Part 2 Kalman and Bayes Filters

LECTURE NOTES NO. 5

This chapter and the following few chapters will discuss Kalman Filter and its advanced algorithms and theory. The basic formula of Kalman Filter is analogous to Recursive Least Squares; both are types of the Prediction-Error Method where estimates are corrected recursively based on prediction error. While RLS is a deterministic estimation algorithm where statistical properties of variables are neither quantified nor used, Kalman Filter exploits statistical properties in an optimal manner. Furthermore, Kalman Filter incorporates a plant dynamic model in its prediction mechanism and thereby allows us to estimate the state of the plant as a state observer. Finally, Kalman Filter is flexible and expandable in many ways. Unlike the Wiener Filter, an alternative optimal filter applicable to linear systems alone, Kalman Filter, even in its original form, is applicable to linear time-varying systems. Its state-space representation is flexible and powerful enough to extend the algorithm to nonlinear systems and non-Gaussian processes. Its multivariate output formulation is effective to integrate multiple sensor modalities, having diverse sampling rates, noise properties, and physical measures. Its uncertainty quantification formula is effective for evaluating and predicting expected "usefulness" of each sensor modality and thereby allows us to optimize a sensing strategy, which has led to the development of Adaptive Sampling techniques and Simultaneous Localization and Mapping (SLAM). All together Kalman Filter and its advanced algorithms have been making significant contributions to today's control and navigation technologies. We will begin with the basic discrete Kalman Filter and the Kalman-Bucy Filter, followed by two major extensions. One is to nonlinear systems, and the other is to non-Gaussian processes.

5. Discrete Kalman Filter

5.1 State Estimation Using Observers

In discrete-time form a linear time-varying, deterministic, dynamical system is represented by

$$x_{t+1} = A_t x_t + B_t u_t \tag{1}$$

where $x_t \in R^{nx1}$ is a n-dimensional state vector, $u_k \in R^{rx1}$ is an input vector, and A_t, B_t are matrices with proper dimensions. Outputs of the system are functions of the state vector and are represented with a ℓ-dimensional vector $y_t \in R^{\ell x1}$:

$$y_t = H_t x_t \tag{2}$$

where $H_t \in R^{\ell xn}$ is an observation matrix.

Given those parameter matrices (A_t, B_t, H_t) and initial conditions of the state variables, one can simulate the system for predicting states and outputs in response to a time sequence of inputs. See Figure 4-1 below. This simulator may not work well when the model parameters are not exactly known; actual outputs observed in the real system will differ from the predicted values.

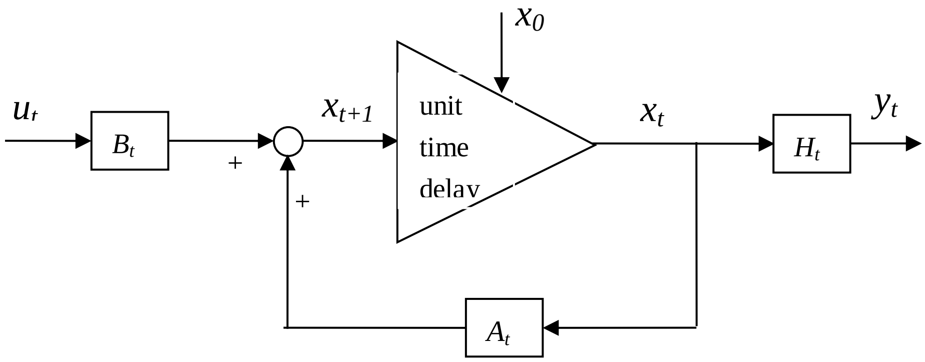


Figure 5-1 Dynamic simulator of deterministic system

A dynamic state observer is a real-time simulator with a feedback mechanism for recursively correcting its estimated state based on the actual outputs measured from the real physical system. See Figure 5-2 below. Note that, unlike a standard feedback control system, the discrepancy between the predicted outputs $\hat{y}_t$ and the actual outputs y_t from the real system are fed back to the ***model*** rather than the real physical system. Using a feedback gain matrix $L_t \in R^{nx\ell}$, the state observer is given by

$$
\begin{aligned}
\hat{x}_{t+1} &= A_t \hat{x}_t + B_t u_t + L_t(y_t - \hat{y}_t) \\
\hat{y}_t &= H_t \hat{x}_t
\end{aligned}
\tag{3}
$$

To differentiate the estimated state from the actual state of the physical system, the estimated state residing in the real-time simulator is denoted $\hat{x}_t$. With this feedback the state of the simulator will follow the actual state of the real system, and thereby estimate the state accurately. If the system is ***observable***, convergence of the estimated state to the actual state can be guaranteed with a proper feedback gain. In other words, a stable observer can forget its initial conditions; regardless of an initial estimated state $\hat{x}_0$, the observer can produce the correct state as it converges. This is Luenberger's State Observer for deterministic systems.

A special case of the above state observer is estimation of constant parameters $\hat{\theta}$. See equation (17) in Chapter 2. Replacing the state transition matrix A_t by the nxn identity matrix and setting inputs to zero leads to a recursive parameter estimation formula in (2-17):

$$
\hat{\theta}(t) = \hat{\theta}(t-1) + \mathrm{K}_t\left(y(t) - \hat{y}(t)\right) \tag{2-17}
$$

The difference from the previous parameter estimation problem is that in state estimation the state makes "state transition" as designated by the state transition matrix A_t and the input matrix B_t driven by an input time sequence. Both recursive parameter estimation and state estimation, however, are analogous; both based on the Prediction-Error-Correction formula.

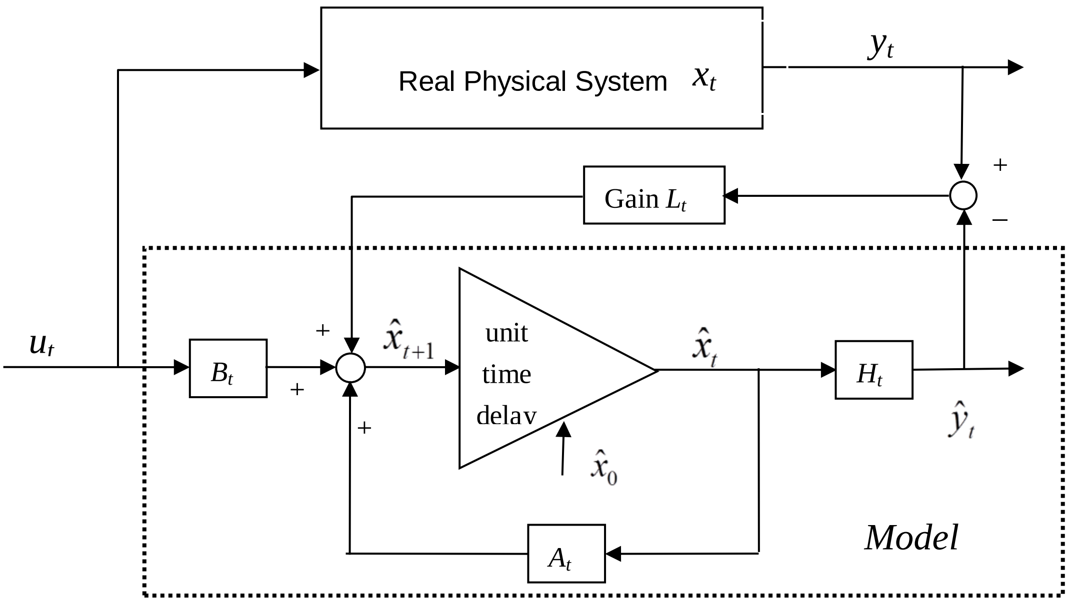


Figure 5-2 Luenberger's state observer for deterministic linear system

Luenberger's state observer is strictly for deterministic systems. In actual systems, sensor signals are to some extent corrupted with noise, and the state transition of the actual process is to some extent disturbed by noise. If stochastic properties of these noise sources are available, state estimation may be performed more effectively than simply using sensor signals as noise-free signals and estimating the sate based on noise-free state transition model. Rudolph Kalman investigated this problem and developed the celebrated Kalman Filter. Surprisingly enough, Kalman did it 10 years before Luenberger published his state observer paper.

To formulate this stochastic state estimation problem we need to use properties of multivariable random processes, which will be summarized in the following section.

5.2 Incorporating Multivariate Random Processes into State Equations

We extend the state equation given by (1) in the previous section to the one as a multivariable random process. Namely, the state $x_t \in R^{nx1}$ is driven not only by the input $u_t \in R^{rx1}$ but also by noise, which is a random process. Let $w_t \in R^{nx1}$ be a multivariable random process, called "Process Noise", driving the state through another matrix $G_t \in R^{nxn}$. The state equation is then given by

$$x_{t+1} = A_t x_t + B_t u_t + G_t w_t \qquad (4)$$

See Figure 5-3. Since the process noise is a random process, the state x_t driven by w_t is a random process. The second term on the right hand side, $B_t u_t$, is a deterministic term. In the following stochastic state estimation, this deterministic part of inputs is not important, since its influence upon the state x_t is completely predictable and hence it can be eliminated without loss of generality. Therefore we often use the following state equation:

$$x_{t+1} = A_t x_t + G_t w_t \qquad (5)$$

The outputs of the system are noisy, as long as they are measured with physical sensors. Let $v_t \in R^{\ell x1}$ be another multivariable random process, called "Measurement Noise". We treat the observed output as the superimposition of measurement noise v_t and the term completely determined by the state variable x_t.

$$y_t = H_t x_t + v_t \tag{6}$$

See Figure 4-3. Since measurement noise v_t is a multivariable random process, the outputs measured with sensor, too, are a multivariable random process.

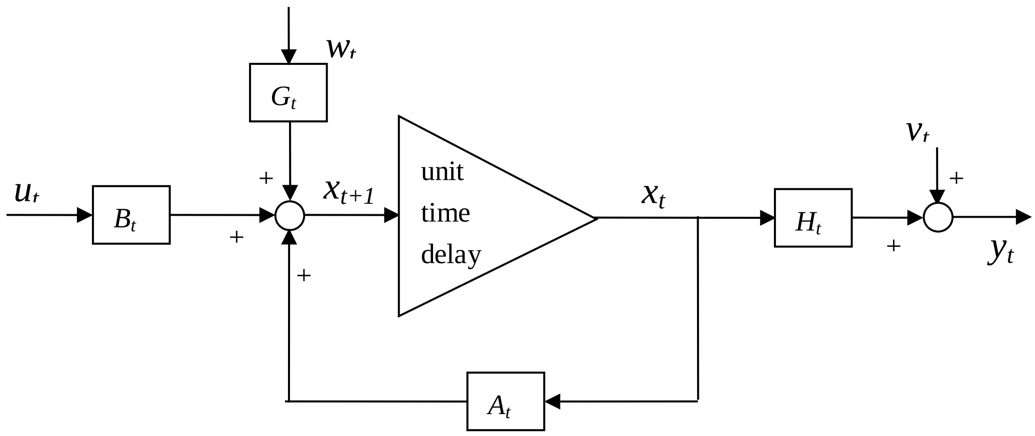


Figure 5-3 State space representation of linear time varying system with process noise and measurement noise

The stochastic properties of the process noise and measurement noise described above are now characterized as multivariable random processes. It is a common practice that the mean of noise is set to zero since, if the means are non-zero, the origins of the state variables and the outputs can be shifted so that the mean of the noise is zero.

$$E[v_t] = 0, \quad E[w_t] = 0 \tag{7}$$

From equation (3-25), the covariance of measurement noise $v_t \in R^{\ell x1}$ is given by

$$C_V(t,s) = E[v_t \cdot v_s^T] \in R^{\ell x \ell} \tag{8}$$

If the noise signals at any two time slices are uncorrelated,

$$C_V(t,s) = E[v_t \cdot v_s^T] = 0, \quad \forall t \neq s \tag{9}$$

the noise is called "White". (We will discuss why this is called *white* later in the power spectrum chapter.) Note that, if $t = s$, the above covariance is that of the first order density.

$$C_V(t) = E[v_t \cdot v_t^T] \tag{10}$$

The diagonal elements of this matrix are variances of the individual output signals. Multiple sensor signals may be correlated. For example, a 2D vision system produces x and y coordinate signals, which may be correlated. Then, it is likely that the off-diagonal elements of the covariance matrix C_v are non-zero.

The process noise can be characterized in the same way. The covariance matrix is then given by:

$$C_W(t,s) = E[w_t \cdot w_s^{\ T}] \in R^{nxn} \tag{11}$$

Furthermore, the correlation between the process noise and the measurement noise may exist, if both are generated in part by the same disturbance source. This can be represented with the covariance matrix given by:

$$C_{WV}(t,s) = E[w_t \cdot v_s^{\ T}] \in R^{nx\ell} \tag{12}$$

Usually the covariance between the process and measurement noises is zero.

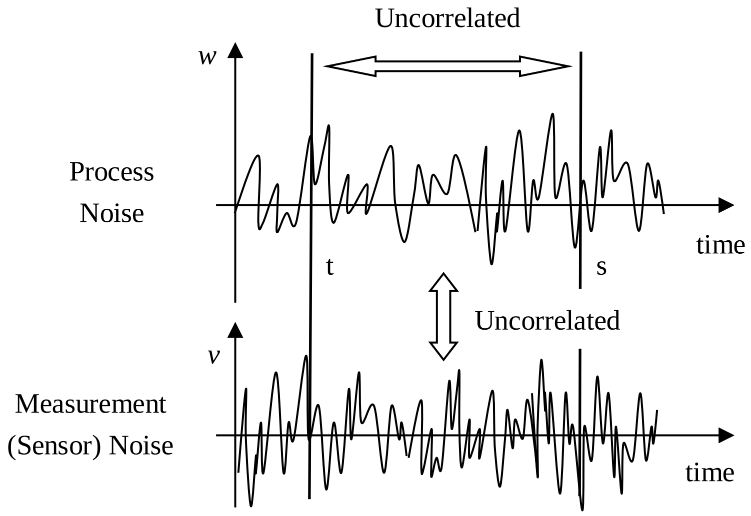


Figure 5-4 Noise characteristics

5.3 Framework of the Discrete-Time Kalman Filter

Consider a dynamical system given by equations (13) and (14),

$$x_{t+1} = A_t x_t + G_t w_t \tag{13}$$

$$y_t = H_t x_t + v_t \tag{14}$$

where $x_t \in R^{nx1}$, $y_t \in R^{\ell x1}$, $w_t \in R^{nx1}$, $v_t \in R^{\ell x1}$, $A_t, G_t \in R^{n\times n}$, and $H_t \in R^{\ell xn}$. Assume that the process noise w_t and the measurement noise v_t have zero mean values,

$E[w_t]=0,$ (15)
$E[v_t]=0.$ (16)

and that they have the following covariance matrices:

$$C_V(t,s) = E[v_t \cdot v_s{}^T] = \begin{cases} 0 & \forall t \neq s \\ R_t & \forall t = s \end{cases} \tag{17}$$

$$C_W(t,s) = E[w_t \cdot w_s{}^T] = \begin{cases} 0 & \forall t \neq s \\ Q_t & \forall t = s \end{cases} \tag{18}$$

$$C_{WV}(t,s) = E[w_t \cdot v_s{}^T] = 0 \qquad \forall t, \forall s \tag{19}$$

where matrix R_t is of $\ell \mathrm{x} \ell$, and is positive definite, and matrix $Q_t \in R^{nxn}$ is positive semi-definite.

Optimal State Estimation Problem

Obtain an optimal estimate of state vector x_t based on measurements y_i, $i = 1,2,\ldots t$, that minimizes the mean squared error:

$$\bar{J}_t = E\left[(\hat{x}_t - x_t)^T (\hat{x}_t - x_t)\right] \tag{20}$$

subject to the state equation (13) and the output equation (14) with white, uncorrelated process and measurement noises of zero mean and the covariant matrices given by equations (15) - (19). (Necessary initial conditions are assumed.)

Rudolf E. Kalman solved this problem around 1960[1].

Kalman Filter: two major points of his seminal work in 1960.

I) If we assume that the optimal filter is linear, then the Kalman filter is the state estimator having the smallest unconditioned error covariance among all linear filters.

II) If we assume that the noise is Gaussian, then the Kalman filter is the optimal minimum variance estimator among all linear *and* non-linear filters.

5.4 The Discrete Kalman Filter as a Linear Optimal Filter

The Discrete Kalman Filter provides a recursive solution to the minimum mean-square error estimation problem described previously. Figure 5-5 depicts the outline of the discrete

[1] It has been debatable to determine who invented so called Kalman Filter. A radar theoretician, Peter Swerling, developed a similar algorithm at Rand Corporation. His seminal paper in 1958 anticipated Kalman Filter.

Kalman filter.

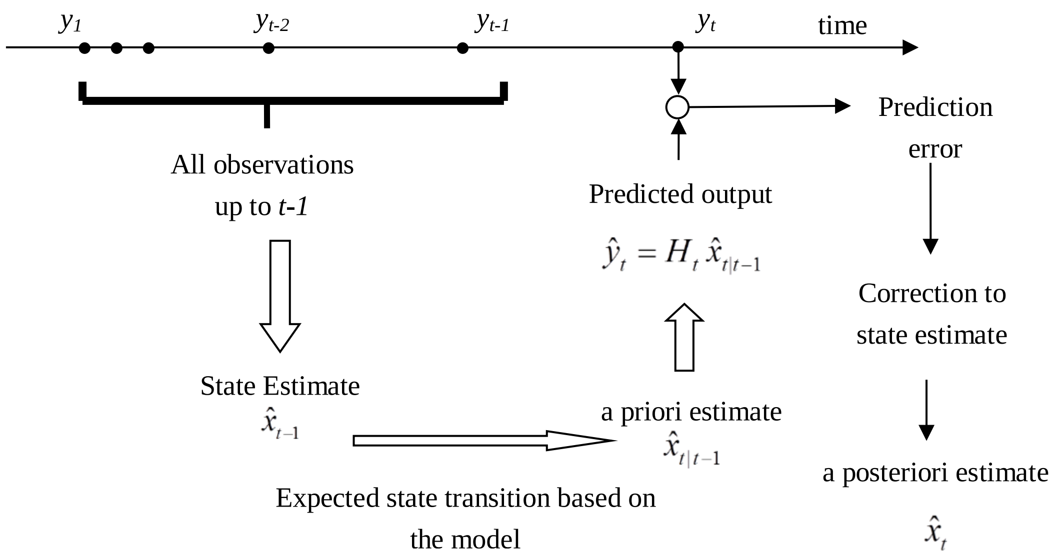


Figure 5-5 Outline of the Kalman filter algorithm

This algorithm consists of three major components:

Expected state transition

From (8), we know how the previous estimate $\hat{x}_{t-1}$ will make a transition

$x_t = A_{t-1}x_{t-1} + G_{t-1}w_{t-1}$; Let's write this as $\hat{x}_{t|t-1}$

Transition from estimated state at time *t-1*, $\hat{x}_{t-1}$

$$\begin{aligned} \hat{x}_{t|t-1} &= E[A_{t-1}\hat{x}_{t-1} + G_{t-1}w_{t-1}] \\ &= A_{t-1}\hat{x}_{t-1} + G_{t-1}E[\cancel{w_{t-1}}] \end{aligned} \tag{21}$$

This estimate $\hat{x}_{t|t-1}$, termed a priori state estimate, provides the expected state based on $\hat{x}_{t-1}$.

This is called "a priori", since it is an estimate before assimilating a new output y_t.

Predicted output

Form (9) and (10)

$$\hat{y}_t = H_t\hat{x}_{t|t-1} \quad \text{Note } E[v_t]=0 \tag{22}$$

Correction of the state estimate

Assimilating a new measurement y_t, we can update the state estimate in proportion to the output prediction error.

$$\hat{x}_t = \hat{x}_{t|t-1} + K_t(y_t - H_t\hat{x}_{t|t-1}) \tag{23}$$

Equation (23) provides a structure of linear filter in recursive form. $K_t \in R^{n\times l}$ is a gain matrix to be optimized so that the mean squared error (expected value of error) of state estimation may be minimized.

A more general form of linear filter is

$$\hat{x}_t = K_{t1}\hat{x}_{t|t-1} + K_{t2}y_t \tag{24}$$

Both (23) and this form provide the same result.

5.5 The Kalman Gain

Consider the error of a posteriori estimate $\hat{x}_t$

$$\begin{aligned} e_t \equiv \hat{x}_t - x_t &= \hat{x}_{t|t-1} + K_t(y_t - H_t\hat{x}_{t|t-1}) - x_t \\ &= \hat{x}_{t|t-1} + K_t(H_t x_t + v_t - H_t\hat{x}_{t|t-1}) - x_t \\ &= (I - K_t H_t)\varepsilon_t + K_t v_t \end{aligned} \tag{25}$$

where ε_t is a priori estimation error, i.e. before assimilating the new measurement y_t.

$$\varepsilon_t \equiv \hat{x}_{t|t-1} - x_t \tag{26}$$

For the following calculation, let us omit the subscript *t* for brevity,

$$\begin{aligned} e_t^T e_t &= [\varepsilon_t - K_t H_t \varepsilon_t + K_t v_t]^T [\varepsilon_t - K_t H_t \varepsilon_t + K_t v_t] \\ &= \varepsilon^T\varepsilon + \varepsilon^T H^T K^T K H\varepsilon - 2\varepsilon^T K H \varepsilon + 2\varepsilon^T K v - 2v^T K^T K H \varepsilon + v^T K^T K v \end{aligned} \tag{27}$$

Let us differentiate the scalar function $e_t^T e_t$ with respect to matrix K by using the following matrix differentiation rules.

i) $$f \equiv \begin{pmatrix} a_1 & a_2 & \dots & a_n \end{pmatrix} \begin{bmatrix} K_{11} & \cdots & K_{1\ell} \\ \vdots & \ddots & \vdots \\ K_{n1} & \cdots & K_{n\ell} \end{bmatrix} \begin{pmatrix} b_1 \\ \vdots \\ b_\ell \end{pmatrix} = \bar{a}^T K \bar{b} \rightarrow \frac{df}{dK} = \left\{ \frac{\partial f}{\partial K_{ij}} \right\} = \{a_i b_j\} = \bar{a}\bar{b}^T \tag{28}$$

….. *Rule 1*

ii) $g = \bar{c}^T K^T K \bar{b}, \quad \bar{b} \in R^{\ell \times 1}, \quad \bar{c} \in R^{\ell \times 1}, \quad K \in R^{n \times \ell}$

$$\frac{dg}{dK} = \left\{ \frac{\partial}{\partial K_{pq}} \sum_{i=1}^{n} \sum_{j=1}^{\ell} \sum_{k=1}^{\ell} K_{ik} c_k K_{ij} b_j \right\} = \left\{ \sum_{j=1}^{\ell} c_q K_{pj} b_j + \sum_{k=1}^{\ell} K_{pk} c_k b_q \right\} = K\bar{b}\bar{c}^T + K\bar{c}\bar{b}^T \tag{29}$$

……*Rule 2*

Using these rules,

$$\begin{aligned}\frac{d}{dK} e_t^T e_t &= \frac{d}{dK} [\underbrace{\varepsilon^T H^T}_{\bar{c}^T} K^T K \underbrace{H\varepsilon}_{\bar{b}} - 2v^T K^T KH\varepsilon + v^T K^T Kv] \leftarrow rule\, 2 \\ &\quad + 2\frac{d}{dK}[\varepsilon^T Kv - \varepsilon^T KH\varepsilon] \leftarrow rule\, 1 \\ &= KH\,\varepsilon\varepsilon^T H^T + KH\varepsilon\varepsilon^T H^T - 2[KH\varepsilon v^T + Kv\varepsilon^T H^T] + 2Kvv^T + 2[\varepsilon v^T - \varepsilon\varepsilon^T H^T]\end{aligned} \tag{30}$$

The necessary condition for the mean squared error of state estimate with respect to the gain matrix K is:

$$\frac{d\bar{J}_t}{dK} = 0 \tag{31}$$

Taking expectation of $e_t^T e_t$, differentiating it w.r.t. K and setting it to zero yield:

$$E\left[KH\varepsilon\varepsilon^T H^T - KH\varepsilon v^T - Kv\varepsilon^T H^T + Kvv^T + \varepsilon v^T - \varepsilon\varepsilon^T H^T\right] = 0 \tag{32}$$

KH can be factored out,

$$KHE[\varepsilon\varepsilon^T]H^T - KHE[\varepsilon v^T] - KE[v\varepsilon^T]H^T + KE[vv^T] + E[\varepsilon v^T] - E[\varepsilon\varepsilon^T]H^T = 0 \tag{33}$$

Examine the term $E[\varepsilon v^T]$ using (26) and (21),

$$\begin{aligned}E[\varepsilon_t v_t^T] &= E[(\hat{x}_{t|t-1} - x_t)v_t^T] \\ &= E[\hat{x}_{t|t-1} v_t^T] - E[x_t v_t^T]\end{aligned}$$

For the second term

$x_t = A \cdot x_{t-1} + w_{t-1}$ $\longleftrightarrow$ Uncorrelated with v_t

$A \cdot x_{t-2} + w_{t-2}$ $\longleftrightarrow$ Uncorrelated with v_t

$$\therefore E[x_t v_t^T] = AE[x_{t-1} v_t^T] + E[w_{t-1} v_t^T] = 0$$

Now note that the state x_t has been driven by the process noise $w_{t-1}, w_{t-2}, \cdots$, which are uncorrelated with the measurement noise v_t. Therefore, the second term vanishes: $E[x_t v_t^T] = 0$.

For the first term $\hat{x}_{t|t-1} = A_{t-1}\hat{x}_{t-1}$

$$\hat{x}_{t-1} = \hat{x}_{t-1|t-2} + K_{t-1}(y_{t-1} - H\hat{x}_{t-1|t-2})$$

$$\to H \cdot x_{t-1} + v_{t-1} \longleftrightarrow \text{Uncorrelated with } v_t$$

$$\to A \cdot x_{t-2} + w_{t-2} \longleftrightarrow \text{Uncorrelated with } v_t$$

In the above expression the state estimate $\hat{x}_{t-1}$ is dependent upon the previous process noise $w_{t-2}, w_{t-3}, \cdots$ as well as on the previous measurement noise $v_{t-1}, v_{t-2}, \cdots$, both of which are uncorrelated with the current measurement noise v_t. Therefore, the first term, too, vanishes.

$$\therefore E[\hat{x}_{t|t-1} v_t^T] = 0$$

Therefore

$$E[\varepsilon_t v_t^T] = 0 \tag{34}$$

Similarly,

$$E[\varepsilon v^T] = E[v\varepsilon^T] = 0 \tag{35}$$

Let us define the error covariance of a priori state estimation

$$P_{t|t-1} \equiv E[\varepsilon_t \varepsilon_t^T] = E[(\hat{x}_{t|t-1} - x_t)(\hat{x}_{t|t-1} - x_t)^T] \tag{36}$$

Substituting (35) and (36) into (33), we can conclude that the optimal gain must satisfy

$$K_t H_t P_{t|t-1} H_t^T + K_t R_t - P_{t|t-1} H_t^T = 0 \tag{37}$$

$$\therefore \quad K_t = P_{t|t-1} H_t^T [H_t P_{t|t-1} H_t^T + R_t]^{-1} \tag{38}$$

This is called the **Kalman Gain**.

5.6 Updating the Error Covariance

The above Kalman gain contains the a priori error covariance $P_{t|t-1}$. This can be updated recursively based on the state transition model.

Define the a posteriori state estimation error covariance

$$P_t = E[(\hat{x}_t - x_t)(\hat{x}_t - x_t)^T] = E[e_t e_t^T] \tag{39}$$

This covariance P_t can be computed in the same way as in the previous section. From (25),

$$P_t = E[((I-KH)\varepsilon + Kv)((I-KH)\varepsilon + Kv)^T]$$
$$= E[(I-KH)\varepsilon\varepsilon^T(I-KH)^T] + E[(I-KH)\varepsilon v^T K^T] + E[Kv\varepsilon^T(I-KH)^T] + E[Kvv^T K^T]$$
$$= (I-KH)E[\varepsilon_t\varepsilon_t^T](I-KH)^T + KE[vv^T]K^T$$
$$\therefore \quad P_t = (I-KH)P_{t|t-1}(I-KH)^T + KR_tK^T \qquad (40)$$

Substituting the Kalman gain (38) into (40) yields

$$P_t = (I - K_tH_t)P_{t|t-1} \qquad (41)$$

Exercise. Derive (41)

Furthermore, based on P_t we can compute $P_{t+1|t}$ by using the state transition equation (8).

Consider

$$\varepsilon_{t+1} = \hat{x}_{t+1|t} - x_{t+1}$$
$$= A_t\hat{x}_t - (A_tx_t + G_tw_t) \qquad (42)$$
$$= A_te_t - G_tw_t$$

From (36)

$$P_{t+1|t} = E[\varepsilon_{t+1}\varepsilon_{t+1}^T]$$
$$= E[(A_te_t - G_tw_t)(A_te_t - G_tw_t)^T] \qquad (43)$$
$$= A_tE[e_te_t^T]A_t^T - G_tE[w_te_t^T]A_t^T - A_tE[e_tw_t^T]G_t^T + G_tE[w_tw_t^T]G_t^T$$

Evaluating $E[w_te_t^T]\, and \quad E[e_tw_t^T]$

$$E[e_tw_t^T] = E[(\hat{x}_t - x_t)w_t^T] = E[\{\hat{x}_{t|t-1} + K_t(y_t - \hat{y}_t)\}w_t^T] - E[x_tw_t^T]$$
$$= E[A_{t-1}\hat{x}_{t-1}w_t^T] + E[K_t(H_tx_t + v_t)w_t^T] - E[K_tH_t\hat{x}_{t|t-1}w_t^T] - E[x_tw_t^T] \qquad (44)$$
$$= A_{t-1}E[\hat{x}_{t-1}w_t^T] + (K_tH_t - I)E[x_tw_t^T] + K_tE[v_tw_t^T] - K_tH_tE[\hat{x}_{t|t-1}w_t^T]$$

The first term: $\hat{x}_{t-1}$ does not depend on w_t, hence vanishes. For the second term, using (8), we can write $E[x_tw_t^T] = E[(A_{t-1}x_{t-1} + G_{t-1}w_{t-1})w_t^T] = 0$ since $E[w_{t-1}w_t^T] = 0$. The third term vanishes since the process noise and measurement noise are not correlated. The last term, too, vanishes, since $\hat{x}_{t|t-1}$ does not include w_t. Therefore, $E[e_tw_t^T] = E[w_te_t^T] = 0$.

$$\therefore \quad P_{t+1|t} = A_tP_tA_t^T + G_tQ_tG_t^T \qquad (45)$$

5.7 The Recursive Calculation Procedure for the Discrete Kalman Filter

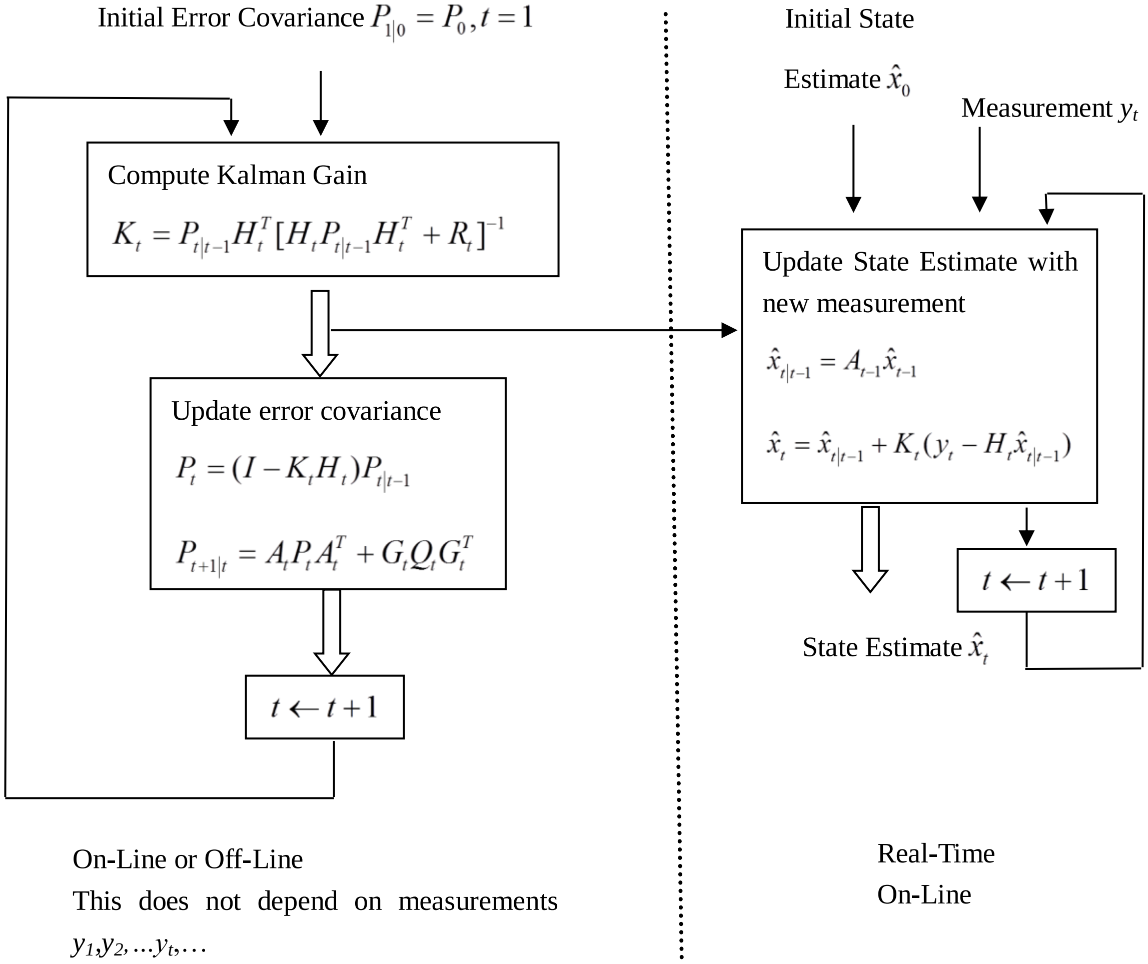


Figure 5-6 Flow chart of computation of Discrete Kalman Filter

5.8 Anatomy of the Discrete Kalman Filter

To better understand the Discrete Kalman Filter, let us consider the following questions:

Question 1: The measurement noise covariance R_t represents the reliability of sensors. How is this property of sensors used in the Kalman gain for correcting (updating) the state estimate?

Question 2: How is the state estimate error covariance P_t used for updating the state estimate?

Question 3: How do the gain K_t and error co-variances P_t and $P_{t|t-1}$ evolve with time? How are they affected by R_t and Q_t?

The Discrete Kalman Filter

Measurement:

$$y_t = H_t x_t + v_t \tag{9}$$

Minimizing the <u>mean</u> squared error

$$\bar{J}_t = E[(\hat{x}_t - x_t)^T (\hat{x}_t - x_t)] \tag{20}$$

Uncorrelated measurement noise

$$E[v_t] = 0, \quad E[v_t v_s^T] = \begin{cases} 0 & t \neq s \\ R_t & t = s \end{cases} \qquad \text{Noise Covariance}$$

Optimal Estimate

$$\hat{x}_t = \hat{x}_{t|t-1} + \underbrace{K_t}_{\text{The Kalman Gain}} \underbrace{\left(y_t - \overset{H_t \hat{x}_{t|t-1}}{\hat{y}_t}\right)}_{\text{Estimation output error}} \tag{23}$$

$$K_t = P_{t|t-1} H_t^T \left(H_t P_{t|t-1} H_t^T + R_t \right)^{-1} \tag{38}$$

Error Covariance update (a priori to a posteriori):

$$P_t = (I - K_t H_t) P_{t|t-1} \tag{41}$$

$$P_t \overset{\Delta}{=} E\left[(\hat{x}_t - x_t)(\hat{x}_t - x_t)^T \right] \quad \text{: a posteriori state estimation error covariance}$$

$$P_{t|t-1} \overset{\Delta}{=} E\left[(\hat{x}_{t|t-1} - x_t)(\hat{x}_{t|t-1} - x_t)^T \right] \quad \text{: a priori state estimation error covariance}$$

Post multiplying $H_t P_{t/t-1} H_t^T + R_t$ to (38),

$$K_t \left(H_t P_{t|t-1} H_t^T + R_t \right) = P_{t|t-1} H_t^T$$

From (41)

$$K_t H_t P_{t|t-1} = P_{t|t-1} - P_t$$

$$\cancel{P_{t|t-1} H_t^T} - P_t H_t^T + K_t R_t = \cancel{P_{t|t-1} H_t^T}$$

$$\therefore K_t R_t = P_t H_t^T \tag{46}$$

The measurement noise covariance R_t is assumed to be non-singular,

$$K_t = P_t H_t^T R_t^{-1} \tag{47}$$

Therefore

$$\hat{x}_t = \hat{x}_{t|t-1} + P_t H_t^T R_t^{-1} \Delta y_t \tag{48}$$

Q1. Without loss of generality, we can write

$$R_t = \begin{bmatrix} \sigma_1^2 & & & 0 \\ & \sigma_2^2 & & \\ & & \ddots & \\ 0 & & & \sigma_l^2 \end{bmatrix} \qquad \Delta y_t = \begin{bmatrix} \Delta y_{t1} \\ \vdots \\ \Delta y_{tl} \end{bmatrix}$$

since if not diagonal we can change the coordinates.

$$\hat{x}_t = \hat{x}_{t|t-1} + P_t H_t^T \begin{bmatrix} \Delta y_{t1}/\sigma_1^2 \\ \vdots \\ \Delta y_{tl}/\sigma_l^2 \end{bmatrix} \tag{49}$$

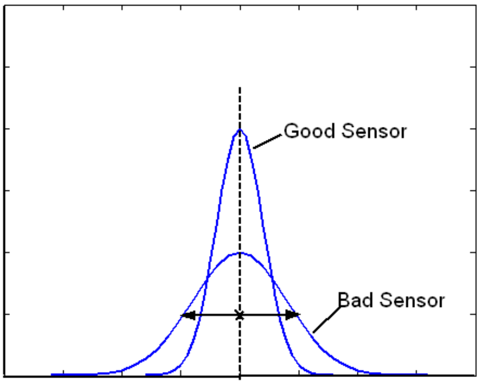


Figure 5-7 Sensor noise variances

Depending on the measurement noise variance, σ_i^2, the error correction term is attenuated; $\Delta y_{ti}/\sigma_i^2$. If the i-th sensor noise is large, i.e. large σ_i^2, the error correction based on that sensor is reduced.

A more general case is depicted in Figure 5-8.

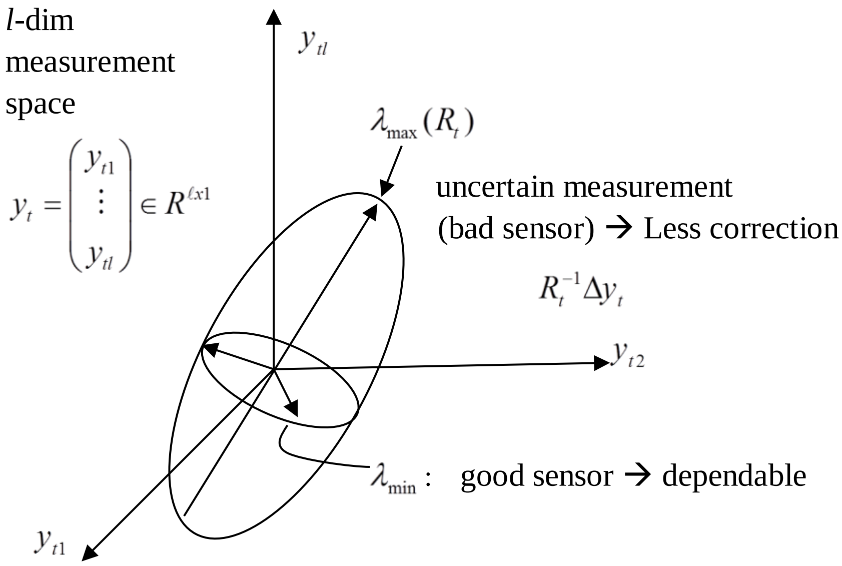
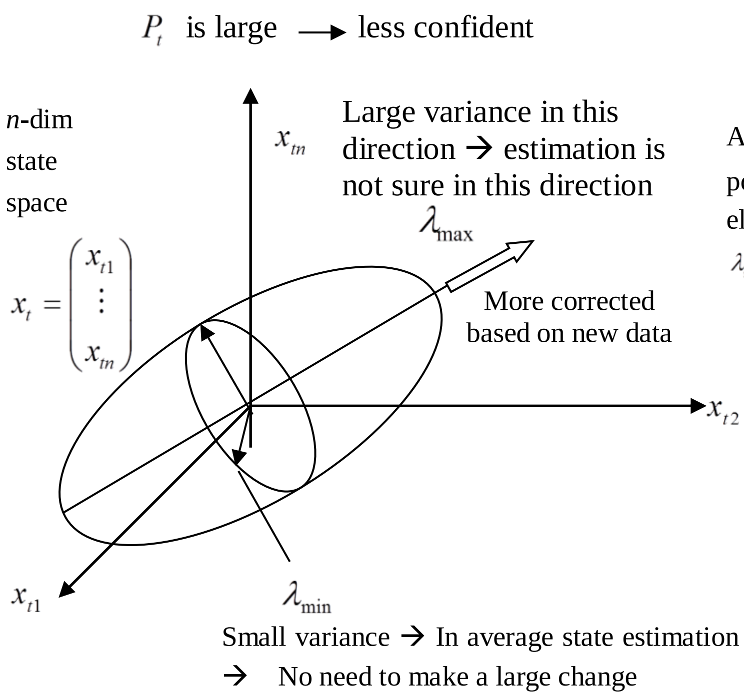


Figure 5-8 Sensor noise covariance

Q2. By definition

$$P_t = E[e_t e_t^T]; \;\; e_t = \hat{x}_t - x_t$$

P_t is the error covariance of a posteriori state estimation. P_t is interpreted as a metric indicating the level of "expected confidence" in state estimation at the t-th stage of correction.



Associated with matrix P_t, which is positive-definite, we can consider an ellipsoid with eigenvalues $\lambda_{\min}(P_t)$, $\lambda_{\max}(P_t)$

$$\hat{x}_t = \hat{x}_{t|t-1} + P_t \Delta q \quad (50)$$

where $\Delta q = H_t^T R_t^{-1} \Delta y_t$

Figure 5-9 Estimation error covariance P_t

The Kalman filter makes a clever trade-off between the intensity of sensor noise and the confidence level of the state estimation that has been made up to the present time; $P_t = E\left[e_t e_t^T\right]$.

Q3: How does the state estimation error covariance change over time? This is a complex problem, and we will discuss dynamic properties and convergence conditions more thoroughly later in the following chapter. In this section we consider only a simple scalar case given by

$$x_{t+1} = x_t + w_t, \quad E\left[w_t w_s\right] = \begin{cases} 0, & \forall t \neq s \\ Q, & \forall t = s \end{cases}$$

$$y_t = x_t + v_t, \quad E\left[v_t v_s\right] = \begin{cases} 0, & \forall t \neq s \\ R, & \forall t = s \end{cases}$$

where $x_t \in \Re^1$, and all the variables are scalar: $R,\ Q,\ P_t, P_{t|t-1} \in \Re^1$.
If we assume convergence of the error co-variances and the Kalman gain, $P_\infty = \lim_{t\to\infty} P_t$, $c = \lim_{t\to\infty} P_{t|t-1}$, and $K_\infty = \lim_{t\to\infty} K_t$, then we can write

$$K_\infty = \frac{c}{c+R}, \quad P_\infty = (1-K_\infty)c, \text{ and } c = P_\infty + Q \quad (51)$$

from (38), (41), and (45). Eliminating K_∞ and P_∞ from the above equations yields

$$c^2 - Qc - QR = 0 \quad (52)$$

which can be solved for c:

$$c = \frac{Q}{2} + \sqrt{\left(\frac{Q}{2}\right)^2 + QR}$$

$$\therefore P_\infty = \sqrt{\left(\frac{Q}{2}\right)^2 + QR} - \frac{Q}{2} \qquad (53)$$

From this result, we can gain some insights how the process noise and the measurement noise affect the error co-variances and the Kalman gain.

- In case the process noise variance is very small: $Q \ll 1$, i.e. a very predictable process, the a priori state estimation error covariance and the Kalman gain converge to zero, $c = \lim_{t\to\infty} P_{t|t-1} \cong 0$, $K_\infty = 0$ along with the a posteriori error covariance, $P_\infty \cong 0$.
- In case the sensor noise variance is very small: $R \ll 1$, i.e. very accurate sensors, the a priori state estimation error covariance converges to the process noise variance: $c = \lim_{t\to\infty} P_{t|t-1} \cong Q$, and the Kalman gain converges to $K_\infty = 1$ with approximately zero a posteriori error covariance $P_\infty \cong 0$.

5.9 Implementation Issues

Kalman filters have been applied to diverse applications since early 60's. Various implementation techniques have also been developed. This section briefly describes a few implementation issues. There are three known failure scenarios in which Kalman filters do not work well:

a) Unobservable or nearly unobservable processes
b) Numerical instability
c) Blind spot

a) Poor observability

This issue can be checked with the well-known observability condition. A more practical and informative method is to examine the error covariance matrix P_t. If the process is poorly observable, the variance associated with some unobservable state variables tends to blow up.

If the process is poorly observable, one should change the sensors, or a new sensor must be added.

b) Numerical instability

Asymmetric covariance: By definition, covariance matrices are symmetry, but numerically they may become asymmetric, leading to divergence in recursive computation. Such an asymmetric covariance often comes from the computation of:

$$P_t = (I - K_t H_t) P_{t|t-1} \qquad (41)$$

where $K_t \in R^{n\times\ell}$ and $H_t \in R^{\ell\times n}$ are not square matrices. Some round off errors yield an asymmetric posteriori covariance P_t, although the a priori covariance $P_{t|t-1}$ is symmetry. To resolve this problem, it is efficient to use Joseph's form (40):

$$P_t = (I - KH)P_{t|t-1}(I - KH)^T + KR_t K^T \tag{54}$$

which is equivalent to (41), as discussed previously. Note that both terms on the right hand side are symmetric matrices.

U-D Factorization: Since the covariance matrix is a real, symmetric, positive-definite matrix, it can be decomposed to the following *U-D* Factorization form:

$$P = UDU^T$$
$$D = \begin{pmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & & \vdots \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & \cdots & d_n \end{pmatrix}, \quad U = \begin{pmatrix} 1 & * & \cdots & * \\ 0 & 1 & * & * \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & 0 & 1 \end{pmatrix} \tag{55}$$

where matrix D is a diagonal matrix while matrix U is an upper triangular matrix. This particular form assures the positive definiteness of the covariance matrix, and implicitly preserves the symmetry of P. Furthermore, if the covariance update formula of Kalman filter is converted to the one in the diagonalized space using the upper triangular matrix U, the dynamic range of computation reduces to 50 % of the original formulation. See more details in Section 9.5 in Brown and Hwang's textbook.

c) Blind spot

Consider another failure scenario. When covariance matrices of both process noise and measurement noise are deemed to be very small, the state estimation error-covariance reduces quickly. This is clear[2] from the covariance propagation and update formulae for discrete Kalman filter:

$$P_{t+1|t} = A_t P_t A_t^T + G_t Q_t G_t^T \text{, and } P_t = (I - K_t H_t)P_{t|t-1}.$$

This implies that the Kalman gain diminishes quickly. Once the Kalman gain diminishes, the subsequent observations are ignored. In other words, the Kalman filter is decoupled from the sensors and the real process. This blind spot problem is often triggered by numerical round off error in computing covariance matrices as well.

[2] This is even clearer in the continuous Kalman Filter to be discussed in the following section. In lieu of the recursive covariance propagation and update formulae, we have the following Riccati differential equation (62):

$$\dot{P} = FP + PF^T - PH^T R^{-1} HP + GQG^T$$

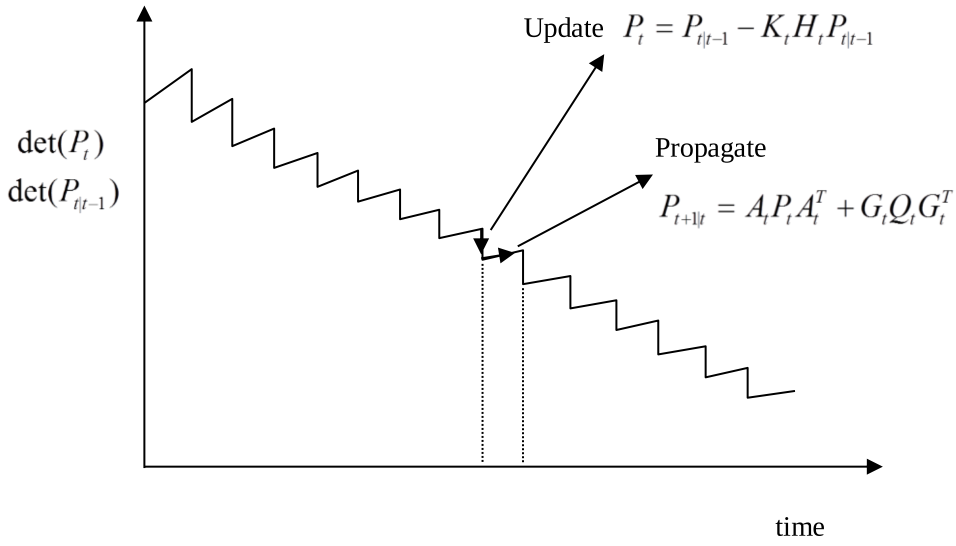


Figure 5-10 Transition of error covariance