X Lecture 10, Nov. 10, 2010

Signal Estimation

In previous lectures, we discussed paramet estimation, where we were concerned with the estimation of parameters that are either static, i.e., not changing with time or slowly changing with time.

In some cases, we are interested in estimatin parameters that change with time. In radar, for example, we may be interested not only in knowing the position and velocity of the target, but, we may also like to track these parameters in different times. A quantity changing with time is called a signal and estimation of a quantity (the value of a parameter) in time is the subject of signal estimation.

Kalman filter

A discrete time dynamic System may be modelled as:

$$\underline{X}_{n+1} = \underline{f}_n(\underline{X}_n, \underline{U}_n) \quad n=0,1,\ldots$$

where the sequence of vectors $\underline{X}_0, \underline{X}_1, \ldots$ represents the state of the system at different times and $\underline{U}_0, \underline{U}_1, \ldots$ is a sequence of vector in $\mathbb{R}^s$ acting on $\{\underline{X}_n\}$ as the input sequence and $\underline{f}_n$, $n=0,1,\ldots$ is a sequence of functions each mapping $\mathbb{R}^m \times \mathbb{R}^s$ to $\mathbb{R}^m$ where m is the dimension of $\{\underline{X}_n\}$.

In addition to the above "state equation" a dynamical system is described by an output equation:

$$\underline{Z}_n = \underline{h}_n(\underline{X}_n) \quad n=0,1,\ldots$$

where $\underline{h}_n : \mathbb{R}^m \to \mathbb{R}^k$ where k is the number of outputs per output vector, i.e., $\underline{Z}_n \in \mathbb{R}^k$.

Example : <u>One-dimensional</u> <u>motion</u>

Assume that we would like to characterize the one dimensional motion of a particle.

Denote its position by $P(t)$ and its velocity by $V(t)$. We have $V(t) = \frac{dP(t)}{dt}$ and $A(t) = \frac{dV(t)}{dt}$.

Let's take a measurement of speed and position of the particle every T_s seconds.

Then

$$P_{n+1} = P(t)\Big|_{t=nT_s+T_s} \cong P_n + T_s V_n$$

if T_s is sufficiently small.

Here $V_n = V(nT_s)$ and $P_n = P(nT_s)$

Similary

$$V(n+1) \cong V_n + T_s A_n$$

where $A_n = A(t)\Big|_{t=nT_s}$ is the acceleration of the particle at n-th sampling time.

Let's take P_n and V_n as the state of the system, i.e.,

$$X_{1,n} = P_n = P(nT_s)$$

$$X_{2,n} = V_n = V(nT_s)$$

Take the accelerat A_n as the activation (input), i.

$$U_n = A_n = A(nT_s)$$

Also, assume that the ~~output~~ is the position, i.e.,

$$Z_n = P_n = X_{1,n}$$

Then, we can write

$$\underline{X}_{n+1} = \underline{F}\,\underline{X}_n + \underline{G}\,U_n \qquad n = 0, 1, \cdots$$

and

$$\underline{Z}_n = H\,\underline{X}_n \qquad n = 0, 1, \cdots$$

where $F = \begin{bmatrix} 1 & T_s \\ 0 & 1 \end{bmatrix}$

$\underline{G}$ is

$$G = \begin{bmatrix} 0 \\ T_s \end{bmatrix}$$

and $H = [1 \quad 0]$

So, in this case $m = 2$, $s = 1$, $k = 1$ and $\underline{f}_n$ and $\underline{h}_n$ are

$$\underline{f}_n(\underline{X}, \underline{U}) = F\,\underline{X} + G\,\underline{U}$$

and

$$\underline{h}_n(\underline{X}) = H\,\underline{X}$$

In many applications, we observe the output of a stochastic system in presence of noise, and we wish to find the state upto time t of the system at time u, i.e., we observe

$$\underline{Y}_n = \underline{Z}_n + \underline{V}_n \qquad n = 0, 1, \ldots, t$$

where $\underline{V}_n$ is the observation noise.

We wish to estimate $\underline{X}_u$.

If $u = t$ then the estimation problem is a filtering problem. For $u < t$ it is called the smoothing problem and for $u > t$, the problem is called a prediction problem.

If we use the mean-squared-error measure, i.e., $E\{\|\underline{X}_u - \hat{\underline{X}}_u\|^2\}$ for state estimates, we know from before that

$$\hat{\underline{X}}_u = E\{\underline{X}_u \mid \underline{Y}_0, \underline{Y}_1, \ldots, \underline{Y}_t\}$$

This is similar to the vector estimation in previous lectures (see chapter IV of the text).

The difference between signal estimation and vector estimation is that, in the case of signal estimation, we are interested in real-time estimation for all time instants. But the amount of data to be used for calculating the conditional expectation grows, the computational complexity grows boundlessly making the unrestricted approach un-feasible. So, we need to impose (or use existing) structure of the model.

One restriction we can impose is to assume that the dynamical system is linear, i.e.,

$$\underline{X}_{n+1} = \underline{F}_n \underline{X}_n + \underline{G}_n \underline{U}_n \qquad n = 0, 1, \cdots$$

and

$$\underline{Y}_n = \underline{H}_n \underline{X}_n + \underline{V}_n \qquad n = 0, 1, \cdots$$

where $\underline{F}_n$ is an $m \times m$ matrix

$\underline{G}_n$ is an $m \times s$ matrix

$\underline{H}_n$ is a $k \times m$ matrix

Another assumption which makes problem simpler by putting more restriction on the problem is that the input sequence $\{\underline{U}_n\}_{n=0}^{\infty}$ and the observation noise $\{\underline{V}_n\}_{n=0}^{\infty}$ are independent zero-mean Gaussian random vectors. We may also assume that the initial state $\underline{X}_0$ is a Gaussian random vector independent of $\{\underline{U}_n\}_{n=0}^{\infty}$ and $\{\underline{V}_n\}_{n=0}^{\infty}$.

With the above two assumptions, the computation of conditional-mean state estimator takes a very simple and computationally attractive form.

Following proposition gives the form for $u = t$ (filtering) and $u = t+1$ (prediction)

Proposition: The Discrete-Time Kalman-Bucy filter.

For a linear stochastic system with $\{\underline{U}_n\}_{n=0}^{\infty}$ and $\{\underline{V}_n\}_{n=0}^{\infty}$ being independen sequence of independent zero-mean Gaussian vectors independent of the initial condition $\underline{X}_0$, the estimates:

$$\hat{\underline{X}}_{t|t} \triangleq E\{\underline{X}_t | \underline{Y}_0^t\}$$

and

$$\hat{\underline{X}}_{t+1|t} \triangleq E\{\underline{X}_{t+1} | \underline{Y}_0^t\}$$

are given as:

$$\hat{\underline{X}}_{t|t} = \hat{X}_{t|t-1} + \underline{K}_t(\underline{Y}_t - H_t \hat{\underline{X}}_{t|t-1}) \qquad t = 0, 1, \cdots \qquad (A)$$

and

$$\hat{\underline{X}}_{t+1|t} = \underline{F}_t \hat{X}_{t|t} \qquad t = 0, 1, \cdots \qquad (B)$$

with the initialization:

$$\hat{X}_{0|-1} = \underline{m}_0 \triangleq E\{\underline{X}_0\}$$

the matrix $\underline{K}_t$ is given as:

$$\underline{K}_t = \Sigma_{t|t-1} \underline{H}_t^T (\underline{H}_t \Sigma_{t|t} \underline{H}_t^T + R_t)^{-1} \qquad (C)$$

with

$$\Sigma_{t|t-1} \triangleq Cov[\underline{X}_t | \cancel{Y}_0^{t-1}]$$

and $R_t \triangleq Cov(\underline{V}_t)$

Note that
$$\hat{X}_{t|t-1} = E\{\underline{X}_t \mid Y_0^{t-1}\}$$

So,
$$\Sigma_{t|t-1} = Cov(\underline{X}_t \mid \underline{Y}_0^{t-1})$$
$$= E\{\|\underline{X}_t - E\{\underline{X}_t \mid \underline{Y}_0^{t-1}\}\|^2\}$$
$$= E\{\|\underline{X}_t - \hat{\underline{X}}_{t|t-1}\|^2\}$$

That is $\Sigma_{t|t-1}$ is the covariance matrix of the <u>prediction</u> <u>error</u> conditioned on $\underline{Y}_0^{t-1}$.
Similarly $\Sigma_{t|t}$ is the covariance of the filtering error: $\Sigma_{t|t} \triangleq Cov(\underline{X}_t \mid Y_0^t)$
$$= E\{\|\underline{X}_t - \hat{X}_{t|t}\|^2\}.$$

$\Sigma_{t|t}$ and $\Sigma_{t+1|t}$ can be computed <u>recursively</u> as follows:
$$\Sigma_{t|t} = \Sigma_{t|t-1} - K_t \underline{H}_t \Sigma_{t|t-1} \qquad t = 0, 1, \ldots, \qquad (C)$$
and
$$\Sigma_{t+1|t} = F_t \Sigma_{t|t} F_t^T + G_t Q_t G_t^T \qquad t = 0, 1, \ldots, \qquad (D)$$
with the initialization $\Sigma_{0|-1} = \Sigma_0 \triangleq Cov(\underline{X}_0)$

and where $\underline{Q}_t$ is the covariance of $\underline{U}_t$, i.e.,

$$Q_t = Cov(\underline{U}_t).$$

Proof:

$$\underline{\hat{X}}_{t+1|t} = E\{\underline{X}_{t+1} | Y_0^t\} = E\{\underline{F}_t \underline{X}_t + G_t \underline{U}_t | Y_0^t\}$$

$$= F_t E\{\underline{X}_t | Y_0^t\} + G_t E\{\underline{U}_t | Y_0^t\}$$

$$= F_t \underline{\hat{X}}_{t|t} + G_t E\{\underline{U}_t | Y_0^t\}$$

Note that Y_0^t is determined by $\underline{X}_0^t$, $\underline{V}_0^t$ or equivalently from $\underline{X}_0$, $\underline{U}_0^{t-1}$ and V_0^t. None of these depend on $\underline{U}_t$. So, $E\{\underline{U}_t | Y_0^t\} = E\{\underline{U}_t\} = \underline{0}$ and, therefore,

$$\boxed{\hat{X}_{t+1|t} = \underline{F}_n \hat{X}_{t|t}, \quad t = 0, 1, \dots,} \qquad (B)$$

We also have:

$$\Sigma_{t+1|t} = Cov(\underline{X}_{t+1} | \underline{Y}_0^t)$$

$$= Cov(F_t \underline{X}_t + G_t \underline{U}_t | Y_0^t)$$

$$= Cov(F_t \underline{X}_t | Y_0^t) + Cov(G_t \underline{U}_t | Y_0^t)$$

$$= Cov(F_t X_t | Y_0^t) + Cov(G_t \underline{U}_t)$$

Using $$\mathrm{Cov}(AX) = A\,\mathrm{Cov}(\underline{X})A^T$$

we get,

$$\boxed{\begin{aligned}\Sigma_{t+1|t} &= F_t\,\mathrm{Cov}(\underline{X}_t \mid Y_0^t)\,F_t^T + G_t\,\mathrm{Cov}(U_t)\,G_t^T \\ &= F_t\,\Sigma_{t|t}F_t^T + G_t\,Q_t\,G_t^T\end{aligned}}$$

(D)

We have now proved the validity of (B) and (D). The other two equations can be proven by induction, i.e., we show they are true for $t=0$ and then we show that they are true for $t=t_0>0$ they are true for $t=t_0-1$.

For $t=0$, the measurement equation is:

$$\underline{Y}_0 = H_0\,\underline{X}_0 + \underline{V}_0$$

Since $\underline{X}_0$ and $\underline{V}_0$ are independent Gaussian vectors, the estimation of $\underline{X}_0$ from $\underline{Y}_0$ is the estimation of $\underline{X}_0 \sim N(\underline{m}_0, \Sigma_0)$ in the presence of noise vector $\underline{V}_0 \sim N(0, R_0)$, i.e.,

$$\hat{\underline{X}}_{0|0} \triangleq E\{\underline{X}_0 \mid \underline{Y}_0\}$$
$$= \underline{m}_0 + \Sigma_0 H_0^T (H_0 \Sigma_0 H_0^T + R_0)^{-1} (\underline{Y}_0 - H_0 \underline{m}_0)$$
$$= \underline{X}_{0|-1} + K_0 (Y_0 - H_0 \hat{\underline{X}}_{0|-1})$$

We have used the following definitions (given in the statement of the proposition):

$$\underline{m}_0 = \hat{X}_{0|-1}, \quad K_0 = K_{0|-1} H_0^T (H_0 \Sigma_{0|-1} H_0^T + R_0)^{-1}$$

and $\Sigma_{0|-1} = \Sigma_0$.

The above is equation (A) for $t=0$.

The error covariance matrix will be:

$$\Sigma_{0|0} = \Sigma_0 - \Sigma_0 H_0^T (H_0 \Sigma_0 H_0^T + R_0)^{-1} H_0 \Sigma_0$$
$$= \Sigma_{0|-1} - K_0 H_0 \Sigma_{0|-1}$$

which is (C) for $t=0$.

Now, assume that (A) and (C) are valid for $t = t_0 - 1$.

Note that $\underline{X}_{t_0}$ and $\underline{Y}_0^{t_0-1}$ are the linear transformatio of the Gaussian vector: $\underline{X}_0$, $\underline{U}_0^{t_0-1}$ and $\underline{V}_0^{t_0-1}$!
So, they are Gaussian and $\underline{X}_{t_0}$ is Conditionally Gaussian given $\underline{Y}_0^{t_0-1}$. The Conditional distribution of $\underline{X}_{t_0}$ given $\underline{Y}_0^{t-1}$ is $N(\hat{\underline{X}}_{t_0|t_0-1}, \Sigma_{t_0|t_0-1})$.

Also $\underline{V}_{t_0}$ is Gaussian and independent of $\underline{Y}_0^{t_0-1}$. So, it is conditionally Gaussian with distribution $N(0, R_{t_0})$. Since $\underline{V}_{t_0}$ is independent of $\underline{X}_0$, $\underline{V}_0^{t_0-1}$ and $\underline{U}_0^{t_0-1}$, it is conditionally (conditioned on $\underline{Y}_0^{t_0-1}$) independent of $\underline{X}_{t_0}$.

So,

$$\underline{Y}_{t_0} = H_{t_0} \underline{X}_{t_0} + \underline{V}_{t_0}$$

given $\underline{Y}_0^{t_0-1}$ is a Gaussian linear equation

Now, if we find $\hat{X}_{t_0|t_0}$ as the expectation of $\underline{X}_{t_0}$ given $\underline{Y}_{t_0}$, we have

$$\hat{X}_{t_0|t_0} = \hat{X}_{t_0|t_0-1} H_{t_0}^T (H_{t_0} \Sigma_{t_0|t_0-1} H_{t_0}^T + R_{t_0})^{-1}$$
$$\times (\underline{Y}_{t_0} - H_{t_0} \hat{X}_{t_0|t_0-1})$$

where the fact that

$$\underline{X}_{t_0} \sim N(\hat{X}_{t_0|t_0-1}, \Sigma_{t_0|t_0-1})$$

when conditioned on $\underline{Y}_0^{t_0-1}$.

Just substituting $\underline{K}_{t_0}$ for

$$K_{t_0|t_0-1} H_{t_0}^T (H_{t_0} \Sigma_{t_0|t_0-1} H_{t_0}^T + R_{t_0})^{-1}$$

we get equation (A)

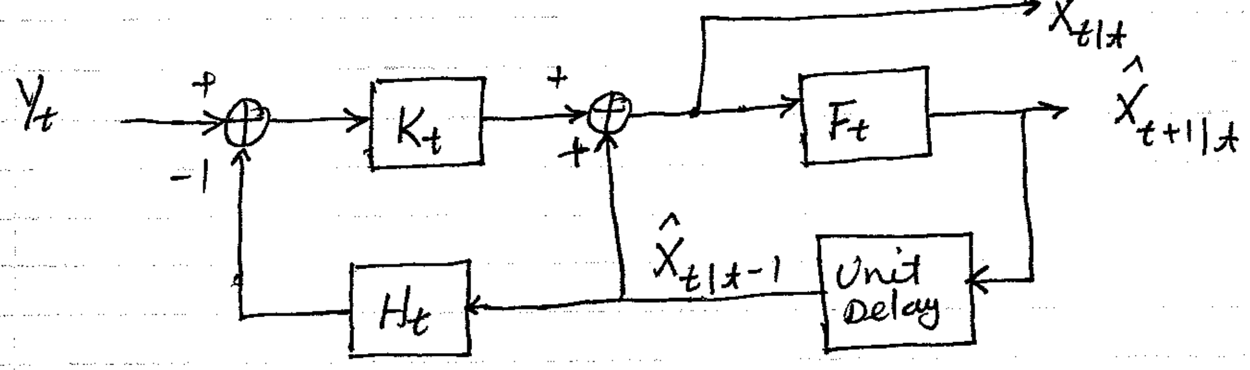
In a similar way, using

$$\Sigma_{t_0|t_0} = \Sigma_{t_0|t_0-1} - \Sigma_{t_0|t_0-1} H_{t_0}^T (H_{t_0} \Sigma_{t_0|t_0-1} H_{t_0}^T + R_{t_0})^{-1}$$

$$\times H_{t_0} \Sigma_{t_0|t_0-1}$$

and replacing for K_t, we get (C).

The Kalman-Bucy estimator's block diagram is:



Note that the <u>recursive</u> structure of this estimator makes it very computationally efficient. Note that although the prediction $\hat{X}_{t+1|t}$ and filtered estimate $\hat{X}_{t|t}$ depend on all previous data, i.e., $\underline{Y}_0^t$, they can be computed only using $\underline{Y}_t$ and the previous prediction $\hat{X}_{t|t-1}$. Therefore, we do not need to store and process at every stage all observations $\{Y_t\}_{t=0}^{\infty}$. We need only stor and update a single vector

$\hat{X}_{t|t-1}$ (the prediction vector). All other parts of the filter including the Kalman gain matrix K_t, are computed from the parameters of the system model.

Looking more closely at the Kalman filter Equations, we see that we deal with two steps:

In step one, the measurement update stage, we do measurement update:

$$\hat{X}_{t|t} = \hat{X}_{t|t-1} + K_t (\underline{Y}_t - H_t \hat{X}_{t|t-1})$$

while in step two, the time update stage, we project the estimate to the next time $(t+1)$:

$$\hat{X}_{t+1|t} = F_t \hat{X}_{t|t} .$$

While time update is performed based on the system's structure, i.e., it is exclusively derived from state equations, the measurement update is done based on the measurement equation.

Innovation Process

Let's look at the measurement update:

$$\hat{\underline{X}}_{t|t} = \hat{X}_{t|t-1} + K_t(\underline{Y}_t - H_t\hat{\underline{X}}_{t|t-1})$$

We observe that the estimate consists of two parts: a) previous estimate (time updated) and b) a correction term, i.e.,

$$K_t(\underline{Y}_t - H_t\hat{\underline{X}}_{t|t-1}).$$

Denot the vector in the second term as:

$$\underline{I}_t = \underline{Y}_t - H_t\hat{X}_{t|t-1}$$

Since

$$\underline{Y}_t = H_t\underline{X}_t + \underline{V}_t$$

we have

$$Y_{t|t-1} \triangleq E\{\underline{Y}_t | \underline{Y}_0^{t-1}\} = H_t E\{X_t | \underline{Y}_0^{t-1}\} + E\{\underline{V}_t | \underline{Y}_0^{t-1}\} = H_t\hat{X}_{t|t-1}$$

(since $\underline{V}_t$ is independent of $\underline{Y}_0^{t-1}$ and is zero-mean)

So,

$$\underline{I}_t = \underline{Y}_t - \hat{\underline{Y}}_{t|t-1}$$

and, this form indicates that $\underline{I}_t$ is an error signal representing the error in prediction of $\underline{Y}_t$ based on $\underline{Y}_0^{t-1}$. This is called <u>residual</u> or <u>innovation</u> signal. The reason is that, we can write $\underline{Y}_t$ as

$$\underline{Y}_t = \hat{\underline{Y}}_{t|t-1} + \underline{I}_t ,$$

that is $\underline{Y}_t$ consists of a part that can be inferred from the past history and an independent <u>new</u> component, $\underline{I}_t$.

It is easy to show that the innovation sequence $\{\underline{I}_t\}_{t=0}^{\infty}$ is a sequence of independent zero-mean Gaussian vectors.

Example: One-dimensional (single variable) Case.

Consider the case $m = k = 1$, i.e., the state is a one-dimensional (scalar), $X_n \in \mathbb{R}$ and there is a one-dimensional output, equivalent observation.

$$X_{n+1} = f X_n + U_n, \quad n = 0, 1, \cdots$$

$$Y_n = h X_n + V_n, \quad n = 0, 1, \cdots$$

where $\{U_n\}_{n=0}^{\infty}$ and $\{V_n\}_0^{\infty}$ are independent sequences of i.i.d. $N(0, q)$ and $N(0, r)$ random variables. Also $X_0 \sim N(m_0, \Sigma_0)$ and f, h, q, r and Σ_0 are scalars.

The estimation recursions are:

$$\hat{X}_{t|t} = \hat{X}_{t|t-1} + K_t (Y_t - h \hat{X}_{t|t-1}) \quad t = 0, 1, \cdots$$

and

$$\hat{X}_{t+1|t} = f \hat{X}_{t|t} \quad t = 0, 1, \cdots$$

The Kalman gain is given by:

$$K_t = \frac{\Sigma_{t|t-1}\, h}{h^2 \Sigma_{t|t-1} + r} = \frac{1}{h} \frac{\Sigma_{t|t-1}}{\Sigma_{t|t-1} + \frac{r}{h^2}}$$

$\Sigma_{t|t-1}$ is the MSE of the estimation of X_t from $\underline{Y}_0^{t-1}$ and $\frac{r}{h^2}$ is a measure of noisiness of observation. The reason is that

$$Y_t = h X_t + V_t \Rightarrow \frac{1}{h} Y_t = X_t + \frac{V_t}{h}$$

The difference between $\frac{1}{h} Y_t$ and X_t has the average squared error equal to the variance of $\frac{V_t}{h}$, i.e., $\frac{r}{h^2}$.

When

$$\Sigma_{t|t-1} \ll \frac{r}{h^2}$$

i.e., when the observation is very noisy, we get $K_t = 0$ and, therefore

$$\hat{X}_{t|t-1} = \hat{X}_{t|t-1}$$

This means that we ignore the observation and rely on past estimate of X_t.

On the other hand when

$$\Sigma_{t|t} \gg \frac{r}{h^2}$$

Then $K_t = \frac{1}{h}$ and we have:

$$\hat{X}_{t|t} = \frac{1}{h} Y_t$$

meaning that, we trust the present observation.