Stochastic Processes

Topic 9

Prediction and Kalman Filtering

nctuee09f

Reading:

- Textbook Sec. 9.2.

Summary

In this topic, I will discuss:

- Review of Conditional Expectation and MMSE
- Linear State Variable Model
- MMSE Prediction
- Kalman Filter

Notation

We will use the following notation rules, unless otherwise noted, to represent symbols during this course.

- Boldface upper case letter to represent **MATRIX**
- Boldface lower case letter to represent **vector**
- Superscript $(\cdot)^T$ and $(\cdot)^H$ to denote transpose and hermitian (conjugate transpose), respectively
- Upper case italic letter to represent *RANDOM VARIABLE*

1 Prediction

(1) Review:

(a) MMSE: (topic 5) We want to estimate a random vector $\mathbf{x}$ based on observation $\mathbf{y}$ using the minimum mean-squared error criterion. The answer is

$$\begin{aligned} \hat{\mathbf{g}}_{mmse}(\mathbf{y}) &= \arg\min_{g(\mathbf{y})} E\left[\left|\left|\mathbf{x} - g(\mathbf{y})\right|\right|^2\right] \\ &= E[\mathbf{x}|\mathbf{y}] \\ &\triangleq \hat{\mathbf{x}}_{mmse} \end{aligned}$$

(b) LMMSE: (topic 5) We want the rule $g(\mathbf{y})$ to be constrained by $g(\mathbf{y}) = \mathbf{A} \cdot \mathbf{y} + \mathbf{b}$. We know

$$\hat{\mathbf{x}}_{lmmse} = \mathbf{m_x} + \mathbf{K_{xy}}\mathbf{K_y}^{-1}\big(\mathbf{y} - \mathbf{m_y}\big).$$

(c) Recall from HW 2, the extra problem.

Let $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$ be collectively jointly Gaussian random vectors.

— If $\mathbf{y}$ and $\mathbf{z}$ are statistically independent, then

$$E[\mathbf{x}|\mathbf{y},\mathbf{z}] = E[\mathbf{x}|\mathbf{y}] + E[\mathbf{x}|\mathbf{z}] - \mathbf{m_x},$$

where $\mathbf{m_x} = E[\mathbf{x}]$.

— If $\mathbf{y}$ and $\mathbf{z}$ are not necessarily statistically independent, then

$$\begin{aligned} E[\mathbf{x}|\mathbf{y},\mathbf{z}] &= E[\mathbf{x}|\mathbf{y},\tilde{\mathbf{z}}] \\ &= E[\mathbf{x}|\mathbf{y}] + E[\mathbf{x}|\tilde{\mathbf{z}}] - \mathbf{m_x}, \end{aligned}$$

where $\tilde{\mathbf{z}} = \mathbf{z} - E[\mathbf{z}|\mathbf{y}]$.

(2) Basic state variable model [1]

— The state variable model is often used to study a ***time-varying*** (dynamic) phenomena embedded in observations or measurements, e.g. time-varying channels appeared in the received signals of a wireless link.

— The model is characterized by a ***state*** vector $\mathbf{x}[k]$ and a ***measurement*** vector $\mathbf{z}[k]$.

State vector:

$$\mathbf{x}[k+1] = \mathbf{A}[k+1,k] \cdot \mathbf{x}[k] + \mathbf{B}[k+1,k] \cdot \mathbf{w}[k] + \mathbf{C}[k+1,k] \cdot \mathbf{u}[k]$$

Measurement vector:

$$\mathbf{z}[k+1] = \mathbf{H}[k+1] \cdot \mathbf{x}[k+1] + \mathbf{v}[k+1].$$

— $\mathbf{w}[k]$ is used to model

$\Rightarrow$ disturbance forces acting on the system

$\Rightarrow$ errors in modeling the system

$\Rightarrow$ probabilistic behavior of the state vector $\mathbf{x}[k]$

— $\mathbf{v}[k]$ is used to model errors/noise in the measurement

— Example in a wireless communication link

In wireless communications, a ***time-varying channel*** vector process $\mathbf{x}[k+1]$ at time instant t_{k+1} is often described by the ***state vector*** to specify it's dynamic evolution from its previous state $\mathbf{x}[k]$ at time instant t_k. At the receiver, the ***received signal*** $\mathbf{z}[k+1]$ is corrupted by noise $\mathbf{v}[k+1]$ and affected by the channel vector $\mathbf{x}[k+1]$.

The other factors in the wireless link such as channel coding or desired and mostly unknown transmitted signal are incorporated in the system matrix $\mathbf{H}[k+1]$. So, typically, the channel vector $\mathbf{x}[k+1]$ and the system matrix $\mathbf{H}[k+1]$ are both unknown to the receiver. However, in the following development, we will assume that we have the knowledge of the system matrix $\mathbf{H}[k+1]$.

We will employ the basic state variable model to establish the MMSE predictor and the Kalman filter in the next part.

Assumptions made to the model here

The model is repeated here for convenience.

State vector:

$$\mathbf{x}[k+1] = \mathbf{A}[k+1,k] \cdot \mathbf{x}[k] + \mathbf{B}[k+1,k] \cdot \mathbf{w}[k] + \mathbf{C}[k+1,k] \cdot \mathbf{u}[k]$$

Measurement vector:

$$\mathbf{z}[k+1] = \mathbf{H}[k+1] \cdot \mathbf{x}[k+1] + \mathbf{v}[k+1].$$

— $\mathbf{u}[k]$ is a known input. The matrices $\mathbf{A}[k+1,k]$, $\mathbf{B}[k+1,k]$, and $\mathbf{C}[k+1,k]$ are also known.
(In general, they may not be available)

— $\mathbf{w}[k]$ and $\mathbf{v}[k]$ are mutually uncorrelated jointly Gaussian white noise sequence with

$$E[\mathbf{w}[i]\mathbf{w}^H[j]] = \mathbf{Q}[i] \cdot \delta_{ij} \quad \text{and} \quad E[\mathbf{v}[i]\mathbf{v}^H[j]] = \mathbf{R}[i] \cdot \delta_{ij}.$$

— The initial state vector $\mathbf{x}[0]$ is a Gaussian vector (all components are jointly Gaussian) with known mean $\mathbf{m}_x[0]$ and covariance matrix $\mathbf{K}_x[0]$.

— The matrix $\mathbf{R}[i]$ is positive definite. (inverse exists)

— $\mathbf{x}[0]$ is statistically independent with $\mathbf{w}[k]$ and $\mathbf{v}[k]$.

Note:
When $\mathbf{x}[0]$, $\mathbf{w}[k]$ and $\mathbf{v}[k]$ are jointly Gaussian, then $\mathbf{z}[k]$, $k = 1, 2, \ldots$ are also jointly Gaussian.

(4) **Prediction**

— <u>**Problem Formulation**</u>

At time instant t_k, we have the measured data vector $\mathbf{y}[k] = \left[\mathbf{z}^T[k], \ldots, \mathbf{z}^T[1]\right]^T$ available. We want to ***predict*** the random vector $\mathbf{x}[k+1]$ that is about to occur at time t_{k+1} based on $\mathbf{y}[k]$ ***using MMSE criterion***.

<u>**Solution**</u>

From lectures in topic 5, we know the solution is

$$\hat{\mathbf{x}}[k+1|k] \triangleq E\left[\mathbf{x}[k+1] \mid \mathbf{y}[k]\right].$$

Based on the state vector in the state variable model and the assumptions made in the above, the predictor can be written as

$$\boxed{\hat{\mathbf{x}}[k+1|k] = \mathbf{A}[k+1,k] \cdot \hat{\mathbf{x}}[k|k] + \mathbf{C}[k+1,k] \cdot \mathbf{u}[k],}$$

where

$$\hat{\mathbf{x}}[k|k] \triangleq E\left[\mathbf{x}[k] \mid \mathbf{y}[k]\right]$$

is the ***filtering*** result at time t_k. At this point, we are not able to proceed until we know what the filtering result $\hat{\mathbf{x}}[k|k]$ is in the the next section. It is clear here that the ***prediction*** and ***filtering*** are closely related.

— Error covariance matrix

The prediction error is defined by

$$\tilde{\mathbf{x}}[k+1|k] \triangleq \mathbf{x}[k+1] - \hat{\mathbf{x}}[k+1|k].$$

Again from the state vector model, it immediately follows that

$$\tilde{\mathbf{x}}[k+1|k] = \mathbf{A}[k+1,k] \cdot \tilde{\mathbf{x}}[k|k] + \mathbf{B}[k+1,k] \cdot \mathbf{w}[k].$$

It's easy to show that the prediction error has zero mean. Then, the prediction error covariance matrix is

$$\begin{aligned} \mathbf{P}[k+1|k] &= E\left[\tilde{\mathbf{x}}[k+1|k]\tilde{\mathbf{x}}^H[k+1|k]\right] \qquad (1) \\ &= \mathbf{A}[k+1,k]\mathbf{P}[k|k]\mathbf{A}^H[k+1,k] + \mathbf{B}[k+1,k]\mathbf{Q}[k]\mathbf{B}^H[k+1,k], \end{aligned}$$

where $\mathbf{P}[k|k]$ is the error covariance matrix of the filtering error $\tilde{\mathbf{x}}[k|k]$.

Remark:

Note that $\hat{\mathbf{x}}[0|0]$ and $\mathbf{P}[0|0]$ initialize the single-stage prediction and its error covariance matrix, respectively. They both are

$$\hat{\mathbf{x}}[0|0] = E[\mathbf{x}[0]|\text{no measurement}] = E[\mathbf{x}[0]] = \mathbf{m}_x[0],$$

$$\mathbf{P}[0|0] = E\left[\tilde{\mathbf{x}}[0|0]\tilde{\mathbf{x}}^H[0|0]\right] = \mathbf{K}_x[0].$$

(5) **Innovations Process**

Now consider the predictor $\hat{\mathbf{z}}[k+1|k]$ of the measurement vector $\mathbf{z}[k+1]$. We know $\hat{\mathbf{z}}[k+1|k] = E\big[\mathbf{z}[k+1]|\mathbf{y}[k]\big]$.

The difference between the new observation $\mathbf{z}[k+1]$ and the predictor $\hat{\mathbf{z}}[k+1|k]$ is often referred to as the innovations process, denoted by

$$\tilde{\mathbf{z}}[k+1|k] \triangleq \mathbf{z}[k+1] - \hat{\mathbf{z}}[k+1|k].$$

Remarks:

— The innovations process $\tilde{\mathbf{z}}[k+1|k]$ and the measurement vector $\mathbf{z}[k+1]$ are causally invertible. That is, we can compute one from the other using physically realizable filter.

— From the measurement vector model, we know

$$\hat{\mathbf{z}}[k+1|k] = \mathbf{H}[k+1]\hat{\mathbf{x}}[k+1|k] \tag{2}$$

$$\tilde{\mathbf{z}}[k+1|k] = \mathbf{H}[k+1]\tilde{\mathbf{x}}[k+1|k] + \mathbf{v}[k+1]. \tag{3}$$

— The innovations process is a zero mean Gaussian ***white*** noise sequence with covariance matrix

$$E[\tilde{\mathbf{z}}[k+1|k]\tilde{\mathbf{z}}^H[k+1|k]] = \mathbf{H}[k+1]\mathbf{P}[k+1|k]\mathbf{H}^H[k+1] + \mathbf{R}[k+1].$$

$\Rightarrow$ Justify that why the innovations process is white!!

— ***Why*** introducing ***innovations process***?

The innovations process and $\mathbf{z}[k+1]$ are ***physically equivalent*** in the sense that one can be computed from the other using physically realizable transformations. The appealing property of being mutually orthogonal (or mutually independent in the jointly Gaussian scenario) makes the algebraic derivations of Kalman filter more manageable by the use of innovations process rather than the newly observed data $\mathbf{z}[k+1]$. The "innovations" provided by $\tilde{\mathbf{z}}[i+1|i]$ promise exactly identical information that the newly measured $\mathbf{z}[k+1]$ data can offer.

2 Kalman Filter

(1) In words, Kalman filter is a ***recursive MMSE state estimator***, which uses all of the measurements up to and including the one made at current time instant.

(2) (**Problem Formulation**)
We want to estimate $\mathbf{x}[k+1]$ based on the stacked measurement vector $\mathbf{y}[k+1] = \left[\mathbf{z}^T[k+1], \mathbf{z}^T[k], \ldots, \mathbf{z}^T[1]\right]^T$.
We know the solution is

$$\hat{\mathbf{x}}[k+1|k+1] = E[\mathbf{x}[k+1]|\mathbf{y}[k+1]].$$

However, as time grows, the computation involved increases as well. Based on the state variable model, we intend to develop a ***recursive*** form of the MMSE estimate.

(3) Here are the results:

— The ***recursive*** mean squared estimator $\hat{\mathbf{x}}[k+1|k+1]$ of $\mathbf{x}[k+1]$ takes the form of ***prediction*** and ***correction*** as [1]

$$\boxed{\hat{\mathbf{x}}[k+1|k+1] = \underbrace{\hat{\mathbf{x}}[k+1|k]}_{\textbf{prediction}} + \underbrace{\overbrace{\mathbf{G}[k+1]}^{\text{Kalman Gain Matrix}} \; \overbrace{\tilde{\mathbf{z}}[k+1|k]}^{\text{Innovations}}}_{\textbf{correction}}}$$

for $k = 1, 2, \ldots$ where $\hat{\mathbf{x}}[0|0] = \mathbf{m}_x[0]$ and $\tilde{\mathbf{z}}[k+1|k] = \mathbf{z}[k+1] - \hat{\mathbf{z}}[k+1|k]$ is the innovations process.

— $\mathbf{G}[k+1]$ is called the Kalman gain matrix.

— The Kalman gain matrix can be carried out by

$$\mathbf{G}[k+1] = \mathbf{P}[k+1|k]\mathbf{H}^H[k+1] \cdot \Big(\mathbf{H}[k+1]\mathbf{P}[k+1|k]\mathbf{H}^H[k+1] + \mathbf{R}[k+1]\Big)^{-1},$$

where $\mathbf{P}[k+1|k]$ is as in (1) and

$$\begin{aligned}
&\mathbf{P}[k+1|k+1] \\
&\quad = \Big(\mathbf{I} - \mathbf{G}[k+1]\mathbf{H}[k+1]\Big) \cdot \mathbf{P}[k+1|k] \cdot \Big(\mathbf{I} - \mathbf{G}[k+1]\mathbf{H}[k+1]\Big)^H \\
&\qquad + \mathbf{G}[k+1]\mathbf{R}[k+1]\mathbf{G}^H[k+1] \\
&\quad = \Big(\mathbf{I} - \mathbf{G}[k+1]\mathbf{H}[k+1]\Big) \cdot \mathbf{P}[k+1|k]
\end{aligned}$$

— The procedure of the Kalman filtering is illustrated in Fig. 1.

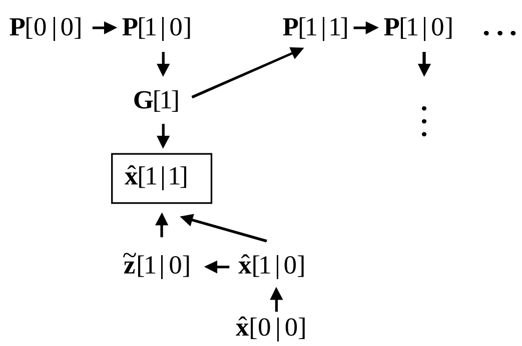

Figure 1: The procedure of the Kalman filtering.

Example: Example 9.2-2 in the textbook on page 583.

Consider the Gauss-Markov signal model

$$X[n] = 0.9X[n-1] + W[n], \quad n \geq 0,$$

with means equal to zero and $\sigma_W^2 = 0.19$. Assume $X[-1] = 0$. The scalar measurement is

$$Y[n] = X[n] + V[n], \quad n \geq 0,$$

with $\sigma_V^2 = 1$.

References

[1] Jerry M. Mendel, *Lessons in Estimation Theory for Signal Processing, Communications, and Control*, Prentice Hall, 1995.