$, as well as the Kalman gain $g[n]$ can all be recursively computed off-line—i.e., prior to data acquisition—since they make no use of the observations. Hence, the only quantities that need to be computed in real-time are the predictions and filtered estimates themselves. Hence, an extraordinarily small number of real-time computations is required to compute each sample of $\hat{x}[n|n]$. Moreover, this number of computations is identical for all n in contrast to the general case which required $\mathcal{O}(n^3)$ computations to obtain each output sample!

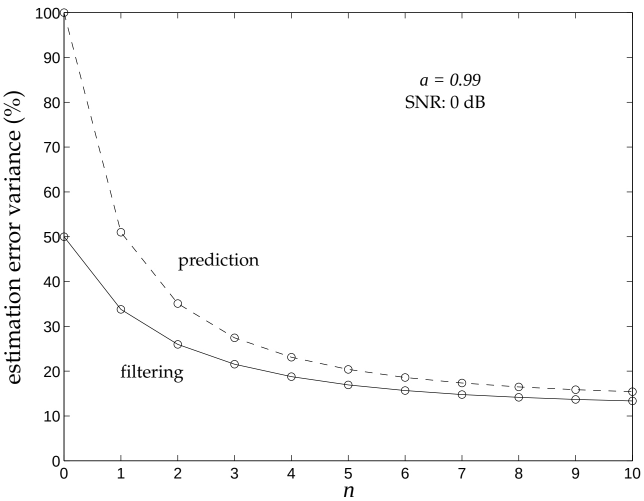

Figure 7.1. Typical prediction and filtering error covariance progression with the Kalman algorithm.

To emphasize this point, and to illustrate some typical behavior of the algorithm, Fig. 7.1 plots the precomputed prediction and filtering error variances (as a percentage of $\mathrm{var}\, \mathsf{x}[n] = \sigma_{\mathsf{v}}^2/(1-a^2)$) for the case in which $a = 0.99$ and $c = 1$. The output noise variance σ_{w}^2 is chosen to correspond to an SNR of 0 dB.

While the class of signal models captured by (7.17) is fairly narrow, some straightforward generalizations lead to a significantly richer class of models and corresponding extensions to the Kalman filter algorithm. As one simple example, it is straightforward to modify the algorithm for the case when the state equation takes the form

$$\mathsf{x}[n+1] = a\mathsf{x}[n] + b\mathsf{v}[n] \tag{7.40}$$

where b is some constant. In particular, since this corresponds to just a scaling of the driving noise $\mathsf{v}[n]$, all we need to do is replace instances of σ_{v}^2 in the algorithm with $b^2\sigma_{\mathsf{v}}^2$. Thus it suffices to replace (7.39g) with

$$\lambda_{\mathsf{e}}[n+1|n] = a^2\, \lambda_{\mathsf{e}}[n|n] + b^2\sigma_{\mathsf{v}}^2. \tag{7.41}$$

Also, when the initial condition is appropriately chosen, the model (7.17) corresponds to the statistics of processes $\mathsf{x}[n]$ and $\mathsf{y}[n]$ being jointly wide-sense stationary. However, a variety of nonstationary processes can be modeled by making the constants $a, b, c, \sigma_{\mathsf{v}}^2, \sigma_{\mathsf{w}}^2$ time-varying. It is remarkably easy to modify the estimation algorithm for this case. In fact, reviewing our derivation reveals that apart from notational differences the derivation in this new case is identical, so we need only replace instances of these constants in (7.39) with the corresponding time-varying quantities $a[n], b[n], c[n], \sigma_{\mathsf{v}}^2[n], \sigma_{\mathsf{w}}^2[n]$.

However, even in the stationary case a considerable richer class of models can be obtained by allowing $\mathsf{x}[n]$ to be a *vector* process of dimension L_{x}, so that the state equation becomes L_{x}-dimensional. For example, with this modification any wide-sense stationary process $\mathsf{y}[n]$ with a rational spectrum can be captured with the statespace model. Similarly, it is straightforward to accommodate the case of observations taking the form of a vector process of dimension L_{y}. We therefore conclude our discussion of Kalman filtering in this addendum with a development of the results for this vector case, following the same sequence of steps we used to obtain the results for the scalar case.

7.3.3 Vector Kalman Filtering

In this section we consider a rather general state space model of the form

$$\mathsf{x}[n+1] = \mathbf{A}[n]\,\mathsf{x}[n] + \mathbf{B}[n]\,\mathsf{v}[n] \tag{7.42a}$$

$$\mathsf{y}[n] = \mathbf{C}[n]\,\mathsf{x}[n] + \mathsf{w}[n]. \tag{7.42b}$$

In this model, $\mathsf{v}[n]$ is a L_{v}-dimensional process, which may be different from the dimension of $\mathsf{x}[n]$. Hence, the matrices $\mathbf{A}[n]$, $\mathbf{B}[n]$ and $\mathbf{C}[n]$ are $L_{\mathsf{x}} \times L_{\mathsf{x}}$, $L_{\mathsf{x}} \times L_{\mathsf{v}}$, and $L_{\mathsf{y}} \times L_{\mathsf{x}}$, respectively. Moreover, the processes $\mathsf{v}[\cdot]$ and $\mathsf{w}[\cdot]$ are zero-mean and white with covariances $\mathbf{\Lambda}_{\mathsf{v}}[n]$ and $\mathbf{\Lambda}_{\mathsf{w}}[n]$, respectively, and the initial condition $\mathsf{x}[0]$ has zero-mean and covariance $\mathbf{\Lambda}_{\mathsf{x}}[0] = \mathbf{\Lambda}$. The random variables $\mathsf{x}[0]$, $\mathsf{v}[\cdot]$, and $\mathsf{w}[\cdot]$ are uncorrelated with one another.

In what follows, we make use of the following notation for the (matrix-valued) cross-covariance function for any pair of vector processes $\mathsf{q}[n]$ and $\mathsf{r}[n]$:

$$\mathbf{K}_{\mathsf{qr}}[n,k] = \operatorname{cov}\left(\mathsf{q}[n], \mathsf{r}[k]\right), \tag{7.43}$$

so that for zero-mean processes we have

$$\mathbf{K}_{\mathsf{qr}}[n,k] = E\left[\mathsf{q}[n]\,\mathsf{r}^{\mathrm{T}}[k]\right]. \tag{7.44}$$

Using this notation, the joint second-order statistics of $\mathsf{v}[n]$ and $\mathsf{w}[n]$ can be expressed in the form

$$\mathbf{K}_{\mathsf{vv}}[n,k] = \mathbf{\Lambda}_{\mathsf{v}}[n]\,\delta[n-k] \tag{7.45}$$

$$\mathbf{K}_{\mathsf{ww}}[n,k] = \mathbf{\Lambda}_{\mathsf{w}}[n]\,\delta[n-k] \tag{7.46}$$

$$\mathbf{K}_{\mathsf{vw}}[n,k] = \mathbf{0}. \tag{7.47}$$

As other notation, we use $\hat{\mathsf{r}}[n|k]$ to denote the linear least-squares estimate of $\mathsf{r}[n]$ based on $\mathsf{y}[0], \mathsf{y}[1], \dots, \mathsf{y}[k]$, and use $\mathbf{\Lambda}_{\mathsf{e}}[n|k]$ to denote the covariance of the error

$$\mathsf{e}[n|k] = \mathsf{x}[n] - \hat{\mathsf{x}}[n|k] \tag{7.48}$$

in the estimate of $\mathsf{x}[n]$ based on $\mathsf{y}[0], \mathsf{y}[1], \dots, \mathsf{y}[k]$, i.e.,

$$\mathbf{\Lambda}_{\mathsf{e}}[n|k] = E\left[\mathsf{e}[n|k]\,\mathsf{e}^{\mathrm{T}}[n|k]\right]. \tag{7.49}$$

The relevant normal (Wiener-Hopf) equations in the vector case follow from our results in Chapter 3 on vector estimation. In particular, the filtered estimates $\hat{\mathsf{x}}[n|n]$ satisfy

$$E\left[\left(\hat{\mathsf{x}}[n|n] - \mathsf{x}[n]\right)\mathsf{y}^{\mathrm{T}}[k]\right] = \mathbf{0}, \qquad k = 0, 1, \dots, n. \tag{7.50}$$

Likewise, the predictions $\hat{\mathbf{x}}[n|n-1]$ satisfy

$$E\left[(\hat{\mathbf{x}}[n|n-1]-\mathbf{x}[n])\,\mathbf{y}^{\mathrm{T}}[k]\right]=\mathbf{0}, \qquad k=0,1,\ldots,n-1. \tag{7.51}$$

Our development is again facilitated by working with the innovations process

$$\mathbf{z}[n]=\mathbf{y}[n]-\hat{\mathbf{y}}[n|n-1], \tag{7.52}$$

which is causally equivalent to $\mathbf{y}[n]$ and satisfies the whiteness condition

$$\mathbf{\Lambda}_{\mathbf{zz}}[n,k]=\mathbf{\Lambda}_{\mathbf{z}}\,\delta[n-k] \tag{7.53}$$

because $\mathbf{z}[n]$ is orthogonal to $\mathbf{y}[0],\mathbf{y}[1],\ldots,\mathbf{y}[n-1]$ and, due to the causal equivalence, also to $\mathbf{z}[0],\mathbf{z}[1],\ldots,\mathbf{z}[n-1]$.

Expressing our estimates in terms of the innovations we obtain our update equation for the filtered estimates. In particular, due to (7.53) we can write

$$\begin{aligned}\hat{\mathbf{x}}[n|n]&=\sum_{k=0}^{n}\mathbf{K}_{xz}[n,k]\,\mathbf{K}_{zz}^{-1}[k,k]\,\mathbf{z}[k]\\&=\hat{\mathbf{x}}[n|n-1]+\mathbf{G}[n]\,\mathbf{z}[n]\end{aligned} \tag{7.54}$$

where

$$\mathbf{G}[n]=\mathbf{K}_{xz}[n,n]\,\mathbf{K}_{zz}^{-1}[n,n] \tag{7.55}$$

and

$$\hat{\mathbf{x}}[n|n-1]=\sum_{k=0}^{n-1}\mathbf{K}_{xz}[n,k]\,\mathbf{K}_{zz}^{-1}[k,k]\,\mathbf{z}[k]. \tag{7.56}$$

Likewise, the filtered error covariance satisfies

$$\begin{aligned}\mathbf{\Lambda}_{\mathbf{e}}[n|n]&=\mathbf{\Lambda}_{\mathbf{x}}[n]-\sum_{k=0}^{n}\mathbf{K}_{\mathbf{xz}}[n,k]\,\mathbf{K}_{\mathbf{zz}}^{-1}[k,k]\,\mathbf{K}_{\mathbf{xz}}^{\mathrm{T}}[n,k]\\&=\mathbf{\Lambda}_{\mathbf{e}}[n|n-1]-\mathbf{G}[n]\,\mathbf{K}_{\mathbf{xz}}^{\mathrm{T}}[n,n]\end{aligned} \tag{7.57}$$

where we have used that

$$\mathbf{\Lambda}_{\mathbf{e}}[n|n-1]=\mathbf{\Lambda}_{\mathbf{x}}[n]-\sum_{k=0}^{n-1}\mathbf{K}_{\mathbf{xz}}[n,k]\,\mathbf{K}_{\mathbf{zz}}^{-1}[k,k]\,\mathbf{K}_{\mathbf{xz}}^{\mathrm{T}}[n,k]. \tag{7.58}$$

Next, using the linearity of the estimation in conjunction with the dynamics (7.42a) and the fact that $\mathbf{v}[n]$ is uncorrelated with both $\mathbf{x}[0]$ and $\mathbf{y}[0],\mathbf{y}[1],\ldots,\mathbf{y}[n]$ we obtain the prediction equation

$$\hat{\mathbf{x}}[n+1|n]=\mathbf{A}[n]\,\hat{\mathbf{x}}[n|n]. \tag{7.59}$$

Then since

$$\begin{aligned}\mathbf{e}[n+1|n]&=(\mathbf{x}[n+1]-\hat{\mathbf{x}}[n+1|n])\\&=\mathbf{A}[n]\;(\mathbf{x}[n]-\hat{\mathbf{x}}[n|n])+\mathbf{B}[n]\,\mathbf{v}[n]\\&=\mathbf{A}[n]\,\mathbf{e}[n|n]+\mathbf{B}[n]\,\mathbf{v}[n]\end{aligned} \tag{7.60}$$

we obtain

$$\mathbf{\Lambda}_{\mathbf{e}}[n+1|n] = \mathbf{A}[n]\,\mathbf{\Lambda}_{\mathbf{e}}[n|n]\,\mathbf{A}^{\mathrm{T}}[n] + \mathbf{B}[n]\,\mathbf{\Lambda}_{\mathbf{v}}\,\mathbf{B}^{\mathrm{T}}[n]. \tag{7.61}$$

Finally, we exploit (7.52) and (7.42b) to obtain expressions for the gain matrix $\mathbf{G}[n]$. In particular, using (7.42b) and the fact that $\mathbf{w}[n]$ is uncorrelated with $\mathbf{y}[0], \mathbf{y}[1], \ldots, \mathbf{y}[n-1]$ we obtain

$$\hat{\mathbf{y}}[n|n-1] = \mathbf{C}[n]\,\hat{\mathbf{x}}[n|n-1] \tag{7.62}$$

which when combined with (7.52) yields

$$\mathbf{z}[n] = \mathbf{C}[n]\,\mathbf{e}[n|n-1] + \mathbf{w}[n]. \tag{7.63}$$

Hence,

$$\mathbf{K}_{\mathbf{xz}}[n,n] = E\left[\mathbf{x}[n]\,\mathbf{z}^{\mathrm{T}}[n]\right] = \mathbf{\Lambda}_{\mathbf{e}}[n|n-1]\,\mathbf{C}^{\mathrm{T}}[n] \tag{7.64}$$

and

$$\begin{aligned}\mathbf{K}_{\mathbf{zz}}[n,n] &= E\left[\mathbf{z}[n]\,\mathbf{z}^{\mathrm{T}}[n]\right] \\ &= \mathbf{C}[n]\,\mathbf{\Lambda}_{\mathbf{e}}[n|n-1]\,\mathbf{C}^{\mathrm{T}}[n] + \mathbf{\Lambda}_{\mathbf{w}}[n],\end{aligned} \tag{7.65}$$

so that the Kalman gain (7.55) becomes

$$\mathbf{G}[n] = \mathbf{\Lambda}_{\mathbf{e}}[n|n-1]\,\mathbf{C}^{\mathrm{T}}[n]\left[\mathbf{C}[n]\,\mathbf{\Lambda}_{\mathbf{e}}[n|n-1]\,\mathbf{C}^{\mathrm{T}}[n] + \mathbf{\Lambda}_{\mathbf{w}}[n]\right]^{-1}. \tag{7.66}$$

Summary

Collecting together the equations we derived, we obtain the following rather general vector version of the recursive Kalman filtering algorithm when $\mathbf{x}[n]$ and $\mathbf{y}[n]$ are related according to (7.42):

1. Initialize the prediction and its associated error variance according to

$$\hat{\mathbf{x}}[0|-1] = \mathbf{0} \tag{7.67a}$$

$$\mathbf{\Lambda}_{\mathbf{e}}[0|-1] = \mathbf{\Lambda}, \tag{7.67b}$$

and let $n = 0$.

2. Compute the Kalman gain matrix

$$\mathbf{G}[n] = \mathbf{\Lambda}_{\mathbf{e}}[n|n-1]\,\mathbf{C}^{\mathrm{T}}[n]\left(\mathbf{C}[n]\,\mathbf{\Lambda}_{\mathbf{e}}[n|n-1]\,\mathbf{C}^{\mathrm{T}}[n] + \mathbf{\Lambda}_{\mathbf{w}}[n]\right)^{-1}. \tag{7.67c}$$

and generate the filtered estimate and its associated error covariance from the corresponding prediction quantities according to

$$\hat{\mathbf{x}}[n|n] = \hat{\mathbf{x}}[n|n-1] + \mathbf{G}[n]\ (\mathbf{y}[n] - \mathbf{C}[n]\,\hat{\mathbf{x}}[n|n-1]) \tag{7.67d}$$

$$\mathbf{\Lambda}_{\mathbf{e}}[n|n] = \mathbf{\Lambda}_{\mathbf{e}}[n|n-1] - \mathbf{G}[n]\,\mathbf{C}[n]\,\mathbf{\Lambda}_{\mathbf{e}}[n|n-1]. \tag{7.67e}$$

3. Generate the next prediction and its associated error covariance from the corresponding filtered quantities according to

$$\hat{\mathbf{x}}[n+1|n] = \mathbf{A}[n]\,\hat{\mathbf{x}}[n|n] \tag{7.67f}$$

$$\mathbf{\Lambda}_{\mathbf{e}}[n+1|n] = \mathbf{A}[n]\,\mathbf{\Lambda}_{\mathbf{e}}[n|n]\mathbf{A}^{\mathrm{T}}[n] + \mathbf{B}[n]\,\mathbf{\Lambda}_{\mathbf{v}}[n]\,\mathbf{B}^{\mathrm{T}}[n]. \tag{7.67g}$$

4. Increment n and go to step 2.

Again, the Kalman gain $\mathbf{G}[n]$ and error covariance matrices can be precomputed and stored in memory, reducing further the number of real-time computations required to obtain each sample of the estimated process. As a final comment, when necessary it is straightforward to accommodate nonzero means in the Kalman algorithm.