2.160 System Identification, Estimation, and Learning

Lecture Notes No. 6

February 24, 2006

4.5.1 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,

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

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

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

….. Rule 1

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

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

…...Rule 2

Using these rules,

$$\frac{d}{dK}e_t^T e_t = \frac{d}{dK}[\underbrace{\varepsilon^T H^T}_{c^T} K^T K \underbrace{H\varepsilon}_{b} - 2v^T K^T KH\varepsilon + v^T K^T Kv] \leftarrow rule\,2$$

$$+2\frac{d}{dK}[\varepsilon^T Kv - \varepsilon^T KH\varepsilon] \leftarrow rule\,1$$

$$= KH\,\varepsilon\varepsilon^T H^T + KH\varepsilon\varepsilon^T H^T - 2[KH\varepsilon v^T + Kv\varepsilon^T H^T] + 2Kvv^T + 2[\varepsilon v^T - \varepsilon\varepsilon^T H^T] \tag{30}$$

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),

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

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

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

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

y_{t-1} → $H \cdot x_{t-1} + v_{t-1}$ ⟷ Uncorrelated with v_t

x_{t-1} → $A \cdot x_{t-2} + w_{t-2}$ ⟷ Uncorrelated with v_t

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

For the second term

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

x_{t-1} → $A \cdot x_{t-2} + w_{t-2}$ ⟷ Uncorrelated with v_t

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

Therefore

$$E[\varepsilon_t v_t^T] = 0 \qquad (34)$$

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: $E[x_t v_t^T] = 0$. In the first term, the previous 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. This leads to

$$E[\varepsilon v^T] = E[v \varepsilon^T] = 0 \qquad (35)$$

Let us define the error covariance of a priori state estimation

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

.

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

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

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

This is called the **Kalman Gain**.

4.5.2 Updating the Error Covariance

The above Kalman gain contains the a priori error covariance $P_{t|t-1}$. This must be updated recursively based on each new measurement and the state transition model. Define the a posteriori state estimation error covariance

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

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

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

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

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

Exercise. Derive (41)

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

Consider

$$
\begin{aligned}
\varepsilon_{t+1} &= \hat{x}_{t+1|t} - x_{t+1} \\
&= A_t\hat{x}_t - (A_t x_t + G_t w_t) \qquad (42) \\
&= A_t e_t - G_t w_t
\end{aligned}
$$

From (36)

$$
\begin{aligned}
P_{t+1|t} &= E[\varepsilon_{t+1}\varepsilon_{t+1}^T] \\
&= E[(A_t e_t - G_t w_t)(A_t e_t - G_t w_t)^T] \qquad (43) \\
&= A_t E[e_t e_t^T]A_t^T - G_t E[w_t e_t^T]A_t^T - A_t E[e_t w_t^T]G_t^T + G_t E[w_t w_t^T]G_t^T
\end{aligned}
$$

Evaluating $E[w_t e_t^T]\, and \quad E[e_t w_t^T]$

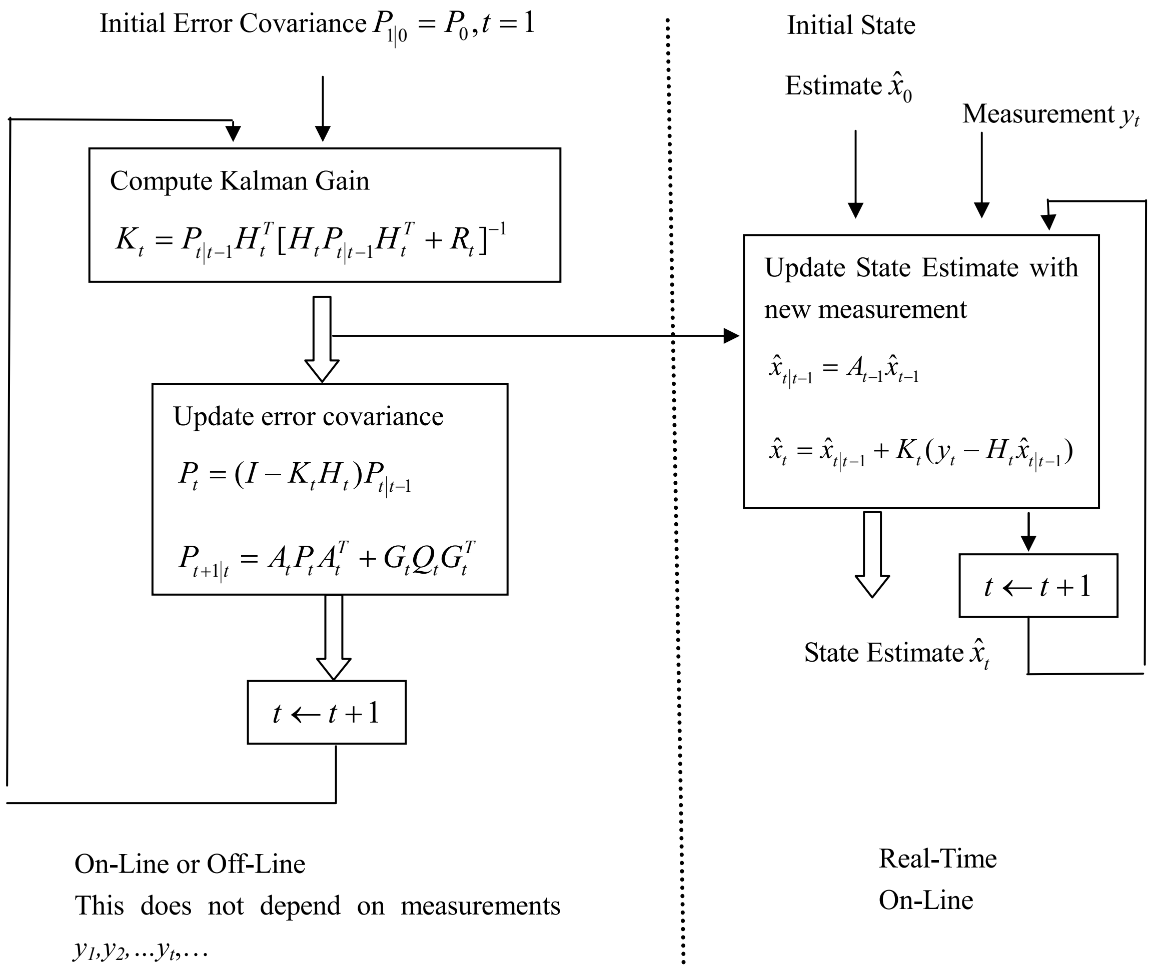
$$
\begin{aligned}
E[e_t w_t^T] &= E[(\hat{x}_t - x_t)w_t^T] = E[\{\hat{x}_{t|t-1} + K_t(y_t - \hat{y}_t)\}w_t^T] - E[x_t w_t^T] \\
&= E[A_{t-1}\hat{x}_{t-1}w_t^T] + E[K_t(H_t x_t + v_t)w_t^T] - E[K_t H_t \hat{x}_{t|t-1}w_t^T] - E[x_t w_t^T] \qquad (44) \\
&= A_{t-1}E[\hat{x}_{t-1}w_t^T] + (K_t H_t - I)E[x_t w_t^T] + K_t E[v_t w_t^T] - K_t H_t E[\hat{x}_{t|t-1}w_t^T]
\end{aligned}
$$

The first term: $\hat{x}_{t-1}$ does not depend on w_t, hence vanishes. For the second term, using (8), we can write $E[x_t w_t^T] = E[(A_{t-1}x_{t-1} + G_{t-1}w_{t-1})w_t^T] = 0$ 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, $E[e_t w_t^T] = E[w_t e_t^T] = 0$.

$$\therefore \quad P_{t+1|t} = A_t P_t A_t^T + G_t Q_t G_t^T \tag{45}$$

4.5.3 The Recursive Calculation Procedure for the Discrete Kalman Filter



4.6 Anatomy of the Discrete Kalman Filter

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

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

Optimal Estimate

$H_t \hat{x}_{t|t-1}$

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

Estimation output error

The Kalman Gain

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

Error Covariance update (a priori to a posteriori):

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

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

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

Questions

Q1: How is the measurement noise covariance R_t used in the Kalman filter for correcting (updating) the state estimate?

R_t … sensor quality

Q2: How is the state estimate error covariance P_t used for updating the state estimate?

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,

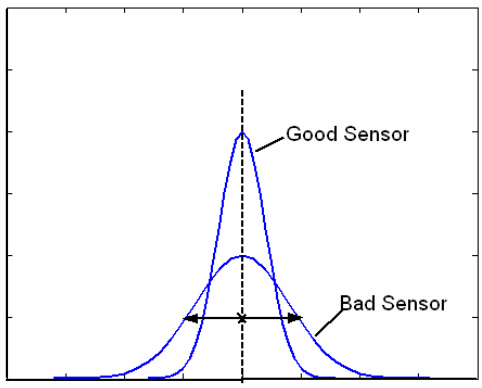
$$K_t = P_t H_t^T R_t^{-1} \tag{47}$$

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.

$$\hat{x}_t = \hat{x}_{t|t-1} + P_t H_t^T \begin{bmatrix} \Delta y_{t1} / \sigma_1^2 \\ \vdots \\ \Delta y_{tl} / \sigma_l^2 \end{bmatrix} \tag{28}$$

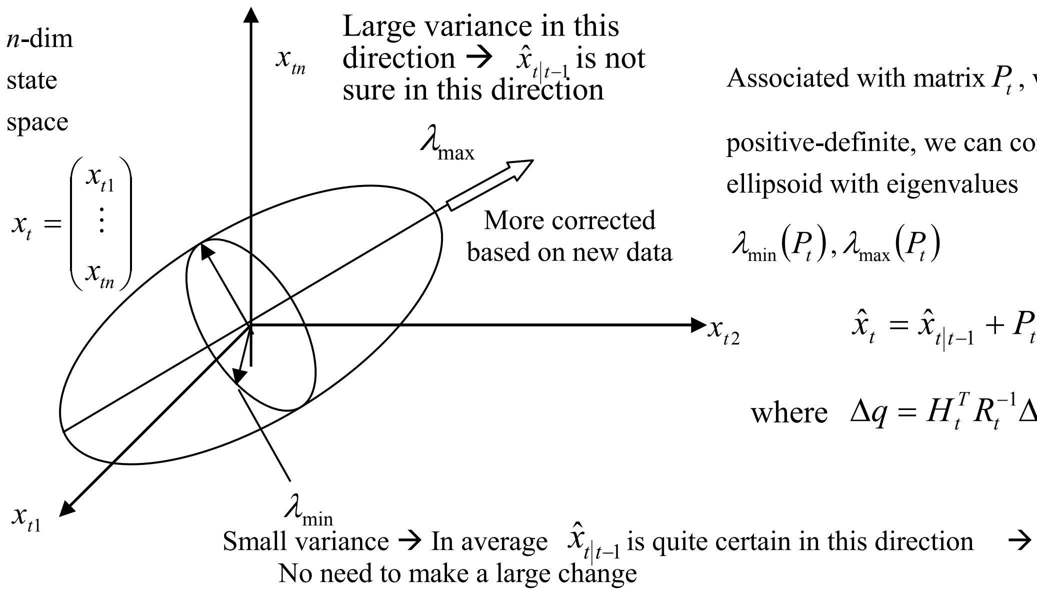
Depending on the measurement noise variance, σ_i^2, the error correction term is attenuated; $\Delta y_{ti} / \sigma_i^2$. If the i-th sensor noise is large, i.e. large σ_i^2, the error correction based on that sensor is reduced.

Q2. By definition

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

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.

P_t is large $\longrightarrow$ less confident

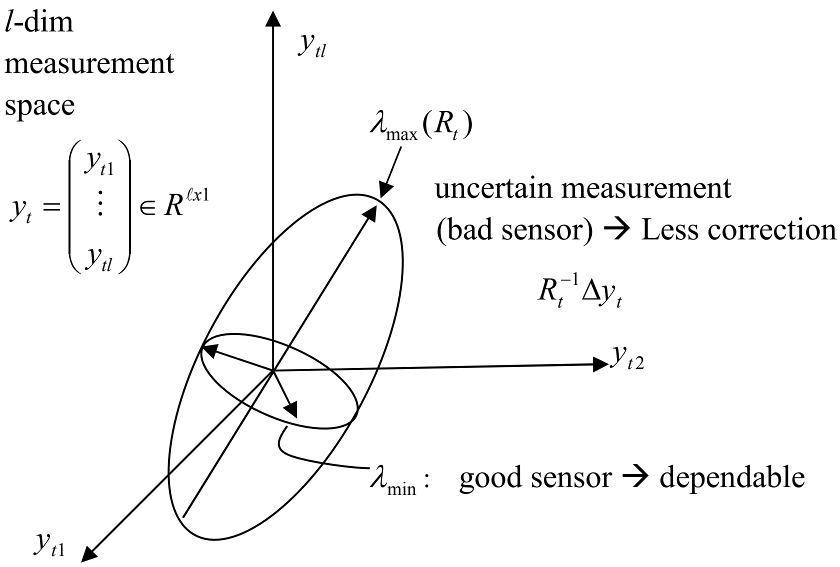


Associated with matrix P_t, which is positive-definite, we can consider an ellipsoid with eigenvalues

$\lambda_{min}(P_t), \lambda_{max}(P_t)$

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

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



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; $P_t = E[e_t e_t^T]$.

How does the state estimation error covariance change over time?