1

KALMAN FILTERS

Simon Haykin

Communications Research Laboratory, McMaster University, Hamilton, Ontario, Canada
(haykin@mcmaster.ca)

1.1 INTRODUCTION

The celebrated *Kalman filter*, rooted in the state-space formulation of linear dynamical systems, provides a recursive solution to the linear optimal filtering problem. It applies to stationary as well as nonstationary environments. The solution is recursive in that each updated estimate of the state is computed from the previous estimate and the new input data, so only the previous estimate requires storage. In addition to eliminating the need for storing the entire past observed data, the Kalman filter is computationally more efficient than computing the estimate directly from the entire past observed data at each step of the filtering process.

In this chapter, we present an introductory treatment of Kalman filters to pave the way for their application in subsequent chapters of the book. We have chosen to follow the original paper by Kalman [1] for the

Kalman Filtering and Neural Networks, Edited by Simon Haykin
ISBN 0-471-36998-5

derivation; see also the books by Lewis [2] and Grewal and Andrews [3]. The derivation is not only elegant but also highly insightful.

Consider a *linear, discrete-time dynamical system* described by the block diagram shown in Figure 1.1. The concept of *state* is fundamental to this description. The *state vector* or simply *state*, denoted by $\mathbf{x}_k$, is defined as the minimal set of data that is sufficient to uniquely describe the unforced dynamical behavior of the system; the subscript k denotes discrete time. In other words, the state is the least amount of data on the past behavior of the system that is needed to predict its future behavior. Typically, the state $\mathbf{x}_k$ is unknown. To estimate it, we use a set of observed data, denoted by the vector $\mathbf{y}_k$.

In mathematical terms, the block diagram of Figure 1.1 embodies the following pair of equations:

1. *Process equation*

$$\mathbf{x}_{k+1} = \mathbf{F}_{k+1,k}\mathbf{x}_k + \mathbf{w}_k, \tag{1.1}$$

where $\mathbf{F}_{k+1,k}$ is the *transition matrix* taking the state $\mathbf{x}_k$ from time k to time $k+1$. The process noise $\mathbf{w}_k$ is assumed to be additive, white, and Gaussian, with zero mean and with covariance matrix defined by

$$E[\mathbf{w}_n\mathbf{w}_k^T] = \begin{cases} \mathbf{Q}_k & \text{for } n = k, \\ 0 & \text{for } n \neq k, \end{cases} \tag{1.2}$$

where the superscript T denotes matrix transposition. The dimension of the state space is denoted by M.

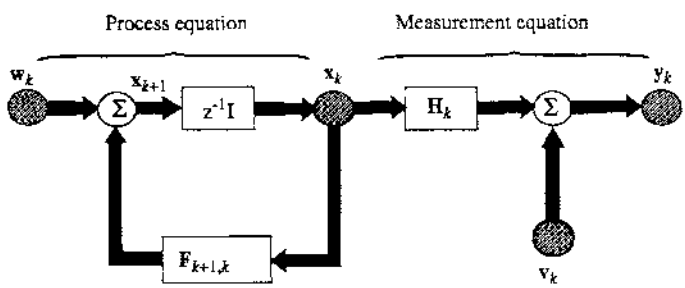

Figure 1.1 Signal-flow graph representation of a linear, discrete-time dynamical system.

2. *Measurement equation*

$$\mathbf{y}_k = \mathbf{H}_k \mathbf{x}_k - \mathbf{v}_k, \tag{1.3}$$

where $\mathbf{y}_k$ is the *observable* at time k and $\mathbf{H}_k$ is the *measurement matrix*. The measurement noise $\mathbf{v}_k$ is assumed to be additive, white, and Gaussian, with zero mean and with covariance matrix defined by

$$E[\mathbf{v}_n \mathbf{v}_k^T] = \begin{cases} \mathbf{R}_k & \text{for } n = k, \\ \mathbf{0} & \text{for } n \neq k. \end{cases} \tag{1.4}$$

Moreover, the measurement noise $\mathbf{v}_k$ is uncorrelated with the process noise $\mathbf{w}_k$. The dimension of the measurement space is denoted by N.

The Kalman filtering problem, namely, the problem of jointly solving the process and measurement equations for the unknown state in an optimum manner may now be formally stated as follows:

- Use the entire observed data, consisting of the vectors $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_k$, to find for each $k \geq 1$ the minimum mean-square error estimate of the state $\mathbf{x}_i$.

The problem is called *filtering* if $i = k$, *prediction* if $i > k$, and *smoothing* if $1 \leq i < k$.

1.2 OPTIMUM ESTIMATES

Before proceeding to derive the Kalman filter, we find it useful to review some concepts basic to optimum estimation. To simplify matters, this review is presented in the context of scalar random variables; generalization of the theory to vector random variables is a straightforward matter. Suppose we are given the observable

$$y_k = x_k + v_k,$$

where x_k is an unknown signal and v_k is an additive noise component. Let $\hat{x}_k$ denote the a posteriori estimate of the signal x_k, given the observations $y_1, y_2, \ldots, y_k$. In general, the estimate $\hat{x}_k$ is different from the unknown

signal x_k. To derive this estimate in an optimum manner, we need a *cost (loss) function* for incorrect estimates. The cost function should satisfy two requirements:

- The cost function is nonnegative.
- The cost function is a nondecreasing function of the *estimation error* $\tilde{x}_k$ defined by

$$\tilde{x}_k = x_k - \hat{x}_k.$$

These two requirements are satisfied by the *mean-square error* defined by

$$\begin{aligned} J_k &= E[(x_k - \hat{x}_k)^2] \\ &= E[\tilde{x}_k^2], \end{aligned}$$

where E is the expectation operator. The dependence of the cost function J_k on time k emphasizes the nonstationary nature of the recursive estimation process.

To derive an optimal value for the estimate $\hat{x}_k$, we may invoke two theorems taken from stochastic process theory [1, 4]:

Theorem 1.1 Conditional mean estimator *If the stochastic processes* $\{x_k\}$ *and* $\{y_k\}$ *are jointly Gaussian, then the optimum estimate* $\hat{x}_k$ *that minimizes the mean-square error* J_k *is the conditional mean estimator:*

$$\hat{x}_k = E[x_k | y_1, y_2, \ldots, y_k].$$

Theorem 1.2 Principle of orthogonality *Let the stochastic processes* $\{x_k\}$ *and* $\{y_k\}$ *be of zero means; that is,*

$$E[x_k] = E[y_k] = 0 \quad \text{for all } k.$$

Then:

(i) *the stochastic process* $\{x_k\}$ *and* $\{y_k\}$ *are jointly Gaussian; or*

(ii) *if the optimal estimate* $\hat{x}_k$ *is restricted to be a linear function of the observables and the cost function is the mean-square error,*

(iii) *then the optimum estimate* $\hat{x}_k$*, given the observables* $y_1, y_2, \ldots, y_k$*, is the orthogonal projection of* x_k *on the space spanned by these observables.*

With these two theorems at hand, the derivation of the Kalman filter follows.

1.3 KALMAN FILTER

Suppose that a measurement on a linear dynamical system, described by Eqs. (1.1) and (1.3), has been made at time k. The requirement is to use the information contained in the new measurement $\mathbf{y}_k$ to update the estimate of the unknown state $\mathbf{x}_k$. Let $\hat{\mathbf{x}}_k^-$ denote a priori estimate of the state, which is already available at time k. With a linear estimator as the objective, we may express the a posteriori estimate $\hat{\mathbf{x}}_k$ as a linear combination of the a priori estimate and the new measurement, as shown by

$$\hat{\mathbf{x}}_k = \mathbf{G}_k^{(1)}\hat{\mathbf{x}}_k^- + \mathbf{G}_k\mathbf{y}_k, \tag{1.5}$$

where the multiplying matrix factors $\mathbf{G}_k^{(1)}$ and $\mathbf{G}_k$ are to be determined. To find these two matrices, we invoke the principle of orthogonality stated under Theorem 1.2. The *state-error vector* is defined by

$$\tilde{\mathbf{x}}_k = \mathbf{x}_k - \hat{\mathbf{x}}_k. \tag{1.6}$$

Applying the principle of orthogonality to the situation at hand, we may thus write

$$E[\tilde{\mathbf{x}}_k\mathbf{y}_i^T] = \mathbf{0} \quad \text{for } i = 1, 2, \ldots, k-1. \tag{1.7}$$

Using Eqs. (1.3), (1.5), and (1.6) in (1.7), we get

$$E[(\mathbf{x}_k - \mathbf{G}_k^{(1)}\hat{\mathbf{x}}_k^- - \mathbf{G}_k\mathbf{H}_k\mathbf{x}_k - \mathbf{G}_k\mathbf{w}_k)\mathbf{y}_i^T] = \mathbf{0} \quad \text{for } i = 1, 2, \ldots, k-1. \tag{1.8}$$

Since the process noise $\mathbf{w}_k$ and measurement noise $\mathbf{v}_k$ are uncorrelated, it follows that

$$E[\mathbf{w}_k\mathbf{y}_i^T] = \mathbf{0}.$$

Using this relation and rearranging terms, we may rewrite Eq. (8) as

$$E[(\mathbf{I} - \mathbf{G}_k\mathbf{H}_k - \mathbf{G}_k^{(1)})\mathbf{x}_k\mathbf{y}_i^T - \mathbf{G}_k^{(1)}(\mathbf{x}_k - \hat{\mathbf{x}}_k^-)\mathbf{y}_i^T] = \mathbf{0}, \tag{1.9}$$

where $\mathbf{I}$ is the identity matrix. From the principle of orthogonality, we now note that

$$E[(\mathbf{x}_k - \hat{\mathbf{x}}_k^-)\mathbf{y}_i^T] = \mathbf{0}.$$

Accordingly, Eq. (1.9) simplifies to

$$(\mathbf{I} - \mathbf{G}_k\mathbf{H}_k - \mathbf{G}_k^{(1)})E[\mathbf{x}_k\mathbf{y}_i^T] = \mathbf{0} \quad \text{for } i = 1, 2, \ldots, k-1. \tag{1.10}$$

For arbitrary values of the state $\mathbf{x}_k$ and observable $\mathbf{y}_i$, Eq. (1.10) can only be satisfied if the scaling factors $\mathbf{G}_k^{(1)}$ and $\mathbf{G}_k$ are related as follows:

$$\mathbf{I} - \mathbf{G}_k\mathbf{H}_k - \mathbf{G}_k^{(1)} = \mathbf{0},$$

or, equivalently, $\mathbf{G}_k^{(1)}$ is defined in terms of $\mathbf{G}_k$ as

$$\mathbf{G}_k^{(1)} = \mathbf{I} - \mathbf{G}_k\mathbf{H}_k. \tag{1.11}$$

Substituting Eq. (1.11) into (1.5), we may express the a posteriori estimate of the state at time k as

$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_k^- + \mathbf{G}_k(\mathbf{y}_k - \mathbf{H}_k\hat{\mathbf{x}}_k^-), \tag{1.12}$$

in light of which, the matrix $\mathbf{G}_k$ is called the *Kalman gain*.

There now remains the problem of deriving an explicit formula for $\mathbf{G}_k$. Since, from the principle of orthogonality, we have

$$E[(\mathbf{x}_k - \hat{\mathbf{x}}_k)\mathbf{y}_k^T] = \mathbf{0}, \tag{1.13}$$

it follows that

$$E[(\mathbf{x}_k - \hat{\mathbf{x}}_k)\hat{\mathbf{y}}_k^T] = \mathbf{0}, \tag{1.14}$$

where $\hat{\mathbf{y}}_k^T$ is an estimate of $\mathbf{y}_k$ given the previous measurement $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_{k-1}$. Define the *innovations* process

$$\tilde{\mathbf{y}}_k = \mathbf{y}_k - \hat{\mathbf{y}}_k. \tag{1.15}$$

The innovation process represents a measure of the "new" information contained in $\mathbf{y}_k$; it may also be expressed as

$$\begin{aligned} \tilde{\mathbf{y}}_k &= \mathbf{y}_k - \mathbf{H}_k \hat{\mathbf{x}}_k^- \\ &= \mathbf{H}_k \mathbf{x}_k + \mathbf{v}_k - \mathbf{H}_k \hat{\mathbf{x}}_k^- \\ &= \mathbf{H}_k \tilde{\mathbf{x}}_k^- + \mathbf{v}_k. \end{aligned} \tag{1.16}$$

Hence, subtracting Eq. (1.14) from (1.13) and then using the definition of Eq. (1.15), we may write

$$E[(\mathbf{x}_k - \hat{\mathbf{x}}_k)\tilde{\mathbf{y}}_k^T] = \mathbf{0}. \tag{1.17}$$

Using Eqs. (1.3) and (1.12), we may express the state-error vector $\mathbf{x}_k - \hat{\mathbf{x}}_k$ as

$$\begin{aligned} \mathbf{x}_k - \hat{\mathbf{x}}_k &= \tilde{\mathbf{x}}_k^- - \mathbf{G}_k(\mathbf{H}_k \tilde{\mathbf{x}}_k^- + \mathbf{v}_k) \\ &= (\mathbf{I} - \mathbf{G}_k \mathbf{H}_k)\tilde{\mathbf{x}}_k^- - \mathbf{G}_k \mathbf{v}_k. \end{aligned} \tag{1.18}$$

Hence, substituting Eqs. (1.16) and (1.18) into (1.17), we get

$$E[\{(\mathbf{I} - \mathbf{G}_k \mathbf{H}_k)\tilde{\mathbf{x}}_k^- - \mathbf{G}_k \mathbf{v}_k\}(\mathbf{H}_k \tilde{\mathbf{x}}_k^- + \mathbf{v}_k)] = \mathbf{0}. \tag{1.19}$$

Since the measurement noise $\mathbf{v}_k$ is independent of the state $\mathbf{x}_k$ and therefore the error $\tilde{\mathbf{x}}_k^-$, the expectation of Eq. (1.19) reduces to

$$(\mathbf{I} - \mathbf{G}_k \mathbf{H}_k) E[\tilde{\mathbf{x}}_k \tilde{\mathbf{x}}_k^{T-}] \mathbf{H}_k^T - \mathbf{G}_k E[\mathbf{v}_k \mathbf{v}_k^T] = \mathbf{0}. \tag{1.20}$$

Define the *a priori covariance matrix*

$$\begin{aligned} \mathbf{P}_k^- &= E[(\mathbf{x}_k - \hat{\mathbf{x}}_k^-)(\mathbf{x}_k - \hat{\mathbf{x}}_k^-)^T] \\ &= E[\tilde{\mathbf{x}}_k^- \cdot \tilde{\mathbf{x}}_k^{T-}]. \end{aligned} \tag{1.21}$$

Then, invoking the covariance definitions of Eqs. (1.4) and (1.21), we may rewrite Eq. (1.20) as

$$(\mathbf{I} - \mathbf{G}_k\mathbf{H}_k)\mathbf{P}_k^-\mathbf{H}_k^T - \mathbf{G}_k\mathbf{R}_k = \mathbf{0}.$$

Solving this equation for $\mathbf{G}_k$, we get the desired formula

$$\mathbf{G}_k = \mathbf{P}_k^-\mathbf{H}_k^T[\mathbf{H}_k\mathbf{P}_k^-\mathbf{H}_k^T + \mathbf{R}_k]^{-1}, \tag{1.22}$$

where the symbol $[\cdot]^{-1}$ denotes the inverse of the matrix inside the square brackets. Equation (1.22) is the desired formula for computing the Kalman gain $\mathbf{G}_k$, which is defined in terms of the a priori covariance matrix $\mathbf{P}_k^-$.

To complete the recursive estimation procedure, we consider the *error covariance propagation*, which describes the effects of time on the covariance matrices of estimation errors. This propagation involves two stages of computation:

1. The a priori covariance matrix $\mathbf{P}_k^-$ at time k is defined by Eq. (1.21). Given $\mathbf{P}_{k-1}^+$, compute the a posteriori covariance matrix $\mathbf{P}_k$, which, at time k, is defined by

$$\begin{aligned}\mathbf{P}_k &= E[\tilde{\mathbf{x}}_k\tilde{\mathbf{x}}_k^T] \\ &= E[(\mathbf{x}_k - \hat{\mathbf{x}}_k)(\mathbf{x}_k - \hat{\mathbf{x}}_k)^T].\end{aligned} \tag{1.23}$$

2. Given the "old" a posteriori covariance matrix, $\mathbf{P}_{k-1}$, compute the "updated" a priori covariance matrix $\mathbf{P}_k^-$.

To proceed with stage 1, we substitute Eq. (1.18) into (1.23) and note that the noise process $\mathbf{v}_k$ is independent of the a priori estimation error $\tilde{x}_k^-$. We thus obtain[†]

$$\begin{aligned}\mathbf{P}_k &= (\mathbf{I} - \mathbf{G}_k\mathbf{H}_k)E[\tilde{\mathbf{x}}_k^-\tilde{\mathbf{x}}_k^{T-}](\mathbf{I} - \mathbf{G}_k\mathbf{H}_k)^T + \mathbf{G}_kE[\mathbf{v}_k\mathbf{v}_k^T]\mathbf{G}_k^T \\ &= (\mathbf{I} - \mathbf{G}_k\mathbf{H}_k)\mathbf{P}_k^-(\mathbf{I} - \mathbf{G}_k\mathbf{H}_k)^T + \mathbf{G}_k\mathbf{R}_k\mathbf{G}_k^T.\end{aligned} \tag{1.24}$$

[†]Equation (1.24) is referred to as the "Joseph" version of the covariance update equation [5].

Expanding terms in Eq. (1.24) and then using Eq. (1.22), we may reformulate the dependence of the a posteriori covariance matrix $\mathbf{P}_k$ on the a priori covariance matrix $\mathbf{P}_k^-$ in the simplified form

$$\begin{aligned}\mathbf{P}_k &= (\mathbf{I} - \mathbf{G}_k\mathbf{H}_k)\mathbf{P}_k^- - (\mathbf{I} - \mathbf{G}_k\mathbf{H}_k)\mathbf{P}_k\,\mathbf{H}_k^T\mathbf{G}_k^T + \mathbf{G}_k\mathbf{R}_k\mathbf{G}_k^T \\ &= (\mathbf{I} - \mathbf{G}_k\mathbf{H}_k)\mathbf{P}_k^- - \mathbf{G}_k\mathbf{R}_k\mathbf{G}_k^T + \mathbf{G}_k\mathbf{R}_k\mathbf{G}_k^T \\ &= (\mathbf{I} - \mathbf{G}_k\mathbf{H}_k)\mathbf{P}_k^-. \end{aligned} \tag{1.25}$$

For the second stage of error covariance propagation, we first recognize that the a priori estimate of the state is defined in terms of the "old" a posteriori estimate as follows:

$$\hat{\mathbf{x}}_k^- = \mathbf{F}_{k,k-1}\hat{\mathbf{x}}_{k-1}. \tag{1.26}$$

We may therefore use Eqs. (1.1) and (1.26) to express the a priori estimation error in yet another form:

$$\begin{aligned}\tilde{\mathbf{x}}_k &= \mathbf{x}_k - \hat{\mathbf{x}}_k^- \\ &= (\mathbf{F}_{k,k-1}\mathbf{x}_{k-1} + \mathbf{w}_{k-1}) - (\mathbf{F}_{k,k-1}\hat{\mathbf{x}}_{k-1}) \\ &= \mathbf{F}_{k,k-1}(\mathbf{x}_{k-1} - \hat{\mathbf{x}}_{k-1}) + \mathbf{w}_{k-1} \\ &= \mathbf{F}_{k,k-1}\tilde{\mathbf{x}}_{k-1} + \mathbf{w}_{k-1}. \end{aligned} \tag{1.27}$$

Accordingly, using Eq. (1.27) in (1.21) and noting that the process noise $\mathbf{w}_k$ is independent of $\tilde{\mathbf{x}}_{k-1}$, we get

$$\begin{aligned}\mathbf{P}_k^- &= \mathbf{F}_{k,k-1}E[\tilde{\mathbf{x}}_{k-1}\tilde{\mathbf{x}}_{k-1}^T]\mathbf{F}_{k,k-1}^T + E[\mathbf{w}_{k-1}\mathbf{w}_{k-1}^T] \\ &= \mathbf{F}_{k,k-1}\mathbf{P}_{k-1}\mathbf{F}_{k,k-1}^T + \mathbf{Q}_{k-1}, \end{aligned} \tag{1.28}$$

which defines the dependence of the a priori covariance matrix $\mathbf{P}_k^-$ on the "old" a posteriori covariance matrix $\mathbf{P}_{k-1}$.

With Eqs. (1.26), (1.28), (1.22), (1.12), and (1.25) at hand, we may now summarize the recursive estimation of state as shown in Table 1.1. This table also includes the initialization. In the absence of any observed data at time $k = 0$, we may choose the initial estimate of the state as

$$\hat{\mathbf{x}}_0 = E[\mathbf{x}_0], \tag{1.29}$$

Table 1.1 Summary of the Kalman filter

State-space model

$$\mathbf{x}_{k+1} = \mathbf{F}_{k+1,k}\mathbf{x}_k + \mathbf{w}_k,$$
$$\mathbf{y}_k = \mathbf{H}_k\mathbf{x}_k + \mathbf{v}_k,$$

where $\mathbf{w}_k$ and $\mathbf{v}_k$ are independent, zero-mean, Gaussian noise processes of covariance matrices $\mathbf{Q}_k$ and $\mathbf{R}_k$, respectively.

Initialization: For $k = 0$, set

$$\hat{\mathbf{x}}_0 = E[\mathbf{x}_0],$$
$$\mathbf{P}_0 = E[(\mathbf{x}_0 - E[\mathbf{x}_0])(\mathbf{x}_0 - E[\mathbf{x}_0])^T].$$

Computation: For $k = 1, 2, \ldots$, compute:

State estimate propagation

$$\hat{\mathbf{x}}_k^- = \mathbf{F}_{k,k-1}\hat{\mathbf{x}}_{k-1}^-;$$

Error covariance propagation

$$\mathbf{P}_k^- = \mathbf{F}_{k,k-1}\mathbf{P}_{k-1}\mathbf{F}_{k,k-1}^T + \mathbf{Q}_{k-1};$$

Kalman gain matrix

$$\mathbf{G}_k = \mathbf{P}_k^-\mathbf{H}_k^T\left[\mathbf{H}_k\mathbf{P}_k^-\mathbf{H}_k^T - \mathbf{R}_k\right]^{-1};$$

State estimate update

$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_k^- + \mathbf{G}_k\left(\mathbf{y}_k - \mathbf{H}_k\hat{\mathbf{x}}_k^-\right);$$

Error covariance update

$$\mathbf{P}_k = (\mathbf{I} - \mathbf{G}_k\mathbf{H}_k)\mathbf{P}_k^-.$$

and the initial value of the a posteriori covariance matrix as

$$\mathbf{P}_0 = E[(\mathbf{x}_0 - E[\mathbf{x}_0])(\mathbf{x}_0 - E[\mathbf{x}_0])^T]. \tag{1.30}$$

This choice for the initial conditions not only is intuitively satisfying but also has the advantage of yielding an *unbiased* estimate of the state $\mathbf{x}_k$.

1.4 DIVERGENCE PHENOMENON: SQUARE-ROOT FILTERING

The Kalman filter is prone to serious numerical difficulties that are well documented in the literature [6]. For example, the a posteriori covariance matrix $\mathbf{P}_k$ is defined as the difference between two matrices $\mathbf{P}_k^-$ and

$\mathbf{G}_k\mathbf{H}_k\mathbf{P}_k^-$; see Eq. (1.25). Hence, unless the numerical accuracy of the algorithm is high enough, the matrix $\mathbf{P}_k$ resulting from this computation may *not* be nonnegative-definite. Such a situation is clearly unacceptable, because $\mathbf{P}_k$ represents a covariance matrix. The unstable behavior of the Kalman filter, which results from numerical inaccuracies due to the use of finite-wordlength arithmetic, is called the *divergence phenomenon*.

A refined method of overcoming the divergence phenomenon is to use numerically stable unitary transformations at every iteration of the Kalman filtering algorithm [6]. In particular, the matrix $\mathbf{P}_k$ is propagated in a square-root form by using the *Cholesky factorization*:

$$\mathbf{P}_k = \mathbf{P}_k^{1/2}\mathbf{P}_k^{T/2}, \tag{1.31}$$

where $\mathbf{P}_k^{1/2}$ is reserved for a lower-triangular matrix, and $\mathbf{P}_k^{T/2}$ is its transpose. In linear algebra, the Cholesky factor $\mathbf{P}_k^{1/2}$ is commonly referred to as the square root of the matrix $\mathbf{P}_k$. Accordingly, any variant of the Kalman filtering algorithm based on the Cholesky factorization is referred to as square-root filtering. The important point to note here is that the matrix product $\mathbf{P}_k^{1/2}\mathbf{P}_k^{T/2}$ is much less likely to become indefinite, because the product of any square matrix and its transpose is always positive-definite. Indeed, even in the presence of roundoff errors, the numerical conditioning of the Cholesky factor $\mathbf{P}_k^{1/2}$ is generally much better than that of $\mathbf{P}_k$ itself.

1.5 RAUCH–TUNG–STRIEBEL SMOOTHER

In Section 1.3, we addressed the optimum linear filtering problem. The solution to the linear prediction problem follows in a straightforward manner from the basic theory of Section 1.3. In this section, we consider the *optimum smoothing problem*.

To proceed, suppose that we are given a set of data over the time interval $0 < k \leq N$. Smoothing is a non-real-time operation in that it involves estimation of the state $\mathbf{x}_k$ for $0 < k \leq N$, using all the available data, past as well as future. In what follows, we assume that the final time N is *fixed*.

To determine the optimum state estimates $\hat{\mathbf{x}}_k$ for $0 < k \leq N$, we need to account for past data $\mathbf{y}_j$ defined by $0 < j \leq k$, and future data $\mathbf{y}_j$ defined by $k < j \leq N$. The estimation pertaining to the past data, which we refer to as *forward filtering* theory, was presented in Section 1.3. To deal with the

issue of state estimation pertaining to the future data, we use *backward filtering*, which starts at the final time N and runs backwards. Let $\hat{\mathbf{x}}_k^f$ and $\hat{\mathbf{x}}_k^b$ denote the state estimates obtained from the forward and backward recursions, respectively. Given these two estimates, the next issue to be considered is how to combine them into an overall smoothed estimate $\hat{\mathbf{x}}_k$, which accounts for data over the entire time interval. Note that the symbol $\hat{\mathbf{x}}_k$ used for the smoothed estimate in this section is not to be confused with the filtered (i.e., a posteriori) estimate used in Section 1.3.

We begin by rewriting the process equation (1.1) as a recursion for decreasing k, as shown by

$$\mathbf{x}_k = \mathbf{F}_{k+1,k}^{-1}\mathbf{x}_{k+1} - \mathbf{F}_{k+1,k}^{-1}\mathbf{w}_k, \tag{1.32}$$

where $\mathbf{F}_{k+1,k}^{-1}$ is the inverse of the transition matrix $\mathbf{F}_{k+1,k}$. The rationale for backward filtering is depicted in Figure 1.2*a*, where the recursion begins at the final time N. This rationale is to be contrasted with that of forward filtering depicted in Figure 1.2*b*. Note that the a priori estimate $\hat{\mathbf{x}}_k^b$ and the a posteriori estimate $\hat{\mathbf{x}}_k^b$ for backward filtering occur to the right and left of time k in Figure 1.2*a*, respectively. This situation is the exact opposite to that occurring in the case of forward filtering depicted in Figure 1.2*b*.

To simplify the presentation, we introduce the two definitions:

$$\mathbf{S}_k = [\mathbf{P}_k^b]^{-1}, \tag{1.33}$$

$$\mathbf{S}_k^- = [\mathbf{P}_k^{b-}]^{-1}, \tag{1.34}$$

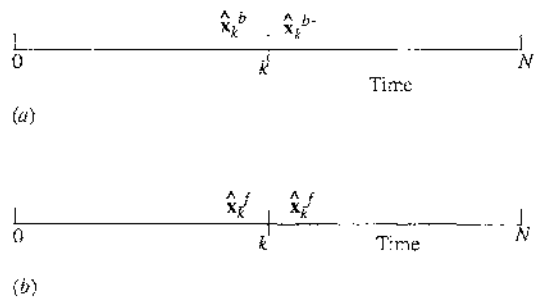

Figure 1.2 Illustrating the smoother time-updates for (*a*) backward filtering and (*b*) forward filtering.

and the two intermediate variables

$$\hat{\mathbf{z}}_k = [\mathbf{P}_k^b]^{-1}\hat{\mathbf{x}}_k^b = \mathbf{S}_k\hat{\mathbf{x}}_k^b, \tag{1.35}$$

$$\hat{\mathbf{z}}_k = [\mathbf{P}_k^b]^{-1}\hat{\mathbf{x}}_k^b = \mathbf{S}_k^-\hat{\mathbf{x}}_k^b . \tag{1.36}$$

Then, building on the rationale of Figure 1.2*a*, we may derive the following updates for the backward filter [2]:

1. *Measurement updates*

$$\mathbf{S}_k = \mathbf{S}_k^- + \mathbf{H}_k\mathbf{R}_k^{-1}\mathbf{H}_k, \tag{1.37}$$

$$\hat{\mathbf{z}}_k = \hat{\mathbf{z}}_k^- + \mathbf{H}_k^T\mathbf{R}_k^{-1}\mathbf{y}_k, \tag{1.38}$$

   where $\mathbf{y}_k$ is the observable defined by the measurement equation (1.3), $\mathbf{H}_k$ is the measurement matrix, and $\mathbf{R}_k^{-1}$ is the inverse of the covariance matrix of the measurement noise $\mathbf{v}_k$.

2. *Time updates*

$$\mathbf{G}_k^b = \mathbf{S}_{k+1}[\mathbf{S}_{k+1} + \mathbf{Q}_k^{-1}]^{-1}, \tag{1.39}$$

$$\mathbf{S}_k^- = \mathbf{F}_{k+1,k}^T(\mathbf{I} - \mathbf{G}_k^b)\mathbf{S}_{k+1}\mathbf{F}_{k+1,k}, \tag{1.40}$$

$$\hat{\mathbf{z}}_k^- = \mathbf{F}_{k+1,k}^T(\mathbf{I} - \mathbf{G}_k^b)\hat{\mathbf{z}}_{k+1}, \tag{1.41}$$

   where $\mathbf{G}_k^b$ is the Kalman gain for backward filtering and $\mathbf{Q}_k^{-1}$ is the inverse of the covariance matrix of the process noise $\mathbf{w}_k$. The backward filter defined by the measurement and time updates of Eqs. (1.37) (1.41) is the *information formulation* of the Kalman filter. The information filter is distinguished from the basic Kalman filter in that it propagates the inverse of the error covariance matrix rather than the error covariance matrix itself.

Given observable data over the interval $0 < k \leq N$ for fixed N, suppose we have obtained the following two estimates:

- The forward a posteriori estimate $\hat{\mathbf{x}}_k^f$ by operating the Kalman filter on data $\mathbf{y}_j$ for $0 < j \leq k$.
- The backward a priori estimate $\hat{\mathbf{x}}_k^b$ by operating the information filter on data $\mathbf{y}_j$ for $k < j \leq N$.

With these two estimates and their respective error covariance matrices at hand, the next issue of interest is how to determine the smoothed estimate $\hat{\mathbf{x}}_k$ and its error covariance matrix, which incorporate the overall data over the entire time interval $0 < k \leq N$.

Recognizing that the process noise $\mathbf{w}_k$ and measurement noise $\mathbf{v}_k$ are independent, we may formulate the error covariance matrix of the a posteriori smoothed estimate $\hat{\mathbf{x}}_k$ as follows:

$$\begin{aligned}\mathbf{P}_k &= [[\mathbf{P}_k^f]^{-1} - [\mathbf{P}_k^{b-}]^{-1}]^{-1} \\ &= [[\mathbf{P}_k^f]^{-1} + \mathbf{S}_k^-]^{-1}. \end{aligned} \tag{1.42}$$

To proceed further, we invoke the *matrix inversion lemma*, which may be stated as follows [7]. Let $\mathbf{A}$ and $\mathbf{B}$ be two positive-definite matrices related by

$$\mathbf{A} = \mathbf{B}^{-1} + \mathbf{C}\mathbf{D}^{-1}\mathbf{C}^T,$$

where $\mathbf{D}$ is another positive-definite matrix and $\mathbf{C}$ is a matrix with compatible dimensions. The matrix inversion lemma states that we may express the inverse of the matrix $\mathbf{A}$ as follows:

$$\mathbf{A}^{-1} = \mathbf{B} - \mathbf{B}\mathbf{C}[\mathbf{D} + \mathbf{C}^T\mathbf{B}\mathbf{C}]^{-1}\mathbf{C}^T\mathbf{B}.$$

For the problem at hand, we set

$$\begin{aligned}\mathbf{A} &= \mathbf{P}_k^{-1}, \\ \mathbf{B} &= \mathbf{P}_k^f, \\ \mathbf{C} &= \mathbf{I}, \\ \mathbf{D} &= [\mathbf{S}_k^-]^{-1}, \end{aligned}$$

where $\mathbf{I}$ is the identity matrix. Then, applying the matrix inversion lemma to Eq. (1.42), we obtain

$$\begin{aligned}\mathbf{P}_k &= \mathbf{P}_k^f - \mathbf{P}_k^f[\mathbf{P}_k^{b-} + \mathbf{P}_k^f]^{-1}\mathbf{P}_k^f \\ &= \mathbf{P}_k^f - \mathbf{P}_k^f\mathbf{S}_k[\mathbf{I} + \mathbf{P}_k^f\mathbf{S}_k]^{-1}\mathbf{P}_k^f. \end{aligned} \tag{1.43}$$

From Eq. (1.43), we find that the a posteriori smoothed error covariance matrix $\mathbf{P}_k$ is smaller than or equal to the a posteriori error covariance

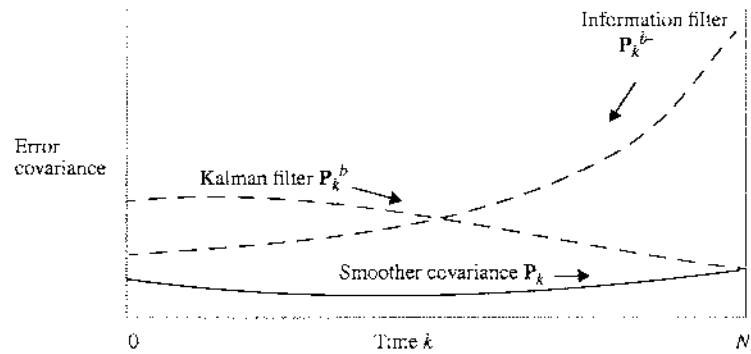

Figure 1.3 Illustrating the error covariance for forward filtering, backward filtering, and smoothing.

matrix $\mathbf{P}_k^f$ produced by the Kalman filter, which is naturally due to the fact that smoothing uses additional information contained in the future data. This point is borne out by Figure 1.3, which depicts the variations of $\mathbf{P}_k$, $\mathbf{P}_k^f$, and $\mathbf{P}_k^b$ with k for a one-dimensional situation.

The a posteriori smoothed estimate of the state is defined by

$$\hat{\mathbf{x}}_k = \mathbf{P}_k([\mathbf{P}_k^f]^{-1}\hat{\mathbf{x}}_k^f + [\mathbf{P}_k^{b-}]^{-1}\hat{\mathbf{x}}_k^{b-}). \tag{1.44}$$

Using Eqs. (1.36) and (1.43) in (1.44) yields, after simplification,

$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_k^f - (\mathbf{P}_k\mathbf{z}_k - \mathbf{G}_k\hat{\mathbf{x}}_k^f), \tag{1.45}$$

where the *smoother gain* is defined by

$$\mathbf{G}_k = \mathbf{P}_k^f\mathbf{S}_k[\mathbf{I} + \mathbf{P}_k^f\mathbf{S}_k^-]^{-1}, \tag{1.46}$$

which is not to be confused with the Kalman gain of Eq. (1.22).

The optimum smoother just derived consists of three components:

- A forward filter in the form of a Kalman filter.
- A backward filter in the form of an information filter.
- A separate smoother, which combines results embodied in the forward and backward filters.

The *Rauch–Tung–Striebel* smoother, however, is more efficient than the three-part smoother in that it incorporates the backward filter and separate

smoother into a single entity [8, 9]. Specifically, the measurement update of the Rauch–Tung–Striebel smoother is defined by

$$\mathbf{P}_k = \mathbf{P}_k^f - \mathbf{A}_k(\mathbf{P}_{k+1}^f - \mathbf{P}_{k+1})\mathbf{A}_k^T, \tag{1.47}$$

where $\mathbf{A}_k$ is the new gain matrix:

$$\mathbf{A}_k = \mathbf{P}_k^f \mathbf{F}_{k+1,k}^T [\mathbf{P}_{k+1}^f]^{-1}. \tag{1.48}$$

The corresponding time update is defined by

$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_k^f + \mathbf{A}_k(\hat{\mathbf{x}}_{k+1} - \hat{\mathbf{x}}_{k+1}^f) \tag{1.49}$$

The Rauch–Tung–Striebel smoother thus proceeds as follows:

1. The Kalman filter is applied to the observable data in a forward manner, that is, $k = 0, 1, 2, \ldots$, in accordance with the basic theory summarized in Table 1.1.
2. The recursive smoother is applied to the observable data in a backward manner, that is, $k = N-1, N-2, \ldots$, in accordance with Eqs. (1.47)–(1.49).
3. The initial conditions are defined by

$$\mathbf{P}_N = \mathbf{P}_N^f, \tag{1.50}$$

$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_k^f. \tag{1.51}$$

Table 1.2 summarizes the computations involved in the Rauch–Tung–Striebel smoother.

1.6 EXTENDED KALMAN FILTER

The Kalman filtering problem considered up to this point in the discussion has addressed the estimation of a state vector in a linear model of a dynamical system. If, however, the model is *nonlinear*, we may extend the use of Kalman filtering through a linearization procedure. The resulting filter is referred to as the *extended Kalman filter* (*EKF*) [10–12]. Such an

Table 1.2 Summary of the Rauch–Tung–Striebel smoother

State-space model

$$\mathbf{x}_{k+1} = \mathbf{F}_{k+1,k}\mathbf{x}_k + \mathbf{w}_k$$
$$\mathbf{y}_k = \mathbf{H}_k\mathbf{x}_k + \mathbf{v}_k$$

where $\mathbf{w}_k$ and $\mathbf{v}_k$ are independent, zero-mean, Gaussian noise processes of covariance matrices $\mathbf{Q}_k$ and $\mathbf{R}_k$, respectively.

Forward filter

Initialization: For $k = 0$, set

$$\hat{\mathbf{x}}_0 = E[\mathbf{x}_0],$$
$$\mathbf{P}_0 = E[(\mathbf{x}_0 - E[\mathbf{x}_0])(\mathbf{x}_0 - E[\mathbf{x}_0])^T].$$

Computation: For $k = 1, 2, \ldots$, compute

$$\hat{\mathbf{x}}_k^{f-} = \mathbf{F}_{k,k-1}\hat{\mathbf{x}}_{k-1}^{f-},$$
$$\mathbf{P}_k^{f-} = \mathbf{F}_{k,k-1}\mathbf{P}_{k-1}^{f-}\mathbf{F}_{k,k-1}^T + \mathbf{Q}_{k-1},$$
$$\mathbf{G}_k^f = \mathbf{P}_k^{f-}\mathbf{H}_k^T[\mathbf{H}_k\mathbf{P}_k^{f-}\mathbf{H}_k^T + \mathbf{R}_k]^{-1},$$
$$\hat{\mathbf{x}}_k^{f-} = \hat{\mathbf{x}}_k^{f-} + \mathbf{G}_k^f(\mathbf{y}_k - \mathbf{H}_k\hat{\mathbf{x}}_k^{f-}).$$

Recursive smoother

Initialization: For $k = N$, set

$$\mathbf{P}_N = \mathbf{P}_N^f,$$
$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_k^f.$$

Computation: For $k = N-1, N-2$, compute

$$\mathbf{A}_k = \mathbf{P}_k^f\mathbf{F}_{k+1,k}^T[\mathbf{P}_{k+1}^{f-}]^{-1},$$
$$\mathbf{P}_k = \mathbf{P}_k^f - \mathbf{A}_k(\mathbf{P}_{k+1}^{f-} - \mathbf{P}_{k+1})\mathbf{A}_k^T,$$
$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_k^f - \mathbf{A}_k\left(\hat{\mathbf{x}}_{k+1} - \hat{\mathbf{x}}_{k+1}^f\right).$$

extension is feasible by virtue of the fact that the Kalman filter is described in terms of difference equations in the case of discrete-time systems.

To set the stage for a development of the extended Kalman filter, consider a nonlinear dynamical system described by the state-space model

$$\mathbf{x}_{k+1} = \mathbf{f}(k, \mathbf{x}_k) + \mathbf{w}_k, \tag{1.52}$$

$$\mathbf{y}_k = \mathbf{h}(k, \mathbf{x}_k) + \mathbf{v}_k, \tag{1.53}$$

where, as before, $\mathbf{w}_k$ and $\mathbf{v}_k$ are independent zero-mean white Gaussian noise processes with covariance matrices $\mathbf{R}_k$ and $\mathbf{Q}_k$, respectively. Here, however, the functional $\mathbf{f}(k, \mathbf{x}_k)$ denotes a nonlinear transition matrix function that is possibly time-variant. Likewise, the functional $\mathbf{h}(k, \mathbf{x}_k)$ denotes a *nonlinear measurement matrix* that may be time-variant, too.

The basic idea of the extended Kalman filter is to *linearize* the state-space model of Eqs. (1.52) and (1.53) at each time instant around the most recent state estimate, which is taken to be either $\hat{\mathbf{x}}_k$ or $\hat{\mathbf{x}}_k^-$, depending on which particular functional is being considered. Once a linear model is obtained, the standard Kalman filter equations are applied.

More explicitly, the approximation proceeds in two stages.

Stage 1 The following two matrices are constructed:

$$\mathbf{F}_{k+1,k} = \left.\frac{\partial \mathbf{f}(k, \mathbf{x})}{\partial \mathbf{x}}\right|_{\mathbf{x}=\hat{\mathbf{x}}_k}, \tag{1.54}$$

$$\mathbf{H}_k = \left.\frac{\partial \mathbf{h}(k, \mathbf{x}_k)}{\partial \mathbf{x}}\right|_{\mathbf{x}=\hat{\mathbf{x}}_k^-}. \tag{1.55}$$

That is, the ijth entry of $\mathbf{F}_{k+1,k}$ is equal to the partial derivative of the ith component of $\mathbf{F}(k, \mathbf{x})$ with respect to the jth component of $\mathbf{x}$. Likewise, the ijth entry of $\mathbf{H}_k$ is equal to the partial derivative of the ith component of $\mathbf{H}(k, \mathbf{x})$ with respect to the jth component of $\mathbf{x}$. In the former case, the derivatives are evaluated at $\hat{\mathbf{x}}_k$, while in the latter case, the derivatives are evaluated at $\hat{\mathbf{x}}_k^-$. The entries of the matrices $\mathbf{F}_{k+1,k}$ and $\mathbf{H}_k$ are all known (i.e., computable), by having $\hat{\mathbf{x}}_k$ and $\hat{\mathbf{x}}_k^-$ available at time k.

Stage 2 Once the matrices $\mathbf{F}_{k+1,k}$ and $\mathbf{H}_k$ are evaluated, they are then employed in a *first-order Taylor approximation* of the nonlinear functions $\mathbf{F}(k, \mathbf{x}_k)$ and $\mathbf{H}(k, \mathbf{x}_k)$ around $\hat{\mathbf{x}}_k$ and $\hat{\mathbf{x}}_k^-$, respectively. Specifically, $\mathbf{F}(k, \mathbf{x}_k)$ and $\mathbf{H}(k, \mathbf{x}_k)$ are approximated as follows

$$\mathbf{F}(k, \mathbf{x}_k) \approx \mathbf{F}(\mathbf{x}, \hat{\mathbf{x}}_k) + \mathbf{F}_{k+1,k}(\mathbf{x}, \hat{\mathbf{x}}_k), \tag{1.56}$$

$$\mathbf{H}(k, \mathbf{x}_k) \approx \mathbf{H}(\mathbf{x}, \hat{\mathbf{x}}_k^-) + \mathbf{H}_{k+1,k}(\mathbf{x}, \hat{\mathbf{x}}_k^-). \tag{1.57}$$

With the above approximate expressions at hand, we may now proceed to approximate the nonlinear state equations (1.52) and (1.53) as shown by, respectively,

$$\mathbf{x}_{k+1} \approx \mathbf{F}_{k+1,k}\mathbf{x}_k + \mathbf{w}_k + \mathbf{d}_k,$$

$$\mathbf{y}_k \approx \mathbf{H}_k\mathbf{x}_k + \mathbf{v}_k,$$

where we have introduced two new quantities:

$$\bar{\mathbf{y}}_k = \mathbf{y}_k - \{\mathbf{h}(\mathbf{x}, \hat{\mathbf{x}}_k^-) - \mathbf{H}_k \hat{\mathbf{x}}_k^-\}, \tag{1.58}$$

$$\mathbf{d}_k = \mathbf{f}(\mathbf{x}, \hat{\mathbf{x}}_k) - \mathbf{F}_{k+1,k}\hat{\mathbf{x}}_k. \tag{1.59}$$

The entries in the term $\bar{\mathbf{y}}_k$ are all known at time k, and, therefore, $\bar{\mathbf{y}}_k$ can be regarded as an observation vector at time n. Likewise, the entries in the term $\mathbf{d}_k$ are all known at time k.

Table 1.3 Extended Kalman filter

State-space model

$$\mathbf{x}_{k+1} = \mathbf{f}(k, \mathbf{x}_k) + \mathbf{w}_k,$$
$$\mathbf{y}_k = \mathbf{h}(k, \mathbf{x}_k) + \mathbf{v}_k,$$

where $\mathbf{w}_k$ and $\mathbf{v}_k$ are independent, zero mean, Gaussian noise processes of covariance matrices $\mathbf{Q}_k$ and $\mathbf{R}_k$, respectively.

Definitions

$$\mathbf{F}_{k+1,k} = \frac{\partial \mathbf{f}(k, \mathbf{x})}{\partial \mathbf{x}}\Big|_{\mathbf{x}=\mathbf{x}_k},$$

$$\mathbf{H}_k = \frac{\partial \mathbf{h}(k, \mathbf{x})}{\partial \mathbf{x}}\Big|_{\mathbf{x}=\mathbf{x}_k^-}.$$

Initialization: For $k = 0$, set

$$\hat{\mathbf{x}}_0 = E[\mathbf{x}_0],$$
$$\mathbf{P}_0 = E[(\mathbf{x}_0 - E[\mathbf{x}_0])(\mathbf{x}_0 - E[\mathbf{x}_0])^T].$$

Computation: For $k = 1, 2, \ldots$, compute:

State estimate propagation

$$\hat{\mathbf{x}}_k^- = \mathbf{f}(k, \hat{\mathbf{x}}_{k-1});$$

Error covariance propagation

$$\mathbf{P}_k^- = \mathbf{F}_{k,k-1}\mathbf{P}_{k-1}\mathbf{F}_{k,k-1}^T + \mathbf{Q}_{k-1};$$

Kalman gain matrix

$$\mathbf{G}_k = \mathbf{P}_k^-\mathbf{H}_k^T\left[\mathbf{H}_k\mathbf{P}_k^-\mathbf{H}_k^T + \mathbf{R}_k\right]^{-1};$$

State estimate update

$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_k^- + \mathbf{G}_k\mathbf{y}_k - \mathbf{h}(k, \hat{\mathbf{x}}_k^-);$$

Error covariance update

$$\mathbf{P}_k = (\mathbf{I} - \mathbf{G}_k\mathbf{H}_k)\mathbf{P}_k^-.$$

Given the linearized state-space model of Eqs. (1.58) and (1.59), we may then proceed and apply the Kalman filter theory of Section 1.3 to derive the extended Kalman filter. Table 1.2 summarizes the recursions involved in computing the extended Kalman filter.

1.7 SUMMARY

The basic Kalman filter is a linear, discrete-time, finite-dimensional system, which is endowed with a recursive structure that makes a digital computer well suited for its implementation. A key property of the Kalman filter is that it is the *minimum mean-square (variance) estimator of the state* of a linear dynamical system.

The Kalman filter, summarized in Table 1.1, applies to a linear dynamical system, the state space model of which consists of two equations:

- The process equation that defines the evolution of the state with time.
- The measurement equation that defines the observable in terms of the state.

The model is stochastic owing to the additive presence of process noise and measurement noise, which are assumed to be Gaussian with zero mean and known covariance matrices.

The Rauch Tung Striebel smoother, summarized in Table 1.2, builds on the Kalman filter to solve the optimum smoothing problem in an efficient manner. This smoother consists of two components: a forward filter based on the basic Kalman filter, and a combined backward filter and smoother.

Applications of Kalman filter theory may be extended to nonlinear dynamical systems, as summarized in Table 1.3. The derivation of the extended Kalman filter hinges on linearization of the nonlinear state-space model on the assumption that deviation from linearity is of first order.

REFERENCES

[1] R.E. Kalman, "A new approach to linear filtering and prediction problems," *Transactions of the ASME, Ser. D, Journal of Basic Engineering*, **82**, 34–45 (1960).

[2] F.H. Lewis, *Optical Estimation with an Introduction to Stochastic Control Theory*. New York: Wiley, 1986.

[3] M.S. Grewal and A.P. Andrews, *Kalman Filtering: Theory and Practice*. Englewood Cliffs, NJ: Prentice-Hall, 1993.

[4] H.L. Van Trees, *Detection, Estimation, and Modulation Theory*, Part I. New York: Wiley, 1968.

[5] R.S. Bucy and P.D. Joseph, *Filtering for Stochastic Processes, with Applications to Guidance*. New York: Wiley, 1968.

[6] P.G. Kaminski, A.E. Bryson, Jr., and S.F. Schmidt, "Discrete square root filtering: a survey of current techniques," *IEEE Transactions on Automatic Control*, **16**, 727–736 (1971).

[7] S. Haykin, *Adaptive Filter Theory*, 3rd ed. Upper Saddle River, NJ: Prentice-Hall, 1996.

[8] H.E. Rauch, "Solutions to the linear smoothing problem," *IEEE Transactions on Automatic Control*, **11**, 371–372 (1963).

[9] H.E. Rauch, F. Tung, and C.T. Striebel, "Maximum likelihood estimates of linear dynamic systems," *AIAA Journal*, **3**, 1445–1450 (1965).

[10] A.H. Jazwinski, *Stochastic Processes and Filtering Theory*. New York: Academic Press, 1970.

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

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