# Lecture Notes No. 5

**February 22, 2006**

## 4. Kalman Filtering

## 4.1 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 $\ell$-dimensional vector $y_t \in R^{\ell x1}$:

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

where $H_t \in R^{\ell x n}$ 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.
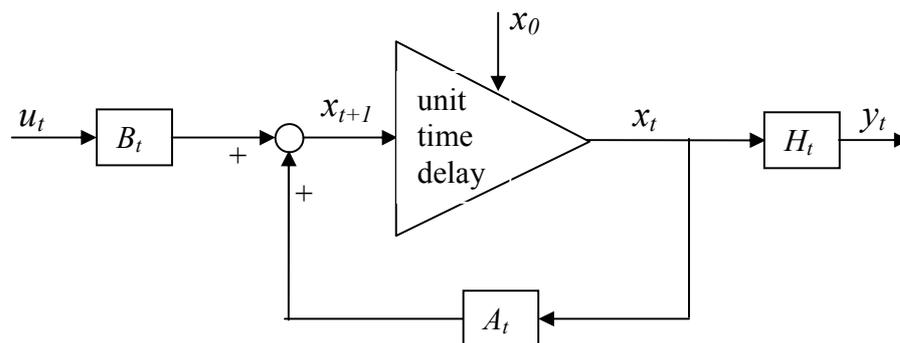


Figure 4-1 Dynamic simulator of
deterministic system

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 4-2 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

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

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.
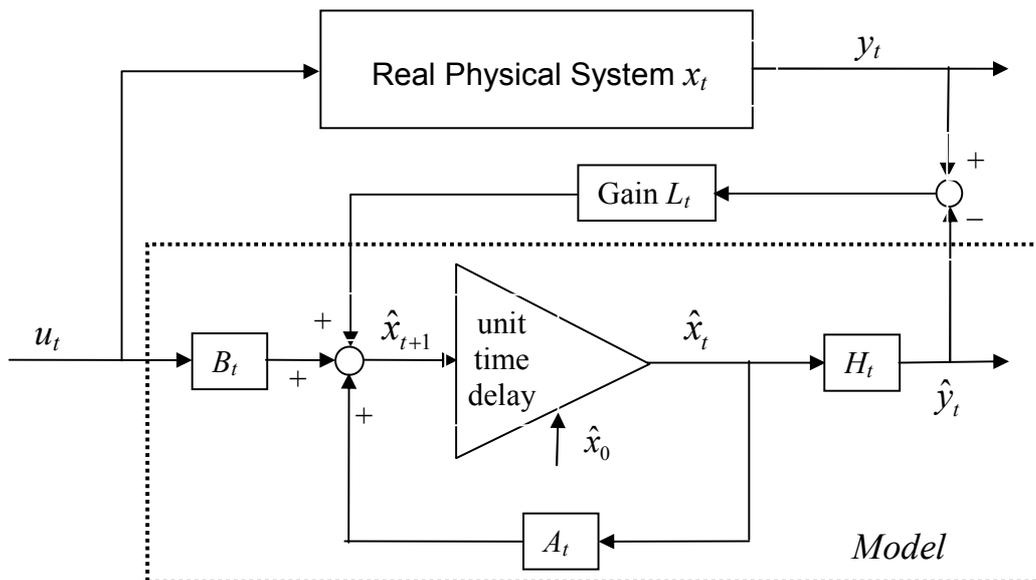


Figure 4-2 Luenberger's state observer for
deterministic linear system

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):

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

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.

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.

## 4.2 Multivariate Random Processes

Let us revisit Section 3.2 and consider the ensemble of time profiles again. Each of these waveforms is a realization of random process X(t). (Such a random waveform can be seen in your oscilloscope, when you increase the gain of a ground signal. The ensemble of waveforms may be considered as a large collection of the same oscilloscope. If you have hundreds of the same type of oscilloscopes, you have a collection of diverse ground noise waveforms. They come from the same ground signal, but the waveforms are all different.) Now that stochastic properties of a single random process have been described with the first and second order densities and auto-correlation functions, let us consider multiple random processes, say X(t), Y(t), and Z(t), and characterize their properties. (In the oscilloscope analogy, you can think about two channels of signals, say channel 1 and channel 2, shown on a single oscilloscope display. If there are a large number of the same type of oscilloscopes displaying both channels, an ensemble of the two random processes can be generated.)

A Multivariable random process :  A vectorial random proces
State variables = $n$-dimential vector
Observable outputs with multiple sensors

First-order density   $f_{XYZ}(x_t, y_t, z_t)$

Covariance: Ensemble mean

$$C_{XYZ}(t) = E\left[\begin{pmatrix} X(t) - m_x(t) \\ Y(t) - m_Y(t) \\ Z(t) - m_Z(t) \end{pmatrix}\begin{pmatrix} X(t) - m_x(t) & Y(t) - m_Y(t) & Z(t) - m_Z(t) \end{pmatrix}\right] \quad (4)$$

If  $m_x = m_y = m_z = 0$

$$C_{XYZ}(t) = \begin{bmatrix} E[X^2(t)] & E[X(t)Y(t)] & E[X(t)Z(t)] \\ E[X(t)Y(t)] & E[Y^2(t)] & E[Y(t)Z(t)] \\ E[X(t)Z(t)] & E[Y(t)Z(t)] & E[Z^2(t)] \end{bmatrix} \qquad (5)$$
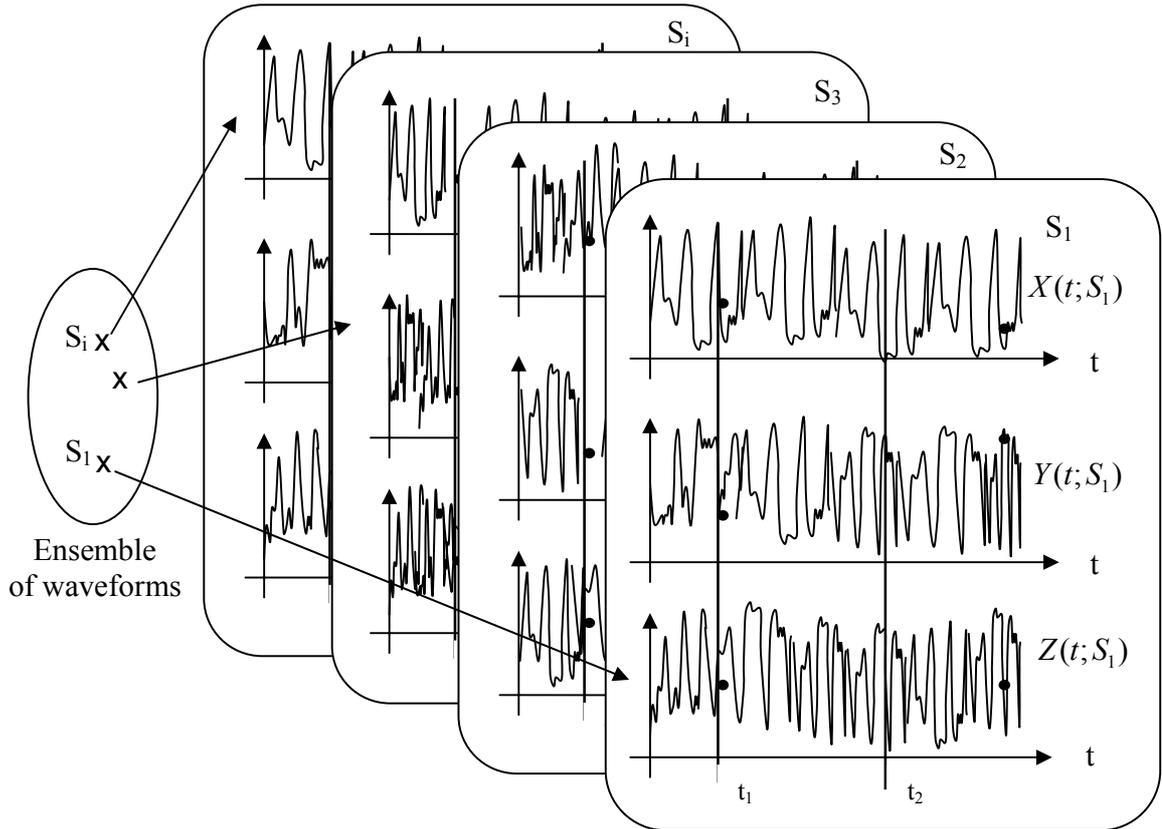


Figure 4-3 Multivariable random processes

Second-order density:

　　　Taking two time slices, $t_1$ and $t_2$, as shown in the above figure, the joint probability density is given by: $f_{X_1 Y_1 Z_1 X_2 Y_2 Z_2}(x_{t_1}, y_{t_1}, z_{t_1}, x_{t_2}, y_{t_2}, z_{t_2})$

　　　If $m_x = m_y = m_z = 0$, the covariance is given by

$$C_{XYZ}(t_1, t_2) = \begin{bmatrix} E[X(t_1)X(t_2)] & E[X(t_1)Y(t_2)] & E[X(t_1)Z(t_2)] \\ E[Y(t_1)X(t_2)] & E[Y(t_1)Y(t_2)] & E[Y(t_1)Z(t_2)] \\ E[Z(t_1)X(t_2)] & E[Z(t_1)Y(t_2)] & E[Z(t_1)Z(t_2)] \end{bmatrix} \qquad (6)$$

Note that the first order covariance in equation (5) can be viewed as a special case of the second order covariance in equation (6); $t_1 = t_2 = t$.

## 4.3 State-Space Modeling of Random Processes

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

$$x_{t+1} = A_t x_t + B_t u_t + G_t w_t \qquad (7)$$

See Figure 4-4. 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:

$$x_{t+1} = A_t x_t + G_t w_t \qquad (8)$$

The outputs of the system are noisy, as long as they are measured with sensors. Let $v_t \in R^{\ell x1}$ be another multivariable random process, called "Measurement Noise", super-imposed on a deterministic part of the output, the part completely determined by the sate variable $x_t$.

$$y_t = H_t x_t + v_t \qquad (9)$$

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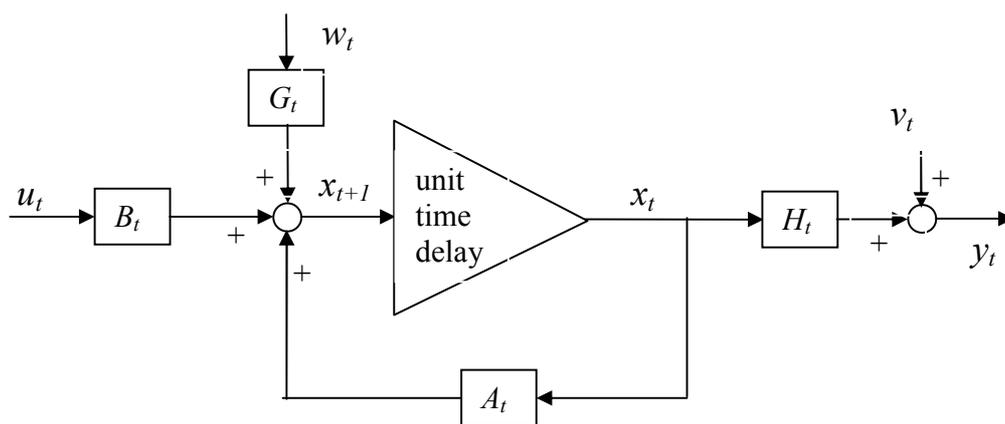


Figure 4-4  State space representation of linear time varying system with process noise and measurement noise

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.

$$E[v_t] = 0$$
$$E[w_t] = 0$$

(10)

From equation (6), the covariance of measurement noise $v_t \in R^{\ell x1}$ is given by

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

(11)

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

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

(12)

the noise is called "White". (We will discuss why this is called white later in the following chapter.) Note that, if $t=s$, the above covariance is that of the first order density, i.e. an auto-covariance.

$$C_V(t) = E[v_t \cdot v_t^T]$$

(13)

The diagonal elements of this matrix are variances of the individual output signals. If those outputs are coupled through the state variables and the measurement matrix $H_t$, (see equation (9)), 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:

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

(14)

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:

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

(15)

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


## 4.4 Framework of the Kalman Filter

Consider a dynamical system given by equations (8) and (9),

$$x_{t+1} = A_t x_t + G_t w_t \tag{8}$$

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

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,

$$E[w_t] = 0, \ E[v_t] = 0. \tag{16}$$

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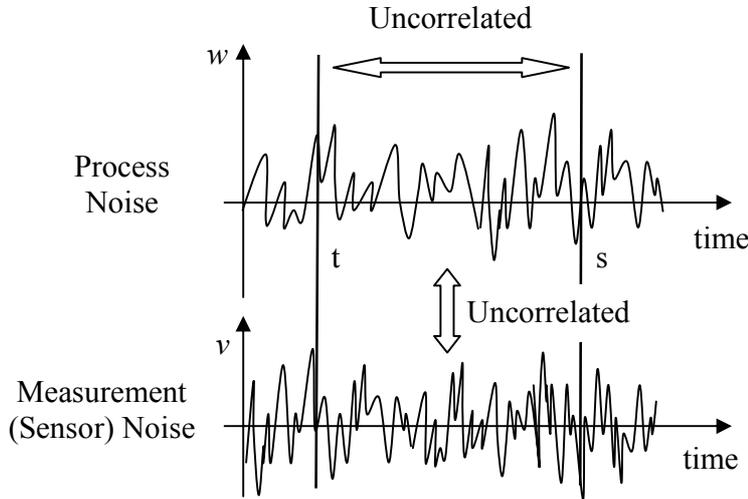


Figure 4-5 Noise characteristics

**Optimal State Estimation Problem**

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

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

7

subject to the state equation (8) and the output equation (9) with white, uncorrelated process and measurement noises of zero mean and the covariant matrices given by equations (16) - (18). (Necessary initial conditions are assumed.)

Rudolf E. Kalman solved this problem around 1960.

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.

## 4.5 The Discrete Kalman Filter as a Linear Optimal Filter

Figure 4-6 depicts the outline of the discrete Kalman filter,
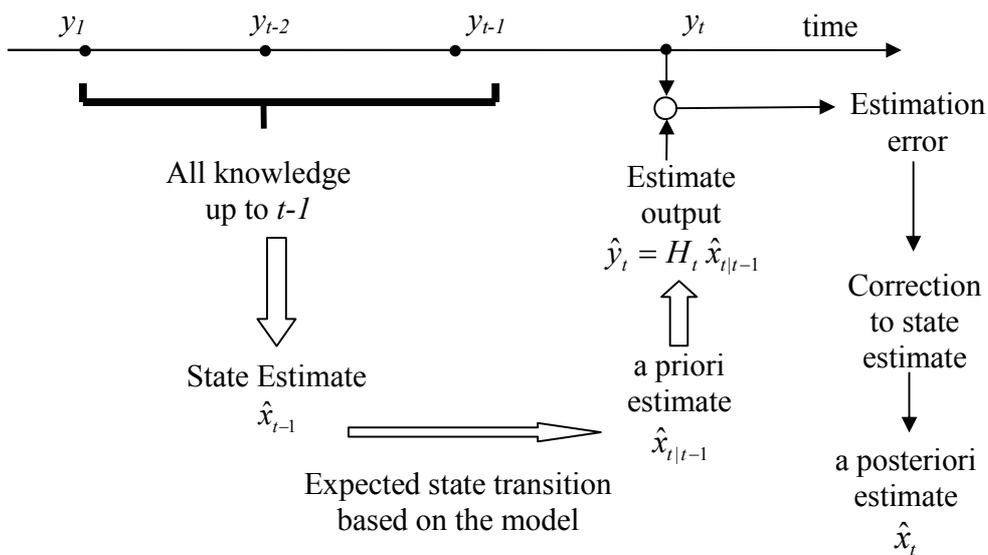


Figure 4-6 Outline of the Kalman filter algorithm

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

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

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

$$\hat{x}_{t|t-1} = E[A_{t-1}\hat{x}_{t-1} + G_{t-1}w_{t-1}]$$
$$= A_{t-1}\hat{x}_{t-1} + G_{t-1}E[w_{t-1}]$$

(21)

Expected state based on $\hat{x}_{t-1}$

*Estimated output*

Form (9) and (10)

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

(22)

*Correction of the state estimate*

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

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

(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

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

(24)

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

9