Lecture 23: Kalman Filtering

1 Linear Dynamic Systems

Contd. from last class (Lecture-22)

A **very important special case** of signal estimation is a "Linear Dynamical System driven by Gaussian noise/controls". Here the state and observation equations are of the following form,

$$\begin{aligned} \underline{X}_{t+1} &= \mathbf{F}_t \underline{X}_t + \mathbf{G}_t \underline{U}_t, \\ \underline{Y}_t &= \mathbf{H}_t \underline{X}_t + \underline{V}_t, \end{aligned} \tag{1}$$

where $\{U_t\}$ and $\{V_t\}$ are independent sequences of independent, zero-mean Gaussian vectors, independent of the initial condition $\underline{X}_0$. Also, $\underline{X}_0 \sim \mathcal{N}(\underline{m}_0, \mathbf{\Sigma}_0)$, with $Cov(U_t) = Q_t, Cov(V_t) = R_t$.

- Filtering : Estimate $\underline{X}_t$ given $(\underline{Y}_0, ..., \underline{Y}_t) \equiv \underline{Y}_0^t$.
- Prediction : Estimate $\underline{X}_{t+1}$ given $(\underline{Y}_0, ..., \underline{Y}_t) \equiv \underline{Y}_0^t$.

The *criterion* to be minimized is the mean square error of estimator $\underline{\hat{X}}_{t|t}$, i.e., $\mathbb{E}[||\underline{\hat{X}}_{t|t} - \underline{X}_t||_2^2]$.

Conceptual Solution Parametric Estimation: Let $\mathbf{Y} = (\underline{Y}_0, ..., \underline{Y}_t) = \underline{Y}_0^t \sim f(y|\theta)$, and, $\Theta \equiv \underline{X}_0^t \equiv (\underline{X}_0, ..., \underline{X}_t)$. Prior on Θ is the distribution of $\underline{X}_0^t$ induced by the distribution of $\underline{X}_0$ and $\{\underline{U}_n\}_{n=0}^t$,

$$\mathbf{Y} \sim f(y|\theta) \tag{2}$$

$$\mathbf{Y} = \mathbf{H}\Theta + \mathbf{V} \tag{3}$$

$$\begin{bmatrix} \underline{Y_0} \\ \vdots \\ \vdots \\ \underline{Y_t} \end{bmatrix} = \begin{bmatrix} H_0 & & & \\ & H_1 & & \\ & & \ddots & \\ & & & H_t \end{bmatrix} \begin{bmatrix} \underline{X_0} \\ \vdots \\ \vdots \\ \underline{X_t} \end{bmatrix} + \begin{bmatrix} \underline{V_0} \\ \vdots \\ \vdots \\ \underline{V_t} \end{bmatrix} \tag{4}$$

and loss function $L(g(\theta), \mathbf{W}(\mathbf{Y})) = ||\theta_t - \mathbf{W}(\mathbf{Y}))||_2^2$. From Bayesian estimation, the best estimator is $\mathbb{E}[g(\theta)|\mathbf{Y}] = \mathbb{E}[\underline{X}_t|\underline{Y}_0...\underline{Y}_t]$.

Theorem 1.1. *Discrete time Kalman-Bucy Filter For the linear dynamic system represented by the set of eqns.* (1)*, the optimal squared error estimates,*

$$\begin{aligned}\underline{\hat{X}}_{t|t} &:= \mathbb{E}[\underline{X}_t|\underline{Y}_0^t],\\ \underline{\hat{X}}_{t+1|t} &:= \mathbb{E}[\underline{X}_{t+1}|\underline{Y}_0^t],\end{aligned}$$

obey the following recursions,

$$\underline{\hat{X}}_{t|t} = \underline{\hat{X}}_{t|t-1} + \boldsymbol{K}_t(\underline{Y}_t - \boldsymbol{H}_t\underline{\hat{X}}_{t|t-1}),\ t = 0, 1, ... \tag{5}$$

$$\underline{\hat{X}}_{t+1|t} = \boldsymbol{F}\underline{\hat{X}}_{t|t}, \tag{6}$$

with initialization,

$$\begin{aligned}\underline{\hat{X}}_{0|-1} &= \underline{m}_0 = \mathbb{E}[\underline{X}_0],\\ \boldsymbol{K}_t &:= \boldsymbol{\Sigma}_{t|t-1}\boldsymbol{H}_t^T(\boldsymbol{H}_t\boldsymbol{\Sigma}_{t|t-1}\boldsymbol{H}_t^T + \boldsymbol{R}_t)^{-1},\end{aligned}$$

where,

$$\begin{aligned}\boldsymbol{\Sigma}_{0|-1} &:= \boldsymbol{\Sigma}_0,\\ \boldsymbol{\Sigma}_{t|t-1} &:= Cov(\underline{X}_t|\underline{Y}_0^t) = Cov(\underbrace{(\underline{X}_t - \underline{\hat{X}}_{t|t-1})}_{prediction\ error}|\underline{Y}_0^t).\end{aligned}$$

More over, the covariance matrices satisfy the recursion,

$$\boldsymbol{\Sigma}_{t|t} = \boldsymbol{\Sigma}_{t|t-1} - \boldsymbol{K}_t\boldsymbol{H}_t\boldsymbol{\Sigma}_{t|t-1}, \tag{7}$$

$$\boldsymbol{\Sigma}_{t+1|t} = \boldsymbol{F}_t\boldsymbol{\Sigma}_{t|t}\boldsymbol{F}_t^T + \boldsymbol{G}_t\boldsymbol{Q}_t\boldsymbol{G}_t^T, \tag{8}$$

Remark 1. K-B Filter gives a sequential rule to output estimates.

Proof. First, Let us prove (6), (8),

$$\begin{aligned}\underline{\hat{X}}_{t+1|t} &= \mathbb{E}[\underline{X}_{t+1}|\underline{Y}_0^t]\\ &= \mathbb{E}[\mathbf{F}_t\underline{X}_t + \mathbf{G}_t\underline{U}_t|\underline{Y}_0^t]\\ &= \mathbf{F}_t\mathbb{E}[\underline{X}_t|\underline{Y}_0^t] + \mathbf{G}_t\mathbb{E}[\underline{U}_t|\underline{Y}_0^t]\\ &= \mathbf{F}_t\underline{\hat{X}}_{t|t}\\ \boldsymbol{\Sigma}_{t+1|t} &= Cov(\underline{X}_{t+1}|\underline{Y}_0^t)\\ &= Cov(\mathbf{F}_t\underline{X}_t + \mathbf{G}_t\underline{U}_t|\underline{Y}_0^t)\\ &= Cov(\mathbf{F}_t\underline{X}_t|\underline{Y}_0^t) + Cov(\mathbf{G}_t\underline{U}_t|\underline{Y}_0^t)\\ &= \mathbf{F}_tCov(\underline{X}_t|\underline{Y}_0^t)\mathbf{F}_t^T + \mathbf{G}_tCov(\underline{U}_t|\underline{Y}_0^t)\mathbf{G}_t^T\\ &= \mathbf{F}_t\boldsymbol{\Sigma}_{t|t}\mathbf{F}_t^T + \mathbf{G}_t\mathbf{Q}_t\mathbf{G}_t^T\end{aligned}$$

Now, consider the proof of (5):

Lemma 1.2. *Suppose $A \in \mathbb{R}^n, B \in \mathbb{R}$ are jointly Gaussian random vectors with $\mathbb{E}[\mathbb{A}] = \underline{\mu}_A, \mathbb{E}[B] = \underline{\mu}_B$, $Cov(A) = \mathbf{\Sigma}_A$, $Cov(B) = \mathbf{\Sigma}_B$, $Cov(A, B) = \mathbf{\Sigma}_{AB} = \mathbf{\Sigma}_{BA}^T = \mathbb{E}[(A - \underline{\mu}_A)(B - \underline{\mu}_B)^T]$. Then the conditional distribution of B given A is (multivariate) Gaussian, with mean,*

$$\underline{\mu}_{B|A} = \underline{\mu}_B + \mathbf{\Sigma}_{BA} + \mathbf{\Sigma}_A^{-1}(A - \underline{\mu}_A)$$

and covariance,

$$\mathbf{\Sigma}_{B|A} = \mathbf{\Sigma}_B - \mathbf{\Sigma}_{BA}\mathbf{\Sigma}_A^{-1}\mathbf{\Sigma}_{AB}$$

Consider, $\underline{Y}_t = \mathbf{H}_t \underline{X}_t + \underline{V}_t$, we note that

1. $\underline{X}_t$ is conditionally Gaussian given $\underline{Y}_0^{t-1}$ ($\underline{X}_t$ &$\underline{Y}_0^{t-1}$ are linear transformations of $\underline{X}_0, \{\underline{U}_n\}_n, \{\underline{V}_n\}_n$).
2. Given $\underline{Y}_0^{t-1}$, $\underline{X}_t \sim \mathcal{N}(\hat{\underline{X}}_{t|t-1}, \mathbf{\Sigma}_{t|t-1})$.
3. $\underline{V}_t$ is Gaussian.
4. Given $\underline{Y}_0^{t-1}, \underline{V}_t \perp\!\!\!\perp \underline{X}_t$, because $(\underline{V}_t \perp\!\!\!\perp (\underline{X}_o, \underline{U}_0^{t-1}\underline{V}_0^{t-1}))$.

Therefore, using Lemma. 1.2, given $\underline{Y}_0^{t-1}$, the conditional distribution of $\underline{X}_t$ given $\underline{Y}_t$ is Gaussian with mean,

$$\begin{aligned}\underline{\mu}_B + \mathbf{\Sigma}_{BA}\mathbf{\Sigma}_A^{-1}(A - \underline{\mu}_A) = \hat{\underline{X}}_{t|t-1} + \mathbb{E}[(\underline{X}_t - \hat{\underline{X}}_{t|t-1})(\underline{Y}_t - \mathbb{E}[\underline{Y}_t|\underline{Y}_0^{t-1}])^T] \\ Cov(\underline{Y}_t|\underline{Y}_0^{t-1})^{-1}(\underline{Y}_t - \mathbb{E}[\underline{Y}_t|\underline{Y}_0^{t-1}]),\end{aligned}$$

i.e.,

$$\begin{aligned}\hat{\underline{X}}_{t|t} = \mathbb{E}[\underline{X}_t|\underline{Y}_0^t] &= \hat{\underline{X}}_{t|t-1} + \mathbf{\Sigma}_{t|t-1}\mathbf{H}_t^T(\mathbf{H}_t\Sigma_{t|t-1}\mathbf{H}_t^T + \mathbf{R}_t)^{-1}(\underline{Y}_t - \mathbf{H}_t\hat{\underline{X}}_{t|t-1}) \\ &= \hat{\underline{X}}_{t|t-1} + \mathbf{K}_t(\underline{Y}_t - \mathbf{H}_t\hat{\underline{X}}_{t|t-1}),\end{aligned}$$

and similarly for the conditional covariance of $\underline{X}_t$ given $\underline{Y}_t$, $\underline{Y}_0^{t-1}$. □

Note 1. The KB Filter computes both an estimate $(\hat{\underline{X}}_{t|t}, \hat{\underline{X}}t + 1|t)$ and its mean square error(MSE) given covariance $(\mathbf{\Sigma}_{t|t}\mathbf{\Sigma}_{t+1|t})$.

Note 2. The sequence of conditional covariance $\{\mathbf{\Sigma}_{t|t}\}_{t=0}^{\infty}$ does not depend on the observations $\{\underline{Y}_t\}_{t=0}^{\infty}$.

Consider the vector $\underline{I}_t = \underline{Y}_t - \mathbf{H}_t\hat{\underline{X}}_{t|t-1}$. $\underline{I}_t$ is a "correction" term representing error in predicting the observations, called the innovation at time t. <u>Claim:</u> $\{\underline{I}_t\}_{t=0}^{\infty}$ is a sequence of zero mean independent Gaussian vectors.

Proof. Since $\underline{Y}_t$ and $\mathbf{H}_t\hat{\underline{X}}_{t|t-1}$ are Gaussians.

$$\begin{aligned}
\mathbb{E}[\underline{I}_t] &= \mathbb{E}\left[\underline{Y}_t - \mathbf{H}_t\hat{\underline{X}}_{t|t-1}\right] \\
&= \mathbb{E}[\underline{Y}_t - \mathbf{H}_t\mathbb{E}\left[\underline{X}_t|\underline{Y}_0^{t-1}\right]] \\
&= \mathbb{E}[\mathbf{H}_t\underline{X}_t - \mathbf{H}_t\mathbb{E}[\underline{X}_t|\underline{Y}_0^{t-1}]] \\
&= \mathbb{E}\left[\mathbb{E}\left[\left(\mathbf{H}_t\underline{X}_t - \mathbf{H}_t\mathbb{E}[\underline{X}_t|\underline{Y}_0^{t-1}]\right)\middle|\underline{Y}_0^{t-1}\right]\right] \\
&= 0.
\end{aligned}$$

$\forall s < t$, we have

$$\begin{aligned}
\mathbb{E}[\underline{I}_t\underline{I}_s^T] &= \mathbb{E}[[\underline{I}_t\underline{I}_s^T|\underline{Y}_0^s]] \\
&= \mathbb{E}[\mathbb{E}[\underline{I}_t|\underline{Y}_0^s]\underline{I}_s^T], \\
&= 0
\end{aligned}$$

where, $\mathbb{E}[\underline{I}_t|\underline{Y}_0^s] = \mathbb{E}[(\underline{Y}_t - \mathbf{H}_t\hat{\underline{X}}_{t|t-1})|\underline{Y}_0^s] = 0.$ □