5

LINEAR OPTIMAL FILTERS AND PREDICTORS

Prediction is difficult—especially of the future.
—attributed to Niels Henrik David Bohr (1885–1962)

5.1 CHAPTER FOCUS

5.1.1 Estimation Problem

This is the problem of estimating the state of a linear stochastic system by using measurements that are linear functions of the state.

We suppose that stochastic systems can be represented by the types of plant and measurement models (for continuous and discrete time) shown as Equations 5.1–5.6 in Table 5.1, with dimensions of the vector and matrix quantities as shown in Table 5.2. The symbols $\Delta(k-\ell)$ and $\delta(t-s)$ stand for the *Kronecker delta function* and the *Dirac delta function* (actually, a *generalized* function), respectively.

The measurement and plant noise v_k and w_k are assumed to be zero-mean random processes, and the initial value x_0 is a variate with known mean x_0 and known covariance matrix P_0. Although the noise sequences w_k and v_k are assumed to be uncorrelated, the derivation in Section 5.5 will remove this restriction and modify the estimator equations accordingly.

The objective will be to find an estimate of the n state vector x_k represented by $\hat{x}_k$, a linear function of the measurements $z_i, \ldots, z_k$, that minimizes the weighted mean-squared error

Kalman Filtering: Theory and Practice Using MATLAB®, Fourth Edition.
Mohinder S. Grewal and Angus P. Andrews.

TABLE 5.1 Linear Plant and Measurement Models

Model	Continuous Time	Discrete Time	Equation Number
Plant	$\dot{x}(t) = F(t)x(t) + w(t)$	$x_k = \Phi_{k-1}x_{k-1} + w_{k-1}$	5.1
Measurement	$z(t) = H(t)x(t) + v(t)$	$z_k = H_k x_k + v_k$	5.2
Plant noise	$\mathrm{E}\langle w(t)\rangle = 0$	$\mathrm{E}\langle w_k\rangle = 0$	5.3
	$\mathrm{E}\langle w(t)w^{\mathrm{T}}(s)\rangle = \delta(t-s)Q(t)$	$\mathrm{E}\langle w_k w_i^{\mathrm{T}}\rangle = \Delta(k-i)Q_k$	5.5
Observation noise	$\mathrm{E}\langle v(t)\rangle = 0$	$\mathrm{E}\langle v_k\rangle = 0$	5.5
	$\mathrm{E}\langle v(t)v^{\mathrm{T}}(s)\rangle = \delta(t-s)R(t)$	$\mathrm{E}\langle v_k v_i^{\mathrm{T}}\rangle = \Delta(k-i)R_k$	5.6

TABLE 5.2 Dimensions of Arrays in Linear Model

Symbol	Dimensions	Symbol	Dimensions
x, w	$n \times 1$	Φ, Q	$n \times n$
z, v	$\ell \times 1$	H	$\ell \times n$
R	$\ell \times \ell$	Δ, δ	Scalar

$$\mathrm{E}\langle [x_k - \hat{x}_k]^{\mathrm{T}} M [x_k - \hat{x}_k]\rangle,$$

where M is any *symmetric nonnegative-definite weighting matrix*.

5.1.2 Main Points to Be Covered

5.1.2.1 Linear Least-Mean-Squared Estimation Problem We are now prepared to derive the mathematical forms of optimal linear estimators for the states of linear stochastic systems defined in the previous chapters. This is called the *linear quadratic* (LQ) estimation problem. The dynamic systems are linear and the performance cost functions are quadratic (least-mean-squared estimation error). We have already seen in Chapter 3 that the solution does not depend on the random processes being Gaussian.

5.1.2.2 Filtering, Prediction, and Smoothing There are three general types of estimators for the LQ problem:

- *Predictors* use observations *strictly prior* to the time that the state of the dynamic system is to be estimated:

$$t_{\mathrm{obs}} < t_{\mathrm{est}}.$$

- *Filters* use observations *up to and including* the time that the state of the dynamic system is to be estimated:

$$t_{\text{obs}} \leq t_{\text{est}}.$$

- *Smoothers* use observations *beyond* the time that the state of the dynamic system is to be estimated:

$$t_{\text{obs}} > t_{\text{est}}.$$

5.1.2.3 Kalman gain A straightforward and simple approach using the orthogonality principle is used in the derivation[1] of *estimators*. These estimators will have *minimum variance* and be *unbiased* and *consistent*.

Two easier derivations of the Kalman gain formula are also provided, one starting with the Gaussian maximum-likelihood (ML) estimator and the other starting with the linear least-mean-square (LMS) estimator.

5.1.2.4 Unbiased Estimators The Kalman filter can be characterized as an algorithm for computing the conditional mean and covariance of the probability distribution of the state of a linear stochastic system with uncorrelated random process and measurement noise. The conditional mean is the unique *unbiased* estimate. It is propagated in feedback form by a system of linear differential equations or by the corresponding discrete-time equations. The conditional covariance is propagated by a nonlinear differential equation or its discrete-time equivalent. This implementation automatically minimizes the expected risk associated with any quadratic loss function of the estimation error.

5.1.2.5 Performance Properties of Optimal Estimators The statistical performance of the estimator can be predicted a priori (i.e., before it is actually used) by solving the nonlinear differential (or difference) equations used in computing the optimal feedback gains of the estimator. These are called *Riccati equations*,[2] and the behavior of their solutions can be shown analytically in the most trivial cases. These equations also provide a means for verifying the proper performance of the actual estimator when it is running.

[1]For more mathematically oriented derivations, consult any of the references such as Anderson and Moore [1], Bozic [2], Brammer and Siffling [3], Brown [4], Bryson and Ho [5], Bucy and Joseph [6], Catlin [7], Chui and Chen [8], Gelb et al. [9], Jazwinski [10], Kailath [11], Maybeck [12, 13], Mendel [14, 15], Nahi [16], Ruymgaart and Soong [17], and Sorenson [18].

[2]Named in 1763 by Jean le Rond D'Alembert (1717–1783) for Count Jacopo Francesco Riccati (1676–1754), who had studied a second-order scalar differential equation [19], although not the form that we have here [20, 21]. Kalman gives credit to Richard S. Bucy for showing him that the Riccati differential equation is analogous to spectral factorization for defining optimal gains. The Riccati equation also arises naturally in the problem of separation of variables in ordinary differential equations and in the transformation of two-point boundary-value problems to initial-value problems [22].

5.2 KALMAN FILTER

5.2.1 Observational Update Problem for System State Estimator

Suppose that a measurement has been made at time t_k and that the information it provides is to be applied in updating the estimate of the state x of a stochastic system at time t_k. It is assumed that the measurement is linearly related to the state by an equation of the form $z_k = Hx_k + v_k$, where H is the *measurement sensitivity matrix* and v_k is the *measurement noise*.

5.2.2 Estimator in Linear Form

The optimal linear estimate is equivalent to the general (nonlinear) optimal estimator if the variates x and z are jointly Gaussian (see Section 5.8.1). Therefore, it suffices to seek an updated estimate $\hat{x}_{k(+)}$—based on the observation z_k—that is a *linear* function of the a priori estimate and the measurement z:

$$\hat{x}_{k(+)} = K_k^1 \hat{x}_{k(-)} + \overline{K}_k z_k, \tag{5.7}$$

where $\hat{x}_{k(-)}$ is the a priori estimate of x_k and $\hat{x}_{k(+)}$ is the a posteriori value of the estimate.

5.2.3 Solving for the Kalman Gain

The matrices K_k^1 and $\overline{K}_k$ are as yet unknown. We seek those values of K_k^1 and $\overline{K}_k$ such that the new estimate $\hat{x}_{k(+)}$ will satisfy the orthogonality principle of Section 4.8.2. This orthogonality condition can be written in the form

$$\mathrm{E}\langle [x_k - \hat{x}_{k(+)}] z_i^{\mathrm{T}} \rangle = 0, \quad i = 1, 2, \ldots, k-1, \tag{5.8}$$

$$\mathrm{E}\langle [x_k - \hat{x}_{k(+)}] z_k^{\mathrm{T}} \rangle = 0. \tag{5.9}$$

If one substitutes the formula for x_k from Equation 5.1 (in Table 5.1) and for $\hat{x}_{k(+)}$ from Equation 5.7 into Equation 5.8, then one will observe from Equations 5.1 and 5.2 that the data $z_1, \ldots, z_k$ do not involve the noise term w_k. Therefore, because the random sequences w_k and v_k are uncorrelated, it follows that $\mathrm{E}\langle w_k z_i^{\mathrm{T}} \rangle = 0$ for $1 \le i \le k$. (See Problem 5.6.)

Using this result, one can obtain the following relation:

$$\mathrm{E}\langle [\Phi_{k-1} x_{k-1} + w_{k-1} - K_k^1 \hat{x}_{k(-)} - \overline{K}_k z_k] z_i^{\mathrm{T}} \rangle = 0, \quad i = 1, \ldots, k-1. \tag{5.10}$$

But because $z_k = H_k x_k + v_k$, Equation 5.10 can be rewritten as

$$\mathrm{E}\langle [\Phi_{k-1} x_{k-1} - K_k^1 \hat{x}_{k(-)} - \overline{K}_k H_k x_k - \overline{K}_k v_k] z_i^{\mathrm{T}} \rangle = 0, \quad i = 1, \ldots, k-1. \tag{5.11}$$

We also know that Equations 5.8 and 5.9 hold at the previous step, that is,

$$\mathrm{E}\langle[x_{k-1} - \hat{x}_{(k-1)(+)}]z_i^{\mathrm{T}}\rangle = 0, \quad i = 1, \ldots, k-1,$$

and

$$\mathrm{E}\langle v_k z_i^{\mathrm{T}}\rangle, = 0, \quad i = 1, \ldots, k-1.$$

Then Equation 5.11 can be reduced to the form

$$\begin{aligned}
\Phi_{k-1}\mathrm{E}\langle x_{k-1}z_i^{\mathrm{T}}\rangle - K_k^1\mathrm{E}\langle\hat{x}_{k(-)}z_i^{\mathrm{T}}\rangle - \overline{K}_k H_k\Phi_{k-1}\mathrm{E}\langle x_{k-1}z_i^{\mathrm{T}}\rangle - \overline{K}_k\mathrm{E}\langle v_k z_i^{\mathrm{T}}\rangle &= 0,\\
\Phi_{k-1}\mathrm{E}\langle x_{k-1}z_i^{\mathrm{T}}\rangle - K_k^1\mathrm{E}\langle\hat{x}_{k(-)}z_i^{\mathrm{T}}\rangle - \overline{K}_k H_k\Phi_{k-1}\mathrm{E}\langle x_{k-1}z_i^{\mathrm{T}}\rangle &= 0,\\
\mathrm{E}\langle[x_k - \overline{K}_k H_k x_k - K_k^1 x_k] - K_k^1(\hat{x}_{k(-)} - x_k)z_i^{\mathrm{T}}\rangle &= 0,\\
[I - K_k^1 - \overline{K}_k H_k]\mathrm{E}\langle x_k z_i^{\mathrm{T}}\rangle &= 0. \qquad (5.12)
\end{aligned}$$

Equation 5.12 can be satisfied for any given x_k if

$$K_k^1 = I - \overline{K}_k H_k. \qquad (5.13)$$

Clearly, this choice of K_k^1 causes Equation 5.7 to satisfy a portion of the condition given by Equation 5.8, which was derived in Section 5.8. The choice of $\overline{K}_k$ is such that Equation 5.9 is satisfied.

Let the errors

$$\tilde{x}_{k(+)} \stackrel{\Delta}{=} \hat{x}_{k(+)} - x_k, \qquad (5.14)$$

$$\tilde{x}_{k(-)} \stackrel{\Delta}{=} \hat{x}_{k(-)} - x_k, \qquad (5.15)$$

$$\begin{aligned}
\tilde{z}_k &\stackrel{\Delta}{=} \hat{z}_{k(-)} - z_k\\
&= H_k\hat{x}_{k(-)} - z_k. \qquad (5.16)
\end{aligned}$$

Vectors $\tilde{x}_{k(+)}$ and $\tilde{x}_{k(-)}$ are the estimation errors after and before updates, respectively.[3]

The parameter $\hat{x}_k$ depends linearly on x_k, which depends linearly on z_k. Therefore, from Equation 5.9,

$$\mathrm{E}\langle[x_k - \hat{x}_{k(+)}]\tilde{z}_{k(-)}^{\mathrm{T}}\rangle = 0 \qquad (5.17)$$

and also (by subtracting Equation 5.9 from Equation 5.17)

$$\mathrm{E}\langle[x_k - \hat{x}_{k(+)}]\tilde{z}_k^{\mathrm{T}}\rangle = 0. \qquad (5.18)$$

[3]The symbol ~ is officially called a *tilde* but often called a *squiggle*.

Substitute for $x_k, \hat{x}_{k(+)}$ and $\tilde{z}_k$ from Equations 5.1, 5.7, and 5.16, respectively. Then

$$\mathrm{E}\langle[\Phi_{k-1}x_{k-1} + w_{k-1} - K^1_{k(-)} - \overline{K}_k z_k][H_k\hat{x}_{k(-)} - z_k]^{\mathrm{T}}\rangle = 0.$$

However, by the system structure

$$\mathrm{E}\langle w_k z_k^{\mathrm{T}}\rangle = \mathrm{E}\langle w_k \hat{x}^{\mathrm{T}}_{k(+)}\rangle = 0,$$
$$\mathrm{E}\langle[\Phi_{k-1}x_{k-1} - K^1_k\hat{x}_{k(-)} - \overline{K}_k z_k][H_k\hat{x}_{k(-)} - z_k]^{\mathrm{T}}\rangle = 0.$$

Substituting for K^1_k, z_k, and $\tilde{x}_{k(-)}$ and using the fact that $\mathrm{E}\tilde{x}_{k(-)}v_k^{\mathrm{T}} = 0$, this last result can be modified as follows:

$$\begin{aligned}
0 &= \mathrm{E}\langle[\Phi_{k-1}x_{k-1} - \hat{x}_{k(-)} + \overline{K}_k H_k\hat{x}_{k(-)} - \overline{K}_k H_k x_k - \overline{K}_k v_k] \\
&\qquad [H_k\hat{x}_{k(-)} - H_k x_k - v_k]^{\mathrm{T}}\rangle \\
&= \mathrm{E}\langle[(x_k - \hat{x}_{k(-)}) - \overline{K}_k H_k(x_k - \hat{x}_{k(-)}) - \overline{K}_k v_k][H_k\tilde{x}_{k(-)} - v_k]^{\mathrm{T}}\rangle \\
&= \mathrm{E}\langle[(-\tilde{x}_{k(-)} + \overline{K}_k H_k\tilde{x}_{k(-)} - \overline{K}_k v_k][H_k\tilde{x}_{k(-)} - v_k]^{\mathrm{T}}\rangle.
\end{aligned}$$

By definition, the a priori covariance (the error covariance matrix before the update) is

$$P_{k(-)} = \mathrm{E}\langle\tilde{x}_{k(-)}\tilde{x}^{\mathrm{T}}_{k(-)}\rangle.$$

It satisfies the equation

$$[I - \overline{K}_k H_k]P_{k(-)}H_k^{\mathrm{T}} - \overline{K}_k R_k = 0,$$

and, therefore, the Kalman gain can be expressed as

$$\overline{K}_k = P_{k(-)}H_k^{\mathrm{T}}[H_k P_{k(-)}H_k^{\mathrm{T}} + R_k]^{-1}, \tag{5.19}$$

which is the solution we seek for the gain as a function of the a priori covariance.

One can derive a similar formula for the a posteriori covariance (the error covariance matrix after update), which is defined as

$$P_{k(+)} = \mathrm{E}\langle[\tilde{x}_{k(+)}\tilde{x}^{\mathrm{T}}_{k(+)}]\rangle. \tag{5.20}$$

By substituting Equation 5.13 into Equation 5.7, one obtains the equations

$$\begin{aligned}
\hat{x}_{k(+)} &= (I - \overline{K}_k H_k)\hat{x}_{k(-)} + \overline{K}_k z_k, \\
\hat{x}_{k(+)} &= \hat{x}_{k(-)} + \overline{K}_k[z_k - H_k\hat{x}_{k(-)}].
\end{aligned} \tag{5.21}$$

Subtract x_k from both sides of the latter equation to obtain the equations

$$\begin{aligned}\hat{x}_{k(+)} - x_k &= \hat{x}_{k(-)} + \overline{K}_k H_k x_k + \overline{K}_k v_k - \overline{K}_k H_k \hat{x}_{k(-)} - x_k,\\ \tilde{x}_{k(+)} &= \tilde{x}_{k(-)} - \overline{K}_k H_k \tilde{x}_{k(-)} + \overline{K}_k v_k,\\ \tilde{x}_{k(+)} &= (I - \overline{K}_k H_k)\tilde{x}_{k(-)} + \overline{K}_k v_k. \end{aligned} \tag{5.22}$$

By substituting Equation 5.22 into Equation 5.20 and noting that $\mathrm{E}\langle \tilde{x}_{k(-)} v_k^{\mathrm{T}} \rangle = 0$, one obtains

$$\begin{aligned}P_{k(+)} &= \mathrm{E}\langle [I - \overline{K}_k H_k]\tilde{x}_{k(-)}\tilde{x}_{k(-)}^{\mathrm{T}}[I - \overline{K}_k H_k]^{\mathrm{T}} + \overline{K}_k v_k v_k^{\mathrm{T}} \overline{K}_k^{\mathrm{T}} \rangle\\ &= (I - \overline{K}_k H_k) P_{k(-)} (I - \overline{K}_k H_k)^{\mathrm{T}} + \overline{K}_k R_k \overline{K}_k^{\mathrm{T}}. \end{aligned} \tag{5.23}$$

This last equation is the so-called "Joseph form" of the covariance update equation derived by Bucy and Joseph [6]. By substituting for $\overline{K}_k$ from Equation 5.19, it can be put into the following forms:

$$\begin{aligned}P_{k(+)} &= P_{k(-)} - \overline{K}_k H_k P_{k(-)}\\ &\quad - P_{k(-)} H_k^{\mathrm{T}} \overline{K}_k^{\mathrm{T}} + \overline{K}_k H_k P_{k(-)} H_k^{\mathrm{T}} \overline{K}_k^{\mathrm{T}} + \overline{K}_k R_k \overline{K}_k^{\mathrm{T}}\\ &= (I - \overline{K}_k H_k) P_{k(-)} - P_{k(-)} H_k^{\mathrm{T}} \overline{K}_k^{\mathrm{T}}\\ &\quad + \underbrace{\overline{K}_k (H_k P_{k(-)} H_k^{\mathrm{T}} + R_k)}_{P_{k(-)} H_k^{\mathrm{T}}} \overline{K}_k^{\mathrm{T}}\\ &= (I - \overline{K}_k H_k) P_{k(-)}, \end{aligned} \tag{5.24}$$

the last of which is the one most often used in computation. This implements the effect that *conditioning on the measurement* has on the covariance matrix of estimation uncertainty.

Error covariance extrapolation models the effects of time on the covariance matrix of estimation uncertainty, which is reflected in the a priori values of the covariance and state estimates,

$$\begin{aligned}P_{k(-)} &= \mathrm{E}\langle \tilde{x}_{k(-)} \tilde{x}_{k(-)}^{\mathrm{T}} \rangle,\\ \hat{x}_{k(-)} &= \Phi_{k-1} \hat{x}_{k-1(+)}, \end{aligned} \tag{5.25}$$

respectively. Subtract x_k from both sides of the last equation to obtain the equations

$$\begin{aligned}\hat{x}_{k(-)} - x_k &= \Phi_{k-1}\hat{x}_{k-1(+)} - x_k,\\ \tilde{x}_{k(-)} &= \Phi_{k-1}[\hat{x}_{k-1(+)} - x_{k-1}] - w_{k-1}\\ &= \Phi_{k-1}\tilde{x}_{k-1(+)} - w_{k-1} \end{aligned}$$

for the propagation of the estimation error, $\tilde{x}$. Postmultiply it by $\tilde{x}_k^{\mathrm{T}}(-)$ (on both sides of the equation) and take the expected values. Use the fact that $\mathrm{E}\tilde{x}_{k-1}\ w_{k-1}^{\mathrm{T}} = 0$ to obtain the results

$$\begin{aligned} P_{k(-)} &\stackrel{\text{def}}{=} \mathrm{E}\langle \tilde{x}_{k(-)}\tilde{x}_{k(-)}^{\mathrm{T}}\rangle \\ &= \Phi_{k-1}\mathrm{E}\langle \tilde{x}_{k-1(+)}\tilde{x}_{k-1(+)}^{\mathrm{T}}\rangle\Phi_{k-1}^{\mathrm{T}} + \mathrm{E}\langle w_{k-1}\ w_{k-1}^{\mathrm{T}}\rangle \\ &= \Phi_{k-1}P_{k-1}^{(+)}\Phi_{k-1}^{\mathrm{T}} + Q_{k-1}, \end{aligned} \tag{5.26}$$

which gives the a priori value of the covariance matrix of estimation uncertainty as a function of the previous a posteriori value.

5.2.4 Kalman Gain from Gaussian Maximum Likelihood

The original derivation of the Kalman gain makes the fewest possible assumptions but requires the most mathematical rigor to obtain the most general result. The essential elements of that proof have been covered in the previous sections and Chapter 4.

An alternative derivation using the *linear Gaussian maximum-likelihood estimator* (*LGMLE* or *GMLE*) makes much more restrictive assumptions about the distribution of the state vector and measurements, but the resulting formula for the Kalman gain is the same. This derivation was introduced after some instructors observed symptoms resembling post traumatic stress disorder among students struggling with the more rigorous derivation.

In essence, this approach uses the mean μ_x and the information matrix $Y_{xx} = P_{xx}^{-1}$ as parameters for Gaussian distributions, then dismisses the Gaussian normalization factors to allow Y_{xx} to represent measurements that would render it singular.

The resulting functions are no longer probability functions, because their integrals are no longer 1 (one) and are not necessarily finite. They are called *Gaussian likelihood functions*, and they have properties similar to those of Gaussian probability distributions for joint and independent likelihoods. However, because integrals of likelihood functions are not necessarily defined, *expectation* can no longer be used to characterize likelihoods the way it characterizes probability measures. Least-mean-squared estimation error is not defined for likelihood functions, so it is replaced by *maximum likelihood* as a criterion for optimal estimation.

The means and information matrices of Gaussian likelihood functions are the parameters used in deriving the Kalman gain.

Example 5.1 (Combining Independent Gaussian Likelihoods) Consider the two-dimensional Gaussian likelihood function with a singular information matrix,

$$Y_a = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix},$$

in which case the 2D Gaussian likelihood function will have a shape like that shown in Figure 5.1(a), with the direction of "no information" shown by the double-ended arrow indicating the direction of the zero eigenvector of Y_{xx}.

This is a situation that could not be represented by a Gaussian probability density function, because its integral is not finite. The likelihood function, on the other hand, can represent the fact that there is no information in one direction. In essence, the eigenvalues of the information matrix represent *how much* information is available in the direction of the corresponding eigenvectors.

This also illustrates a case in which there is no unique location of the maximum of the likelihood function. It achieves its maximum along an entire infinite line.

The 2D Gaussian likelihood shown in Figure 5.1(b) represents the situation with information matrix

$$Y_b = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix},$$

in which case the direction of the zero eigenvector is orthogonal to that shown in Figure 5.1(a).

If these two likelihood functions are *independent,* in the sense that their joint likelihood is the point-wise product of the individual likelihoods, then their joint likelihood will be as illustrated in Figure 5.1(c). This resembles a 2D Gaussian likelihood function—and indeed it is. The formula for the Kalman gain can be derived by solving for the mean μ_c and information matrix Y_c of the likelihood shown in Figure 5.1(c) as a function of the means and information matrices of the likelihoods shown in Figure 5.1(a) and (b).

This alternative derivation of the formula for the Kalman gain matrix $\overline{K}_k$ uses the analogies shown in Figure 5.2 between the variable parameters of Kalman filtering ($\hat{x}$, P_{xx}), Gaussian probability densities (μ_x, P_{xx}), and Gaussian likelihood functions (μ_x, Y_{xx}).

5.2.4.1 Gaussian Maximum-Likelihood Estimation The theoretical basis for ML estimation was given its present form by Ronald A. Fisher (1890–1962) in the period 1912–1930 [23]. This work could draw upon on a body of theory going back more than a century [24–27].

An early application was for finding parameters of a distribution, given samples [26]. For Gaussian distributions, this is equivalent to finding the mean and covariance of a distribution. When Gaussian ML is formulated as a recursive estimator, this is analogous to the Kalman filtering problem if one equates the mean with the estimate, the covariance with the mean-squared estimation error, and samples with measurements. This analogy also provides a simple and direct derivation of the Kalman gain, resulting in the same formula derived by Kalman [28].

Gaussian Likelihoods and Log-Likelihoods The *Gaussian probability density function* for $x \in \mathcal{N}(\mu_x, P_{xx})$ is

$$p(x, \mu_x, P_{xx}) = \frac{1}{\sqrt{2\pi \det P_{xx}}} \exp\left(-\frac{1}{2}(x-\mu_x)^{\mathrm{T}} P_{xx}^{-1}(x-\mu_x)\right), \qquad (5.27)$$

where μ_x is the mean of X and P_{xx} is the covariance (second central moment).

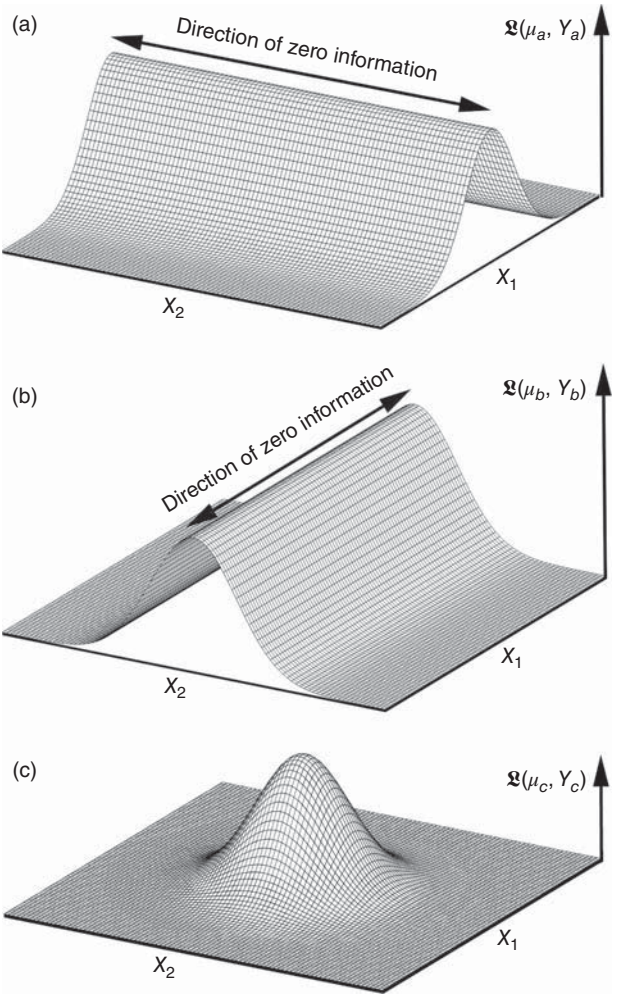

Figure 5.1 Products of Gaussian likelihoods.

The *Gaussian likelihood function* equivalent to Equation 5.27 would be of the same form

$$\mathcal{L}(x,\ \mu_x,\ Y_{xx}) = c\ \exp\ \left(-\frac{1}{2}(x-\mu_x)^{\mathrm{T}} Y_{xx}(x-\mu_x)\right), \tag{5.28}$$

except that is using the *information matrix* $Y_{xx} = P_{xx}^{-1}$ as a parameter. The scaling constant $c > 0$ plays no role in estimation, other than to allow such scaling to be ignored. The information matrix Y_{xx}, on the other hand, adds some modeling capabilities that are unavailable in probability modeling.

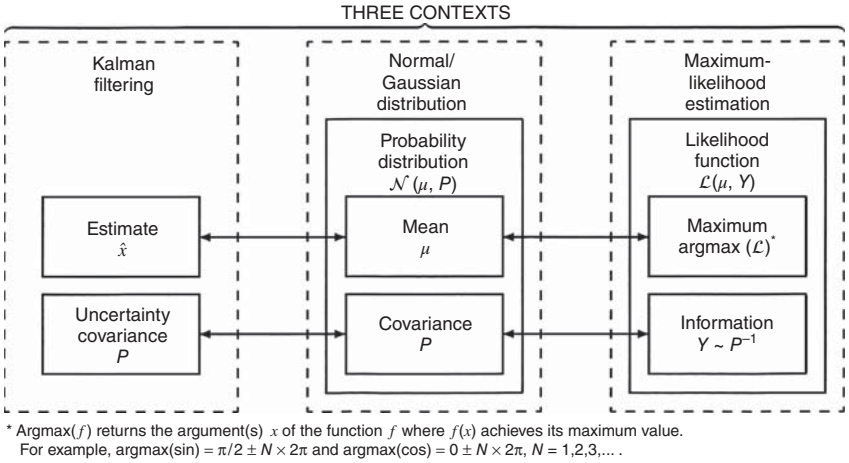

Figure 5.2 Analogous concepts in three different contexts.

Most of the useful work of Gaussian likelihoods is done using their logarithms, or *log-likelihoods*:

$$\log\ [\mathcal{L}(x,\ \mu_x,\ Y_{xx})] = \log(c) - \frac{1}{2}(x-\mu_x)^{\mathrm{T}} Y_{xx}(x-\mu_x). \tag{5.29}$$

Because the Gaussian likelihoods are always positive valued and the logarithm is a monotonically increasing function of positive numbers, maximizing the log-likelihood is equivalent to maximizing the likelihood.

5.2.4.2 Information Matrices All covariance matrices in Kalman filtering are symmetric and positive definite, because the variances of estimated quantities are never absolutely zero. Even fundamental constants of physics have associated error variances, and the value of π represented in finite-precision arithmetic has some roundoff error.

Consequently, all covariance matrices P_{xx} in Kalman filtering will have a matrix inverse $Y_{xx} = P_{xx}^{-1}$, the corresponding information matrix. In Kalman filtering, because P_{xx} is always symmetric and positive definite, the corresponding information matrix Y_{xx} will also be symmetric and positive definite. In fact, they have the same eigenvectors, and the corresponding eigenvalues of Y_{xx} will be the reciprocals of those of P_{xx}.

In ML estimation, however, the information matrices Y_{xx} are only symmetric and *nonnegative definite* (i.e., with zero eigenvalues possible) and, therefore, *not necessarily invertible.*

Using the information matrix in place of the covariance matrix in Gaussian likelihood functions allows us to model what estimation theorists would call "flat priors," a condition under which prior assumptions have no influence on the ultimate estimate. This cannot be done using covariance matrices, because it would require that some

eigenvalues be infinite. It can be done using information matrices by allowing them to have zero eigenvalues whose eigenvectors represent linear combinations of the state space in which there is zero information. For example, information matrices can be used to represent the information in a measurement, and the dimension of which may be less than the dimension of the state vector.

Example 5.2 (GNSS Pseudorange Measurement) Global navigation satellite navigation systems (GNSS) use accurate onboard clocks and precise time stamping of transmitted signals, so that receivers with accurate synchronized clocks can determine the propagation time delay, and from that an estimate of the distance the signal has traveled between the transmitting antenna and the receiving antenna. The estimate ρ is called a *pseudorange* because it also includes errors due to the receiver clock and atmospheric propagation delays.

Each pseudorange is then a one-dimensional measurement of the location of the receiver antenna relative to the satellite antenna, the location of which is known. If u is a unit vector in the direction from the transmitting antenna to the receiver antenna, then the partial derivative of the pseudorange measurement with respect to receiver antenna location x (a 3-vector) will be the 1×3 vector

$$\frac{\partial \rho}{\partial x} = u^{\mathrm{T}}.$$

The information that this measurement adds to the navigation solution (3D position) is represented by the 3×3 *information matrix*

$$Y_{xx} = uR^{-1}u^{\mathrm{T}},$$

where R is the mean-squared "sensor noise," the error in ρ due to signal processing noise, etc. This information matrix has rank 1, with one nonzero eigenvector u and two zero eigenvectors orthogonal to it.

Singular Value Decompositions of Information Matrices Singular value decomposition (svd) is used as a diagnostic tool for characterizing certain properties of matrices, and it offers insights into how information matrices behave.

The svd of an arbitrary $m \times n$ real matrix M is a factoring in the form $M = LDR$, where L is an $m \times m$ unitary matrix, R is an $n \times n$ unitary matrix, and D is an $m \times n$ matrix with nonnegative values down its main diagonal—with zeros elsewhere. The columns of L are called the *left eigenvectors* of M, the rows of R are called the *right eigenvectors,* and the main diagonal elements of D are called the *singular values* of M.

Because an information matrix Y_{xx} is symmetric and nonnegative definite, its svd will have the same left and right eigenvectors, which will be real vectors. In that

case, $L = R^{\mathrm{T}} = V$ and its svd can be expressed in terms of its eigenvalues λ_i (all nonnegative) and corresponding real eigenvectors e_i as

$$Y_{xx} = \sum_i \lambda_i e_i e_i^{\mathrm{T}} \tag{5.30}$$

$$= V \operatorname{diag}(\lambda) V^{\mathrm{T}} \tag{5.31}$$

$$V = \begin{bmatrix} e_1 & e_2 & e_3 & \cdots & e_n \end{bmatrix} \tag{5.32}$$

$$\operatorname{diag}(\lambda) = \begin{bmatrix} \lambda_1 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0 & \cdots & 0 \\ 0 & 0 & \lambda_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_n \end{bmatrix}, \tag{5.33}$$

The eigenvalues λ_i in the svds have the largest eigenvalue first, followed by the rest in order of decreasing value:

$$\lambda_1 \geq \lambda_2 \geq \lambda_2 \geq \cdots \geq \lambda_n.$$

Consequently, if Y_{xx} has rank $n - r$, the last r eigenvalues will be zero.

Example 5.3 (Singular Value Decompositions of 2×2 Information Matrices) The 2×2 information matrices used in Example 5.1 and their corresponding svds and eigenvalue–eigenvector decompositions can be summarized as

	Y	=	L	×	D	×	R	λ_1	e_1	λ_2	e_2
(a)	$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$	=	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$		$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$		$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	1	$\begin{bmatrix} 1 \\ 0 \end{bmatrix}$	0	$\begin{bmatrix} 0 \\ 1 \end{bmatrix}$
(b)	$\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$	=	$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$		$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$		$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$	1	$\begin{bmatrix} 0 \\ 1 \end{bmatrix}$	0	$\begin{bmatrix} 1 \\ 0 \end{bmatrix}$
(c)	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	=	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$		$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$		$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	1	$\begin{bmatrix} 1 \\ 0 \end{bmatrix}$	1	$\begin{bmatrix} 0 \\ 1 \end{bmatrix}$

Moore–Penrose generalized matrix inverse The *Moore–Penrose generalized inverse* of Y_{xx} can be defined in terms of its svd as

$$Y_{xx}^{\dagger} = \sum_{\lambda_i \neq 0} \lambda_i^{-1} e_i e_i^{\mathrm{T}}, \tag{5.34}$$

which is always symmetric and nonnegative definite and of the same rank as Y_{xx}.

5.2.4.3 Formulas for Joint Independent Likelihoods Two probability distributions are called *statistically independent* if and only if their joint probability is the product of the individual probabilities. The same is true for likelihoods.

Following the notation of Figure 5.1, let the joint likelihood function $\mathcal{L}_c(x,\ \mu_c,\ Y_c)$ of two independent Gaussian likelihoods $\mathcal{L}_a(x,\ \mu_a,\ Y_a)$ and $\mathcal{L}_b(x,\ \mu_b,\ Y_b)$ be represented by the product of the two likelihood functions:

$$c_c \exp\left(-\frac{1}{2}\left(x-\mu_c\right)^{\mathrm{T}} Y_c (x-\mu_c)\right)$$

$$= \mathcal{L}_c(x, \mu_c, Y_c) \tag{5.35}$$

$$= \mathcal{L}_a(x, \mu_a, Y_a) \times \mathcal{L}_b(x, \mu_b, Y_b) \tag{5.36}$$

$$= c_a \exp\left(-\frac{1}{2}\left(x-\mu_a\right)^{\mathrm{T}} Y_a (x-\mu_a)\right) \times c_b \exp\left(-\frac{1}{2}\left(x-\mu_b\right)^{\mathrm{T}} Y_b (x-\mu_b)\right) \tag{5.37}$$

$$= c_a c_b \exp\left(-\frac{1}{2}\left(x-\mu_a\right)^{\mathrm{T}} Y_a (x-\mu_a) - \frac{1}{2}(x-\mu_b)^{\mathrm{T}} Y_b (x-\mu_b)\right). \tag{5.38}$$

Taking the logarithm of both sides and differentiating once and twice with respect to x will yield the following sequence of equations:

$$\log\ (c_c) - \frac{1}{2}(x-\mu_c)^{\mathrm{T}} Y_c (x-\mu_c)$$

$$= \log\ (c_a) + \log\ (c_b) - \frac{1}{2}(x-\mu_a)^{\mathrm{T}} Y_a (x-\mu_a)$$

$$-\frac{1}{2}(x-\mu_b)^{\mathrm{T}} Y_b (x-\mu_b) \tag{5.39}$$

$$Y_c(x-\mu_c) = Y_a(x-\mu_a) + Y_b(x-\mu_b) \tag{5.40}$$

$$Y_c = Y_a + Y_b, \tag{5.41}$$

the last line of which says that *information is additive.* Setting $x = 0$ in the next-to-last line yields the equation

$$Y_c \mu_c = Y_a \mu_a + Y_b \mu_b. \tag{5.42}$$

Equations 5.41 and 5.42, with appropriate substitution of Kalman filtering variables, are all that is needed to solve for the Kalman gain.

5.2.4.4 Solving for the Kalman Gain

Substitutions The following substitutions will be made in Equations 5.41 and 5.42:

$$\left.\begin{array}{rcll} \mu_a & = & \hat{x}_{k(-)} & \text{a priori estimate} \\ Y_a & = & P_{k(-)}^{-1} & \text{a priori information} \\ \mu_b & = & H_k^{\dagger} z_k & \text{measurement mean} \\ Y_b & = & H_k^{\mathrm{T}} R_k^{-1} H_k & \text{measurement information} \\ \mu_c & = & \hat{x}_{k(+)} & \text{a posteriori estimate} \\ Y_c & = & P_{k(+)}^{-1} & \text{a posteriori information,} \end{array}\right\} \tag{5.43}$$

where $z_k = H_k \hat{x}_{k(-)} + v_k$ is the measurement, H_k is the measurement sensitivity matrix, v_k is the noise on the measurement, and R_k is the covariance of v_k.

Solving for the Covariance Update With the substitutions of Equation 5.43, Equation 5.41 becomes

$$P_{k(+)}^{-1} = P_{k(-)}^{-1} + H_k^{\mathrm{T}} R_k^{-1} H_k, \tag{5.44}$$

where $P_{k(-)}^{-1}$ is the a priori state information and $H_k^{\mathrm{T}} R_k^{-1} H_k = Y_b$ is the information in the kth measurement z_k.

An Inverse Matrix Modification Formula We can use the general formula for the inverse of a matrix sum due to Duncan [29] (among others)

$$(A^{-1} + BC^{-1}D)^{-1} = A - AB(C + DAB)^{-1}DA. \tag{5.45}$$

More Substitutions With the substitutions

$$\left.\begin{array}{rcll} A^{-1} & = & Y_a, & \text{the a priori information matrix for } \hat{x} \\ A & = & P_{k(-)}, & \text{the a priori covariance matrix for } \hat{x} \\ B & = & H_k^{\mathrm{T}}, & \text{transpose of the measurement sensitivity matrix} \\ C & = & R_k, & \text{covariance of measurement noise } v_k \\ D & = & H_k, & \text{the measurement sensitivity matrix,} \end{array}\right\} \tag{5.46}$$

Equation 5.45 becomes

$$\left.\begin{array}{rcll} P_{k(+)} & = & Y_c^{-1} & \\ & = & \left(Y_a + H_k^{\mathrm{T}} R_k^{-1} H_k\right)^{-1} & \text{(Eq. 5.44)} \\ & = & Y_a^{-1} - Y_a^{-1} H_k^{\mathrm{T}} (H_k Y_a^{-1} H_k^{\mathrm{T}} + R_k)^{-1} H_k Y_A^{-1} & \text{(Eq. 5.45)} \\ & = & P_{k(-)} - \underbrace{P_{k(-)} H_k{}^{\mathrm{T}} (H_k P_{k(-)} H_k{}^{\mathrm{T}} + R_k)^{-1}}_{\overline{K}_k} H_k P_{k(-)}, & \text{(Eq. 5.43)} \end{array}\right\} \tag{5.47}$$

where the expression labeled "$\overline{K}_k$" will also be found in the derivation of the update of the estimate in the following.

Solving for the Estimate Update Equation 5.42 with substitutions from Equation 5.43 will have the form

$$
\begin{aligned}
\hat{x}_{k(+)} &= \mu_c && \text{(Eq. 5.43)}\\
&= Y_c^{-1}(Y_a\mu_a + Y_b\mu_b) && \text{(Eq. 5.42)}\\
&= P_{k(+)}\left[P_{k(-)}^{-1}\hat{x}_{k(-)} + H_k^{\mathrm{T}}R_k^{-1}H_kH_k^{\dagger}z_k\right] && \text{(Eq. 5.43)}\\
&= [P_{k(-)} - P_{k(-)}H_k^{\mathrm{T}}(H_kP_{k(-)}H_k^{\mathrm{T}} + R_k)^{-1}H_kP_{k(-)}] && \text{(Eq. 5.47)}\\
&\quad \times [P_{k(-)}^{-1}\hat{x}_{k(-)} + H_k^{\mathrm{T}}R_k^{-1}H_kH_k^{\dagger}z]\\
&= [I - P_{k(-)}H_k^{\mathrm{T}}(H_kP_{k(-)}H_k^{\mathrm{T}} + R_k)^{-1}H_k]\\
&\quad \times [\hat{x}_{k(-)} + P_{k(-)}H_k^{\mathrm{T}}R_k^{-1}H_kH_k^{\dagger}z]\\
&= \hat{x}_{k(-)} + P_{k(-)}H_k^{\mathrm{T}}(H_kP_{k(-)}H_k^{\mathrm{T}} + R_k)^{-1}\\
&\quad \times \{[(H_kP_{k(-)}H_k^{\mathrm{T}} + R_k)R_k^{-1} - H_kP_{k(-)}\,H_k^{\mathrm{T}}R_k^{-1}]z_k - H_k\hat{x}_{k(-)}\}\\
&= \hat{x}_{k(-)} + P_{k(-)}H_k^{\mathrm{T}}(H_kP_{k(-)}H_k^{\mathrm{T}} + R_k)^{-1}\\
&\quad \times \{[H_kP_{k(-)}H_k^{\mathrm{T}}R_k{}^{-1} + I - H_kP_{k(-)}H_k^{\mathrm{T}}R_k{}^{-1}]z_k - H_k\hat{x}_{k(-)}\}\\
&= \hat{x}_{k(-)} + \underbrace{\{P_{k(-)}H_k^{\mathrm{T}}(H_kP_{k(-)}H_k^{\mathrm{T}} + R_k)^{-1}\}}_{\overline{K}_k}[z_k - H_k\hat{x}_{k(-)}].
\end{aligned}
$$

The Kalman Gain The last equation of the above can now be rewritten as two equations,

$$\hat{x}_{k(+)} = \hat{x}_{k(-)} + \overline{K}_k[z_k - H_k\hat{x}_{k(-)}] \tag{5.48}$$

$$\overline{K}_k = P_{k(-)}H_k{}^{\mathrm{T}}(H_kP_{k(-)}H_k^{\mathrm{T}} + R_k)^{-1}, \tag{5.49}$$

the first of which models the estimation update in feedback form, with the Kalman gain given by the second equation.

5.2.4.5 Other Argmax Estimators The linear Gaussian ML estimator was used above as an alternative derivation of the Kalman gain formula. In this case, the resulting estimate (argmax of the likelihood function) is the mode, mean, and median of the underlying Gaussian distribution.

There is an alternative class of estimators called *maximum a posteriori probability* (MAP) estimators which use Bayes Rule to compute the argmax of the a posteriori probability density function to select the value of the variable to be estimated at which its probability density is greatest (maximum mode). These estimators are applicable to a more general class of problems (including non-Gaussian and nonlinear) than the Kalman filter, but they tend to have computational complexities that would eliminate them from consideration for real-time practical implementations as filters. They are used for some nonlinear and nonreal-time applications, however. (See, e.g., Bain and Crisan [30] or Crassidis and Jenkins [31].)

5.2.5 Kalman Gain from Recursive Linear LMS Estimator

This is another low stress derivation of the Kalman gain, starting from the linear LMS estimator in recursive form. It makes fewer assumptions than the derivation from Gaussian ML estimation and uses just a bit of matrix arithmetic.

5.2.5.1 Linear Least Mean Squares If the error vector v in the least-squares problem has a known covariance R,

$$z = Hx + v \tag{5.50}$$

$$R \overset{\text{def}}{=} \underset{v}{\mathrm{E}}\langle vv^{\mathrm{T}}\rangle \tag{5.51}$$

with svd

$$R = U_R D_R U_R^{\mathrm{T}} \tag{5.52}$$

as a product of orthogonal (U) and diagonal (D) matrices, then its inverse

$$R^{-1} = U_R D_R^{-1} U_R^{\mathrm{T}} \tag{5.53}$$

has a symmetric matrix square root

$$S_R \overset{\text{def}}{=} U_R D_R^{-1/2} U_R^{\mathrm{T}} \tag{5.54}$$

$$S_R^2 = U_R D_R^{-1/2} \underbrace{U_R^{\mathrm{T}} U_R}_{I} D_R^{-1/2} U_R^{\mathrm{T}} \tag{5.55}$$

$$= U_R D_R^{-1/2} D_R^{-1/2} U_R^{\mathrm{T}} \tag{5.56}$$

$$= U_R D_R^{-1} U_R^{\mathrm{T}} \tag{5.57}$$

$$= R^{-1}. \tag{5.58}$$

Consequently, the rescaled least-squares problem

$$\underbrace{S_R z}_{z^\star} = \underbrace{S_R H}_{H^\star} x + \underbrace{S_R v}_{v^\star} \tag{5.59}$$

has error covariance

$$\begin{aligned}
R^\star &\stackrel{\text{def}}{=} \mathop{\mathrm{E}}_{v}\langle v^\star v^{\star\mathrm{T}}\rangle && (5.60)\\
&= \mathop{\mathrm{E}}_{v}\langle S_R v v^{\mathrm{T}} S_R^{\mathrm{T}}\rangle && (5.61)\\
&= S_R \mathop{\mathrm{E}}_{v}\langle v v^{\mathrm{T}}\rangle S_R^{\mathrm{T}} && (5.62)\\
&= (U_R D_R^{-1/2} U_R^{\mathrm{T}}) R (U_R D_R^{-1/2} U_R^{\mathrm{T}}) && (5.63)\\
&= (U_R D_R^{-1/2} U_R^{\mathrm{T}})(U_R D_R U_R^{\mathrm{T}})(U_R D_R^{-1/2} U_R^{\mathrm{T}}) && (5.64)\\
&= U_R D_R^{-1/2} \underbrace{U_R^{\mathrm{T}} U_R}_{I} D_R \underbrace{U_R^{\mathrm{T}} U_R}_{I} D_R^{-1/2} U_R^{\mathrm{T}} && (5.65)\\
&= U_R \underbrace{D_R^{-1/2} D_R D_R^{-1/2}}_{I} U_R^{\mathrm{T}} && (5.66)\\
&= U_R U_R^{\mathrm{T}} && (5.67)\\
&= I. && (5.68)
\end{aligned}$$

That is, the rescaled error covariance is an identity matrix, meaning that the individual scalar errors are uncorrelated and all have the same variance (1).

In that case, the rescaled least-squares problem (Equation 5.59)

$$S_R z = S_R H x + v^\star \tag{5.69}$$

has solution

$$\begin{aligned}
\hat{x} &= [H^{\star T} H^\star]^{-1} H^{\star T} z^\star && (5.70)\\
&= [(S_R H)^{\mathrm{T}} (S_R H)]^{-1} (S_R H)^{\mathrm{T}} S_R z && (5.71)\\
&= [(S_R H)^{\mathrm{T}} (S_R H)]^{-1} (S_R H)^{\mathrm{T}} S_R z && (5.72)\\
&= [H^{\mathrm{T}} S_R^{\mathrm{T}} S_R H]^{-1} H^{\mathrm{T}} S_R^{\mathrm{T}} S_R z && (5.73)\\
&= [H^{\mathrm{T}} R^{-1} H]^{-1} H^{\mathrm{T}} R^{-1} z, && (5.74)
\end{aligned}$$

provided that R and the *information matrix* (the analog of the gramian in least-squares estimation)

$$Y \stackrel{\text{def}}{=} H^{\mathrm{T}} R^{-1} H \tag{5.75}$$

are nonsingular, in which case

$$Y^{-1} = P, \tag{5.76}$$

the covariance of estimation uncertainty with all errors independent and having equal variance. That $\hat{x}$ is the LMS estimate.

5.2.5.2 Recursive LMS If the observation vector z can be partitioned into subvectors

$$\mathcal{Z}_k = \begin{bmatrix} z_1 \\ z_2 \\ z_3 \\ \vdots \\ z_k \end{bmatrix} \tag{5.77}$$

such that the associated additive noise subvectors are uncorrelated,

$$\underset{v}{\mathrm{E}}\langle v_i v_j^{\mathrm{T}} \rangle = \begin{cases} 0, & i \neq j \\ R_i, & i = j, \end{cases} \tag{5.78}$$

then the matrices H and R can be similarly partitioned as

$$\mathcal{H}_k = \begin{bmatrix} H_1 \\ H_2 \\ H_3 \\ \vdots \\ H_k \end{bmatrix} \text{ and } \mathcal{R}_k = \begin{bmatrix} R_1 & 0 & 0 & \cdots & 0 \\ 0 & R_2 & 0 & \cdots & 0 \\ 0 & 0 & R_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & R_k \end{bmatrix}$$

and the associated information and covariance matrices

$$Y_k = \sum_{\ell=1}^{k} H_\ell^{\mathrm{T}} R_\ell^{-1} H_\ell \tag{5.79}$$

$$= Y_{k-1} + H_k^{\mathrm{T}} R_k^{-1} H_k \tag{5.80}$$

$$P_k = Y_k^{-1} \tag{5.81}$$

$$= \{P_{k-1}^{-1} + H_k^{\mathrm{T}} R_k^{-1} H_k\}^{-1} \tag{5.82}$$

to which one can apply the general "modified matrix" inversion formula 5.45 with

$$A = P_{k-1}^{-1} \tag{5.83}$$

$$B = H_k^{\mathrm{T}} \tag{5.84}$$

$$C = -R_k \tag{5.85}$$

$$D = H_k \tag{5.86}$$

to obtain

$$P_k = P_{k-1} - \underbrace{P_{k-1}H_k^{\mathrm{T}}[R_k + H_k P_{k-1} H_k^{\mathrm{T}}]^{-1}}_{\overline{K}_k} H_k P_{k-1}, \tag{5.87}$$

the formula for covariance matrix measurement updates of the Kalman filter, and in which you might recognize the underbraced expression as our quarry:

$$\overline{K}_k = P_{k-1}H_k^{\mathrm{T}}[R_k + H_k P_{k-1} H_k^{\mathrm{T}}]^{-1}.$$

However, it remains to be proven that this value provides the optimal linear gain for recursive LMSs estimation.

Recursive LMS Estimate The corresponding recursive update of the estimate will be

$$\hat{x}_k = P_k \left\{ \sum_{\ell=1}^{k} H_\ell^{\mathrm{T}} R_\ell^{-1} z_\ell \right\} \tag{5.88}$$

$$= \{P_{k-1} - \overline{K}_k H_k P_{k-1}\} \left\{ \sum_{\ell=1}^{k-1} H_\ell^{\mathrm{T}} R_\ell^{-1} z_\ell + H_k^{\mathrm{T}} R_k^{-1} z_k \right\} \tag{5.89}$$

$$= \{I - \overline{K}_k H_k\} \left\{ P_{k-1} \sum_{\ell=1}^{k-1} H_\ell^{\mathrm{T}} R_\ell^{-1} z_\ell + P_{k-1} H_k^{\mathrm{T}} R_k^{-1} z_k \right\} \tag{5.90}$$

$$= \{I - \overline{K}_k H_k\} \{\hat{x}_{k-1} + P_{k-1} H_k^{\mathrm{T}} R_k^{-1} z_k\} \tag{5.91}$$

$$= x_{k-1} - \overline{K}_k H_k x_{k-1} + \{I - \overline{K}_k H_k\} P_{k-1} H_k^{\mathrm{T}} R_k^{-1} z_k \tag{5.92}$$

$$= x_{k-1} - \overline{K}_k H_k x_{k-1} + \{P_{k-1} H_k^{\mathrm{T}} R_k^{-1} - \overline{K}_k H_k P_{k-1} H_k^{\mathrm{T}} R_k^{-1}\} z_k \tag{5.93}$$

$$= x_{k-1} - \overline{K}_k H_k x_{k-1} + \mathcal{X}_k z_k, \tag{5.94}$$

where

$$\mathcal{X}_k \stackrel{\text{def}}{=} P_{k-1} H_k^{\mathrm{T}} \underbrace{\{I - [R_k + H_k P_{k-1} H_k^{\mathrm{T}}]^{-1} H_k P_{k-1} H_k^{\mathrm{T}}\} R_k^{-1}}_{\mathcal{Y}_k} \tag{5.95}$$

$$\mathcal{Y}_k \stackrel{\text{def}}{=} \{I - [R_k + H_k P_{k-1} H_k^{\mathrm{T}}]^{-1} H_k P_{k-1} H_k^{\mathrm{T}}\} R_k^{-1} \tag{5.96}$$

$$[R_k + H_k P_{k-1} H_k^{\mathrm{T}}]\mathcal{Y}_k = \{R_k + H_k P_{k-1} H_k^{\mathrm{T}} - H_k P_{k-1} H_k^{\mathrm{T}}\} R_k^{-1} \tag{5.97}$$

$$= \{R_k\} R_k^{-1} \tag{5.98}$$

$$= I \tag{5.99}$$

$$\mathcal{Y}_k = [R_k + H_k P_{k-1} H_k^{\mathrm{T}}]^{-1} \tag{5.100}$$

$$\mathcal{K}_k = P_{k-1}H_k^{\mathrm{T}}\mathcal{Y}_k \tag{5.101}$$

$$= P_{k-1}H_k^{\mathrm{T}}[R_k + H_k P_{k-1}H_k^{\mathrm{T}}]^{-1} \tag{5.102}$$

$$\stackrel{\text{def}}{=} \overline{K}_k, \tag{5.103}$$

so that Equation 5.94 becomes

$$\hat{x}_k = \hat{x}_{k-1} - \overline{K}_k H_k \hat{x}_{k-1} + \overline{K}_k z_k \tag{5.104}$$

$$= \hat{x}_{k-1} + \overline{K}_k(z_k - H_k\hat{x}_{k-1}). \tag{5.105}$$

That is, the formula for the recursive LMS estimate in Equation 5.105 is exactly the same as the Kalman measurement update formula, and with the same Kalman gain, as given by Equation 5.88.

This completes the derivation of the Kalman gain from the recursive LMS estimator.

5.2.6 Summary of Equations for the Discrete-Time Kalman Estimator

The equations derived in the previous section are summarized in Table 5.3. In this formulation of the filter equations, G has been combined with the plant covariance by multiplying G_{k-1} and G_{k-1}^{T}, for example,

$$Q_{k-1} = G_{k-1}\mathrm{E}\langle w_{k-1}w_{K-1}^{\mathrm{T}}\rangle G_{k-1}^{\mathrm{T}}$$
$$= G_{k-1}\overline{Q}_{k-1}G_{k-1}^{\mathrm{T}}.$$

The relation of the filter to the system is illustrated in the block diagram of Figure 5.3.

The basic steps of the computational procedure for the discrete-time Kalman estimator are as follows:

1. Compute $P_{k(-)}$ using $P_{(k-1)(+)}$, Φ_{k-1}, and Q_{k-1}.

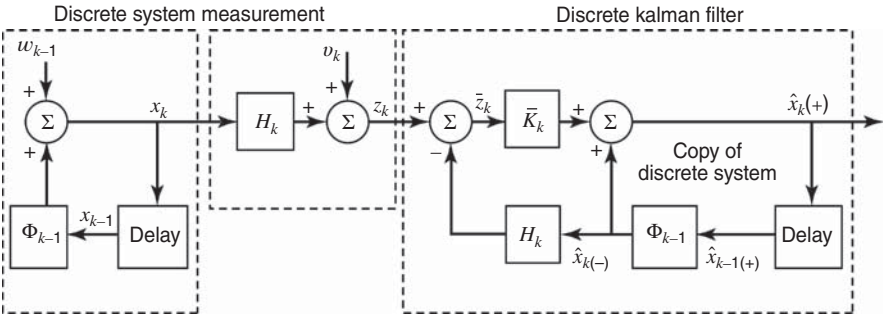

Figure 5.3 Block diagram of system, measurement model, and discrete-time Kalman filter.

TABLE 5.3 Discrete-Time Kalman Filter Equations

System dynamic model	$x_k = \Phi_{k-1}x_{k-1} + w_{k-1}$ $w_k \sim \mathcal{N}(0, Q_k)$
Measurement model	$z_k = H_k x_k + v_k$ $v_k \sim \mathcal{N}(0, R_k)$
Initial conditions	$\mathrm{E}\langle x_0 \rangle = \hat{x}_0$ $\mathrm{E}\langle \tilde{x}_0 \tilde{x}_0^{\mathrm{T}} \rangle = P_0$
Independence assumption	$\mathrm{E}\langle w_k v_j^{\mathrm{T}} \rangle = 0$ for all k and j
State estimate extrapolation (Equation 5.25)	$\hat{x}_{k(-)} = \Phi_{k-1}\hat{x}_{(k-1)(+)}$
Error covariance extrapolation (Equation 5.26)	$P_{k(-)} = \Phi_{k-1}P_{(k-1)(+)}\Phi_{k-1}^{\mathrm{T}} + Q_{k-1}$
State estimate observational update (Equation 5.21)	$\hat{x}_{k(+)} = \hat{x}_{k(-)} + \overline{K}_k[z_k - H_k\hat{x}_{k(-)}]$
Error covariance update (Equation 5.24)	$P_{k(+)} = [I - \overline{K}_k H_k]P_{k(-)}$
Kalman gain matrix (Equation 5.19)	$\overline{K}_k = P_{k(-)}H_k^{\mathrm{T}}[H_k P_{k(-)}H_k^{\mathrm{T}} + R_k]^{-1}$

2. Compute $\overline{K}_k$ using $P_{k(-)}$ (computed in step 1), H_k, and R_k.
3. Compute $P_{k(+)}$ using $\overline{K}_k$ (computed in step 2) and $P_{k(-)}$ (from step 1).
4. Compute successive values of $\hat{x}_{k(+)}$ recursively using the computed values of $\overline{K}_k$ (from step 3), the given initial estimate $\hat{x}_0$, and the input data z_k.

Step 4 of the Kalman filter implementation (computation of $\hat{x}_{k(+)}$) can be implemented only for state vector propagation where simulator or real data sets are available. An example of this is given in Section 5.11.

In the design trade-offs, the covariance matrix update (steps 1 and 3) should be checked for symmetry and positive definiteness. Failure to attain either condition is a sign that something is wrong—either a program "bug" or an ill-conditioned problem. In order to overcome ill-conditioning, another equivalent expression for $P_{k(+)}$ is called the *Joseph form*,[4] as shown in Equation 5.23:

$$P_{k(+)} = [I - \overline{K}_k H_k]P_{k(-)}[I - \overline{K}_k H_k]^{\mathrm{T}} + \overline{K}_k R_k \overline{K}_k^{\mathrm{T}}.$$

Note that the right-hand side of this equation is the summation of two symmetric matrices. The first of these is positive definite and the second is nonnegative definite, thereby making $P_{k(+)}$ a positive-definite matrix.

[4] After Bucy and Joseph [6].

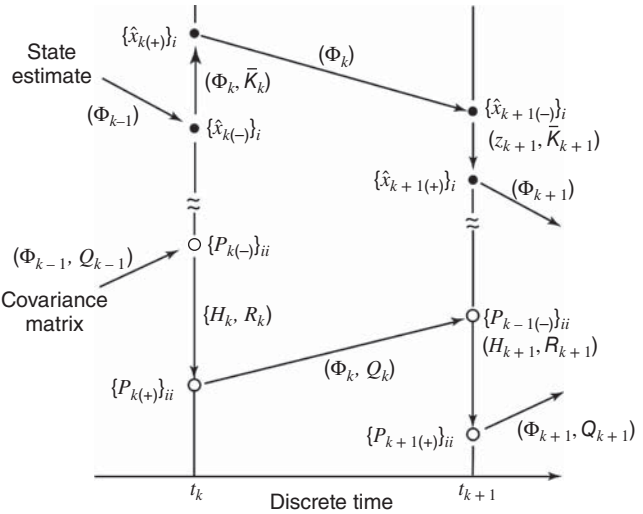

Figure 5.4 Representative sequence of values of filter variables in discrete time.

There are many other forms[5] for $\overline{K}_k$ and $P_{k(+)}$ that might not be as useful for robust computation. It can be shown that state vector update, Kalman gain, and error covariance equations represent an asymptotically stable system, and therefore, the estimate of state $\hat{x}_k$ becomes independent of the initial estimate $\hat{x}_0, P_0$ as k is increased.

Figure 5.4 shows a typical time sequence of values assumed by the ith component of the estimated state vector (plotted with solid circles) and its corresponding variance of estimation uncertainty (plotted with open circles). The arrows show the successive values assumed by the variables, with the annotation (in parentheses) on the arrows indicating which input variables define the indicated transitions. Note that each variable assumes two distinct values at each discrete time: its a priori value corresponding to the value *before* the information in the measurement is used, and the a posteriori value corresponding to the value *after* the information is used.

Typical behavior of successive values of the diagonal elements of P_k, the covariance matrix of estimation uncertainty, is illustrated in Figure 5.5, which displays the "sawtooth" pattern between the a priori ($P_{k(-)}$) and a posteriori ($P_{k(+)}$) values of P_k. As a rule, for the same value of k, the diagonal values of $P_{k(+)}$ tend to be less than or equal to those of $P_{k(-)}$.

The filter would be said to *converge* if $P \to 0$ as $k \to \infty$, but this will not happen unless $Q = 0$ or $\Phi = 0$. An example of convergence with $Q = 0$ is shown in Figure 5.6, in which case the true state variable is a constant ($\Phi = 1$) and convergence is exponential. The state variable in this example is a scalar and the $\pm 1\sigma$ values bracketing the estimate are calculated from the square root of P. The separation of the

[5]Some of the alternative forms for computing $\overline{K}_k$ and $P_{k(+)}$ can be found in Jazwinski [10], Kailath [11], and Sorenson [32].

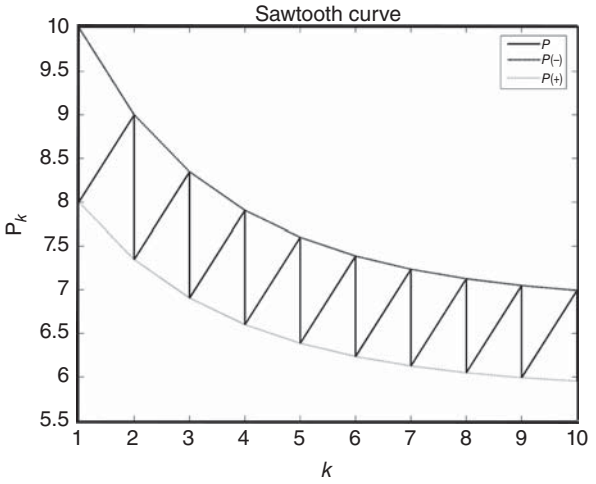

Figure 5.5 Sawtooth pattern of P_k diagonal values.

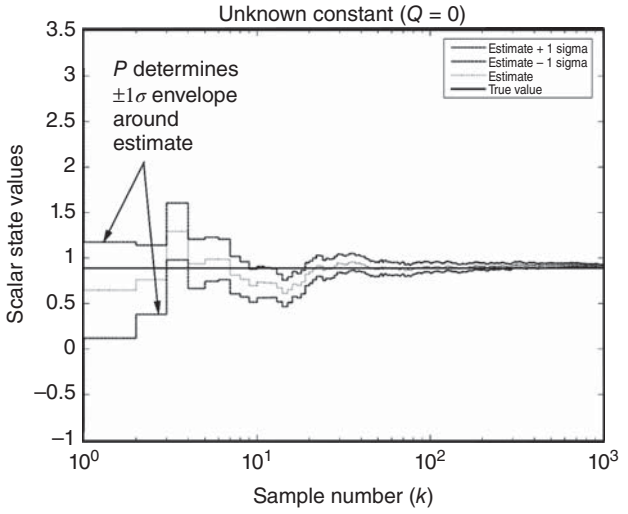

Figure 5.6 Convergence of P_k.

$\pm 1\sigma$ values is (slowly) converging to zero, which indicates that the filter is, indeed, converging.

In the more general case, P may still converge to some nonzero steady-state value.

Example 5.4 (Discrete-Time Implementation with Numerical Values) Let the scalar system dynamics and observations be given by the following equations:

$$x_k = x_{k-1} + w_{k-1}, \quad z_k = x_k + v_k,$$

$$\mathrm{E}\langle v_k \rangle = \mathrm{E}\langle w_k \rangle = 0,$$

$$\mathrm{E}\langle v_{k_1} v_{k_2} \rangle = 2\Delta(k_2 - k_1), \mathrm{E}\langle w_{k_1} w_{k_2} \rangle = \Delta(k_2 - k_1),$$

$$z_1 = 2, z_2 = 3,$$

$$\mathrm{E}\langle x(0) \rangle = \hat{x}_0 = 1,$$

$$\mathrm{E}\langle [x(0) - \hat{x}_0]^2 \rangle = P_0 = 10.$$

The objective is to find $\hat{x}_3$ and the steady-state covariance matrix P_∞. One can use the equations in Table 5.3 with

$$\Phi = 1 = H, Q = 1, R = 2,$$

for which

$$\boxed{P_{k(-)} = P_{(k-1)(+)} + 1},$$

$$\bar{K}_k = \frac{P_{k(-)}}{P_{k(-)} + 2} = \frac{P_{(k-1)(+)} + 1}{P_{(k-1)(+)} + 3},$$

$$P_{k(+)} = \left[1 - \frac{P_{(k-1)(+)} + 1}{P_{(k-1)(+)} + 3}\right](P_{(k-1)(+)} + 1),$$

$$\boxed{P_{k(+)} = \frac{2(P_{(k-1)(+)} + 1)}{P_{(k-1)(+)} + 3}},$$

$$\boxed{\hat{x}_{k(+)} = \hat{x}_{(k-1)(+)} + \bar{K}_k(z_k - \hat{x}_{(k-1)(+)})}.$$

Let

$$P_{k(+)} = P_{(k-1)(+)} = P \text{ (steady-state covariance)},$$

$$P = \frac{2(P+1)}{P+3},$$

$$P^2 + P - 2 = 0,$$

$$P = 1, \text{ positive-definite solution.}$$

For $k = 1$,

$$\hat{x}_1(+) = \hat{x}_0 + \frac{P_0 + 1}{P_0 + 3}(2 - \hat{x}_0) = 1 + \frac{11}{13}(2 - 1) = \frac{24}{13}.$$

Following is a table for the various values of the Kalman filter:

k	$P_{k(-)}$	$P_{k(+)}$	$\overline{K}_k$	$\hat{x}_{k(+)}$
1	11	$\frac{22}{13}$	$\frac{11}{13}$	$\frac{24}{13}$
2	$\frac{35}{13}$	$\frac{70}{61}$	$\frac{35}{61}$	$\frac{153}{61}$

5.2.7 Treating Vector Measurements with Uncorrelated Errors as Scalars

In many (if not most) applications with vector-valued measurement z, the corresponding matrix R of measurement noise covariance is a diagonal matrix, meaning that the individual components of v_k are uncorrelated. For those applications, it is advantageous to consider the components of z as independent scalar measurements, rather than as a vector measurement. The principal advantages are as follows:

1. *Reduced Computation Time.* The number of arithmetic computations required for processing an ℓ-vector z as ℓ successive scalar measurements is significantly less than the corresponding number of operations for vector measurement processing. It is demonstrated in Chapter 7 that the number of computations for the vector implementation grows as ℓ^3, whereas that of the scalar implementation grows only as ℓ.
2. *Improved Numerical Accuracy.* Avoiding matrix inversion in the implementation of the covariance equations (by making the expression $HPH^{\mathrm{T}} + R$ a scalar) improves the robustness of the covariance computations against roundoff errors.

The filter implementation in these cases requires ℓ iterations of the observational update equations using the rows of H as measurement "matrices" (with row dimension equal to 1) and the diagonal elements of R as the corresponding (scalar) measurement noise covariance. The updating can be implemented iteratively as the following equations:

$$\overline{K}_k^{[i]} = \frac{1}{H_k^{[i]} P_k^{[i-1]} H_k^{[i]\mathrm{T}} + R_k^{[i]}} P_k^{[i-1]} H_k^{[i]\mathrm{T}},$$

$$P_k^{[i]} = P_k^{[i-1]} - \overline{K}_k^{[i]} H_k^{[i]} P_k^{[i-1]},$$

$$\hat{x}_k^{[i]} = \hat{x}_k^{[i-1]} + \overline{K}_k^{[i]} [\{z_k\}_i - H_k^{[i]} \hat{x}_k^{[i-1]}],$$

for $i = 1, 2, 3, \ldots, \ell$, using the initial values

$$P_k^{[0]} = P_{k(-)}, \quad \hat{x}_k^{[0]} = \hat{x}_{k(-)};$$

intermediate variables

$$R_k^{[i]} = i\text{th diagonal element of the } \ell \times \ell \text{ diagonal matrix } R_k,$$
$$H_k^{[i]} = i\text{th row of the } \ell \times n \text{ matrix } H_k;$$

and final values

$$P_k^{[\ell]} = P_{k(+)}, \quad \hat{x}_k^{[\ell]} = \hat{x}_{k(+)}.$$

Example 5.5 (Serial Processing of Vector Measurements) Consider the measurement update problem with

$$\hat{x}_{k-} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \quad P_{k-} = \begin{bmatrix} 4 & 1 \\ 1 & 9 \end{bmatrix}, \quad H_k = \begin{bmatrix} 0 & 2 \\ 3 & 0 \end{bmatrix}, \quad R_k = \begin{bmatrix} 1 & 0 \\ 0 & 4 \end{bmatrix}, \quad z_k = \begin{bmatrix} 3 \\ 4 \end{bmatrix}.$$

Because R is diagonal, the two components of the measurement have independent errors and they can be processed serially, one at a time, as though they were two scalar measurements with mean-squared measurement uncertainties

$$R_1 = 1, \quad R_2 = 4,$$

and measurement sensitivity matrices

$$H_1 = \begin{bmatrix} 0 & 2 \end{bmatrix}, \quad H_2 = \begin{bmatrix} 3 & 0 \end{bmatrix}.$$

Table 5.4 shows the numerical calculations involved in processing both measurements simultaneously as a vector or as two independent scalar measurements. The implementation equations (left column) include some partial results in square brackets ($[\cdot]$) that are reused to reduce the computational effort required.

The final results are exactly the same by either route, although intermediate results do differ on the two-pass serial processing route. Note, in particular, the following differences:

1. There are intermediate values for the estimated state vector ($\hat{x}_{k+/2}$) and covariance matrix ($P_{k+/2}$), based on using just the first measurement vector component.
2. The *expected value* $\hat{z}_2 = H_2 \hat{x}_{k+/2}$ on the second pass, based on the intermediate estimate $\hat{x}_{k+/2}$ from using the first scalar measurement, is not the same as the second component of $\hat{z}$ when the measurement is processed as a vector.
3. The covariance matrix P used in computing the Kalman gain $\overline{K}$ on the second pass has the value of $P_{k+/2}$ from the first pass.
4. The two serial Kalman gain vectors bear little resemblance to the two columns of Kalman gain matrix for vector-valued measurements.

TABLE 5.4 Calculations for Example 5.5

Implementation Equations	Vector Measurements	Scalar Measurements	
		First Measurement	Second Measurement
$\hat{z} = H\hat{x}_{k(-)}$	$\begin{bmatrix}4\\3\end{bmatrix}$	4	105/37*
$\tilde{z} = z - \hat{z}$	$\begin{bmatrix}-1\\1\end{bmatrix}$	−1	43/37
$[HP]$	$\begin{bmatrix}2 & 18\\12 & 3\end{bmatrix}$	$[2 \quad 18]$	$\begin{bmatrix}\frac{432}{37} & \frac{3}{37}\end{bmatrix}$*
$[[HP]H^{\mathrm{T}} + R]$	$\begin{bmatrix}37 & 6\\6 & 40\end{bmatrix}$	37	1444/37
$[[HP]H^{\mathrm{T}} + R]^{-1}$	$\begin{bmatrix}\frac{10}{361} & -\frac{3}{722}\\-\frac{3}{722} & \frac{37}{1444}\end{bmatrix}$	1/37	37/1444
$\overline{K} = (HP)^{\mathrm{T}}[HPH^{\mathrm{T}} + R]^{-1}$	$\begin{bmatrix}\frac{2}{361} & \frac{108}{361}\\\frac{351}{722} & \frac{3}{1444}\end{bmatrix}$	$\begin{bmatrix}\frac{2}{37}\\\frac{18}{37}\end{bmatrix}$	$\begin{bmatrix}\frac{108}{361}\\\frac{3}{1444}\end{bmatrix}$
$\hat{x}_{k(+)} = \hat{x}_{k(-)} + \overline{K}\tilde{z}$	$\begin{bmatrix}\frac{467}{361}\\\frac{2189}{1444}\end{bmatrix}$	†	$\begin{bmatrix}\frac{467}{361}\\\frac{2189}{1444}\end{bmatrix}$
$\hat{x}_{k(+)/2}$†		$\begin{bmatrix}\frac{35}{37}\\\frac{56}{37}\end{bmatrix}$	
$P_{k(+)} = P_{k(-)} - \overline{K}[HP]$	$\begin{bmatrix}\frac{144}{361} & \frac{1}{361}\\\frac{1}{361} & \frac{351}{1444}\end{bmatrix}$	†	$\begin{bmatrix}\frac{144}{361} & \frac{1}{361}\\\frac{1}{361} & \frac{351}{1444}\end{bmatrix}$
$P_{k(+)/2}$†		$\begin{bmatrix}\frac{144}{37} & 1/37\\1/37 & \frac{9}{37}\end{bmatrix}$	

*Note that $\hat{z}$ on the second pass is based on $\hat{x}_{k(+)/2}$ from the first pass, not $\hat{x}_{k(-)}$, and P on the second pass is $P_{k(+)/2}$ from the first pass.

†Values generated on the first pass are intermediate results.

5. Using the measurement vector components one at a time as independent scalar measurements avoids taking matrix inverses, which significantly reduces the overall amount of computation required.
6. Although writing out the results in this way may make it appear that the two-pass approach takes more computation, the opposite is true. In fact, the advantage of multi-pass processing of measurement vector components increases with vector length.

5.2.8 Using the Covariance Equations for Design Analysis

It is important to remember that the Kalman gain and error covariance equations are independent of the actual observations. The covariance equations alone are all that is required for characterizing the performance of a proposed sensor system before it is actually built. At the beginning of the design phase of a measurement and estimation system, when neither real nor simulated data are available, just the covariance calculations can be used to obtain preliminary indications of estimator performance. Covariance calculations consist of solving the estimator equations with steps 1–3 of the previous subsection, repeatedly. These covariance calculations will involve the plant noise covariance matrix Q, measurement noise covariance matrix R, state-transition matrix Φ, measurement sensitivity matrix H, and initial covariance matrix P_0—all of which must be known for the designs under consideration.

5.3 KALMAN–BUCY FILTER

Kalman and Bucy published the continuous-time equivalent of the Kalman filter [33] shortly after the Kalman filter first appeared [28]. It provides a good heuristic model for engineers who prefer to "think continuously, but act discretely." These are people who are more comfortable thinking about dynamics in continuous time, in terms of derivatives. The Kalman–Bucy filter allows them to develop a model in continuous time until they feel confident with it, then switch over to the equivalent Kalman filter model in discrete time.

Analogous to the discrete-time case, the continuous-time random process $x(t)$ and the observation $z(t)$ are given by

$$\dot{x}(t) = F(t)x(t) + G(t)w(t), \tag{5.106}$$

$$z(t) = H(t)x(t) + v(t), \tag{5.107}$$

$$\mathrm{E}w(t) = \mathrm{E}v(t) = 0,$$

$$\mathrm{E}w(t_1)w^{\mathrm{T}}(t_2) = Q(t)\delta(t_2 - t_1), \tag{5.108}$$

$$\mathrm{E}v(t_1)v^{\mathrm{T}}(t_2) = R(t)\delta(t_2 - t_1), \tag{5.109}$$

$$\mathrm{E}w(t)v^{\mathrm{T}}(\eta) = 0, \tag{5.110}$$

where $F(t)$, $G(t)$, $H(t)$, $Q(t)$, and $R(t)$ are $n \times n$, $n \times n$, $\ell \times n$, $n \times n$, and $\ell \times \ell$ matrices, respectively. The term $\delta(t_2 - t_1)$ is the Dirac delta. The covariance matrices Q and R are positive definite.

It is desired to find the estimate of n state vector $x(t)$ represented by $\hat{x}(t)$, which is a linear function of the measurements $z(t)$, $0 \leq t \leq T$ that minimizes the scalar equation

$$\mathrm{E}[x(t) - \hat{x}(t)]^{\mathrm{T}} M[x(t) - \hat{x}(t)], \tag{5.111}$$

where M is a symmetric positive-definite matrix.

The initial estimate and covariance matrix are $\hat{x}_0$ and P_0.

This section provides a formal derivation of the continuous-time Kalman estimator. A rigorous derivation can be achieved by using the orthogonality principle as in the discrete-time case. In view of the main objective (to obtain efficient and practical estimators), less emphasis is placed on continuous-time estimators.

Let Δt be the time interval $[t_k - t_{k-1}]$. As shown in Chapters 2 and 4, the following relationships are obtained:

$$\Phi(t_k, t_{k-1}) = \Phi_k = I + F(t_{k-1})\Delta t + 0(\Delta t^2),$$

where $0(\Delta t^2)$ consists of terms with powers of Δt greater than or equal to 2. For measurement noise

$$R_k = \frac{R(t_k)}{\Delta t},$$

and for process noise

$$Q_k = G(t_k)Q(t_k)G^{\mathrm{T}}(t_k)\Delta t.$$

Equations 5.24 and 5.26 can be combined. By substituting the above relations, one can get the result

$$\begin{aligned} P_{k(-)} &= [I + F(t)\Delta t][I - \overline{K}_{k-1}H_{k-1}]P_{k-1(-)} \\ &\quad \times [I + F(t)\Delta t]^{\mathrm{T}} + G(t)Q(t)G^{\mathrm{T}}(t)\Delta t, \end{aligned} \tag{5.112}$$

$$\begin{aligned} \frac{P_{k(-)} - P_{k-1(-)}}{\Delta t} &= F(t)P_{k-1(-)} + P_{k-1(-)}F^{\mathrm{T}}(t) \\ &\quad + G(t)Q(t)G^{\mathrm{T}}(t) - \frac{\overline{K}_{k-1}H_{k-1}P_{k-1(-)}}{\Delta t} \\ &\quad - F(t)\overline{K}_{k-1}H_{k-1}P_{k-1(-)}F^{\mathrm{T}}(t)\Delta t \\ &\quad + \text{higher order terms}. \end{aligned} \tag{5.113}$$

The Kalman gain of Equation 5.19 becomes, in the limit,

$$\lim_{\Delta t\to 0}\left[\frac{\overline{K}_{k-1}}{\Delta t}\right] = \lim_{\Delta t\to 0}\{P_{k-1(-)}H_{k-1}^{\mathrm{T}}[H_{k-1}P_{k-1(-)}H_{k-1}^{\mathrm{T}}\Delta t + R(t)]^{-1}\}$$
$$= PH^{\mathrm{T}}R^{-1} = \overline{K}(t). \tag{5.114}$$

Substituting Equation 5.114 in 5.113 and taking the limit as $\Delta t \to 0$, one obtains the desired result

$$\dot{P}(t) = F(t)P(t) + P(t)F^{\mathrm{T}}(t) + G(t)Q(t)G^{\mathrm{T}}(t)$$
$$- P(t)H^{\mathrm{T}}(t)R^{-1}(t)H(t)P(t) \tag{5.115}$$

with $P(t_0)$ as the initial condition. This is called the *matrix Riccati differential equation*. Methods for solving it will be discussed in Section 5.8. The differential equation can be rewritten by using the identity

$$P(t)H^{\mathrm{T}}(t)R^{-1}(t)R(t)R^{-1}(t)H(t)P(t) = \overline{K}(t)R(t)\overline{K}^{\mathrm{T}}(t)$$

to transform Equation 5.115 to the form

$$\dot{P}(t) = F(t)P(t) + P(t)F^{\mathrm{T}}(t) + G(t)Q(t)G^{\mathrm{T}}(t) - \overline{K}(t)R(t)\overline{K}^{\mathrm{T}}(t). \tag{5.116}$$

In a similar manner, the state vector update equation can be derived from Equations 5.21 and 5.25 by taking the limit as $\Delta t \to 0$ to obtain the differential equation for the estimate:

$$\hat{x}_{k(+)} = \Phi_{k-1}\hat{x}_{k-1(+)} + \overline{K}[z_k - H_k\Phi_{k-1}x_{k-1(+)}] \tag{5.117}$$
$$\approx [I + F\ \Delta t]\ \hat{x}_{k-1(+)} + \overline{K}_k[z_k - H_k(I + F\ \Delta t)\hat{x}_{k-1(+)}] \tag{5.118}$$
$$\dot{\hat{x}}(t_k) = \lim_{\Delta t\to 0}\frac{x_{k(+)} - x_{k-1(+)}}{\Delta t} \tag{5.119}$$
$$= \lim_{\Delta t\to 0}\left[F\hat{x}_{k-1(+)}\frac{\overline{K}_k}{\Delta t}\left(z_k - H_k\hat{x}_{k-1(+)} - H_kF_k\ \Delta t\ \hat{x}_{k-1(+)}\right)\right] \tag{5.120}$$
$$\dot{\hat{x}}(t_k) = \lim_{\Delta t\to 0}\frac{x_{k(+)} - x_{k-1(+)}}{\Delta t} \tag{5.121}$$
$$= \lim_{\Delta t\to 0}\left[F\hat{x}_{k-1(+)}\frac{\overline{K}_k}{\Delta t}\left(z_k - H_k\hat{x}_{k-1(+)} - H_kF_k\ \Delta t\ \hat{x}_{k-1(+)}\right)\right] \tag{5.122}$$
$$= F(t)\hat{x}(t) + \overline{K}(t)[z(t) - H(t)\hat{x}(t)] \tag{5.123}$$

with initial condition $\hat{x}(0)$. Equations 5.114, 5.116, and 5.123 define the continuous-time Kalman estimator, which is also called the *Kalman–Bucy filter* [28, 33–35].

5.4 OPTIMAL LINEAR PREDICTORS

5.4.1 Prediction as Filtering

Prediction is equivalent to filtering when the measurement data are not available or are unreliable. In such cases, the Kalman gain matrix $\overline{K}_k$ is forced to be zero. Hence, Equations 5.21, 5.25, and 5.123 become

$$\hat{x}_{k(+)} = \Phi_{k-1}\hat{x}_{(k-1)(+)} \tag{5.124}$$

and

$$\dot{\hat{x}}(t) = F(t)\hat{x}(t). \tag{5.125}$$

Previous values of the estimates will become the initial conditions for the above equations.

5.4.2 Accommodating Missing Data

It sometimes happens in practice that measurements that had been scheduled to occur over some time interval ($t_{k_1} < t \leq t_{k_2}$) are, in fact, unavailable or unreliable. The estimation accuracy will suffer from the missing information, but the filter can continue to operate without modification. One can continue using the prediction algorithm given in Section 5.4 to continually estimate x_k for $k > k_1$ using the last available estimate $\hat{x}_{k_1}$ until the measurements again become useful (after $k = k_2$).

It is unnecessary to perform the observational update, because there is no information on which to base the conditioning. In practice, the filter is often run with the measurement sensitivity matrix $H = 0$ so that, in effect, the only update performed is the temporal update.

5.5 CORRELATED NOISE SOURCES

5.5.1 Correlation between Plant and Measurement Noise

We want to consider the extensions of the results given in Sections 5.2 and 5.3, allowing correlation between the two noise processes. Let the correlation be given by

$$\mathrm{E}\langle w_{k_1} v_{k_2}^{\mathrm{T}} \rangle = C_k \Delta(k_2 - k_1), \quad \text{for the discrete-time case,}$$

$$\mathrm{E}\langle w(t_1) v^{\mathrm{T}}(t_2) \rangle = C(t)\delta(t_2 - t_1), \quad \text{for the continuous-time case.}$$

For this extension, the discrete-time estimators have the same initial conditions and state estimate extrapolation and error covariance extrapolation equations. However, the measurement update equations in Table 5.3 have been modified as

$$\overline{K}_k = [P_{k(-)}H_k^{\mathrm{T}} + C_k][H_k P_{k(-)} H_k^{\mathrm{T}} + R_k + H_k C_k + C_k^{\mathrm{T}} H^{\mathrm{T}}]^{-1},$$

$$P_{k(+)} = P_{k(-)} - \overline{K}_k[H_k P_{k(-)} + C_k^{\mathrm{T}}],$$
$$\hat{x}_{k(+)} = \hat{x}_{k(-)} + \overline{K}_k[z_k - H_k \hat{x}_{k(-)}].$$

Similarly, the continuous-time estimator algorithms can be extended to include the correlation. Equation 5.114 is changed as follows [36, 37]:

$$\overline{K}(t) = [P(t)H^{\mathrm{T}}(t) + C(t)]R^{-1}(t).$$

5.5.2 Time-Correlated Measurements

Correlated measurement noise v_k can be modeled by a shaping filter driven by white noise (see Section 5.5). Let the measurement model be given by

$$z_k = H_k x_k + v_k,$$

where

$$v_k = A_{k-1} v_{k-1} + \eta_{k-1} \tag{5.126}$$

and η_k is zero-mean white noise.

Equation 5.1 is augmented by Equation 5.126, and the new state vector $X_k = [x_k \;\; v_k]^{\mathrm{T}}$ satisfies the difference equation:

$$X_k = \begin{bmatrix} x_k \\ -- \\ v_k \end{bmatrix} = \begin{bmatrix} \Phi_{k-1} & 0 \\ --- & --- \\ 0 & A_{k-1} \end{bmatrix} \begin{bmatrix} x_{k-1} \\ -- \\ v_{k-1} \end{bmatrix} + \begin{bmatrix} w_{k-1} \\ --- \\ \eta_{k-1} \end{bmatrix},$$
$$z_k = [H_k \vdots I]X_k.$$

The measurement noise is zero, $R_k = 0$. The estimator algorithm will work as long as $H_k P_{k(-)} H_k^{\mathrm{T}} + R_k$ is invertible. Details of numerical difficulties of this problem (when R_k is singular) are given in Chapter 7.

For continuous-time estimators, the augmentation does not work because $\overline{K}(t) = P(t)H^{\mathrm{T}}(t)R^{-1}(t)$ is required. Therefore, $R^{-1}(t)$ must exist. Alternate techniques are required. For detailed information, see Gelb et al. [9].

5.6 RELATIONSHIPS BETWEEN KALMAN AND WIENER FILTERS

The Wiener filter is defined for stationary systems in continuous time, and the Kalman filter is defined for either stationary or nonstationary systems in either discrete time or continuous time, but with finite-state dimension. To demonstrate the connections on problems satisfying both sets of constraints, take the continuous-time Kalman–Bucy estimator equations of Section 5.3, letting F, G, and H be constants, the noises be stationary (Q and R constant), and the filter reach steady state (P constant). That is,

as $t \to \infty$, then $\dot{P}(t) \to 0$. The Riccati differential equation from Section 5.3 becomes the algebraic Riccati equation

$$0 = FP(\infty) + P(\infty)F^{\mathrm{T}} + GQG^{\mathrm{T}} - P(\infty)H^{\mathrm{T}}R^{-1}HP(\infty)$$

for continuous-time systems. The positive-definite solution of this algebraic equation is the steady-state value of the covariance matrix, $[P(\infty)]$. The Kalman–Bucy filter equation in steady state is then

$$\dot{\hat{x}}(t) = F\hat{x} + \overline{K}(\infty)[z(t) - H\hat{x}(t)].$$

Take the Laplace transform of both sides of this equation, assuming that the initial conditions are equal to zero, to obtain the following transfer function:

$$[sI - F + \overline{K}H]\hat{x}(s) = \overline{K}z(s),$$

where the Laplace transforms $\mathcal{L}\hat{x}(t) = \hat{x}(s)$ and $\mathcal{L}z(t) = z(s)$. This has the solution

$$\hat{x}(s) = [sI - F + \overline{K}H]^{-1}\overline{K}z(s),$$

where the steady-state gain

$$\overline{K} = P(\infty)H^{\mathrm{T}}R^{-1}.$$

This transfer function represents the steady-state Kalman–Bucy filter, which is identical to the Wiener filter [12].

5.7 QUADRATIC LOSS FUNCTIONS

The Kalman filter minimizes *any* quadratic loss function of estimation error. Just the fact that it is *unbiased* is sufficient to prove this property, but saying that the estimate is unbiased is equivalent to saying that $\hat{x} = \mathrm{E}\langle x\rangle$. That is, the estimated value is the *mean* of the probability distribution of the state.

5.7.1 Quadratic Loss Functions of Estimation Error

A *loss function* or *penalty function*[6] is a real-valued function of the outcome of a random event. A loss function reflects the *value* of the outcome. Value concepts can be somewhat subjective. In gambling, for example, your perceived loss function for the outcome of a bet may depend upon your personality and current state of winnings, as well as on how much you have riding on the bet.

[6]These are concepts from decision theory, which includes estimation theory. The theory might have been built just as well on more optimistic concepts, such as "gain functions," "benefit functions," or "reward functions," but the nomenclature seems to have been developed by pessimists. This focus on the negative aspects of the problem is unfortunate, and you should not allow it to dampen your spirit.

5.7.1.1 Loss Functions of Estimates In estimation theory, the perceived loss is generally a function of *estimation error* (the difference between an estimated function of the outcome and its actual value), and it is generally a monotonically increasing function of the absolute value of the estimation error. In other words, bigger errors are valued less than smaller errors.

5.7.1.2 Quadratic Loss Functions If x is a real n-vector (variate) associated with the outcome of an event and $\hat{x}$ is an estimate of x, then a quadratic loss function for the estimation error $\hat{x} - x$ has the form

$$L(\hat{x} - x) = (\hat{x} - x)^{\mathrm{T}} M (\hat{x} - x), \tag{5.127}$$

where M is a *symmetric positive-definite matrix*. One may as well assume that M is symmetric, because the skew-symmetric part of M does not influence the quadratic loss function. The reason for assuming positive definiteness is to assure that the loss is zero only if the error is zero, and loss is a monotonically increasing function of the absolute estimation error.

5.7.2 Expected Value of a Quadratic Loss Function

5.7.2.1 Loss and Risk The expected value of loss is sometimes called *risk*. It will be shown that the expected value of a quadratic loss function of the estimation error $\hat{x} - x$ is a quadratic function of $\hat{x} - \mathrm{E}\langle x \rangle$, where $\mathrm{E}\langle \hat{x} \rangle = \mathrm{E}\langle x \rangle$. This demonstration will depend upon the following identities:

$$\hat{x} - x = (\hat{x} - \mathrm{E}\langle x \rangle) - (x - \mathrm{E}\langle x \rangle), \tag{5.128}$$

$$\mathop{\mathrm{E}}_{x}\langle x - \mathrm{E}\langle x \rangle \rangle = 0, \tag{5.129}$$

$$\begin{aligned} \mathop{\mathrm{E}}_{x}\langle (x - \mathrm{E}\langle x \rangle)^{\mathrm{T}} M (x - \mathrm{E}\langle x \rangle) \rangle & \\ = \mathop{\mathrm{E}}_{x}\langle \mathrm{trace}[(x - \mathrm{E}\langle x \rangle)^{\mathrm{T}} M (x - \mathrm{E}\langle x \rangle)] \rangle & \quad (5.130) \\ = \mathop{\mathrm{E}}_{x}\langle \mathrm{trace}[M (x - \mathrm{E}\langle x \rangle)(x - \mathrm{E}\langle x \rangle)^{\mathrm{T}}] \rangle & \quad (5.131) \\ = \mathrm{trace}[M \mathop{\mathrm{E}}_{x}\langle (x - \mathrm{E}\langle x \rangle)(x - \mathrm{E}\langle x \rangle)^{\mathrm{T}}] & \quad (5.132) \\ = \mathrm{trace}[MP], & \quad (5.133) \end{aligned}$$

$$P \overset{\mathrm{def}}{=} \mathop{\mathrm{E}}_{x}\langle (x - \mathrm{E}\langle x \rangle)(x - \mathrm{E}\langle x \rangle)^{\mathrm{T}} \rangle. \tag{5.134}$$

5.7.2.2 Risk of a Quadratic Loss Function In the case of the quadratic loss function defined above, the expected loss (risk) will be

$$\begin{aligned} \mathcal{R}(\hat{x}) &= \mathop{\mathrm{E}}_{x}\langle L(\hat{x} - x) \rangle & (5.135) \\ &= \mathop{\mathrm{E}}_{x}\langle (\hat{x} - x)^{\mathrm{T}} M (\hat{x} - x) \rangle & (5.136) \end{aligned}$$

$$= \underset{x}{\mathrm{E}}\langle[(\hat{x} - \mathrm{E}\langle x\rangle) - (x - \mathrm{E}\langle x\rangle)]^{\mathrm{T}} M[(\hat{x} - \mathrm{E}\langle x\rangle) - (x - \mathrm{E}\langle x\rangle)]\rangle \tag{5.137}$$

$$= \underset{x}{\mathrm{E}}\langle(\hat{x} - \mathrm{E}\langle x\rangle)^{\mathrm{T}} M(\hat{x} - \mathrm{E}\langle x\rangle) + (x - \mathrm{E}\langle x\rangle)^{\mathrm{T}} M(x - \mathrm{E}\langle x\rangle)\rangle$$
$$- \underset{x}{\mathrm{E}}\langle(\hat{x} - \mathrm{E}\langle x\rangle)^{\mathrm{T}} M(x - \mathrm{E}\langle x\rangle) + (x - \mathrm{E}\langle x\rangle)^{\mathrm{T}} M(\hat{x} - \mathrm{E}\langle x\rangle)\rangle \tag{5.138}$$

$$= (\hat{x} - \mathrm{E}\langle x\rangle)^{\mathrm{T}} M(\hat{x} - \mathrm{E}\langle x\rangle) + \underset{x}{\mathrm{E}}\langle(x - \mathrm{E}\langle x\rangle)^{\mathrm{T}} M(x - \mathrm{E}\langle x\rangle)\rangle$$
$$- (\hat{x} - \mathrm{E}\langle x\rangle)^{\mathrm{T}} M_{x}\langle(x - \mathrm{E}\langle x\rangle)\rangle - \underset{x}{\mathrm{E}}\langle(x - \mathrm{E}\langle x\rangle)\rangle^{\mathrm{T}} M(\hat{x} - \mathrm{E}\langle x\rangle) \tag{5.139}$$

$$= (\hat{x} - \mathrm{E}\langle x\rangle)^{\mathrm{T}} M(\hat{x} - \mathrm{E}\langle x\rangle) + \mathrm{trace}[MP], \tag{5.140}$$

which is a quadratic function of $\hat{x} - \mathrm{E}\langle x\rangle$ with the added nonnegative[7] constant trace$[MP]$.

5.7.3 Unbiased Estimates and Quadratic Loss

The *estimate* $\hat{x} = \mathrm{E}\langle x\rangle$ minimizes the expected value of any *positive-definite quadratic loss function*. From the above derivation,

$$\mathcal{R}(\hat{x}) \geq \mathrm{trace}[MP] \tag{5.141}$$

and

$$\mathcal{R}(\hat{x}) = \mathrm{trace}[MP] \tag{5.142}$$

only if

$$\hat{x} = \mathrm{E}\langle x\rangle, \tag{5.143}$$

where it has been assumed only that the mean $\mathrm{E}\langle x\rangle$ and covariance $\underset{x}{\mathrm{E}}\langle(x - \mathrm{E}\langle x\rangle)(x - \mathrm{E}\langle x\rangle)^{\mathrm{T}}\rangle$ are defined for the probability distribution of x. This demonstrates the utility of quadratic loss functions in estimation theory: They always lead to the mean as the estimate with minimum expected loss (risk).

5.7.3.1 Unbiased Estimates An estimate $\hat{x}$ is called *unbiased* if the expected estimation error $\underset{x}{\mathrm{E}}\langle\hat{x} - x\rangle = 0$. What has just been shown is that an unbiased estimate minimizes the expected value of any quadratic loss function of estimation error.

5.8 MATRIX RICCATI DIFFERENTIAL EQUATION

The need to solve the Riccati equation is perhaps the greatest single cause of anxiety and agony on the part of people faced with implementing a Kalman filter. This section presents a brief discussion of solution methods for the Riccati *differential* equation for the Kalman–Bucy filter. An analogous treatment of the discrete-time problem for the Kalman filter is presented in the next section. A more thorough treatment of the Riccati equation can be found in the book by Bittanti et al. [20].

[7]Recall that M and P are symmetric and nonnegative definite, and the matrix trace of any product of symmetric nonnegative definite matrices is nonnegative.

5.8.1 Transformation to a Linear Equation

The Riccati differential equation was first studied in the eighteenth century as a nonlinear scalar differential equation, and a method was derived for transforming it to a linear matrix differential equation. That same method works when the dependent variable of the original Riccati differential equation is a matrix. That solution method is derived here for the matrix Riccati differential equation of the Kalman–Bucy filter. An analogous solution method for the discrete-time matrix Riccati equation of the Kalman filter is derived in the next section.

5.8.1.1 Matrix Fractions A matrix product of the sort AB^{-1} is called a *matrix fraction*, and a representation of a matrix M in the form

$$M = AB^{-1}$$

will be called a *fraction decomposition* of M. The matrix A is the *numerator* of the fraction, and the matrix B is its *denominator*. It is necessary that the matrix denominator be nonsingular.

5.8.1.2 Linearization by Fraction Decomposition The Riccati differential equation is nonlinear. However, a fraction decomposition of the covariance matrix results in a linear differential equation for the numerator and denominator matrices. The numerator and denominator matrices will be functions of time, such that the product $A(t)B^{-1}(t)$ satisfies the matrix Riccati differential equation and its boundary conditions.

5.8.1.3 Derivation By taking the derivative of the matrix fraction $A(t)B^{-1}(t)$ with respect to t and using the fact[8] that

$$\frac{\mathrm{d}}{\mathrm{d}t}B^{-1}(t) = -B^{-1}(t)\dot{B}(t)B^{-1}(t),$$

one can arrive at the following decomposition of the matrix Riccati differential equation, where GQG^{T} has been reduced to an equivalent Q:

$$\begin{aligned}
&\dot{A}(t)B^{-1}(t) - A(t)B^{-1}(t)\dot{B}(t)B^{-1}(t) \\
&\qquad = \frac{d}{dt}\{A(t)B^{-1}(t)\} && (5.144)\\
&\qquad = \frac{d}{dt}P(t) && (5.145)\\
&\qquad = F(t)P(t) + P(t)F^{\mathrm{T}}(t) \\
&\qquad\quad - P(t)H^{\mathrm{T}}(t)R^{-1}(t)H(t)P(t) + Q(t) && (5.146)
\end{aligned}$$

[8]This formula is derived in Appendix B (in the Wiley web site).

$$= F(t)A(t)B^{-1}(t) + A(t)B^{-1}(t)F^{\mathrm{T}}(t) - A(t)B^{-1}(t)H^{\mathrm{T}}(t)R^{-1}(t)H(t)A(t)B^{-1}(t) + Q(t), \tag{5.147}$$

$$\dot{A}(t) - A(t)B^{-1}(t)\dot{B}(t) = F(t)A(t) + A(t)B^{-1}(t)F^{\mathrm{T}}(t)B(t) - A(t)B^{-1}(t)H^{\mathrm{T}}(t)R^{-1}(t)H(t)A(t) + Q(t)B(t), \tag{5.148}$$

$$\dot{A}(t) - A(t)B^{-1}(t)\{\dot{B}(t)\} = F(t)A(t) + Q(t)B(t) - A(t)B^{-1}(t) \times \{H^{\mathrm{T}}(t)R^{-1}(t)H(t)A(t) - F^{\mathrm{T}}(t)B(t)\}, \tag{5.149}$$

$$\dot{A}(t) = F(t)A(t) + Q(t)B(t), \tag{5.150}$$

$$\dot{B}(t) = H^{\mathrm{T}}(t)R^{-1}(t)H(t)A(t) - F^{\mathrm{T}}(t)B(t), \tag{5.151}$$

$$\frac{d}{dt}\begin{bmatrix} A(t) \\ B(t) \end{bmatrix} = \begin{bmatrix} F(t) & Q(t) \\ H^{\mathrm{T}}(t)R^{-1}(t)H(t) & -F^{\mathrm{T}}(t) \end{bmatrix} \begin{bmatrix} A(t) \\ B(t) \end{bmatrix}. \tag{5.152}$$

The last equation is a linear first-order matrix differential equation. The dependent variable is a $2n \times n$ matrix, where n is the dimension of the underlying state variable.

5.8.1.4 Hamiltonian Matrix This is the name[9] given to the matrix

$$\Psi(t) = \begin{bmatrix} F(t) & Q(t) \\ H^{\mathrm{T}}(t)R^{-1}(t)H(t) & -F^{\mathrm{T}}(t) \end{bmatrix} \tag{5.153}$$

of the matrix Riccati differential equation.

5.8.1.5 Boundary Constraints The initial values of $A(t)$ and $B(t)$ must also be constrained by the initial value of $P(t)$. This is easily satisfied by taking $A(t_0) = P(t_0)$ and $B(t_0) = I$, the identity matrix.

5.8.2 Time-Invariant Problem

In the time-invariant case, the Hamiltonian matrix Ψ is also time invariant. As a consequence, the solution for the numerator A and denominator B of the matrix fraction can be represented in matrix form as the product

$$\begin{bmatrix} A(t) \\ B(t) \end{bmatrix} = e^{\Psi t} \begin{bmatrix} P(0) \\ I \end{bmatrix},$$

where $e^{\Psi t}$ is a $2n \times 2n$ matrix.

[9] After the Irish mathematician and physicist William Rowan Hamilton (1805–1865).

5.8.3 Scalar Time-Invariant Problem

For this problem, the numerator A and denominator B of the matrix fraction AB^{-1} will be scalars, but Ψ will be a 2×2 matrix. We will here show how its exponential can be obtained in closed form. This will illustrate an application of the linearization procedure, and the results will serve to illuminate properties of the solutions—such as their dependence on initial conditions and on the scalar parameters *F, H, R*, and *Q*.

5.8.3.1 Linearizing the Differential Equation The scalar time-invariant Riccati differential equation and its linearized equivalent are

$$\dot{P}(t) = FP(t) + P(t)F^{\mathrm{T}} - P(t)HR^{-1}HP(t) + Q,$$
$$\begin{bmatrix} \dot{A}(t) \\ \dot{B}(t) \end{bmatrix} = \begin{bmatrix} F & Q \\ H^{\mathrm{T}}R^{-1}H & -F^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} A(t) \\ B(t) \end{bmatrix},$$

respectively, where the symbols *F, H, R*, and *Q* represent scalar parameters (constants) of the application, t is a free (independent) variable, and the dependent variable P is constrained as a function of t by the differential equation. One can solve this equation for P as a function of the free variable t and as a function of the parameters *F, H, R*, and *Q*.

5.8.3.2 Fundamental Solution of Linear Time-Invariant Differential Equation The linear time-invariant differential equation has the general solution

$$\begin{bmatrix} A(t) \\ B(t) \end{bmatrix} = e^{\Psi t} \begin{bmatrix} P(0) \\ 1 \end{bmatrix}, \tag{5.154}$$
$$\Psi = \begin{bmatrix} F & Q \\ \dfrac{H^2}{R} & -F \end{bmatrix}.$$

This matrix exponential will now be evaluated by using the characteristic vectors of Ψ, which are arranged as the column vectors of the matrix

$$M = \begin{bmatrix} \dfrac{-Q}{F+\phi} & \dfrac{-Q}{F-\phi} \\ 1 & 1 \end{bmatrix}, \quad \phi = \sqrt{F^2 + \frac{H^2 Q}{R}},$$

with inverse

$$M^{-1} = \begin{bmatrix} \dfrac{-H^2}{2\phi R} & \dfrac{H^2 Q}{2H^2Q + 2F^2R - 2F\phi R} \\ \dfrac{H^2}{2\phi R} & \dfrac{H^2 Q}{2H^2Q + 2F^2R + 2F\phi R} \end{bmatrix},$$

by which it can be diagonalized as

$$M^{-1}\Psi M = \begin{bmatrix} \lambda^2 & 0 \\ 0 & \lambda_1 \end{bmatrix},$$
$$\lambda_2 = -\frac{H^2Q + F^2R}{\phi R}, \quad \lambda_1 = \frac{H^2Q + F^2R}{\phi R},$$

with the characteristic values of Ψ along its diagonal. The exponential of the diagonalized matrix, multiplied by t, will be

$$e^{M^{-1}\Psi Mt} = \begin{bmatrix} e^{\lambda_2 t} & 0 \\ 0 & e^{\lambda_1 t} \end{bmatrix}.$$

Using this, one can write the fundamental solution of the linear homogeneous time-invariant equation as

$$\begin{aligned} e^{\Psi t} &= \sum_{k=0}^{\infty} \frac{1}{k!} t^k \Psi^k \\ &= M\left(\sum_{k=0}^{\infty} \frac{1}{k!}[M^{-1}\Psi M]^k\right) M^{-1} \\ &= Me^{M^{-1}\Psi Mt}M^{-1} \\ &= M\begin{bmatrix} e^{\lambda_2 t} & 0 \\ 0 & e^{\lambda_1 t} \end{bmatrix} M^{-1} \\ &= \frac{1}{2e^{\phi t}\phi}\begin{bmatrix} \phi(\psi(t)+1) + F(\psi(t)-1) & Q(1-\psi(t)) \\ \dfrac{H^2(\psi(t)-1)}{R} & F(1-\psi(t)) + \phi(1+\psi(t)) \end{bmatrix}, \\ \psi(t) &= e^{2\phi t} \end{aligned}$$

and the solution of the linearized system as

$$\begin{aligned} \begin{bmatrix} A(t) \\ B(t) \end{bmatrix} &= e^{\Psi t}\begin{bmatrix} P(0) \\ 1 \end{bmatrix} \\ &= \frac{1}{2e^{\phi t}\phi}\begin{bmatrix} P(0)[\phi(\psi(t)+1) + F(\psi(t)-1)] - \dfrac{Q(\psi(t)-1)}{R^2} \\ \dfrac{P(0)H^2(\psi(t)-1)}{R} - \phi(\psi(t)+1) - F(\psi(t)-1) \end{bmatrix}. \end{aligned}$$

5.8.3.3 General Solution of Scalar Time-Invariant Riccati Equation The general solution formula may now be composed from the previous results as

$$\begin{aligned} P(t) &= A(t)/B(t) \\ &= \frac{\mathcal{N}_P(t)}{\mathcal{D}_P(t)}, \end{aligned} \tag{5.155}$$

$$\begin{aligned} \mathcal{N}_P(t) &= R[P(0)(\phi + F) + Q] + R[P(0)(\phi - F) - Q]e^{-2\phi t} \\ &= R\left[P(0)\left(\sqrt{F^2 + \frac{H^2Q}{R}} + F\right) + Q\right] \\ &\quad + R\left[P(0)\left(\sqrt{F^2 + \frac{H^2Q}{R}} - F\right) - Q\right]e^{-2\phi t}, \end{aligned} \tag{5.156}$$

$$\begin{aligned} \mathcal{D}_P(t) &= [H^2P(0) + R(\phi - F)] - [H^2P(0) - R(F + \phi)]e^{-2\phi t} \\ &= \left[H^2P(0) + R\left(\sqrt{F^2 + \frac{H^2Q}{R}} - F\right)\right] \\ &\quad - \left[H^2P(0) - R\left(\sqrt{F^2 + \frac{H^2Q}{R}} + F\right)\right]e^{-2\phi t}. \end{aligned} \tag{5.157}$$

5.8.3.4 Singular Values of Denominator The denominator $\mathcal{D}_P(t)$ can easily be shown to have a zero for t_0 such that

$$e^{-2\phi t_0} = 1 + 2\frac{R}{H^2} \times \frac{H^2[P(0)\phi + Q] + FR(\phi - F)}{H^2P^2(0) - 2FRP(0) - QR}.$$

However, it can also be shown that $t_0 < 0$ if

$$P(0) > -\frac{R}{H^2}(\phi - F),$$

which is a nonpositive lower bound on the initial value. This poses no particular difficulty, however, since $P(0) \geq 0$ anyway. (We will see in the next section what would happen if this condition were violated.)

5.8.3.5 Boundary Values Given the above formulas for $P(t)$, its numerator $N(t)$, and its denominator $D(t)$, one can easily show that they have the following limiting values:

$$\lim_{t\to 0} N_P(t) = 2P(0)R\sqrt{F^2 + \frac{H^2Q}{R}},$$

$$\lim_{t\to 0} D_P(t) = 2R\sqrt{F^2 + \frac{H^2Q}{R}},$$

$$\lim_{t\to 0} P(t) = P(0),$$

$$\lim_{t\to\infty} P(t) = \frac{R}{H^2}\left(F + \sqrt{F^2 + \frac{H^2Q}{R}}\right). \tag{5.158}$$

5.8.4 Parametric Dependence of the Scalar Time-Invariant Solution

The previous solution of the scalar time-invariant problem will now be used to illustrate its dependence on the parameters F, H, R, Q, and $P(0)$. There are two fundamental algebraic functions of these parameters that will be useful in characterizing the behavior of the solutions: the asymptotic solution as $t \to \infty$ and the time constant of decay to this steady-state solution.

5.8.4.1 Decay Time Constant The only time-dependent terms in the expression for $P(t)$ are those involving $e^{-2\phi t}$. The fundamental decay time constant of the solution is then the algebraic function

$$\tau(F, H, R, Q) = 2\sqrt{F^2 + \frac{H^2Q}{R}} \tag{5.159}$$

of the problem parameters. Note that this function does not depend upon the initial value of P.

5.8.4.2 Asymptotic and Steady-State Solutions The asymptotic solution of the scalar time-invariant Riccati differential equation as $t \to \infty$ is given in Equation 5.158. It should be verified that this is also the solution of the corresponding *steady-state* differential equation

$$\dot{P} = 0,$$

$$P^2(\infty)H^2R^{-1} - 2FP(\infty) - Q = 0,$$

which is also called the *algebraic*[10] Riccati equation. This quadratic equation in $P(\infty)$ has two solutions, expressible as algebraic functions of the problem parameters:

$$P(\infty) = \frac{FR \pm \sqrt{H^2QR + F^2R^2}}{H^2}.$$

The two solutions correspond to the two values for the signum ($\pm$). There is no cause for alarm, however. The solution that agrees with Equation 5.158 is the nonnegative one. The other solution is nonpositive. We are only interested in the nonnegative solution, because the variance P of uncertainty is, by definition, nonnegative.

[10] So called because it is an algebraic equation, not a differential equation. That is, it is constructed from the operations of algebra, not those of the differential calculus. The term by itself is ambiguous in this usage, however, because there are two entirely different forms of the algebraic Riccati equation. One is derived from the Riccati *differential* equation, and the other is derived from the *discrete-time* Riccati equation. The results are both algebraic equations, but they are significantly different in structure.

5.8.4.3 Dependence on Initial Conditions For the scalar problem, the initial conditions are parameterized by $P(0)$. The dependence of the solution on its initial value is not continuous everywhere, however. The reason is that there are two solutions to the steady-state equation. The nonnegative solution is *stable* in the sense that initial conditions sufficiently near to it converge to it asymptotically. The nonpositive solution is *unstable* in the sense that infinitesimal perturbations of the initial condition cause the solution to diverge from the nonpositive steady-state solution and converge, instead, to the nonnegative steady-state solution.

5.8.4.4 Convergent and Divergent Solutions The *eventual* convergence of a solution to the nonnegative steady-state value may pass through infinity to get there. That is, the solution may initially *diverge*, depending on the initial values. This type of behavior is shown in Figure 5.7, which is a multiplot of solutions to an example of the Riccati equation with

$$F = 0, \quad H = 1, \quad R = 1, \quad Q = 1,$$

for which the corresponding continuous-time algebraic (quadratic) Riccati equation

$$\begin{aligned} \dot{P}(\infty) &= 0, \\ 2FP(\infty) - \frac{[P(\infty)H]^2}{R} + Q &= 0, \\ 1 - [P(\infty)]^2 &= 0 \end{aligned}$$

has the two solutions $P(\infty) = \pm 1$. The Riccati differential equation has the closed-form solution

$$P(t) = \frac{e^{2t}[1 + P(0)] - [1 - P(0)]}{e^{2t}[1 + P(0)] + [1 - P(0)]}$$

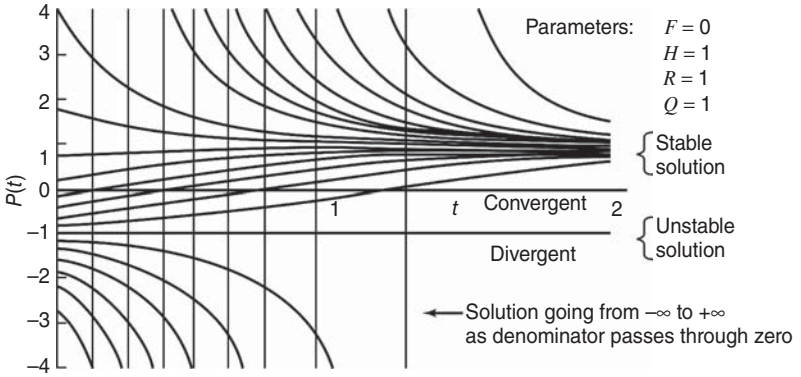

Figure 5.7 Solutions of the scalar time-invariant Riccati equation.

in terms of the initial value $P(0)$. Solutions of the initial-value problem with different initial values are plotted over the time interval $0 \le t \le 2$. All solutions except the one with $P(0) = -1$ appear to converge eventually to $P(\infty) = 1$, but those that disappear off the bottom of the graph diverge to $-\infty$, then converge to $P(\infty) = +1$ from the top of the graph. The vertical lines on the plot are the times at which solutions starting with $P(0) < -1$ pass through $P(t) = \pm\infty$ on their way to $P(\infty) = 1$. This phenomenon occurs at the zeros of the denominator in the expression for $P(t)$, which occur at time

$$t^* = \log_e \sqrt{\frac{P(0) - 1}{P(0) + 1}}$$

for $P(0) < -1$. Solutions with $P(0) > -1$ converge without this discontinuous behavior.

5.8.4.5 Convergent and Divergent Regions The line at $P = -1$ in Figure 5.7 separates initial values into two regions, characterized by the stability of solutions to the initial-value problem. Solutions with initial values above that line converge to the positive steady-state solution. Solutions starting below that line diverge.

5.8.5 Convergence Issues

It is usually risky to infer properties of high order systems from those of lower order. However, the following general trends are apparent in the behavior of the closed-form solution of the scalar time-invariant Riccati differential equation:

1. The solution eventually converges exponentially to the nonnegative steady-state solution. The decay time constant varies as $(F^2 + H^2Q/R)^{1/2}$, which increases with $|F|$, $|H|$, and Q and decreases as R increases (for $R > 0$ and $Q > 0$).
2. Solutions are not uniformly exponentially convergent, however. The initial value does not influence the *asymptotic* decay rate, but it can influence the initial response. In particular, convergence for initial values nearer the unstable steady-state solution is hampered initially.
3. The stable asymptotic solution is

$$P(\infty) = \frac{R}{H^2}\left(F + \sqrt{F^2 + \frac{H^2Q}{R}}\right),$$

which is influenced by both the sign *and* magnitude of F but only by the magnitudes of H, R, and Q.

Stability properties of general (higher order) systems have been proved by Potter [38].

5.8.5.1 Even Unstable Dynamic Systems Have Convergent Riccati Equations Note that the corresponding equation for the variance of the *state*

$$\frac{d}{dt}P(t) = FP + PF^{\mathrm{T}} + Q$$

has the general solution

$$P(t) = \frac{(e^{2Ft} - 1)Q}{2F} + e^{2Ft}P(0)$$

in the scalar case. This dynamic system is unstable if $F > 0$, because the solution $P(t) \to +\infty$ as $t \to \infty$. However, the corresponding Riccati equation (with the conditioning term) approaches a finite limit.

5.8.6 Closed-Form Solution of the Algebraic Riccati Equation

We have seen in the previous subsections the difficulty of obtaining a solution of the general Riccati differential equation in "closed form" (i.e., as a formula in the parameters of the model), even for the simplest (scalar) problem. The following example illustrates the difficulty of obtaining closed-form solutions for the *algebraic* Riccati equation, as well, for a simple model.

Example 5.6 (Solving the Algebraic Riccati Equation in Continuous Time for the Harmonic Resonator Problem) The problem is to characterize the asymptotic uncertainty in estimating the state (position and velocity) of a damped harmonic resonator driven by white noise, given noisy measurements of position. The system model for this problem has been derived in Examples 2.2, 2.3, 2.6, 2.7, 3.9, 3.10, and 3.11. The resulting algebraic Riccati equation for this problem in continuous time has the form

$$\begin{aligned}
0 &= FP + PF^{\mathrm{T}} - PH^{\mathrm{T}}R^{-1}HP + Q,\\
F &= \begin{bmatrix} 0 & 1 \\ -\dfrac{1+\omega^2\tau^2}{\tau^2} & \dfrac{-2}{\tau} \end{bmatrix},\\
H &= \begin{bmatrix} 1 & 0 \end{bmatrix},\\
Q &= \begin{bmatrix} 0 & 0 \\ 0 & q \end{bmatrix},
\end{aligned}$$

which is equivalent to the three scalar equations

$$\begin{aligned}
0 &= -p_{11}^2 + 2Rp_{12},\\
0 &= -R(1+\omega^2\tau^2)p_{11} - 2R\tau p_{12} - \tau^2 p_{11}p_{12} + R\tau^2 p_{22},\\
0 &= -\tau^2 p_{12}^2 - 2R(1+\omega^2\tau^2)p_{12} - 4R\tau p_{22} + Rq\tau^2.
\end{aligned}$$

The first and last of these can be solved as linear equations in the variables p_{12} and p_{22}

$$p_{12} = \frac{p_{11}^2}{2R},$$
$$p_{22} = \frac{Rq\tau^2 - \tau^2 p_{12}^2 - 2R(1+\omega^2\tau^2)p_{12}}{4R\tau}$$

in terms of p_{11}. Substitution of these expressions into the middle scalar equation yields the following quartic equation in p_{11}:

$$0 = \tau^3 p_{11}^4 + 8R\tau^2 p_{11}^3 + 20R^2\tau(5+\omega^2\tau^2)p_{11}^2 + 16R^3(1+\omega^2\tau^2)p_{11} - 4R^3 q\tau^3.$$

This may appear to be a relatively simple quartic equation, but its solution is a rather laborious and tedious process. It has four solutions, only one of which yields a nonnegative covariance matrix P:

$$p_{11} = \frac{R(1-b)}{\tau},$$
$$p_{12} = \frac{R(1-b)^2}{2\tau^2},$$
$$p_{22} = \frac{R}{\tau^3}(-6+2 \quad \omega^2\tau^2 - 4a + (4+a)b),$$
$$a = \sqrt{(1+\omega^2\tau^2)^2 + \frac{q\tau^4}{R}}, \quad b = \sqrt{2(1-\omega^2\tau^2+a)}.$$

Because there is no general formula for solving higher order polynomial equations (i.e., beyond quartic), this relatively simple example is at the limit of complexity for finding closed-form solutions to algebraic Riccati equations by purely algebraic means. Beyond this relatively low level of complexity, it is necessary to employ numerical solution methods. Numbers do not always provide us as much insight into the characteristics of the solution as formulas do, but they are all we can get for most problems of practical significance.

5.8.7 Newton–Raphson Solution of the Algebraic Riccati Differential Equation

The *Newton–Raphson solution* of n differentiable functional equations

$$0 = f_1(x_1, x_2, x_3, \ldots, x_n),$$
$$0 = f_2(x_1, x_2, x_3, \ldots, x_n),$$
$$0 = f_3(x_1, x_2, x_3, \ldots, x_n),$$

$$\vdots$$

$$0 = f_n(x_1, x_2, x_3, \ldots, x_n)$$

in n unknowns $x_1, x_2, x_3, \ldots, x_n$ is the iterative vector procedure

$$x \leftarrow x - \mathcal{F}^{-1} f(x) \tag{5.160}$$

using the vector and matrix variables

$$x = \begin{bmatrix} x_1 & x_2 & x_3 & \cdots & x_n \end{bmatrix}^{\mathrm{T}},$$

$$f(x) = \begin{bmatrix} f_1(x) & f_2(x) & f_3(x) & \cdots & f_n(x) \end{bmatrix}^{\mathrm{T}},$$

$$\mathcal{F} = \begin{bmatrix} \dfrac{\partial f_1}{\partial x_1} & \dfrac{\partial f_1}{\partial x_2} & \dfrac{\partial f_1}{\partial x_3} & \cdots & \dfrac{\partial f_1}{\partial x_n} \\ \dfrac{\partial f_2}{\partial x_1} & \dfrac{\partial f_2}{\partial x_2} & \dfrac{\partial f_2}{\partial d_3} & \cdots & \dfrac{\partial f_2}{\partial x_n} \\ \dfrac{\partial f_3}{\partial x_1} & \dfrac{\partial f_3}{\partial x_2} & \dfrac{\partial f_3}{\partial x_3} & \cdots & \dfrac{\partial f_3}{\partial x_n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial f_n}{\partial x_1} & \dfrac{\partial f_n}{\partial x_2} & \dfrac{\partial f_n}{\partial x_3} & \cdots & \dfrac{\partial f_n}{\partial x_n} \end{bmatrix}.$$

Application of this vector-oriented procedure to matrix equations is generally done by "vectorizing" the matrix of unknowns and using Kronecker products to "matricize" F from what would otherwise be four-dimensional data structures. However, the general approach does not take advantage of the symmetry constraints in the matrix Riccati differential equation. There are two such constraints: one on the symmetry of the Riccati equation itself and the other on the symmetry of the solution, P. Therefore, in solving the steady-state $n \times n$ matrix Riccati differential equation, there are effectively only $n(n+1)/2$ independent scalar equations in $n(n+1)/2$ scalar unknowns. The $n(n+1)/2$ scalar unknowns can be taken as the upper-triangular elements of P, and the $n(n+1)/2$ scalar equations can be taken as those equating upper-triangular terms of the matrix equation. We will first describe the equations by which the matrix equation and the matrix unknown can be vectorized and then show the form that the variables of the Newton–Raphson solution will take for this vectorization.

5.8.7.1 Vectorizing Formulas If one lets the indices i and j stand for the row and column, respectively, of the terms in the matrix Riccati equation, then the respective elements of the *upper-triangular* parts of the matrix equation can be vectorized by the single index p, where

$$1 \le j \le n,$$

$$1 \le i \le j,$$
$$p = \frac{1}{2}j(j-1) + i,$$
$$1 \le p \le \frac{1}{2}n(n+1).$$

Similarly, the upper-triangular part of P can be mapped into a singly subscripted array x with index q, according to the rules

$$1 \le \ell \le n,$$
$$1 \le k \le \ell,$$
$$q = \frac{1}{2}\ell(\ell-1) + k,$$
$$1 \le q \le \frac{1}{2}n(n+1),$$

whereby $P_{k\ell}$ is mapped into x_q.

5.8.7.2 Values of Variables for Newton–Raphson Solution of Steady-State Matrix Riccati Differential Equation The solution is an implementation of the recursion formula 5.74 with

$$f_p = Z_{ij}, \tag{5.161}$$
$$Z = FP + PF^{\mathrm{T}} - PH^{\mathrm{T}}R^{-1}HP + Q, \tag{5.162}$$
$$x_q = P_{k\ell}, \tag{5.163}$$
$$p = \frac{1}{2}j(j-1) + i, \tag{5.164}$$
$$q = \frac{1}{2}\ell(\ell-1) + k, \tag{5.165}$$
$$\begin{aligned}\mathcal{F}_{pq} &= \frac{\partial f_p}{\partial x_q} \\ &= \frac{\partial Z_{ij}}{\partial P_{k\ell}} \\ &= \Delta_{j\ell}\mathcal{S}_{ik} + \Delta_{ik}\mathcal{S}_{j\ell},\end{aligned} \tag{5.166}$$
$$\mathcal{S} = F - PH^{\mathrm{T}}R^{-1}H, \tag{5.167}$$
$$\Delta_{ab} \overset{\text{def}}{=} \begin{cases} 1 & \text{if} \quad a = b \\ 0 & \text{if} \quad a \ne b. \end{cases} \tag{5.168}$$

The least obvious of these is Equation 5.166, which will now be derived.

5.8.7.3 "Dot" Notation for Row and Column Submatrices For any matrix M, let the notation $M_{\cdot j}$ (with a dot $(\cdot)$ where the row index should be) stand for the jth column of M. When this notation is applied to the identity matrix $I, I_{\cdot j}$ will equal a column vector with 1 in the jth row and zeros elsewhere. As a vector, it has the property that

$$MI_{\cdot j} = M_{\cdot j}$$

for any conformable matrix M.

5.8.7.4 Matrix Partial Derivatives With this notation, one can write matrix partial derivatives as follows:

$$\frac{\partial P}{\partial P_{k\ell}} = I_{\cdot k} I_{\cdot \ell}^{\mathrm{T}}, \tag{5.169}$$

$$\frac{\partial Z}{\partial P_{k\ell}} = F\frac{\partial P}{\partial P_{k\ell}} + \frac{\partial P}{\partial P_{k\ell}}F^{\mathrm{T}} - \frac{\partial P}{\partial P_{k\ell}}H^{\mathrm{T}}R^{-1}HP - PH^{\mathrm{T}}R^{-1}H\frac{\partial P}{\partial P_{k\ell}} \tag{5.170}$$

$$= FI_{\cdot k}I_{\cdot \ell}^{\mathrm{T}} + I_{\cdot k}I_{\cdot \ell}^{\mathrm{T}}F^{\mathrm{T}} - I \cdot kI_{\cdot \ell}^{\mathrm{T}}H^{\mathrm{T}}R^{-1}HP - PH^{\mathrm{T}}R^{-1}H\frac{\partial \partial}{\partial P_{k\ell}} \tag{5.171}$$

$$= F_{\cdot k}I_{\cdot \ell}^{\mathrm{T}} + I_{\cdot k}F_{\cdot \ell}^{\mathrm{T}} - I_{\cdot k}M_{\cdot \ell}^{\mathrm{T}} - M_{\cdot k}I_{\cdot \ell}^{\mathrm{T}} \tag{5.172}$$

$$= (F - M)_{\cdot k}I_{\cdot \ell}^{\mathrm{T}} + I_{\cdot k}(F - M)_{\cdot \ell}^{\mathrm{T}} \tag{5.173}$$

$$= S_{\cdot k}I_{\cdot \ell}^{\mathrm{T}} + I_{\cdot k}S_{\cdot \ell}^{\mathrm{T}}, \tag{5.174}$$

$$S = F - M, \tag{5.175}$$

$$M = PH^{\mathrm{T}}R^{-1}H. \tag{5.176}$$

Note that on the right-hand side of Equation 5.174, the first term $(S_{\cdot k}I_{\cdot \ell}^{\mathrm{T}})$ has only one nonzero column—the ℓth column. Similarly, the other term $(I_{\cdot k}S_{\cdot \ell}^{\mathrm{T}})$ has only one nonzero row—its kth row. Consequently, the element in the ith row and jth column of this matrix will be the expression given in Equation 5.166. This completes its derivation.

5.8.7.5 Computational Complexity The number of floating-point operations per iteration for this solution method is dominated by the inversion of the $n(n+1)/2 \times n(n+1)/2$ matrix $\mathcal{F}$, which requires somewhat more than $n^6/8$ flops.

5.8.8 MacFarlane–Potter–Fath Eigenstructure Method

5.8.8.1 Steady-State Solution of Time-Invariant Matrix Riccati Differential Equation It was discovered independently by MacFarlane [39], Potter [40], and Fath [41] that the solution $P(\infty)$ of the continuous-time form of the steady-state matrix Riccati differential equation can be expressed in the form

$$P(\infty) = AB^{-1},$$

$$\begin{bmatrix} A \\ B \end{bmatrix} = \begin{bmatrix} e_{i_1} & e_{i_2} & e_{i_3} & \cdots & e_{i_n} \end{bmatrix},$$

where the matrices A and B are $n \times n$ and the $2n$-vectors e_{i_k} are characteristic vectors of the continuous-time system Hamiltonian matrix

$$\Psi_c = \begin{bmatrix} F & Q \\ H^{\mathrm{T}} R^{-1} H & -F^{\mathrm{T}} \end{bmatrix}.$$

This can be formalized in somewhat greater generality as a lemma.

Lemma 5.1 If A and B are $n \times n$ matrices such that B is nonsingular and

$$\Psi_c \begin{bmatrix} A \\ B \end{bmatrix} = \begin{bmatrix} A \\ B \end{bmatrix} D \tag{5.177}$$

for an $n \times n$ matrix D, then $P = AB^{-1}$ satisfies the steady-state matrix Riccati differential equation

$$0 = FP + PF^{\mathrm{T}} - PH^{\mathrm{T}} R^{-1} HP + Q.$$

Proof Equation 5.177 can be written as two equations,

$$AD = FA + QB, \quad BD = H^{\mathrm{T}} R^{-1} HA - F^{\mathrm{T}} B.$$

If one multiplies both of these on the right by B^{-1} and the last of these on the left by AB^{-1}, one obtains the equivalent equations

$$\begin{aligned} ADB^{-1} &= FAB^{-1} + Q, \\ ADB^{-1} &= AB^{-1} H^{\mathrm{T}} R^{-1} HAB^{-1} - AB^{-1} F^{\mathrm{T}}, \end{aligned}$$

or, taking the differences of the left-hand sides and substituting P for AB^{-1},

$$0 = FP + PF^{\mathrm{T}} - PH^{\mathrm{T}} R^{-1} HP + Q,$$

which was to be proved.

In the case that A and B are formed in this way from n characteristic vectors of Ψ_c, the matrix D will be a diagonal matrix of the corresponding characteristic values. (Check it out for yourself.) Therefore, to obtain the steady-state solution of the matrix Riccati differential equation by this method, it suffices to find n characteristic vectors of Ψ_c such that the corresponding B-matrix is nonsingular. (As will be shown in the next section, the same trick works for the discrete-time matrix Riccati equation.)

5.9 MATRIX RICCATI EQUATION IN DISCRETE TIME

5.9.1 Linear Equations for Matrix Fraction Propagation

The representation of the covariance matrix as a matrix fraction is also sufficient to transform the nonlinear discrete-time Riccati equation for the estimation uncertainty into a linear form. The discrete-time problem differs from the continuous-time problem in two important aspects:

1. The numerator and denominator matrices will be propagated by a $2n \times 2n$ transition matrix and not by differential equations. The approach is otherwise similar to that for the continuous-time Riccati equation, but the resulting $2n \times 2n$ state-transition matrix for the recursive updates of the numerator and denominator matrices is a bit more complicated than the coefficient matrix for the linear form of the continuous-time matrix Riccati equation.
2. There are two distinct values of the discrete-time covariance matrix at any discrete-time step—the a priori value and the a posteriori value. The a priori value is of interest in computing Kalman gains, and the a posteriori value is of interest in the analysis of estimation uncertainty.

The linear equations for matrix fraction propagation of the a priori covariance matrix are derived below. The method is then applied to obtain a closed-form solution for the scalar time-invariant Riccati equation in discrete time and to a method for exponential speedup of convergence to the asymptotic solution.

5.9.2 Matrix Fraction Propagation of the *a priori* Covariance

Lemma 5.2 If the state-transition matrices Φ_k are nonsingular and

$$P_{k(-)} = A_k B_k^{-1} \tag{5.178}$$

is a nonsingular matrix solution of the discrete-time Riccati equation at time t_k, then

$$P_{k+1}(-) = A_{k+1} B_{k+1}^{-1} \tag{5.179}$$

is a solution at time t_{k+1}, where

$$\begin{bmatrix} A_{k+1} \\ B_{k+1} \end{bmatrix} = \begin{bmatrix} Q_k & I \\ I & 0 \end{bmatrix} \begin{bmatrix} \Phi_k^{-\mathrm{T}} & 0 \\ 0 & \Phi_k \end{bmatrix} \begin{bmatrix} H_k^{\mathrm{T}} R_k^{-1} H_k & I \\ I & 0 \end{bmatrix} \begin{bmatrix} A_k \\ B_k \end{bmatrix} \tag{5.180}$$

$$= \begin{bmatrix} \Phi_k + Q_k \Phi_k^{-\mathrm{T}} H_k^{\mathrm{T}} R_k^{-1} H_k & Q_k \Phi_k^{-\mathrm{T}} \\ \Phi_k^{-\mathrm{T}} H_k^{\mathrm{T}} R_k^{-1} H_k & \Phi_k^{-\mathrm{T}} \end{bmatrix} \begin{bmatrix} A_k \\ B_k \end{bmatrix}. \tag{5.181}$$

Proof The following annotated sequence of equalities starts with the product $A_{k+1}B_{k+1}^{-1}$ as defined and proves that it equals P_{k+1}:

$$
\begin{aligned}
A_{k+1}B_{k+1}^{-1} &= \{[\Phi_k + Q_k\Phi_k^{-\mathrm{T}}H_k^{-\mathrm{T}}R_k^{-1}H_k]A_k + Q_k\Phi_k^{-\mathrm{T}}B_k\} \\
&\quad \times \{\Phi_k^{-\mathrm{T}}[H_k^{\mathrm{T}}R_k^{-1}H_kA_kB_k^{-1} + I]B_k\}^{-1} \qquad \text{(definition)} \\
&= \{[\Phi_k + Q_k\Phi_k^{-\mathrm{T}}H_k^{\mathrm{T}}R_k^{-1}H_k]A_k + Q_k\Phi_k^{-\mathrm{T}}B_k\} \\
&\quad \times B_k^{-1}\{H_k^{\mathrm{T}}R_k^{-1}H_kA_kB_k^{-1} + I\}^{-1}\Phi_k^{\mathrm{T}} \qquad \text{(factor } B_k\text{)} \\
&= \{[\Phi_k + Q_k\Phi_k^{-\mathrm{T}}H_k^{\mathrm{T}}R_k^{-\mathrm{T}}H_k]A_kB_k^{-1} + Q_k\Phi_k^{-\mathrm{T}}\} \\
&\quad \times \{H_k^{\mathrm{T}}R_k^{-1}H_kA_kB_k^{-1} + I\}^{-1}\Phi_k^{\mathrm{T}} \qquad \text{(distribute } B_k\text{)} \\
&= \{[\Phi_k + Q_k\Phi_k^{-\mathrm{T}}H_k^{\mathrm{T}}R_k^{-1}H_k]P_{k(-)} + Q_k\Phi_k^{-\mathrm{T}}\} \\
&\quad \times \{H_k^{\mathrm{T}}R_k^{-1}H_kP_{k(-)} + I\}^{-1}\Phi_k^{\mathrm{T}} \qquad \text{(definition)} \\
&= \{\Phi_kP_{k(-)} + Q_k\Phi_k^{-\mathrm{T}}[H_k^{\mathrm{T}}R_k^{-1}H_kP_{k(-)} + I]\} \\
&\quad \times \{H_k^{\mathrm{T}}R_k^{-\mathrm{T}}H_kP_{k(-)} + I\}^{-1}\Phi_k^{\mathrm{T}} \qquad \text{(regroup)} \\
&= \Phi_kP_{k(-)}\{H_k^{\mathrm{T}}R_k^{-\mathrm{T}}H_kP_{k(-)} + I\}^{-1}\Phi_k^{\mathrm{T}} + Q_k\Phi_k^{-\mathrm{T}}\Phi_k^{\mathrm{T}} \quad \text{(distribute)} \\
&= \Phi_k\{H_k^{\mathrm{T}}R_k^{-1}H_k + P_k^{-1}\}^{-1}\Phi_k^{\mathrm{T}} + Q_k \\
&= \Phi_k\{P_{k(-)} - P_{k(-)}H_k^{\mathrm{T}}[H_kP(-)H_k^{\mathrm{T}} + R_k]^{-1} \\
&\quad \times H_kP_{k(-)}\}\Phi_k^{\mathrm{T}} + Q_k \qquad \text{(Hemes)} \\
&= P_{k+1(-)} \qquad \text{(Riccati),}
\end{aligned}
$$

where the "Hemes inversion formula" is given in Appendix B on the companion Wiley web site. This completes the proof.

This lemma is used below to derive a closed-form solution for the steady-state Riccati equation in the scalar time-invariant case and in Chapter 8 to derive a fast iterative solution method for the matrix time-invariant case.

5.9.3 Closed-Form Solution of the Scalar Time-Invariant Case

Because this case can be solved in closed form, it serves to illustrate the application of the linearization method derived above.

5.9.3.1 Characteristic Values and Vectors The linearization will yield the following 2×2 transition matrix for the numerator and denominator matrices representing the covariance matrix as a matrix fraction:

$$\Psi = \begin{bmatrix} Q_k & I \\ I & 0 \end{bmatrix} \begin{bmatrix} \Phi_k^{-\mathrm{T}} & 0 \\ 0 & \Phi_k \end{bmatrix} \begin{bmatrix} H_k^{\mathrm{T}} R_k^{-1} H_k & I \\ I & 0 \end{bmatrix}$$
$$= \begin{bmatrix} \Phi + \dfrac{H^2 Q}{\Phi R} & \dfrac{Q}{\Phi} \\ \dfrac{H^2}{\Phi R} & \dfrac{1}{\Phi} \end{bmatrix}.$$

This matrix has characteristic values

$$\lambda_1 = \frac{H^2 Q + R(\Phi^2 + 1) + \sigma}{2\Phi R}, \quad \lambda_2 = \frac{H^2 Q + R(\Phi^2 + 1) - \sigma}{2\Phi R},$$
$$\sigma = \sigma_1 \sigma_2,$$
$$\sigma_1 = \sqrt{H^2 Q + R(\Phi + 1)^2}, \quad \sigma_2 = \sqrt{H^2 Q + R(\Phi - 1)^2},$$

with ratio

$$\rho = \frac{\lambda_2}{\lambda_1}$$
$$= \frac{\psi - [H^2 Q + R(\Phi^2 + 1)]\sigma}{2\Phi^2 R^2}$$
$$\leq 1,$$
$$\psi = [H^2 Q + R(\Phi^2 + 1)]^2 - 2R^2\Phi^2$$
$$= H^4 Q^2 + 2H^2 QR + 2H^2\Phi^2 QR + R^2 + \Phi^4 R^2.$$

The corresponding characteristic vectors are the column vectors of the matrix

$$M = \begin{bmatrix} \dfrac{-2QR}{H^2 QR(\Phi^2 - 1) + \sigma} & \dfrac{-2QR}{H^2 QR(\Phi^2 - 1) - \sigma} \\ 1 & 1 \end{bmatrix},$$

the inverse of which is

$$M^{-1} = \begin{bmatrix} -\dfrac{H^2}{\sigma_2 \sigma_1} & \dfrac{H^2 Q - R + \Phi^2 R + \sigma_2 \sigma_1}{2\sigma_2 \sigma_1} \\ \dfrac{H^2}{\sigma_2 \sigma_1} & \dfrac{-(H^2 Q) + R - \Phi^2 R + \sigma_2 \sigma_1}{2\sigma_2 \sigma_1} \end{bmatrix}$$
$$= \frac{1}{4QR\sigma_1 \sigma_2} \begin{bmatrix} \tau_1 \tau_2 & 2QR\tau_1 \\ -\tau_1 \tau_2 & -2QR\tau_2 \end{bmatrix},$$
$$\tau_1 = H^2 Q + R(\Phi^2 - 1) + \sigma, \quad \tau_2 = H^2 Q + R(\Phi^2 - 1) - \sigma.$$

5.9.3.2 Closed-Form Solution This will have the form

$$P_k = A_k B_K^{-\mathrm{T}}$$

for

$$\begin{bmatrix} A_k \\ B_k \end{bmatrix} = \Psi^k \begin{bmatrix} P_0 \\ 1 \end{bmatrix} = M \begin{bmatrix} \lambda_1^k & 0 \\ 0 & \lambda_2^k \end{bmatrix} M^{-1} \begin{bmatrix} P_0 \\ 1 \end{bmatrix}.$$

This can be expressed in the form

$$P_k = \frac{(P_0\tau_2 + 2QR) - (P_0\tau_1 + 2QR)\rho^k}{(2H^2P_0 - \tau_1) - (2H^2P_0 - \tau_2)\rho^k},$$

which is similar in structure to the closed-form solution for the scalar time-invariant Riccati differential equation. In both cases, the solution is a ratio of linear functions of an exponential time function. In the discrete-time case, the discrete-time power ρ^k serves essentially the same function as the exponential function $e^{-2\phi t}$ in the closed-form solution of the differential equation. Unlike the continuous-time solution, however, this discrete-time solution can "skip over" zeros of the denominator.

5.9.4 MacFarlane–Potter–Fath Eigenstructure Method

5.9.4.1 Steady-State Solution of Time-Invariant Discrete-Time Matrix Riccati Equation The method presented in Section 5.8.8 for the steady-state solution of the time-invariant matrix Riccati differential equation (i.e., in continuous time) also applies to the Riccati equation in discrete time. As before, it is formalized as a lemma.

Lemma 5.3 If A and B are $n \times n$ matrices such that B is nonsingular and

$$\Psi_d \begin{bmatrix} A \\ B \end{bmatrix} = \begin{bmatrix} A \\ B \end{bmatrix} D \tag{5.182}$$

for an $n \times n$ nonsingular matrix D, then $P_\infty = AB^{-1}$ satisfies the steady-state discrete-time matrix Riccati equation

$$P_\infty = \Phi\{P_\infty - P_\infty H^{\mathrm{T}}[HP_\infty H^{\mathrm{T}} + R]^{-1} HP_\infty\}\Phi^{\mathrm{T}} + Q.$$

Proof If $P_k = AB^{-1}$, then it was shown in Lemma 5.2 that $P_{k+1} = \acute{A}\acute{B}^{-1}$, where

$$\begin{bmatrix} \acute{A} \\ \acute{B} \end{bmatrix} = \begin{bmatrix} (\Phi_k + Q_k\Phi_k^{-\mathrm{T}}H_k^{\mathrm{T}}R_k^{-1}H_k) & Q_k\Phi_k^{-\mathrm{T}} \\ \Phi_k^{-\mathrm{T}}H_k^{\mathrm{T}}R_k^{-1}H_k & \Phi_k^{-\mathrm{T}} \end{bmatrix} \begin{bmatrix} A \\ B \end{bmatrix}$$

$$= \Psi_d \begin{bmatrix} A \\ B \end{bmatrix}$$

$$= \begin{bmatrix} A \\ B \end{bmatrix} D$$

$$= \begin{bmatrix} AD \\ BD \end{bmatrix}.$$

Consequently,

$$\begin{aligned} P_{k+1} &= \acute{A}\acute{B}^{-1} \\ &= (AD)(BD)^{-1} \\ &= ADD^{-1}B^{-1} \\ &= AB^{-1} \\ &= P_k. \end{aligned}$$

That is, AB^{-1} is a steady-state solution, which was to be proved.

In practice, A and B are formed from n characteristic vectors of Ψ_d. The matrix D will be a diagonal matrix of the corresponding nonzero characteristic values.

5.10 MODEL EQUATIONS FOR TRANSFORMED STATE VARIABLES

The question to be addressed here is: What happens to the Kalman filter model equations when the state variables and measurement variables are redefined by linear transformations? The answer to this question can be derived as a set of formulas, giving the new model equations in terms of the parameters of "old" model equations and the linear transformations relating the two sets of variables. In Chapter 8, these formulas will be used to simplify the model equations.

5.10.1 Linear Transformations of State Variables

These are changes of variables by which the "new" state and measurement variables are linear combinations of the respective old state and measurement variables. Such transformations can be expressed in the form

$$\acute{x}_k = A_k x_k, \tag{5.183}$$

$$\acute{z}_k = B_k H_k, \tag{5.184}$$

where x and z are the old variables and $\acute{x}$ and $\acute{z}$ are the new state vector and measurement, respectively.

5.10.1.1 Matrix Constraints One must further assume that for each discrete-time index k, A_k is a *nonsingular* $n \times n$ matrix. The requirements on B_k are less stringent. One need only assume that it is conformable for the product $B_k H_k$, that is, that B_k is a matrix with ℓ columns. The dimension ℓ of $\acute{z}_k$ is arbitrary and can depend on k.

5.10.2 New Model Equations

With the above assumptions, the corresponding state, measurement, and state uncertainty covariance equations of the Kalman filter model are transformed to the following forms:

$$\acute{x}_{k+1} = \acute{\Phi}_k \acute{x}_k + \acute{\omega}_k, \tag{5.185}$$

$$\acute{z} = \acute{H}_k \acute{x}_k + \acute{\upsilon}_k, \tag{5.186}$$

$$\acute{P}_{k(+)} = \acute{P}_{k(-)} - \acute{P}_{k(-)} \acute{H}_k [\acute{H}_k \acute{P}_{k(-)} \acute{H}_k^{\mathrm{T}} + \acute{R}_k] \acute{H}_k \acute{P}_{k(-)}, \tag{5.187}$$

$$\acute{P}_{k+1(-)} = \acute{\Phi}_k \acute{P}_{k(+)} \acute{\Phi}_k^{\mathrm{T}} + \acute{Q}_k, \tag{5.188}$$

where the new model parameters are

$$\acute{\Phi}_k = A_k \Phi_k A_k^{-1}, \tag{5.189}$$

$$\acute{H}_k = B_k H_k A_k^{-1}, \tag{5.190}$$

$$\acute{Q}_k = \mathrm{E}\langle \acute{\omega}_k \acute{\omega}_k^{\mathrm{T}} \rangle \tag{5.191}$$

$$= A_k Q_k A_k^{\mathrm{T}}, \tag{5.192}$$

$$\acute{R} = \mathrm{E}\langle \acute{\upsilon}_k \acute{\upsilon}_k^{\mathrm{T}} \rangle \tag{5.193}$$

$$= B_k R_k B_k^{\mathrm{T}}, \tag{5.194}$$

and the new state estimation uncertainty covariance matrices are

$$\acute{P}_{k(\pm)} = A_k P_{k(\pm)} A_k^{\mathrm{T}}. \tag{5.195}$$

5.11 SAMPLE APPLICATIONS

The Kalman filter has been applied to inertial navigation [6, 42, 43], sensor calibration [44], radar tracking [8], manufacturing [18], economics [12], signal processing [18], and freeway traffic modeling [45]—to cite a few examples. This section shows some applications of the programs provided on the companion web site. A simple example of a second-order underdamped oscillator is given here to illustrate the application of the equations in Table 5.3. This harmonic oscillator is an approximation of a longitudinal dynamics of an aircraft short period [46].

Example 5.7 (Resonator Tracking) Consider a linear, underdamped, second-order system with displacement $x_1(t)$, rate $x_2(t)$, damping ratio ζ and (undamped) natural frequency of 5 rad/s, and constant driving term of 12.0 with additive white noise $w(t)$ normally distributed. The second-order continuous-time dynamic equation

$$\ddot{x}_1(t) + 2\zeta w\dot{x}_1(t) + \omega^2 x_1(t) = 12 + w(t)$$

can be written in state-space form via state-space techniques of Chapter 2:

$$\begin{bmatrix}\dot{x}_1(t)\\ \dot{x}_2(t)\end{bmatrix} = \begin{bmatrix}0 & 1\\ -\omega^2 & -2\zeta\omega\end{bmatrix}\begin{bmatrix}x_1(t)\\ x_2(t)\end{bmatrix} + \begin{bmatrix}0\\ 1\end{bmatrix} w(t) + \begin{bmatrix}0\\ 12\end{bmatrix}.$$

The observation equation is

$$z(t) = x_1(t) + v(t).$$

One can generate a trajectory of 100 simulated data points with plant noise and measurement noise equal to zero using the following initial condition and parameter values:

$$\begin{bmatrix}x_1(0)\\ x_2(0)\end{bmatrix} = \begin{bmatrix}0 & \text{ft}\\ 0 & \text{ft/s}\end{bmatrix},$$

$$P(0) = \begin{bmatrix}2 & 0\\ 0 & 2\end{bmatrix},$$

$$Q = 4.47(\text{ft/s})^2, \quad R = 0.01(\text{ft})^2,$$

$$\zeta = 0.2, \quad \omega = 5 \quad \text{rad/s}.$$

Equations 5.21, 5.24, 5.25, and 5.26 were programmed in MATLAB® software on a PC (see Appendix A) to estimate $\hat{x}_1(t)$ and $\hat{x}_2(t)$. Figure 5.8 shows the resulting estimates of position and velocity using the noise-free data generated from simulating the above second-order equation. The estimated and actual values are identical in this case. Figure 5.9 shows the corresponding RMS uncertainties in position and velocity (Figure 5.9(a)), correlation coefficient between position and velocity (Figure 5.9(b)), and Kalman gains (Figure 5.9(c)). These results were generated from the accompanying MATLAB program exam57.m described in Appendix A with sample interval equal to 1 s.

Example 5.8 (Radar Tracking) This example is that of a pulsed *radar tracking system*. In this system, radar pulses are sent out and return signals are processed by the Kalman filter in order to determine the position of maneuvering airborne objects [47]. This example's equations are drawn from IEEE papers [48, 49].

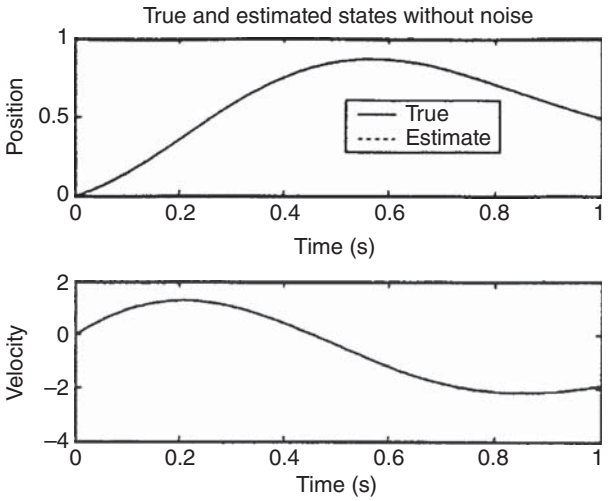

Figure 5.8 Estimated position (ft) and velocity (ft/s) versus time (s).

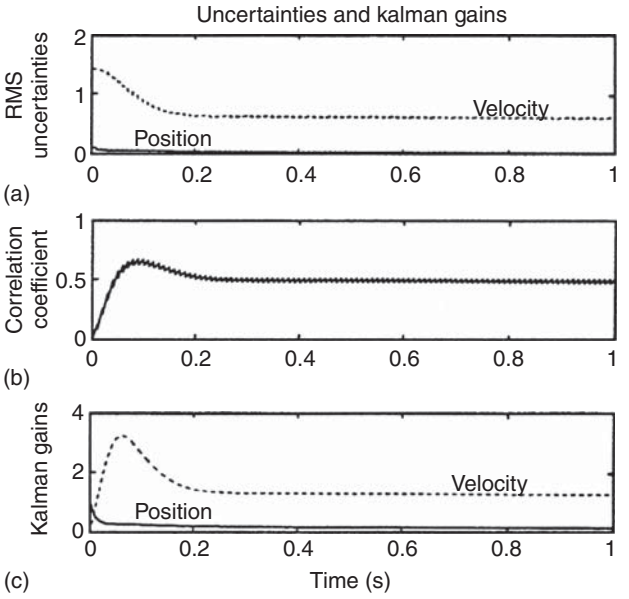

Figure 5.9 RMS uncertainties, position and velocity, correlation coefficient, and Kalman gain.

Difference equations of dynamics equations in state-space formulation are

$$x_k = \begin{bmatrix} 1 & T & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & \rho & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & T & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & \rho \end{bmatrix} x_{k-1} + \begin{bmatrix} 0 \\ 0 \\ w_{k-1}^1 \\ 0 \\ 0 \\ w_{k-1}^2 \end{bmatrix}.$$

The discrete-time observation equation is given by

$$z_k = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix} x_k + \begin{bmatrix} v_k^1 \\ v_k^2 \end{bmatrix},$$

where

$x_k^{\mathrm{T}} = [r_k \quad \dot{r}_k \quad U_k^1 \quad \theta_k \quad \dot{\theta}_k \quad U_k^2]$

r_k is the range of the vehicle at time k

$\dot{r}_k$ is the range rate of the vehicle at time k

U_k^1 is the maneuvering-correlated state noise

θ_k is the bearing of the vehicle at time k

$\dot{\theta}_k$ is the bearing rate of the vehicle at time k

U_k^2 is the maneuvering-correlated state noise

T is the sampling period in seconds

$w_{\mathrm{k}}^{\mathrm{T}} = [w_k^1 w_k^2]$ is the zero-mean white-noise sequences and covariance of σ_1^2 and σ_2^2, respectively

$v_k^{\mathrm{T}} = [v_k^1 v_k^2]$ is the sensor zero-mean white-noise sequence and covariance of σ_r^2 and σ_θ^2, respectively, and w_k and v_k are uncorrelated:

$$\rho = \text{correlation coefficient} = \frac{\mathrm{E}[U_k U_{k-1}]}{\sigma_m^2} = \begin{cases} 1 - \lambda T, & T \le \dfrac{1}{\lambda}, \\ 0, & T > \dfrac{1}{\lambda}, \end{cases}$$

where σ_m^2 is the maneuver variance and λ the inverse of average maneuver duration.

The shaping filter for whitening the maneuver noise is given by

$$U_k^1 = \rho U_{k-1}^1 + w_{k-1}^1,$$

which drives the range rate ($\dot{r}_k$) state of the vehicle, and

$$U_k^2 = \rho U_{k-1}^2 + w_{k-1}^2,$$

which drives the bearing rate (θ_k) state of the vehicle. The derivation of the discrete-time shaping filter is given in Section 5.5 with examples. The range, range rate, bearing, and bearing rate equations have been augmented by the shaping filter equations. The dimension of the state vector is increased from 4×1 to 6×1.

Covariance and gain plots for this system are produced using the Kalman filter program of Appendix A. The following initial covariance (P_0), plant noise (Q), and measurement noise (R) are used to generate the covariance results:

$$P_0 = \begin{bmatrix} \sigma_r^2 & \frac{\sigma_r^2}{T} & 0 & 0 & 0 & 0 \\ \frac{\sigma_r^2}{T} & \frac{2\sigma_r^2}{T^2} + \sigma_1^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_1^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_\theta^2 & \frac{\sigma_\theta^2}{T} & 0 \\ 0 & 0 & 0 & \frac{\sigma_\theta^2}{T} & \frac{2\sigma_\theta^2}{T^2} + \sigma_2^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma_2^2 \end{bmatrix}, \quad Q = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_1^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma_2^2 \end{bmatrix},$$

$$R = \begin{bmatrix} \sigma_r^2 & 0 \\ 0 & \sigma_\theta^2 \end{bmatrix},$$

Here $\rho = 0.5$ and $T = 5, 10, 15$ s, respectively. Also,

$$\sigma_r^2 = (1000 \text{ m})^2, \quad \sigma_\theta^2 = (0.017\text{rad})^2,$$
$$\sigma_1^2 = (103/3)^2, \quad \sigma_2^2 = 1.3 \times 10^{-8}.$$

Some parts of this example are discussed in Reference 50. Results of covariances and Kalman gain plots are shown in Figures 5.10–5.12. Convergence of the diagonal elements of the covariance matrix is shown in these figures for intervals (5, 10, 15 s). Selected Kalman gain values are shown in the following figures for various values of sampling times. These results were generated using the accompanying MATLAB program exam58.m described in Appendix A.

5.12 SUMMARY

5.12.1 Points to Remember

The optimal linear estimator is equivalent to the general (nonlinear) optimal estimator if the random processes x and z are jointly normal. Therefore, the equations for the discrete-time and continuous-time linear optimal estimators can be derived by using the orthogonality principle of Chapter 4. The discrete-time estimator (Kalman filter) has been derived and described, including its implementation equations and

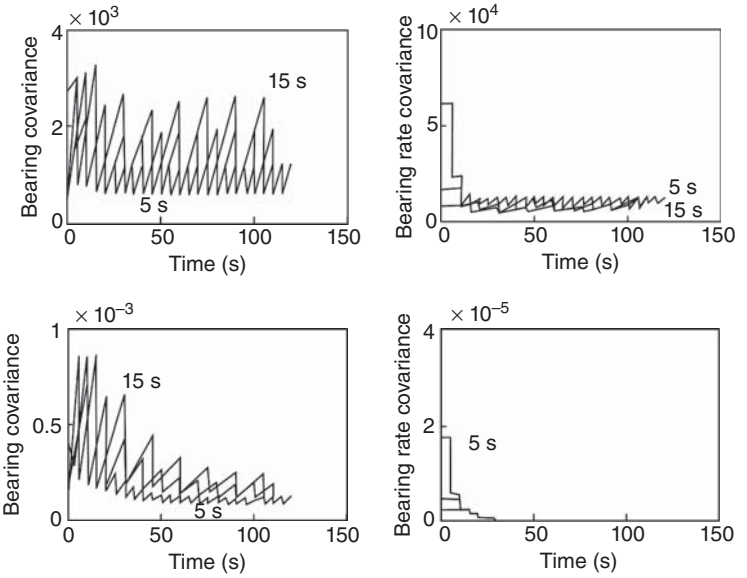

Figure 5.10 Covariances.

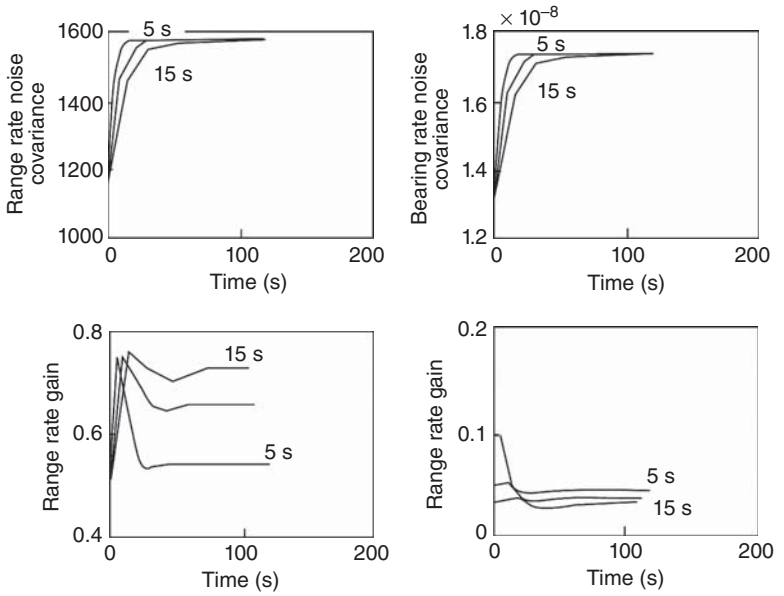

Figure 5.11 Covariances and Kalman gains.

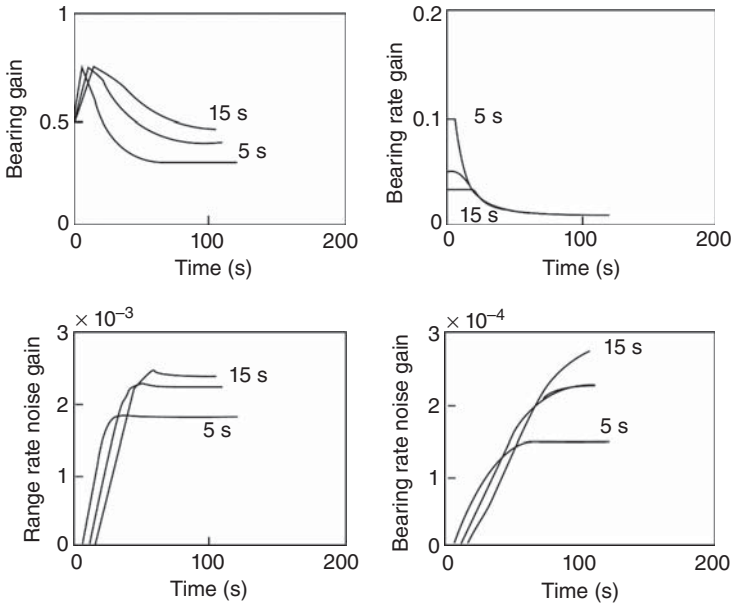

Figure 5.12 Kalman gains.

block diagram description. The continuous-time estimator (Kalman–Bucy filter) is also described.

Three different derivations of the Kalman gain have been given. A fourth derivation—using Newton's method—has been given by Humpherys et al [51].

Prediction is equivalent to filtering when measurements (system outputs) are not available. Implementation equations for continuous-time and discrete-time predictors have been given, and the problem of missing data has been discussed in detail. The estimator equations for the case that there is correlation between plant and measurement noise sources and correlated measurement errors were discussed. Relationships between stationary continuous-time and Kalman filter and Wiener filters were covered.

Methods for solving matrix Riccati differential equations have been included. Examples discussed include the applications of the Kalman filter to (i) estimating the state (phase and amplitude) of a harmonic oscillator and (ii) a discrete-time Kalman filter implementation of a five-dimensional radar tracking problem.

The discrete-time Kalman filter is recursive, in that it turns old values into new values. It is essentially a recycler of estimates and covariances.

An estimator for the state of a dynamic system at time t, using measurements made *after* time t, is called a *smoother* (next chapter).

There is nothing in the Kalman filter derivation that depends on the underlying probability distributions being Gaussian. Kalman recognized that, assuming only linearity, the mean will remain the least-mean-squared error estimator and second central moment will be the covariance matrix of estimation error. Nothing in the

derivation depends on the higher order moments of the underlying distributions, so long as everything is linear.

5.12.2 Important Equations to Remember

5.12.2.1 Kalman Filter The *discrete-time model* for a linear stochastic system has the form

$$\begin{aligned} x_k &= \Phi_{k-1}x_{k-1} + G_{k-1}w_{k-1}, \\ z_k &= H_k x_k + v_k, \end{aligned}$$

where the zero-mean uncorrelated random processes $\{w_k\}$ and $\{v_k\}$ have covariances Q_k and R_k, respectively, at time t_k. The corresponding *Kalman filter equations* have the form

$$\begin{aligned} \hat{x}_{k(-)} &= \Phi_{k-1}\hat{x}_{(k-1)(+)}, \\ P_{k(-)} &= \Phi_{k-1}P_{(k-1)(+)}\Phi_{k-1}^{\mathrm{T}} + G_{k-1}Q_{k-1}G_{k-1}^{\mathrm{T}}, \\ \hat{x}_{k(+)} &= \hat{x}_{k(-)} + \overline{K}_k(z_k - H_k\hat{x}_{k(-)}), \\ \overline{K}_k &= P_{k(-)}H_k^{\mathrm{T}}(H_kP_{k(-)}H_k^{\mathrm{T}} + R_k)^{-1}, \\ P_{k(+)} &= P_{k(-)} - \overline{K}_kH_kP_{k(-)}, \end{aligned}$$

where the $(-)$ indicates the a priori values of the variables (before the information in the measurement is used) and the $(+)$ indicates the a posteriori values of the variables (after the information in the measurement is used). The variable $\overline{K}$ is the Kalman gain.

5.12.2.2 Kalman–Bucy Filter The *continuous-time model* for a linear stochastic system has the form

$$\begin{aligned} \frac{d}{dt}x(t) &= F(t)x(t) + G(t)w(t), \\ z(t) &= H(t)x(t) + v(t), \end{aligned}$$

where the zero-mean uncorrelated random processes $\{w(t)\}$ and $\{v(t)\}$ have covariances $Q(t)$ and $R(t)$, respectively, at time t. The corresponding *Kalman–Bucy filter* equations for the estimate $\hat{x}$ of the state variable x, given the output signal z, has the form

$$\begin{aligned} \frac{d}{dt}\hat{x}(t) &= F(t)\hat{x}(t) + \overline{K}(t)[z(t) - H(t)\hat{x}(t)], \\ \overline{K}(t) &= P(t)H^{\mathrm{T}}(t)R^{-1}(t), \\ \frac{d}{dt}P(t) &= F(t)P(t) + P(t)F^{\mathrm{T}}(t) - \overline{K}(t)R(t)\overline{K}^{\mathrm{T}}(t) + G(t)Q(t)G^{\mathrm{T}}(t). \end{aligned}$$

PROBLEMS

5.1 A scalar discrete-time random sequence x_k is given by

$$x_{k+1} = 0.5x_k + w_k,$$
$$\mathrm{E}x_0 = 0, \mathrm{E}x_0^2 = 1, \mathrm{E}w_k^2 = 1, \mathrm{E}w_k = 0,$$

where w_k is the white noise. The observation equation is given by

$$z_k = x_k + v_k,$$

$\mathrm{E}v_k = 0, \mathrm{E}v_k^2 = 1$, and v_k is also white noise. The terms x_0, w_k, and v_k are all random. Derive a (nonrecursive) expression for

$$\mathrm{E}[x_2|z_0, z_1, z_2].$$

5.2 For the system given in Problem 5.1:
(a) Write the discrete-time Kalman filter equations.
(b) Provide the correction necessary if z_2 was not received.
(c) Derive the loss in terms of the estimate $\hat{x}_3$ due to missing z_2.
(d) Derive the filter for $k \to \infty$ (steady state).
(e) Repeat (d) when every other observation is missed.

5.3 Prove that any Gaussian likelihood function $\mathcal{L}(x,\ \mu_x,\ Y_{xx})$ achieves its maximum at the set of points

$$\underset{x}{\text{argmax}}\, \mathcal{L}(x, \mu_x,\ Y_{xx}) = \left\{ \mu_x + \sum_{\lambda_i=0} a_i e_i \ \mid\ a_i \in \Re \right\},$$

where the summation is over the zero eigenvectors of Y_{xx}. That is, if no eigenvalue of Y_{xx} is zero, the maximum occurs at μ_x. However, if Y_{xx} has zero eigenvalues, the same maximum value is attained at μ_x plus any linear combination of the corresponding zero eigenvectors of Y_{xx}.

5.4 In a single-dimension example of a radar tracking an object by means of track while scan, measurements of the continuous-time target trajectory at some discrete times are made. The process and measurement models are given by

$$\dot{x}(t) = -0.5x(t) + w(t), \quad z_{kT} = x_{kT} + v_{kT},$$

where T is the intersampling interval (assume 1 s for simplicity):

$$\mathrm{E}v_k = \mathrm{E}w(t) = 0,$$

$$\mathrm{E}w(t_1)w(t_2) = 1\delta(t_2 - t_1),$$

$$\mathrm{E}v_{k_1\mathrm{T}}v_{k_2\mathrm{T}} = 1\Delta(k_2 - k_1),$$

$$\mathrm{E}\langle v_k w^{\mathrm{T}}\rangle = 0.$$

Derive the minimum mean-squared-filter of $x(t)$ for all t.

5.5 In Problem 5.3, the measurements are received at discrete times and each measurement is spread over some nonzero time interval (radar beam width nonzero). The measurement equation of Problem 5.4 can be modified to

$$z_{kT+\eta} = x_{kT+\eta} + v_{kT+\eta},$$

where

$$k = 0, 1, 2, \ldots, 0 \le \eta \le \eta_0.$$

Let $T = 1\mathrm{s}$, $\eta_0 = 0.1$ (radar beam width) and $v(t)$ be a zero-mean white Gaussian process with covariance equal to 1. Derive the minimum mean-squared filter of $x(t)$ for all t.

5.6 Prove the condition in the discussion following Equation 5.9 that $\mathrm{E}w_k z_i^{\mathrm{T}} = 0$ for $i = 1, \ldots k$ when w_k and v_k are uncorrelated and white.

5.7 In Example 5.8, use white noise as a driving input to range rate ($\dot{r}_k$) and bearing rate ($\dot{\theta}_k$) equations instead of colored noise. This reduces the dimension of the state vector from 6×1 to 4×1. Formulate the new observation equation. Generate the covariance and Kalman gain plots for the same values of $P_0, Q, R, \sigma_r^2, \sigma_\theta^2, \sigma_1^2$, and σ_2^2.

5.8 For the same problem as Problem 5.7, obtain values of the plant covariance Q for the four-state model such the associated mean-squared estimation uncertainties for range, range rate, bearing, and bearing rate are within 5–10% of those for the six-state model. (*Hint:* This should be possible because the plant noise is used to model the effects of linearization errors, discretization errors, and other unmodeled effects or approximations. This type of suboptimal filtering will be discussed further in Chapter 9.)

5.9 For the estimation problem modeled by the equations

$$\begin{aligned} x_k &= x_{k-1} + w_{k-1}, \\ w_k &\sim \mathcal{N}(0, 30) \text{ and white}, \\ z_k &= x_k + v_k, \\ v_k &\sim \mathcal{N}(0, 20) \text{ and white}, \\ P_0 &= 150, \end{aligned}$$

calculate the values of $P_{k(+)}, P_{k(-)}$, and $\overline{K}_k$ for $k = 1, 2, 3, 4$ and $P_{\infty(+)}$ (the steady-state value).

5.10 *Parameter estimation problem.* Let x be a zero-mean Gaussian random variable with covariance P_0, and let $z_k = x + v_k$ be an observation of x with noise $v_k \sim \mathcal{N}(0, R)$.

(a) Write the recursion equations for $P_{k(+)}, P_{k(-)}$, $\overline{K}_k$, and $\hat{x}_k$.

(b) What is the value of x_1 if $R = 0$?

(c) What is the value of x_1 if $R = +\infty$?

(d) Explain the results of (b) and (c) in terms of measurement uncertainty.

5.11 Assume a stochastic system in continuous time modeled by the equations

$$\begin{aligned} \dot{x}(t) &= -x(t) + w(t), \\ w(t) &\sim \mathcal{N}(0, 30), \\ z(t) &= x(t) + v(t), \\ v(t) &\sim \mathcal{N}(0, 20). \end{aligned}$$

(a) Derive the values of the mean-squared estimation error $P(t)$ and Kalman gain $\overline{K}(t)$ for time $t = 1, 2, 3, 4$.

(b) Solve for the steady-state value of P.

5.12 Show that the matrices P_k and $P(t)$ of Equations 5.23 and 5.116 are *symmetric*. That is, $P_k^{\mathrm{T}} = P_k$ and $P^{\mathrm{T}}(t) = P(t)$.

5.13 Derive the observability matrices for Examples 5.7 and 5.8 to determine whether these systems are observable.

5.14 A vector discrete-time random sequence x_k is given by

$$\begin{aligned} x_k &= \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} x_{k-1} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} w_{k-1}, \\ w_k &\sim \mathcal{N}(0, 1) \text{ and white.} \end{aligned}$$

The observation equation is given by

$$\begin{aligned} z_k &= [1 \quad 0] x_k + v_k, \\ v_k &\sim \mathcal{N}[0, 2 + (-1)^k] \text{ and white.} \end{aligned}$$

Calculate the values of $P_{k(+)}, P_{k(-)}$ and $\overline{K}_k$ for $k = 1, \ldots, 10$ and $P_{\infty(+)}$ (the steady-state value) with

$$P_0 = \begin{bmatrix} 10 & 0 \\ 0 & 10 \end{bmatrix}.$$

5.15 Calculate the values of $P_{k(+)}$, $\overline{K}_k$, and $\hat{x}_{k(+)}$ by serial processing of a vector measurement:

$$\hat{x}_{k(-)} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \; P_{k(-)} = \begin{bmatrix} 4 & 1 \\ 1 & 9 \end{bmatrix}, \; H_k = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, \; R_k = \begin{bmatrix} 3 & 0 \\ 0 & 4 \end{bmatrix}, \; z_k = \begin{bmatrix} 3 \\ 4 \end{bmatrix}.$$

REFERENCES

[1] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*, Prentice-Hall, Englewood Cliffs, NJ, 1979.

[2] S. M. Bozic, *Digital and Kalman Filtering: An Introduction to Discrete-Time Filtering and Optimal Linear Estimation*, John Wiley & Sons, Inc., New York, 1979.

[3] K. Brammer and G. Siffling, *Kalman–Bucy Filters*, Artech House, Norwood, MA, 1989.

[4] R. G. Brown, *Introduction to Random Signal Analysis and Kalman Filtering*, John Wiley & Sons, Inc., New York, 1983.

[5] A. E. Bryson Jr. and Y.-C. Ho, *Applied Optimal Control*, Blaisdell, Waltham, MA, 1969.

[6] R. S. Bucy and P. D. Joseph, *Filtering for Stochastic Processes with Applications to Guidance*, American Mathematical Society, Chelsea Publishing, Providence, RI, 2005.

[7] D. E. Catlin, *Estimation, Control, and the Discrete Kalman Filter*, Springer-Verlag, New York, 1989.

[8] C. K. Chui and G. Chen, *Kalman Filtering with Real-Time Applications*, Springer-Verlag, New York, 1987.

[9] A. Gelb, J. F. Kasper Jr., R. A. Nash Jr., C. F. Price, and A. A. Sutherland Jr., *Applied Optimal Estimation*, MIT Press, Cambridge, MA, 1974.

[10] A. H. Jazwinski, *Stochastic Processes and Filtering Theory*, Dover Publishers, Mineola, NY, 2009 (reprint of 1970 edition by Academic Press).

[11] T. Kailath, *Lectures on Weiner and Kalman Filtering*, Springer-Verlag, New York, 1981.

[12] P. S. Maybeck, *Stochastic Models, Estimation, and Control*, Vol. 1, Academic Press, New York, 1979.

[13] P. S. Maybeck, *Stochastic Models, Estimation, and Control*, Vol. 2, Academic Press, New York, 1982.

[14] J. M. Mendel, *Lessons in Digital Estimation Techniques*, Prentice-Hall, Englewood Cliffs, NJ, 1987.

[15] J. M. Mendel, "Kalman filtering and other digital estimation techniques," *IEEE Individual Learning Package*, IEEE, New York, 1987.

[16] N. E. Nahi, *Estimation Theory and Applications*, John Wiley & Sons, Inc., New York, 1969; reprinted by Krieger, Melbourne, FL, 1975.

[17] P. A. Ruymgaart and T. T. Soong, *Mathematics of Kalman–Bucy Filtering*, Springer-Verlag, New York, 1988.

[18] H. W. Sorenson, Ed., *Kalman Filtering: Theory and Application*, IEEE Press, New York, 1985.

[19] J. F. Riccati, "Animadversationnes in aequationes differentiales secundi gradus," *Acta Eruditorum Quae Lipside Publicantur Supplementa*, Vol. 8, pp. 66–73, 1724.

[20] S. Bittanti, A. J. Laub, and J. C. Willems, Eds., *The Riccati Equation*, Springer-Verlag, Berlin, 1991.

[21] M.-A. Poubelle, I. R. Petersen, M. R. Gevers, and R. R. Bitmead, "A miscellany of results on an equation of Count J. F. Riccati," *IEEE Transactions on Automatic Control*, Vol. AC - 31, pp. 651–654, 1986.

[22] L. Dieci, "Numerical integration of the differential Riccati equation and some related issues," *SIAM Journal of Numerical Analysis*, Vol. 29, pp. 781–815, 1992.

[23] J. Aldrich, "R. A. Fisher and the making of maximum likelihood 1912–1922," *Statistical Science*, Vol. 12. No. 3, pp. 162–176, 2004.

[24] A. Hald, *A History of Mathematical Statistics from 1750 to 1930*, John Wiley & Sons, Inc., New York, 1998.

[25] A. Hald, "On the history of maximum likelihood in relation to inverse probability and least squares," *Statistical Science*, Vol. 14, No. 2, pp. 214–222, 1999.

[26] J. W. Pratt, "F. Y. Edgeworth and R. A. Fisher on the efficiency of maximum likelihood estimation," *Annals of Statistics*, Vol. 4, No. 3, pp. 501–514, 1976.

[27] S. M. Stigler, "The epic story of maximum likelihood," *Statistical Science*, Vol. 22, No. 4, pp. 598–620, 2007.

[28] R. E. Kalman, "A new approach to linear filtering and prediction problems," *ASME Journal of Basic Engineering*, Vol. 82, pp. 34–45, 1960.

[29] W. J. Duncan, "Some devices for the solution of large sets of simultaneous linear equations (with an appendix on the reciprocation of partitioned matrices)," *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, Series 7*, Vol. 35, pp. 660–670, 1944.

[30] A. Bain and D. Crisan, *Fundamentals of Stochastic Filtering*, Springer, New York, 2009.

[31] J. L. Crassidis and J. L. Junkins, *Optimal Estimation of Dynamic Systems*, Chapman and Hall (CRC Press), Boca Raton, FL, 2004.

[32] H. W. Sorenson, "Kalman filtering techniques," in *Advances in Control Systems* Vol. 3 (C. T. Leondes, Ed.), Academic Press, Baltimore, MD, pp. 219–292, 1966.

[33] R. E. Kalman and Richard S. Bucy, "New results in linear filtering and prediction theory," *ASME Journal of Basic Engineering, Series D*, Vol. 83, pp. 95–108, 1961.

[34] H. Kwaknernaak and R. Sivan, *Linear Optimal Control Systems*, John Wiley & Sons, Inc., New York, 1972.

[35] R. E. Kalman, "New methods in Wiener filtering," in *Proceeding of the First Symposium on Engineering Applications of Random Function Theory and Probability*, J. L. Bogdanoff and F. Kozin, Eds., John Wiley & Sons, Inc., New York, pp. 270–388, 1963.

[36] A. E. Bryson Jr. and D. E. Johansen, "Linear filtering for time-varying systems using measurements containing colored noise," *IEEE Transactions on Automatic Control*, Vol. AC - 10, pp. 4–10, 1965.

[37] D. Slepian, "Estimation of signal parameters in the presence of noise," *IRE Transactions on Information Theory*, Vol. IT - 3, pp. 68–69, 1954.

[38] J. E. Potter, A Matrix Equation Arising in Statistical Estimation Theory, Report No. CR - 270, National Aeronautics and Space Administration, New York, 1965.

[39] A. G. J. MacFarlane, "An eigenvector solution of the optimal linear regulator," *Journal of Electronic Control*, Vol. 14, pp. 643–654, 1963.

[40] J. E. Potter, "Matrix quadratic solutions," *SIAM Journal of Applied Mathematics*, Vol. 14, pp. 496–501, 1966.

[41] A. F. Fath, "Computational aspects of the linear optimal regulator problem," *IEEE Transactions on Automatic Control*, Vol. AC - 14, pp. 547–550, 1969.

[42] S. F. Schmidt, "Applications of state-space methods to navigation problems," in *Advances in Control Systems* (C. T. Leondes, Ed.), Vol. 3, Academic Press, New York, pp. 293–340, 1966.

[43] M. S. Grewal, "Application of Kalman filtering to the calibration and alignment of inertial navigation systems," in *Proceedings of PLANS '86—Position, Location and Navigation Symposium*, Las Vegas, NV, Nov. 4–7, 1986, IEEE, New York, 1986.

[44] M. S. Grewal and R. S. Miyasako, "Gyro compliance estimation in the calibration and alignment of inertial measurement units," in *Proceedings of Eighteenth Joint Services Conference on Data Exchange for Inertial Systems*, San Diego, CA, Oct. 28–30, 1986, IEEE Joint Services Data Exchange, San Diego, CA, 1986.

[45] M. S. Grewal and H. J. Payne, "Identification of parameters in a freeway traffic model," *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. SMC - 6, pp. 176–185, 1976.

[46] J. H. Blakelock, *Automatic Control of Aircraft and Missiles*, John Wiley & Sons, Inc., New York, 1965.

[47] T. R. Benedict and G. W. Bordner, "Synthesis of an optimal set of radar track-while-scan smoothing equations," *IEEE Transactions on Automatic Control*, Vol. AC - 7, pp. 27–32, 1962.

[48] R. A. Singer, "Estimating optimal tracking filter performance for manned maneuvering targets," *IEEE Transactions on Aerospace and Electronic Systems*, Vol. AES - 6, pp. 473–483, 1970.

[49] P. S. Maybeck, J. G. Reid, and R. N. Lutter, "Application of an extended Kalman filter to an advanced fire control system," in *Proceedings of the IEEE Conference on Decision and Control*, New Orleans, 1977, pp. 1192–1195.

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

[51] J. Humpherys, P. Redd, and J. West, "A fresh look at the Kalman filter," *SIAM Review*, Vol. 54, No, 4, pp. 801–823, 2012.