CHAPTER 5

The discrete-time Kalman filter

The Kalman filter in its various forms is clearly established as a fundamental tool for analyzing and solving a broad class of estimation problems.

—Leonard McGee and Stanley Schmidt [McG85]

This chapter forms the heart of this book. The earlier chapters were written only to provide the foundation for this chapter, and the later chapters are written only to expand and generalize the results given in this chapter.

As we will see in this chapter, the Kalman filter operates by propagating the mean and covariance of the state through time. Our approach to deriving the Kalman filter will involve the following steps.

1. We start with a mathematical description of a dynamic system whose states we want to estimate.

2. We implement equations that describe how the mean of the state and the covariance of the state propagate with time. These equations, derived in Chapter 4, themselves form a dynamic system.

3. We take the dynamic system that describes the propagation of the state mean and covariance, and implement the equations on a computer. These equations form the basis for the derivation of the Kalman filter because:

Optimal State Estimation, First Edition. By Dan J. Simon

ISBN 0471708585

   (a) The mean of the state is the Kalman filter estimate of the state.

   (b) The covariance of the state is the covariance of the Kalman filter state estimate.

4. Every time that we get a measurement, we update the mean and covariance of the state. This is similar to what we did in Chapter 3 where we used measurements to recursively update our estimate of a constant.

In Section 5.1, we derive the equations of the discrete-time Kalman filter. This includes several different-looking, but mathematically equivalent forms. Various books and papers that deal with Kalman filters present the filter equations in ways that appear very different from one another. It is not always obvious, but these different formulations are actually mathematically equivalent, and we will see this in Section 5.1. (Sections 9.1, 10.5.1, and 11.1 also derive alternate but equivalent formulations of the Kalman filter equations.) In Section 5.2, we will examine some of the theoretical properties of the Kalman filter. One remarkable aspect of the Kalman filter is that it is optimal in several different senses, as we will see in Section 5.2. In Section 5.3, we will see how the Kalman filter equations can be written with a single time update equation. Section 5.4 presents a way to obtain a closed-form equation for the time-varying Kalman filter for a scalar time-invariant system, and a way to quickly compute the steady-state Kalman filter. Section 5.5 looks at some situations in which the Kalman filter is unstable or gives state estimates that are not close to the true state. We will also look at some ways that instability and divergence can be corrected in the Kalman filter.

5.1 DERIVATION OF THE DISCRETE-TIME KALMAN FILTER

Suppose we have a linear discrete-time system given as follows:

$$\begin{aligned} x_k &= F_{k-1}x_{k-1} + G_{k-1}u_{k-1} + w_{k-1} \\ y_k &= H_k x_k + v_k \end{aligned} \tag{5.1}$$

The noise processes $\{w_k\}$ and $\{v_k\}$ are white, zero-mean, uncorrelated, and have known covariance matrices Q_k and R_k, respectively:

$$\begin{aligned} w_k &\sim (0, Q_k) \\ v_k &\sim (0, R_k) \\ E[w_k w_j^T] &= Q_k \delta_{k-j} \\ E[v_k v_j^T] &= R_k \delta_{k-j} \\ E[v_k w_j^T] &= 0 \end{aligned} \tag{5.2}$$

where δ_{k-j} is the Kronecker delta function; that is, $\delta_{k-j} = 1$ if $k = j$, and $\delta_{k-j} = 0$ if $k \neq j$. Our goal is to estimate the state x_k based on our knowledge of the system dynamics and the availability of the noisy measurements $\{y_k\}$. The amount of information that is available to us for our state estimate varies depending on the particular problem that we are trying to solve. If we have all of the measurements up to and including time k available for use in our estimate of x_k, then we can form an *a posteriori* estimate, which we denote as $\hat{x}_k^+$. The "+" superscript denotes that the estimate is *a posteriori*. One way to form the *a posteriori* state estimate is to compute the expected value of x_k conditioned on all of the measurements up to and including time k:

$$\hat{x}_k^+ = E[x_k|y_1, y_2, \cdots, y_k] = \textit{a posteriori} \text{ estimate} \tag{5.3}$$

If we have all of the measurements before (but not including) time k available for use in our estimate of x_k, then we can form an *a priori* estimate, which we denote as $\hat{x}_k^-$. The "–" superscript denotes that the estimate is *a priori*. One way to form the *a priori* state estimate is to compute the expected value of x_k conditioned on all of the measurements before (but not including) time k:

$$\hat{x}_k^- = E[x_k|y_1, y_2, \cdots, y_{k-1}] = \textit{a priori} \text{ estimate} \tag{5.4}$$

It is important to note that $\hat{x}_k^-$ and $\hat{x}_k^+$ are both estimates of the same quantity; they are both estimates of x_k. However, $\hat{x}_k^-$ is our estimate of x_k *before* the measurement y_k is taken into account, and $\hat{x}_k^+$ is our estimate of x_k *after* the measurement y_k is taken into account. We naturally expect $\hat{x}_k^+$ to be a better estimate than $\hat{x}_k^-$, because we use more information to compute $\hat{x}_k^+$:

$$\begin{aligned} \hat{x}_k^- &= \text{estimate of } x_k \text{ before we process the measurement at time } k \\ \hat{x}_k^+ &= \text{estimate of } x_k \text{ after we process the measurement at time } k \end{aligned} \tag{5.5}$$

If we have measurements after time k available for use in our estimate of x_k, then we can form a *smoothed* estimate. One way to form the smoothed state estimate is to compute the expected value of x_k conditioned on all of the measurements that are available:

$$\hat{x}_{k|k+N} = E[x_k|y_1, y_2, \cdots, y_k, \cdots, y_{k+N}] = \text{smoothed estimate} \tag{5.6}$$

where N is some positive integer whose value depends on the specific problem that is being solved. If we want to find the best prediction of x_k more than one time step ahead of the available measurements, then we can form a *predicted* estimate. One way to form the predicted state estimate is to compute the expected value of x_k conditioned on all of the measurements that are available:

$$\hat{x}_{k|k-M} = E[x_k|y_1, y_2, \cdots, y_{k-M}] = \text{predicted estimate} \tag{5.7}$$

where M is some positive integer whose value depends on the specific problem that is being solved. The relationship between the *a posteriori*, *a priori*, smoothed, and predicted state estimates is depicted in Figure 5.1.

In the notation that follows, we use $\hat{x}_0^+$ to denote our initial estimate of x_0 before any measurements are available. The first measurement is taken at time $k = 1$. Since we do not have any measurements available to estimate x_0, it is reasonable to form $\hat{x}_0^+$ as the expected value of the initial state x_0:

$$\hat{x}_0^+ = E(x_0) \tag{5.8}$$

We use the term P_k to denote the covariance of the estimation error. P_k^- denotes the covariance of the estimation error of $\hat{x}_k^-$, and P_k^+ denotes the covariance of the estimation error of $\hat{x}_k^+$:

$$\begin{aligned} P_k^- &= E[(x_k - \hat{x}_k^-)(x_k - \hat{x}_k^-)^T] \\ P_k^+ &= E[(x_k - \hat{x}_k^+)(x_k - \hat{x}_k^+)^T] \end{aligned} \tag{5.9}$$

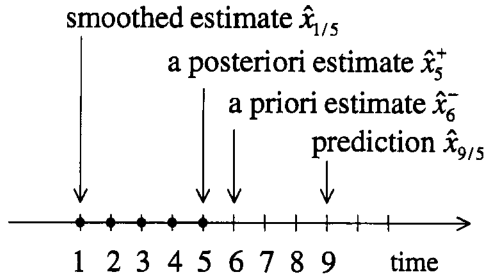

Figure 5.1 Time line showing the relationship between the *a posteriori*, *a priori*, smoothed, and predicted state estimates. In this figure, we suppose that we have received measurements at times up to and including $k = 5$. An estimate of the state at $k < 5$ is called a smoothed estimate. An estimate of the state at $k = 5$ is called the *a posteriori* estimate. An estimate of the state at $k = 6$ is called the *a priori* estimate. An estimate of the state at $k > 6$ is called the prediction.

These relationships are depicted in Figure 5.2. The figure shows that after we process the measurement at time $(k-1)$, we have an estimate of x_{k-1} (denoted $\hat{x}_{k-1}^+$) and the covariance of that estimate (denoted P_{k-1}^+). When time k arrives, before we process the measurement at time k we compute an estimate of x_k (denoted $\hat{x}_k^-$) and the covariance of that estimate (denoted P_k^-). Then we process the measurement at time k to refine our estimate of x_k. The resulting estimate of x_k is denoted $\hat{x}_k^+$, and its covariance is denoted P_k^+.

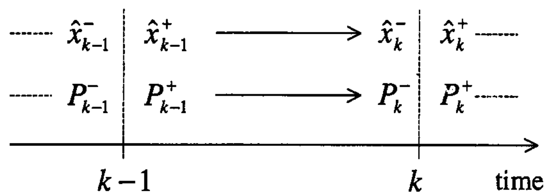

Figure 5.2 Timeline showing *a priori* and *a posteriori* state estimates and estimation-error covariances.

We begin the estimation process with $\hat{x}_0^+$, our best estimate of the initial state x_0. Given $\hat{x}_0^+$, how should we compute $\hat{x}_1^-$? We want to set $\hat{x}_1^- = E(x_1)$. But note that $\hat{x}_0^+ = E(x_0)$, and recall from Equation (4.2) how the mean of x propagates with time: $\bar{x}_k = F_{k-1}\bar{x}_{k-1} + G_{k-1}u_{k-1}$. We therefore obtain

$$\hat{x}_1^- = F_0\hat{x}_0^+ + G_0u_0 \tag{5.10}$$

This is a specific equation that shows how to obtain $\hat{x}_1^-$ from $\hat{x}_0^+$. However, the reasoning can be extended to obtain the following more general equation:

$$\hat{x}_k^- = F_{k-1}\hat{x}_{k-1}^+ + G_{k-1}u_{k-1} \tag{5.11}$$

This is called the time update equation for $\hat{x}$. From time $(k-1)^+$ to time k^-, the state estimate propagates the same way that the mean of the state propagates. This makes sense intuitively. We do not have any additional measurements available to help us update our state estimate between time $(k-1)^+$ and time k^-, so we should just update the state estimate based on our knowledge of the system dynamics.

Next we need to compute the time update equation for P, the covariance of the state estimation error. We begin with P_0^+, which is the covariance of our initial estimate of x_0. If we know the initial state perfectly, then $P_0^+ = 0$. If we have absolutely no idea of the value of x_0, then $P_0^+ = \infty I$. In general, P_0^+ represents the uncertainty in our initial estimate of x_0:

$$\begin{aligned} P_0^+ &= E[(x_0 - \bar{x}_0)(x_0 - \bar{x}_0)^T] \\ &= E[(x_0 - \hat{x}_0^+)(x_0 - \hat{x}_0^+)^T] \end{aligned} \tag{5.12}$$

Given P_0^+, how can we compute P_1^-? Recall from Equation (4.4) how the covariance of the state of a linear discrete-time system propagates with time: $P_k = F_{k-1}P_{k-1}F_{k-1}^T + Q_{k-1}$. We therefore obtain

$$P_1^- = F_0 P_0^+ F_0^T + Q_0 \tag{5.13}$$

This is a specific equation that shows how to obtain P_1^- from P_0^+. However, the reasoning can be extended to obtain the following more general equation:

$$P_k^- = F_{k-1} P_{k-1}^+ F_{k-1}^T + Q_{k-1} \tag{5.14}$$

This is called the time-update equation for P.

We have derived the time-update equations for $\hat{x}$ and P. Now we need to derive the measurement-update equations for $\hat{x}$ and P. Given $\hat{x}_k^-$, how should we compute $\hat{x}_k^+$? The quantity $\hat{x}_k^-$ is an estimate of x_k, and the quantity $\hat{x}_k^+$ is also an estimate of x_k. The only difference between $\hat{x}_k^-$ and $\hat{x}_k^+$ is that $\hat{x}_k^+$ takes the measurement y_k into account. Recall from the recursive least squares development in Section 3.3 that the availability of the measurement y_k changes the estimate of a constant x as follows:

$$\begin{aligned} K_k &= P_{k-1}H_k^T(H_k P_{k-1} H_k^T + R_k)^{-1} \\ &= P_k H_k^T R_k^{-1} \\ \hat{x}_k &= \hat{x}_{k-1} + K_k(y_k - H_k \hat{x}_{k-1}) \\ P_k &= (I - K_k H_k) P_{k-1} (I - K_k H_k)^T + K_k R_k K_k^T \\ &= (P_{k-1}^{-1} + H_k^T R_k^{-1} H_k)^{-1} \\ &= (I - K_k H_k) P_{k-1} \end{aligned} \tag{5.15}$$

where $\hat{x}_{k-1}$ and P_{k-1} are the estimate and its covariance *before* the measurement y_k is processed, and $\hat{x}_k$ and P_k are the estimate and its covariance *after* the measurement y_k is processed. In this chapter, $\hat{x}_k^-$ and P_k^- are the estimate and its covariance before the measurement y_k is processed, and $\hat{x}_k^+$ and P_k^+ are the estimate and its covariance after the measurement y_k is processed. These relationships are shown in Table 5.1.[1]

We can now generalize from the formulas for the estimation of a constant in Section 3.3, to the measurement update equations required in this section. In

[1]We need to use minus and plus superscripts on $\hat{x}_k$ and P_k in order to distinguish between quantities before y_k is taken into account, and quantities after y_k is taken into account. In Chapter 3, we did not need superscripts because x was a constant.

Table 5.1 Relationships between estimates and covariances in Sections 3.3 and 5.1

Section 3.3 Least squares estimation		Section 5.1 Kalman filtering
$\hat{x}_{k-1}$ = estimate before y_k is processed	$\Longrightarrow$	$\hat{x}_k^-$ = *a priori* estimate
P_{k-1} = covariance before y_k is processed	$\Longrightarrow$	P_k^- = *a priori* covariance
$\hat{x}_k$ = estimate after y_k is processed	$\Longrightarrow$	$\hat{x}_k^+$ = *a posteriori* estimate
P_k = covariance after y_k is processed	$\Longrightarrow$	P_k^+ = *a posteriori* covariance

Equation (5.15), we replace $\hat{x}_{k-1}$ with $\hat{x}_k^-$, we replace P_{k-1} with P_k^-, we replace $\hat{x}_k$ with $\hat{x}_k^+$, and we replace P_k with P_k^+. This results in

$$\begin{aligned}
K_k &= P_k^- H_k^T (H_k P_k^- H_k^T + R_k)^{-1} \\
&= P_k^+ H_k^T R_k^{-1} \\
\hat{x}_k^+ &= \hat{x}_k^- + K_k (y_k - H_k \hat{x}_k^-) \\
P_k^+ &= (I - K_k H_k) P_k^- (I - K_k H_k)^T + K_k R_k K_k^T \\
&= \left[(P_k^-)^{-1} + H_k^T R_k^{-1} H_k \right]^{-1} \\
&= (I - K_k H_k) P_k^-
\end{aligned} \tag{5.16}$$

These are the measurement-update equations for $\hat{x}_k$ and P_k. The matrix K_k in the above equations is called the Kalman filter gain.

The discrete-time Kalman filter

Here we summarize the discrete-time Kalman filter by combining the above equations into a single algorithm.

1. The dynamic system is given by the following equations:

$$\begin{aligned}
x_k &= F_{k-1} x_{k-1} + G_{k-1} u_{k-1} + w_{k-1} \\
y_k &= H_k x_k + v_k \\
E(w_k w_j^T) &= Q_k \delta_{k-j} \\
E(v_k v_j^T) &= R_k \delta_{k-j} \\
E(w_k v_j^T) &= 0
\end{aligned} \tag{5.17}$$

2. The Kalman filter is initialized as follows:

$$\begin{aligned}
\hat{x}_0^+ &= E(x_0) \\
P_0^+ &= E[(x_0 - \hat{x}_0^+)(x_0 - \hat{x}_0^+)^T]
\end{aligned} \tag{5.18}$$

3. The Kalman filter is given by the following equations, which are computed for each time step $k = 1, 2, \cdots$:

$$\begin{aligned}
P_k^- &= F_{k-1} P_{k-1}^+ F_{k-1}^T + Q_{k-1} \\
K_k &= P_k^- H_k^T (H_k P_k^- H_k^T + R_k)^{-1}
\end{aligned}$$

$$
\begin{aligned}
&= P_k^+ H_k^T R_k^{-1} \\
\hat{x}_k^- &= F_{k-1}\hat{x}_{k-1}^+ + G_{k-1}u_{k-1} = \text{ \textit{a priori} state estimate} \\
\hat{x}_k^+ &= \hat{x}_k^- + K_k(y_k - H_k\hat{x}_k^-) = \text{ \textit{a posteriori} state estimate} \\
P_k^+ &= (I - K_kH_k)P_k^-(I - K_kH_k)^T + K_kR_kK_k^T \\
&= \left[(P_k^-)^{-1} + H_k^T R_k^{-1} H_k\right]^{-1} \\
&= (I - K_kH_k)P_k^-
\end{aligned} \tag{5.19}
$$

The first expression for P_k^+ above is called the Joseph stabilized version of the covariance measurement update equation. It was formulated by Peter Joseph in the 1960s and can be shown to be more stable and robust than the third expression for P_k^+ [Buc68, Cra04] (see Problem 5.2). The first expression for P_k^+ guarantees that P_k^+ will always be symmetric positive definite, as long as P_k^- is symmetric positive definite. The third expression for P_k^+ is computationally simpler than the first expression, but its form does not guarantee symmetry or positive definiteness for P_k^+. The second form for P_k^+ is rarely implemented as written above but will be useful in our derivation of the information filter in Section 6.2.

If the second expression for K_k is used, then the second expression for P_k^+ must be used. This is because the second expression for K_k depends on P_k^+, so we need to use an expression for P_k^+ that does not depend on K_k.

Note that if x_k is a constant, then $F_k = I$, $Q_k = 0$, and $u_k = 0$. In this case, the Kalman filter of Equation (5.19) reduces to the recursive least squares algorithm for the estimation of a constant vector as given in Equation (3.47).

Finally we mention one more important practical aspect of the Kalman filter. We see from Equation (5.19) that the calculation of P_k^-, K_k, and P_k^+ does not depend on the measurements y_k, but depends only on the system parameters F_k, H_k, Q_k, and R_k. That means that the Kalman gain K_k can be calculated offline before the system operates and saved in memory. Then when it comes time to operate the system in real time, only the $\hat{x}_k$ equations need to be implemented in real time. The computational effort of calculating K_k can be saved during real-time operation by precomputing it. If the Kalman filter is implemented in an embedded system with strict computational requirements, this can make the difference between whether or not the system can be implemented in real time. Furthermore, the performance of the filter can be investigated and evaluated before the filter is actually run. This is because P_k indicates the estimation accuracy of the filter, and it can be computed offline since it does not depend on the measurements. In contrast, as we will see in Chapter 13, the filter gain and covariance for nonlinear systems cannot (in general) be computed offline because they depend on the measurements.

5.2 KALMAN FILTER PROPERTIES

In this section, we summarize some of the interesting and important properties of the Kalman filter. Suppose we are given the linear system of Equation (5.17) and we want to find a causal filter that results in a state estimate $\hat{x}_k$. The error between the true state and the estimated state is denoted as $\tilde{x}_k$:

$$\tilde{x}_k = x_k - \hat{x}_k \tag{5.20}$$

Since the state is partly determined by the stochastic process $\{w_k\}$, x_k is a random variable. Since the state estimate is determined by the measurement sequence $\{y_k\}$, which in turn is partly determined by the stochastic process $\{v_k\}$, $\hat{x}_k$ is a random variable. Therefore, $\tilde{x}_k$ is also a random variable.

Suppose we want to find the estimator that minimizes (at each time step) a weighted two-norm of the expected value of the estimation error $\tilde{x}_k$:

$$\min E\left[\tilde{x}_k^T S_k \tilde{x}_k\right] \tag{5.21}$$

where S_k is a positive definite user-defined weighting matrix. If S_k is diagonal with elements $S_k(1), \cdots, S_k(n)$, then the weighted sum is equal to $S_k(1)E[\tilde{x}_k^2(1)] + \cdots + S_k(n)E[\tilde{x}_k^2(n)]$.

- If $\{w_k\}$ and $\{v_k\}$ are Gaussian, zero-mean, uncorrelated, and white, then the Kalman filter is the solution to the above problem.

- If $\{w_k\}$ and $\{v_k\}$ are zero-mean, uncorrelated, and white, then the Kalman filter is the best linear solution to the above problem. That is, the Kalman filter is the best filter that is a linear combination of the measurements. There may be a nonlinear filter that gives a better solution, but the Kalman filter is the best linear filter. It is often asserted in books and papers that the Kalman filter is not optimal unless the noise is Gaussian. However, as our derivation in this chapter has shown, that is simply untrue. Such statements arise from erroneous interpretations of Kalman filter derivations. Even if the noise is not Gaussian, the Kalman filter is still the optimal *linear* filter.

- If $\{w_k\}$ and $\{v_k\}$ are correlated or colored, then the Kalman filter can be modified to solve the above problem. This will be shown in Chapter 7.

- For nonlinear systems, various formulations of nonlinear Kalman filters approximate the solution to the above problem. This will be discussed further in Chapters 13–15.

Recall the measurement update equation from Equation (5.19):

$$\hat{x}_k^+ = \hat{x}_k^- + K_k(y_k - H_k\hat{x}_k^-) \tag{5.22}$$

The quantity $(y_k - H_k\hat{x}_k^-)$ is called the innovations. This is the part of the measurement that contains new information about the state. In Section 10.1, we will prove that the innovations is zero-mean and white with covariance $(H_k P_k^- H_k^T + R_k)$. In fact, the Kalman filter can actually be derived as a filter that whitens the measurement and hence extracts the maximum possible amount of information from the measurement. This was first proposed in [Kai68]. When a Kalman filter is used for state estimation, the innovations can be measured and its mean and covariance can be approximated using statistical methods. If the mean and covariance of the innovations are not as expected, that means something is wrong with the filter. Perhaps the assumed system model is incorrect, or the assumed noise statistics are incorrect. This can be used in real time to verify Kalman filter performance and parameters, and even to adjust Kalman filter parameters in order to improve performance. An application of this idea will be explored in Section 10.2.