- Optimality in the linear Gaussian case and best linear state estimator in the (linear) non-Gaussian case
- Whiteness of the innovations
- The role of the Riccati equation for the state covariance and its relationship with the CRLB
- Stability and steady-state
- Connection with the observability and controllability of the system

Introduce the *likelihood function of a filter*, to be used later in adaptive filtering.

Discuss the issues of

- Consistency of a dynamic estimator and *estimator evaluation*
- Initialization of estimators

and practical ways to make it consistent.

5.2 LINEAR ESTIMATION IN DYNAMIC SYSTEMS — THE KALMAN FILTER

5.2.1 The Dynamic Estimation Problem

Consider a discrete-time linear dynamic system described by a vector difference equation with additive white Gaussian noise that models "unpredictable disturbances." The dynamic (plant) equation is

$$x(k+1) = F(k)x(k) + G(k)u(k) + v(k) \qquad k = 0, 1, \ldots \tag{5.2.1-1}$$

where $x(k)$ is the n_x-dimensional state vector, $u(k)$ is an n_u-dimensional ***known input*** vector (e.g., control or sensor platform motion), and $v(k)$, $k = 0, 1, \ldots$, is the sequence of zero-mean white Gaussian **process noise** (also n_x-vectors) with covariance

$$E[v(k)v(k)'] = Q(k) \tag{5.2.1-2}$$

The measurement equation is

$$z(k) = H(k)x(k) + w(k) \qquad k = 1, \ldots \tag{5.2.1-3}$$

with $w(k)$ the sequence of zero-mean white Gaussian **measurement noise** with covariance

$$E[w(k)w(k)'] = R(k) \tag{5.2.1-4}$$

The matrices F, G, H, Q, and R are assumed *known* and possibly time-varying. In other words, the system can be *time-varying* and the noises *non-stationary*. The initial state $x(0)$, in general unknown, is modeled as a *random variable*, Gaussian distributed with known mean and covariance. The two noise sequences and the initial state are assumed *mutually independent*.

The above constitutes the ***linear Gaussian (LG) assumption.***

In the dynamic equation (5.2.1-1), the process noise term $v(k)$ is sometimes taken as $\Gamma(k)v(k)$ with $v(k)$ an n_v-vector and $\Gamma(k)$ a known $n_x \times n_v$ matrix. Then the covariance matrix of the disturbance in the state equation, which is $Q(k)$ if $v(k)$ enters directly, is to be replaced by

$$E[[\Gamma(k)v(k)][\Gamma(k)v(k)]'] = \Gamma(k)Q(k)\Gamma(k)' \tag{5.2.1-5}$$

The linearity of (5.2.1-1) and (5.2.1-3) leads to the preservation of the Gaussian property of the state and measurements — this is a *Gauss-Markov process.*

The following notation will be used: The conditional mean

$$\boxed{\hat{x}(j|k) \triangleq E[x(j)|Z^k]} \tag{5.2.1-6}$$

where

$$Z^k \triangleq \{z(i), i \leq k\} \tag{5.2.1-7}$$

denotes the sequence of observations available at time k, is the

- ***Estimate of the state*** if $j = k$ (also called filtered value)
- ***Smoothed value of the state*** if $j < k$
- ***Predicted value of the state*** if $j > k$

The ***estimation error*** is defined as

$$\boxed{\tilde{x}(j|k) \triangleq x(j) - \hat{x}(j|k)} \tag{5.2.1-8}$$

The ***conditional covariance matrix*** of $x(j)$ given the data Z^k or the ***covariance associated with the estimate*** (5.2.1-6) is

$$\boxed{P(j|k) \triangleq E[[x(j) - \hat{x}(j|k)][x(j) - \hat{x}(j|k)]'|Z^k] = E[\tilde{x}(j|k)\tilde{x}(j|k)'|Z^k]} \tag{5.2.1-9}$$

Remarks

The smoothed state has also been called recently the ***retrodicted state*** [Drummond93]. The term smoothing is, however, commonly used, even though retrodiction is the correct antonym of prediction. Sometimes the estimated state is called (incorrectly) the smoothed state.

Note that the *covariance of the state* is the same as the *covariance of the estimation error* — this is a consequence of the fact that the estimate is the conditional mean (5.2.1-6). (See also problem 5-3.)

It was shown earlier that the *MMSE criterion* for estimation leads to the *conditional mean* as the *optimal estimate.*

As discussed in Section 3.2, if two vectors are jointly Gaussian, then the probability density of one conditioned on the other is also Gaussian. Thus the conditional mean (5.2.1-6) will be evaluated using this previous result.

The Estimation Algorithm

The estimation algorithm starts with the ***initial estimate*** $\hat{x}(0|0)$ of $x(0)$ and the associated ***initial covariance*** $P(0|0)$, assumed to be available. The second (conditioning) argument 0 stands for Z^0, the ***initial information***. Practical procedures to obtain the initial estimate and initial covariance will be discussed later.

One cycle of the dynamic estimation algorithm — the ***Kalman filter (KF)*** — will thus consist of mapping the estimate

$$\hat{x}(k|k) \triangleq E[x(k)|Z^k] \tag{5.2.1-10}$$

which is the conditional mean of the state at time k (the "current stage") given the observations up to and including time k, and the associated covariance matrix

$$P(k|k) = E[[x(k) - \hat{x}(k|k)][x(k) - \hat{x}(k|k)]'|Z^k] \tag{5.2.1-11}$$

into the corresponding variables at the next stage, namely, $\hat{x}(k+1|k+1)$ and $P(k+1|k+1)$.

This follows from the fact that a Gaussian random variable is *fully characterized* by its first two moments.

The values of past known inputs are subsumed in the conditioning, but (most of the time) will not be shown explicitly.

5.2.2 Dynamic Estimation as a Recursive Static Estimation

The recursion that yields the state estimate at $k+1$ and its covariance can be obtained from the static estimation equations (3.2.1-7) and (3.2.1-8)

$$\hat{x} \triangleq E[x|z] = \bar{x} + P_{xz}P_{zz}^{-1}(z - \bar{z}) \tag{5.2.2-1}$$

$$P_{xx|z} \triangleq E[(x - \hat{x})(x - \hat{x})'|z] = P_{xx} - P_{xz}P_{zz}^{-1}P_{zx} \tag{5.2.2-2}$$

by the following substitutions, indicated below by "$\rightarrow$".

The *prior* (unconditional) expectations from the static case become *prior to the availability of the measurement at time* $k+1$ in the dynamic case, that is, *given the data up to and including* k.

The *posterior* (conditional) expectations become *posterior to obtaining the measurement at time* $k+1$, that is, *given the data up to and including* $k+1$.

The variable to be estimated is the state at $k+1$

$$x \rightarrow x(k+1) \tag{5.2.2-3}$$

Its mean prior to $k+1$ — the ***(one-step) predicted state*** — is

$$\bar{x} \rightarrow \bar{x}(k+1) \triangleq \hat{x}(k+1|k) \triangleq E[x(k+1)|Z^k] \tag{5.2.2-4}$$

Based on the observation (measurement)

$$z \rightarrow z(k+1) \tag{5.2.2-5}$$

with prior mean — the ***predicted measurement***

$$\bar{z} \rightarrow \bar{z}(k+1) \triangleq \hat{z}(k+1|k) \triangleq E[z(k+1)|Z^k] \tag{5.2.2-6}$$

one computes the estimate posterior to $k+1$ — the ***updated state estimate*** (or, just the ***updated state***)

$$\hat{x} \rightarrow \hat{x}(k+1) \triangleq \hat{x}(k+1|k+1) \triangleq E[x(k+1)|Z^{k+1}] \tag{5.2.2-7}$$

Note

Two notations have been used in (5.2.2-4) and (5.2.2-6): with single argument and with two arguments. The single-argument notation — with "overbar" for one-step predicted values and "hat" for updated value — is patterned after the static case. The two-argument notation, which clearly indicates the time index of the conditioning data, is more general and is the preferred one because it applies to smoothing and general prediction as well. Nevertheless, one can find both in the literature.

The Covariances

The prior covariance matrix of the state variable $x(k+1)$ to be estimated — the ***state prediction covariance*** or ***predicted state covariance*** — is

$$P_{xx} \rightarrow \bar{P}(k+1) \triangleq P(k+1|k) \triangleq \text{cov}[x(k+1)|Z^k] = \text{cov}[\tilde{x}(k+1|k)|Z^k] \tag{5.2.2-8}$$

with the last equality following from (5.2.1-9).

The (prior) covariance of the observation $z(k+1)$ — the ***measurement prediction covariance*** — is

$$P_{zz} \rightarrow S(k+1) \triangleq \text{cov}[z(k+1)|Z^k] = \text{cov}[\tilde{z}(k+1|k)|Z^k] \tag{5.2.2-9}$$

The covariance between the variable to be estimated $x(k+1)$ and the observation $z(k+1)$ is

$$P_{xz} \rightarrow \text{cov}[x(k+1), z(k+1)|Z^k] = \text{cov}[\tilde{x}(k+1|k), \tilde{z}(k+1|k)|Z^k] \tag{5.2.2-10}$$

The posterior covariance of the state $x(k+1)$ — the ***updated state covariance*** — is

$$\begin{aligned} P_{xx|z} \rightarrow P(k+1) &\triangleq P(k+1|k+1) = \text{cov}[x(k+1)|Z^{k+1}] \\ &= \text{cov}[\tilde{x}(k+1|k+1)|Z^{k+1}] \end{aligned} \tag{5.2.2-11}$$

Note

Similarly to the single-argument and two-argument notations for the state, $\bar{x}$ and $\hat{x}$, one has the notations $\bar{P}$ and P for the one-step predicted and updated state covariances. As before, the two-argument notations are preferred and will be used in the sequel. However, the flowchart of the estimation algorithm will be given with both notations.

The Filter Gain

The weighting matrix from the estimation ("updating") equation (5.2.2-1) becomes the ***filter gain***

$$P_{xz}P_{zz}^{-1} \rightarrow W(k+1) \triangleq \text{cov}[x(k+1), z(k+1)|Z^k]S(k+1)^{-1} \qquad (5.2.2\text{-}12)$$

Remark

The reasons for the designation of the above as filter gain are as follows:

1. The recursive estimation algorithm is a filter — it reduces the effect of the various noises on the quantity of interest (the state estimate).
2. The quantity (5.2.2-12) multiplies the observation $z(k+1)$, which is the input to the filter; that is, this quantity is a gain.

5.2.3 Derivation of the Dynamic Estimation Algorithm

The ***predicted state*** (5.2.2-4) follows by applying on the state equation (5.2.1-1) the operator of expectation conditioned on Z^k,

$$E[x(k+1)|Z^k] = E[F(k)x(k) + G(k)u(k) + v(k)|Z^k] \qquad (5.2.3\text{-}1)$$

Since the process noise $v(k)$ is *white and zero mean*, this results in

$$\boxed{\hat{x}(k+1|k) = F(k)\hat{x}(k|k) + G(k)u(k)} \qquad (5.2.3\text{-}2)$$

Subtracting the above from (5.2.1-1) yields the ***state prediction error***

$$\tilde{x}(k+1|k) \triangleq x(k+1) - \hat{x}(k+1|k) = F(k)\tilde{x}(k|k) + v(k) \qquad (5.2.3\text{-}3)$$

Note the cancellation of the input $u(k)$ in (5.2.3-3) — it has no effect on the estimation error as long as it is *known*.

The ***state prediction covariance*** (5.2.2-8) is

$$E[\tilde{x}(k+1|k)\tilde{x}(k+1|k)'|Z^k] = F(k)E[\tilde{x}(k|k)\tilde{x}(k|k)'|Z^k]F(k)' + E[v(k)v(k)'] \qquad (5.2.3\text{-}4)$$

which can be rewritten as

$$\boxed{P(k+1|k) = F(k)P(k|k)F(k)' + Q(k)} \tag{5.2.3-5}$$

The cross-terms in (5.2.3-4) are zero due to the fact that $v(k)$ is zero mean and white and, thus, orthogonal to $\tilde{x}(k|k)$.

The ***predicted measurement*** (5.2.2-6) follows similarly by taking the expected value of (5.2.1-3) conditioned on Z^k,

$$E[z(k+1)|Z^k] = E[H(k+1)x(k+1) + w(k+1)|Z^k] \tag{5.2.3-6}$$

Since the measurement noise $w(k+1)$ is zero mean and white, this becomes

$$\boxed{\hat{z}(k+1|k) = H(k+1)\hat{x}(k+1|k)} \tag{5.2.3-7}$$

Subtracting the above from (5.2.1-3) yields the ***measurement prediction error***

$$\tilde{z}(k+1|k) \triangleq z(k+1) - \hat{z}(k+1|k) = H(k+1)\tilde{x}(k+1|k) + w(k+1) \tag{5.2.3-8}$$

The ***measurement prediction covariance*** in (5.2.2-9) follows from (5.2.3-8), in a manner similar to (5.2.3-5), as

$$\boxed{S(k+1) = H(k+1)P(k+1|k)H(k+1)' + R(k+1)} \tag{5.2.3-9}$$

The covariance (5.2.2-10) between the state and measurement is, using (5.2.3-8),

$$\begin{aligned} E[\tilde{x}(k+1|k)\tilde{z}(k+1|k)'|Z^k] \\ &= E\left[\tilde{x}(k+1|k)[H(k+1)\tilde{x}(k+1|k) + w(k+1)]'|Z^k\right] \\ &= P(k+1|k)H(k+1)' \end{aligned} \tag{5.2.3-10}$$

The ***filter gain*** (5.2.2-12) is, using (5.2.3-9) and (5.2.3-10),

$$\boxed{W(k+1) \triangleq P(k+1|k)H(k+1)'S(k+1)^{-1}} \tag{5.2.3-11}$$

Thus the ***updated state estimate*** (5.2.2-7) can be written according to (5.2.2-1) as

$$\boxed{\hat{x}(k+1|k+1) = \hat{x}(k+1|k) + W(k+1)\nu(k+1)} \tag{5.2.3-12}$$

where

$$\boxed{\nu(k+1) \triangleq z(k+1) - \hat{z}(k+1|k) = \tilde{z}(k+1|k)} \tag{5.2.3-13}$$

is called the ***innovation*** or ***measurement residual***. This is the same as (5.2.3-8), but the notation ν will be used in the sequel. Note that in view of this, S is also the ***innovation covariance***.

Finally, the ***updated covariance*** (5.2.2-11) of the state at $k+1$ is, according to (5.2.2-2),

$$
\begin{aligned}
P(k+1|k+1) &= P(k+1|k) - P(k+1|k) \\
&\quad \cdot H(k+1)'S(k+1)^{-1}H(k+1)P(k+1|k) \\
&= [I - W(k+1)H(k+1)]P(k+1|k)
\end{aligned}
\tag{5.2.3-14}
$$

or, in symmetric form,

$$
\boxed{P(k+1|k+1) = P(k+1|k) - W(k+1)S(k+1)W(k+1)'} \tag{5.2.3-15}
$$

Equation (5.2.3-12) is called the ***state update*** equation, since it yields the updated state estimate, and (5.2.3-15) is the ***covariance update*** equation.

Note the similarity between the state update equation (5.2.3-12) and the recursive LS equation (3.4.2-15). The covariance update equation (5.2.3-15) is analogous to (3.4.2-11). The only difference is that in the LS case the prediction to $k+1$ from k is the same as the updated value at k. This follows from the fact that in the LS formulation one deals with a constant parameter while in a dynamic system the state evolves in time.

Alternative Forms for the Covariance Update

Similarly to (3.4.2-6), there is a recursion for the inverse covariance

$$
P(k+1|k+1)^{-1} = P(k+1|k)^{-1} + H(k+1)'R(k+1)^{-1}H(k+1) \tag{5.2.3-16}
$$

Using the matrix inversion lemma, it can be easily shown that (5.2.3-16) is algebraically equivalent to (5.2.3-15). The filter using (5.2.3-16) instead of (5.2.3-15) is known as the ***information matrix filter*** (see Chapter 7).

As in (3.4.2-12), the filter gain (5.2.3-11) has the alternate expression

$$
W(k+1) = P(k+1|k+1)H(k+1)'R(k+1)^{-1} \tag{5.2.3-17}
$$

An alternative form for the covariance update equation (5.2.3-15), which holds for an *arbitrary gain* W (see Problem 5-5), called the ***Joseph form covariance update***, is

$$
\begin{aligned}
P(k+1|k+1) &= [I - W(k+1)H(k+1)]P(k+1|k) \\
&\quad \cdot [I - W(k+1)H(k+1)]' + W(k+1)R(k+1)W(k+1)'
\end{aligned}
\tag{5.2.3-18}
$$

While this is computationally more expensive than (5.2.3-15), it is less sensitive to round-off errors: It will not lead to negative eigenvalues, as (5.2.3-15) is prone to, due to the subtraction present in it. Numerical techniques that reduce the sensitivity to round-off errors are discussed in Chapter 7.

Equation (5.2.3-18) can be also used for evaluation of the ***sensitivity of the filter to an incorrect gain*** (see Section 5.6). This equation can also be used to obtain the optimal gain (see problem 5-9).

Intuitive Interpretation of the Gain

Note from (5.2.3-11) that the *optimal filter gain* is (taking a simplistic "scalar view" of it)

1. "Proportional" to the state prediction variance
2. "Inversely proportional" to the innovation variance

Thus, the gain is

- "Large" if the state prediction is "inaccurate" (has a large variance) and the measurement is "accurate" (has a relatively small variance)
- "Small" if the state prediction is "accurate" (has a small variance) and the measurement is "inaccurate" (has a relatively large variance)

A large gain indicates a "rapid" response to the measurement in updating the state, while a small gain yields a slower response to the measurement. In the frequency domain it can be shown that these properties correspond to a higher/lower ***bandwidth of the filter***.

A filter whose ***optimal gain*** is higher yields less "noise reduction," as one would expect from a filter with a higher bandwidth. This will be seen quantitatively in the next chapter.

Remark

Equations (5.2.3-9) and (5.2.3-15) yield ***filter-calculated covariances***, which are exact if all the modeling assumptions used in the filter derivation hold. In practice this is not always the case and the validity of these filter-calculated estimation accuracies should be tested, as discussed in Section 5.4.

5.2.4 Overview of the Kalman Filter Algorithm

Under the *Gaussian assumption* for the initial state (or initial state error) and all the noises entering into the system, the Kalman filter is the ***optimal MMSE state estimator***. If these random variables are *not Gaussian* and one has only their first two moments, then, in view of the discussion from Section 3.3, the Kalman filter algorithm is the ***best linear state estimator***, that is, the ***LMMSE state estimator***.

The flowchart of one cycle of the Kalman filter is presented in Fig. 5.2.4-1 with the two-argument notations and in Fig. 5.2.4-2 with the one-argument notations. Note that at every stage (cycle) k the entire past is summarized by the *sufficient statistic* $\hat{x}(k|k)$ and the associated covariance $P(k|k)$.

The left-side column represents the true system's evolution from the state at time k to the state at time $k+1$ with the input $u(k)$ and the process noise $v(k)$. The measurement follows from the new state and the noise $w(k+1)$. The known input (e.g., control, platform motion, or sensor pointing) enters (usually) the system with the knowledge of the latest state estimate and is used by the state estimator to obtain the predicted value for the state at the next time.

The state estimation cycle consists of the following:

1. State and measurement prediction (also called ***time update***)
2. State update (also called ***measurement update***)

The state update requires the filter gain, obtained in the course of the covariance calculations. The covariance calculations are *independent* of the state and measurements (and control — assumed to be known) and can, therefore, be performed *offline*.

The Workhorse of Estimation — The Kalman Filter

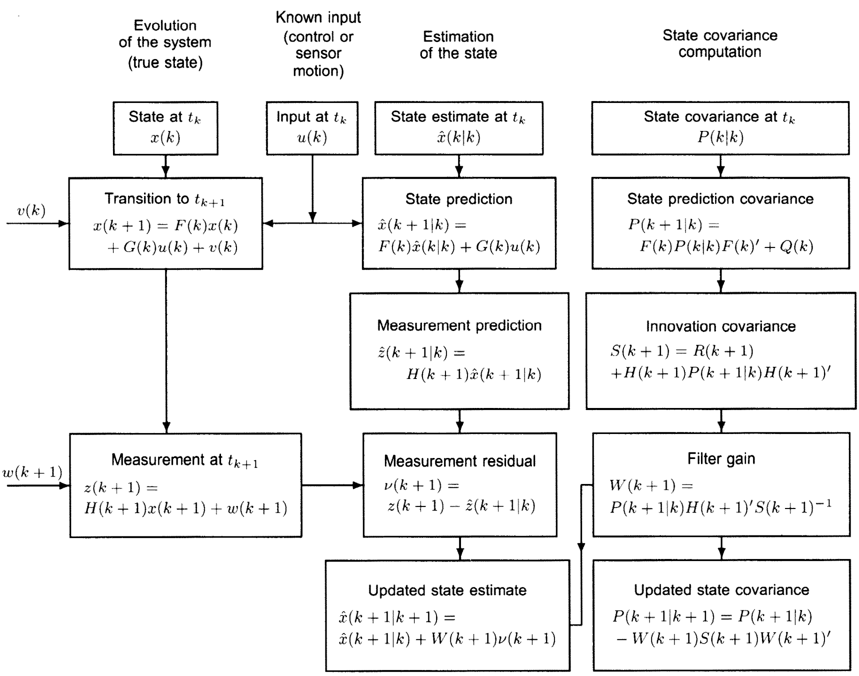

Figure 5.2.4-1: One cycle in the state estimation of a linear system.

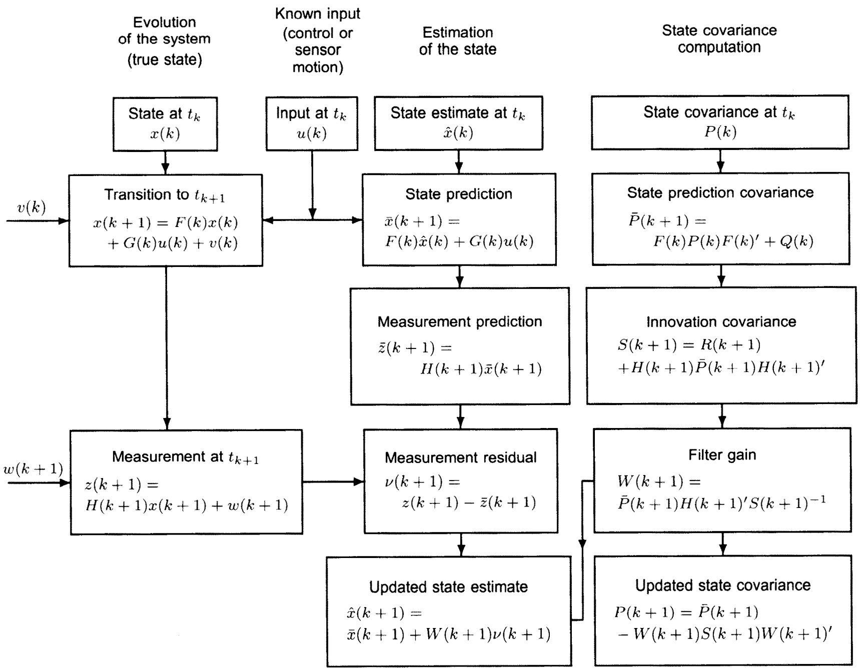

Figure 5.2.4-2: One cycle in the state estimation of a linear system (with single-argument notations).

Summary of the Statistical Assumptions of the Kalman Filter

The initial state has the known mean and covariance

$$E[x(0)|Z^0] = \hat{x}(0|0) \tag{5.2.4-1}$$

$$\text{cov}[x(0)|Z^0] = P(0|0) \tag{5.2.4-2}$$

where Z^0 denotes the initial (prior) information.

The process and measurement noise sequences are *zero mean and white* with *known covariance matrices*

$$E[v(k)] = 0 \tag{5.2.4-3}$$

$$E[v(k)v(j)'] = Q(k)\delta_{kj} \tag{5.2.4-4}$$

$$E[w(k)] = 0 \tag{5.2.4-5}$$

$$E[w(k)w(j)'] = R(k)\delta_{kj} \tag{5.2.4-6}$$

All the above are *mutually uncorrelated*

$$E[x(0)v(k)'] = 0 \qquad \forall k \tag{5.2.4-7}$$

$$E[x(0)w(k)'] = 0 \qquad \forall k \tag{5.2.4-8}$$

$$E[v(k)w(j)'] = 0 \qquad \forall k, j \tag{5.2.4-9}$$

It can be easily shown that under the Gaussian assumption the whiteness and the uncorrelatedness of the noises imply the following:

$$E[v(k)|Z^k] = E[v(k)] = 0 \tag{5.2.4-10}$$

$$E[w(k)|Z^{k-1}] = E[w(k)] = 0 \tag{5.2.4-11}$$

Property (5.2.4-10) was used in (5.2.3-2), while (5.2.4-11) was used in (5.2.3-7).

Remark

The dynamic (plant) equation parameters — the matrices F, G — and the measurement equation parameters — the matrix H — are assumed known.

Computational Requirements

The ***computational requirements*** of the KF are approximately proportional to n^3 where $n = \max(n_x, n_z)$.

Some Extensions

The assumptions of

- White process noise
- White measurement noise
- Uncorrelatedness between the process and the measurement noise sequences

can be relaxed.

An autocorrelated ("colored") noise has to be modeled as the output of a subsystem driven by white noise; that is, it has to be *prewhitened*, as discussed in Chapter 4. For an *autocorrelated process noise*, the state vector has to be augmented to incorporate this subsystem. An example of prewhitening of an autocorrelated process noise is presented in Section 8.2.

The situation where there is *correlation between the two noise sequences* is discussed in Section 8.3. The filter derivation for an *autocorrelated measurement noise*, which can be done without augmenting the state, is presented in Section 8.4.

Discrete-time *smoothing* is presented in Section 8.6.

Continuous-time state estimation as an extension of the discrete-time results is discussed in Chapter 9.

5.2.5 The Matrix Riccati Equation

As pointed out in Section 3.2, the covariance equations in the static MMSE estimation problem are independent of the measurements. Consequently, the covariance equations for the state estimation problem (in a linear dynamic system), derived in Subsection 5.2.3, can be iterated forward offline.

It can be easily shown that the following recursion can be written for the one-step prediction covariance

$$\begin{aligned} P(k+1|k) &= F(k)\{P(k|k-1) - P(k|k-1)H(k)' \\ &\cdot[H(k)P(k|k-1)H(k)' + R(k)]^{-1}H(k)P(k|k-1)\}F(k)' + Q(k) \end{aligned} \tag{5.2.5-1}$$

This is the ***discrete-time (difference) matrix Riccati equation***, or just the ***Riccati equation***. The above follows by substituting (5.2.3-9) and (5.2.3-11) into (5.2.3-15) and substituting the result into (5.2.3-5).

The solution of the above Riccati equation for a time-invariant system converges to a finite ***steady-state covariance*** if

1. The pair $\{F, H\}$ is *completely observable.*

If, in addition,

2. The pair $\{F, C\}$, where $Q \triangleq CC'$ (C is the ***Cholesky factor*** (see Subsection 7.4.2) — a square root of Q), is *completely controllable*,

then the steady-state covariance is a *unique positive definite matrix* — independent of the initial covariance.

The steady-state covariance matrix is the solution of the ***algebraic matrix Riccati equation*** (or just the ***algebraic Riccati equation***)

$$P = F[P - PH'(HPH' + R)^{-1}HP]F' + Q \tag{5.2.5-2}$$

and this yields the ***steady-state gain*** for the Kalman filter.

The interpretation of the above conditions is as follows:

1. The observability condition on the state guarantees a "steady flow" of information about *each* state component — this prevents the uncertainty from becoming unbounded. This condition yields the existence of a (not necessarily unique) steady-state solution for the covariance matrix that is positive definite or positive semidefinite (i.e., with finite positive or nonnegative eigenvalues, respectively).

2. The controllability condition states that the process noise enters into each state component and prevents the covariance of the state from converging to zero. This condition causes the covariance to be positive definite (i.e., all the eigenvalues are positive).

Filter Stability

The convergence of the covariance to a *finite* steady state — that is, *the error becoming a stationary process in the MS sense* — is equivalent to ***filter***

stability in the bounded input bounded output sense.

Stability of the filter does not require the dynamic system to be stable — only the observability condition (1) is required. As indicated above, observability alone does not guarantee uniqueness — the steady-state solution might depend on the initial covariance — but the existence (finiteness) of the solution is the key.

This is particularly important, since the state models used in tracking are *unstable* — they have an integration (from velocity to position). Stability means "bounded input bounded output," and this condition is not satisfied by an integrator — its continuous-time transfer function has a pole at the origin and in discrete time it has a pole at 1.

Remarks

If the state covariance matrix is *positive semidefinite* rather than positive definite, that is, it has some zero eigenvalues that reflect the filter's "belief" that it has "perfectly accurate" estimates of some state components, the gain will be zero for those state components — an *undesirable feature.*

In view of this, in many applications where there is no physical process noise, an ***artificial process noise*** or ***pseudo-noise*** is assumed (i.e., a matrix Q that will lead to condition (2) being satisfied).

The Riccati Equation and the CRLB

The lower bound on the minimum achievable covariance in state estimation is given by the (posterior) CRLB, which was presented for random parameters in (2.7.2-3). Since the state is a random variable, this is applicable to state estimation.

In the linear Gaussian case it can be shown that this is given by the solution of the Riccati equation. In the more general non-Gaussian case, the solution of the Riccati equation is the covariance matrix (actually matrix MSE) associated with the best linear state estimate; however, as discussed in Subsection 3.3.2, a nonlinear estimator can provide better estimates — with covariance smaller than the solution of the Riccati equation.

5.2.6 Properties of the Innovations and the Likelihood Function of the System Model

The Innovations — a Zero-Mean White Sequence

An important property of the ***innovation sequence*** is that it is an *orthogonal sequence*, that is,

$$\boxed{E[\nu(k)\nu(j)'] = S(k)\delta_{kj}} \tag{5.2.6-1}$$

where δ_{kj} is the Kronecker delta function.

This can be seen as follows. Without loss of generality, let $j \leq k - 1$. Use will be made of the smoothing property of the conditional expectations (see Subsection 1.4.12)

$$E[\nu(k)\nu(j)'] = E[E[\nu(k)\nu(j)'|Z^{k-1}]] \tag{5.2.6-2}$$

Note that $\nu(j)$ is a linear combination of the measurements up to j; that is, given Z^{k-1}, it is *not a random variable anymore* and it can thus be taken outside the inner expectation. This yields

$$E[\nu(k)\nu(j)'] = E[E[\nu(k)|Z^{k-1}]\nu(j)'] \tag{5.2.6-3}$$

The inside expectation in (5.2.6-3) is, in view of (5.2.2-6),

$$E[z(k) - \hat{z}(k|k-1)|Z^{k-1}] = 0 \tag{5.2.6-4}$$

and therefore (5.2.6-1) follows for $k \neq j$.

The uncorrelatedness property (5.2.6-1) of the innovations implies that since they are Gaussian, the innovations are independent of each other and thus the innovation sequence is *strictly white*. Without the Gaussian assumption, the innovation sequence is wide sense white.

Thus the innovation sequence is *zero mean and white*.

Remark

Unlike the innovations, the state estimation errors are not white — they are *correlated in time* (see problem 5-11).

The Likelihood Function of the System Model

The joint pdf of the measurements up to k, denoted as

$$Z^k = \{z(j)\}_{j=1}^{k} \tag{5.2.6-5}$$

can be written as

$$p[Z^k] = p[z(k), Z^{k-1}] = p[z(k)|Z^{k-1}]p[Z^{k-1}] = \prod_{i=1}^{k} p[z(i)|Z^{i-1}] \tag{5.2.6-6}$$

where Z^0 is the prior information, shown explicitly only in the expression of (5.2.6-6).

If the above pdfs are Gaussian, then

$$\begin{aligned} p[z(i)|Z^{i-1}] &= \mathcal{N}[z(i); \hat{z}(i|i-1), S(i)] = \mathcal{N}[z(i) - \hat{z}(i|i-1); 0, S(i)] \\ &= \mathcal{N}[\nu(i); 0, S(i)] = p[\nu(i)] \end{aligned} \tag{5.2.6-7}$$

Using (5.2.6-7) in (5.2.6-6) yields

$$\boxed{p[Z^k] = \prod_{i=1}^{k} p[\nu(i)]} \tag{5.2.6-8}$$

that is, the joint pdf of the sequence of measurements Z^k is equal to the product of the marginal pdfs of the corresponding innovations. This shows the *informational equivalence of the measurements and the innovations*.

Since (5.2.6-8) is the joint pdf of Z^k conditioned on the system model (not indicated explicitly), it is the ***likelihood function of the system model.*** This will be used in Chapter 11 to evaluate the "goodness" of models in multiple model adaptive filtering.

5.2.7 The Innovations Representation

The counterpart of the Riccati equation that yields the recursion of the one-step prediction covariance $P(k+1|k)$ is the recursion of the one-step prediction of the state $\hat{x}(k+1|k)$, called the ***innovations representation.***

This is obtained from (5.2.3-2) and (5.2.3-12), without the deterministic input, for simplicity, as

$$\begin{aligned} \hat{x}(k+1|k) &= F(k)\hat{x}(k|k-1) + F(k)W(k)\nu(k) \\ &= F(k)\hat{x}(k|k-1) + W_i(k)[z(k) - H(k)\hat{x}(k|k-1)] \end{aligned} \tag{5.2.7-1}$$

where

$$W_i(k) \triangleq F(k)W(k) \tag{5.2.7-2}$$

is the gain in the innovations representation (sometimes called ambiguously the filter gain).

Equation (5.2.7-1) can also be rewritten as the state equation

$$\hat{x}(k+1|k) = [F(k) - W_i(k)H(k)]\hat{x}(k|k-1) + W_i(k)z(k) \tag{5.2.7-3}$$

with the input being the (nonwhite) sequence $z(k)$ and the output being the innovation

$$\nu(k) = -H(k)\hat{x}(k|k-1) + z(k) \tag{5.2.7-4}$$

This motivates the name innovations representation for the system (5.2.7-3) and (5.2.7-4).