Kalman Filter

Control Theory, Lecture 10

by Sergei Savin

Spring 2023

Content

- Measurement

We can think of a *random variable* $\mathbf{v}$ as a sequence of values $\mathbf{v}_1$, $\mathbf{v}_2$, $\mathbf{v}_3$, ... - sampled from a distribution.

Mean $\bar{\mathbf{v}}$ of a random variable $\mathbf{v}$ is denoted as:

$$\bar{\mathbf{v}} = E[\mathbf{v}] \tag{1}$$

Mean has a number of properties:

$$E[\mathbf{a}] = \mathbf{a}, \qquad \mathbf{a} = \text{const} \tag{2}$$
$$E[\mathbf{x} + \mathbf{y}] = E[\mathbf{x}] + E[\mathbf{y}] \tag{3}$$
$$E[\alpha \mathbf{x}] = \alpha E[\mathbf{x}] \qquad \alpha = \text{const} \tag{4}$$
$$E[\mathbf{A}\mathbf{x}] = \mathbf{A} E[\mathbf{x}] \qquad \mathbf{A} = \text{const} \tag{5}$$

Autocovariance $\mathbf{V} = \mathbf{cov}(\mathbf{v}, \mathbf{v})$ of a random variable $\mathbf{v}$ is defined as:

$$\mathbf{cov}(\mathbf{v}, \mathbf{v}) = E[(\mathbf{v} - E[\mathbf{v}])(\mathbf{v} - E[\mathbf{v}])^\top] \tag{6}$$

To simplify notation in the following sections, we define $\mathbf{cov}(\mathbf{v}) = \mathbf{cov}(\mathbf{v}, \mathbf{v})$. For zero-mean process $E[\mathbf{v}] = 0$ the formula simplifies:

$$\mathbf{cov}(\mathbf{v}) = E[(\mathbf{v})(\mathbf{v})^\top] \tag{7}$$

Autocovariance has a number of properties:

$$\mathbf{cov}(\mathbf{a}) = \mathbf{0}, \qquad \mathbf{a} = \text{const} \tag{8}$$

$$\mathbf{cov}(\mathbf{x} + \mathbf{a}) = \mathbf{cov}(\mathbf{x}), \qquad \mathbf{a} = \text{const} \tag{9}$$

$$\mathbf{cov}(\alpha \mathbf{x}) = \alpha^2 \, \mathbf{cov}(\mathbf{x}) \tag{10}$$

Random variable, 3

A random variable $\mathbf{x}$ with Gaussian distribution can be fully described via its mean $\bar{\mathbf{x}}$ and covariance $\mathbf{X}$:

$$\mathbf{x} \sim \mathcal{N}(\bar{\mathbf{x}}, \mathbf{X}) \tag{11}$$

Let $\mathbf{x}$ be a random variable $\mathbf{x} \sim \mathcal{N}(\bar{\mathbf{x}}, \mathbf{X})$.

Given a constant matrix $\mathbf{M}$ we can define an affine transformation of $\mathbf{x}$:

$$\mathbf{y} = \mathbf{M}\mathbf{x} \tag{12}$$

We can find mean of $\mathbf{y}$:

$$E[\mathbf{y}] = E[\mathbf{M}\mathbf{x}] \tag{13}$$
$$E[\mathbf{y}] = \mathbf{M}E[\mathbf{x}] \tag{14}$$
$$E[\mathbf{y}] = \mathbf{M}\bar{\mathbf{x}} \tag{15}$$

If $\bar{\mathbf{x}} = E[\mathbf{x}] = 0$, then $\bar{\mathbf{y}} = E[\mathbf{y}] = 0$.

Assuming $\bar{\mathbf{x}} = E[\mathbf{x}] = 0$, we get $E[\mathbf{y}] = 0$; with that we can find autocovariance of $\mathbf{y}$:

$$\begin{aligned}
\mathbf{cov}(\mathbf{y}) &= E[(\mathbf{y} - E[\mathbf{y}])(\mathbf{y} - E[\mathbf{y}])^\top] = \\
&= E[\mathbf{y}\mathbf{y}^\top] = \\
&= E[(\mathbf{M}\mathbf{x})(\mathbf{M}\mathbf{x})^\top] = \\
&= E[\mathbf{M}\mathbf{x}\mathbf{x}^\top\mathbf{M}^\top] = \\
&= \mathbf{M}\mathbf{X}\mathbf{M}^\top
\end{aligned}$$

Assume the DT-LTI dynamics takes the form:

$$\mathbf{x}_{i+1} = \mathbf{A}\mathbf{x}_i + \mathbf{B}\mathbf{u}_i + \mathbf{w}_i, \tag{16}$$

where $\mathbf{w} \sim \mathcal{N}(0, \mathbf{Q})$ is *process noise* - random input with Gaussian distribution. We can propose an open-loop observer:

$$\hat{\mathbf{x}}_{i+1} = \mathbf{A}\hat{\mathbf{x}}_i + \mathbf{B}\mathbf{u}_i, \tag{17}$$

where $\hat{\mathbf{x}}$ is state estimate. We can find estimation error $\tilde{\mathbf{x}} = \mathbf{x}_i - \hat{\mathbf{x}}_i$ dynamic:

$$\tilde{\mathbf{x}}_{i+1} = \mathbf{A}\tilde{\mathbf{x}}_i + \mathbf{w}_i \tag{18}$$

Assume you could pick your initial state estimate $\hat{\mathbf{x}}_0$ such that your initial state estimation error $\tilde{\mathbf{x}}_0$ behaves as a random variable sampled from a Gaussian distribution $\tilde{\mathbf{x}}_0 \sim \mathcal{N}(0, \mathbf{P}_0)$.

Knowing mean $E[\tilde{\mathbf{x}}_i]$ we can compute $E[\tilde{\mathbf{x}}_{i+1}]$:

$$E[\tilde{\mathbf{x}}_{i+1}] = E[\mathbf{A}\tilde{\mathbf{x}}_i + \mathbf{w}_i] = \mathbf{A}E[\tilde{\mathbf{x}}_i] \tag{19}$$

Since $E[\tilde{\mathbf{x}}_0] = 0$, we can conclude that $E[\tilde{\mathbf{x}}_i] = 0, \ \forall i$.

State estimation error - covariance

Knowing autocovariance $\mathbf{P}_i$ we can compute $\mathbf{P}_{i+1}$:

$$\begin{aligned}\mathbf{P}_{i+1} &= E[\tilde{\mathbf{x}}_{i+1}\tilde{\mathbf{x}}_{i+1}^\top] = E[(\mathbf{A}\tilde{\mathbf{x}}_i + \mathbf{w}_i)(\mathbf{A}\tilde{\mathbf{x}}_i + \mathbf{w}_i)^\top] = \\ &= E[\mathbf{A}\tilde{\mathbf{x}}_i\tilde{\mathbf{x}}_i^\top\mathbf{A}^\top + \mathbf{A}\tilde{\mathbf{x}}_i\mathbf{w}_i^\top + \mathbf{w}_i\tilde{\mathbf{x}}_i^\top\mathbf{A}^\top + \mathbf{w}_i\mathbf{w}_i^\top]\end{aligned}$$

We can assume that random process $\mathbf{w}$ is uncorrelated with $\tilde{\mathbf{x}}$, meaning that $E[\tilde{\mathbf{x}}_i\mathbf{w}_i^\top] = E[\mathbf{w}_i\tilde{\mathbf{x}}_i^\top] = 0$:

$$\mathbf{P}_{i+1} = E[\mathbf{A}\tilde{\mathbf{x}}_i\tilde{\mathbf{x}}_i^\top\mathbf{A}^\top + \mathbf{w}_i\mathbf{w}_i^\top] = \mathbf{A}\mathbf{P}_i\mathbf{A}^\top + \mathbf{Q}$$

Previously, we computed dynamics of mean and covariance of state estimation error for the case of open-loop observer. But, a stable observer with feedback is obviously preferable. We start by introducing a measurement model:

$$\mathbf{y}_i = \mathbf{H}\mathbf{x}_i + \mathbf{v}_i \tag{20}$$

where $\mathbf{H}$ is a measurement matrix, $\mathbf{y}_i$ is measured output and $\mathbf{v}_i$ is a measurement noise sampled from a Gaussian distribution $\mathbf{v}_i \sim \mathcal{N}(0, \mathbf{R})$.

We can propose the following modification to the observer:

$$\begin{cases} \hat{\mathbf{x}}_{i+1}^{-} = \mathbf{A}\hat{\mathbf{x}}_i + \mathbf{B}\mathbf{u}_i, \\ \hat{\mathbf{x}}_{i+1} = \hat{\mathbf{x}}_{i+1}^{-} + \mathbf{L}_i(\mathbf{y}_i - \mathbf{H}\hat{\mathbf{x}}_{i+1}^{-}) \end{cases} \tag{21}$$

where $\hat{\mathbf{x}}_{i+1}^{-}$ is an *a priori* estimate. We can re-write all this in terms of state estimation error, defining $\tilde{\mathbf{x}}_{i+1}^{-} = \mathbf{x}_{i+1} - \hat{\mathbf{x}}_{i+1}^{-}$:

$$\begin{cases} \tilde{\mathbf{x}}_{i+1}^{-} = \mathbf{A}\tilde{\mathbf{x}}_i + \mathbf{w}_i, \\ \tilde{\mathbf{x}}_{i+1} = \tilde{\mathbf{x}}_{i+1}^{-} - \mathbf{L}_i\mathbf{H}\tilde{\mathbf{x}}_{i+1}^{-} + \mathbf{v}_i \end{cases} \tag{22}$$

We can compute mean dynamics (*propagation*):

$$\begin{cases} E[\tilde{\mathbf{x}}_{i+1}^-] = \mathbf{A} E[\tilde{\mathbf{x}}_i], \\ E[\tilde{\mathbf{x}}_{i+1}] = (\mathbf{I} - \mathbf{L}_i \mathbf{H}) E[\tilde{\mathbf{x}}_{i+1}^-] \end{cases} \tag{23}$$

Since $E[\tilde{\mathbf{x}}_0] = 0$, then $E[\tilde{\mathbf{x}}_1^-] = 0$, and then $E[\tilde{\mathbf{x}}_1] = 0$, and the same for $E[\tilde{\mathbf{x}}_i] = 0$, $E[\tilde{\mathbf{x}}_i^-] = 0$.

We can compute autocovariance dynamics (propagation).
Below is *a priori* estimation error covariance:

$$\begin{aligned}\mathbf{P}_{i+1}^{-} &= E[\tilde{\mathbf{x}}_{i+1}^{-}(\tilde{\mathbf{x}}_{i+1}^{-})^{\top}] = \\ &= \mathbf{A}\mathbf{P}_i\mathbf{A}^{\top} + \mathbf{Q},\end{aligned}$$

With that, we can find *a posteriori* estimation error covariance:

$$\begin{aligned}E[\tilde{\mathbf{x}}_{i+1}\tilde{\mathbf{x}}_{i+1}^{\top}] = E[&(\mathbf{I}-\mathbf{L}_i\mathbf{H})\tilde{\mathbf{x}}_{i+1}^{-}(\tilde{\mathbf{x}}_{i+1}^{-})^{\top}(\mathbf{I}-\mathbf{L}_i\mathbf{H})^{\top}+ \\ &+(\mathbf{I}-\mathbf{L}_i\mathbf{H})\tilde{\mathbf{x}}_{i+1}^{-}\mathbf{v}_i^{\top}+\mathbf{v}_i(\tilde{\mathbf{x}}_{i+1}^{-})^{\top}(\mathbf{I}-\mathbf{L}_i\mathbf{H})^{\top}+\mathbf{v}_i\mathbf{v}_i^{\top}]\end{aligned}$$

Assuming that $\tilde{\mathbf{x}}_{i+1}^{-}$ and $\mathbf{v}_i$ are uncorrelated, we get $E[(\mathbf{I}-\mathbf{L}_i\mathbf{H})\tilde{\mathbf{x}}_{i+1}^{-}\mathbf{v}_i^{\top}] = 0$ and $E[\mathbf{v}_i(\tilde{\mathbf{x}}_{i+1}^{-})^{\top}(\mathbf{I}-\mathbf{L}_i\mathbf{H})^{\top} = 0]$.
With that we simplify:

$$E[\tilde{\mathbf{x}}_{i+1}\tilde{\mathbf{x}}_{i+1}^{\top}] = (\mathbf{I}-\mathbf{L}_i\mathbf{H})\mathbf{P}_{i+1}^{-}(\mathbf{I}-\mathbf{L}_i\mathbf{H})^{\top}+\mathbf{R}$$

How do we pick $\mathbf{L}_i$? We can do it "the same way" as we did with LQR:

$$\mathbf{L}_i = \mathbf{P}_i \mathbf{H}^\top \mathbf{R}^{-1} \tag{24}$$

In practice, it can be better to compute $\mathbf{L}_i$ before we compute $\mathbf{P}_i$. The following allows us to compute $\mathbf{L}_i$ based on $\mathbf{P}_{i-1}$:

$$\mathbf{L}_i = \mathbf{P}_{i-1} \mathbf{H}^\top (\mathbf{H} \mathbf{P}_{i-1} \mathbf{H}^\top + \mathbf{R})^{-1} \tag{25}$$

Further reading

- Simon, D., 2006. Optimal state estimation: Kalman, H infinity, and nonlinear approaches. John Wiley & Sons.

Lecture slides are available via Github, links are on Moodle

You can help improve these slides at:
github.com/SergeiSa/Control-Theory-Slides-Spring-2023

Appendix A

Given a constant vector $\mathbf{c}$ and a constant matrix $\mathbf{M}$ we can define an affine transformation of $\mathbf{x}$:

$$\mathbf{y} = \mathbf{M}\mathbf{x} + \mathbf{c} \tag{26}$$

We can find mean of $\mathbf{y}$:

$$E[\mathbf{y}] = E[\mathbf{M}\mathbf{x} + \mathbf{c}] \tag{27}$$

$$E[\mathbf{y}] = \mathbf{M}E[\mathbf{x}] + \mathbf{c} \tag{28}$$

$$E[\mathbf{y}] = \mathbf{M}\bar{\mathbf{x}} + \mathbf{c} \tag{29}$$

Autocovariance with zero mean

Assuming $E[\mathbf{x}] = 0$, we can find covariance of $\mathbf{y}$:

$$
\begin{aligned}
\mathbf{cov}(\mathbf{y}) = E[(\mathbf{y} - E[\mathbf{y}])(\mathbf{y} - E[\mathbf{y}])^\top] = \\
= E[\mathbf{y}\mathbf{y}^\top + E[\mathbf{y}]E[\mathbf{y}]^\top - \mathbf{y}E[\mathbf{y}]^\top - E[\mathbf{y}]\mathbf{y}^\top] = \\
= E[\mathbf{y}\mathbf{y}^\top + \bar{\mathbf{y}}\bar{\mathbf{y}}^\top - \mathbf{y}\bar{\mathbf{y}}^\top - \bar{\mathbf{y}}\mathbf{y}]^\top] = \\
= E[\mathbf{y}\mathbf{y}^\top] + \bar{\mathbf{y}}\bar{\mathbf{y}}^\top - E[\mathbf{y}]\bar{\mathbf{y}}^\top - \bar{\mathbf{y}}E[\mathbf{y}]^\top = \\
= E[\mathbf{y}\mathbf{y}^\top] + \bar{\mathbf{y}}\bar{\mathbf{y}}^\top - \bar{\mathbf{y}}\bar{\mathbf{y}}^\top - \bar{\mathbf{y}}\bar{\mathbf{y}}^\top \\
= E[(\mathbf{M}\mathbf{x} + \mathbf{c})(\mathbf{M}\mathbf{x} + \mathbf{c})^\top] - \mathbf{c}\mathbf{c}^\top \\
= E[\mathbf{M}\mathbf{x}\mathbf{x}^\top\mathbf{M}^\top + \mathbf{c}\mathbf{c}^\top + \mathbf{M}\mathbf{x}\mathbf{c}^\top + \mathbf{c}\mathbf{x}^\top\mathbf{M}^\top] - \mathbf{c}\mathbf{c}^\top = \\
= \mathbf{M}\mathbf{X}\mathbf{M}^\top + \mathbf{M}\bar{\mathbf{x}}\mathbf{c}^\top + \mathbf{c}\bar{\mathbf{x}}^\top\mathbf{M}^\top = \\
= \mathbf{M}\mathbf{X}\mathbf{M}^\top
\end{aligned}
$$

Without this assumption, the covariance of $\mathbf{y}$ is a little more complicated:

$$\begin{aligned}
\mathbf{cov}(\mathbf{y}) = E[(\mathbf{y} - E[\mathbf{y}])(\mathbf{y} - E[\mathbf{y}])^\top] = \\
= E[\mathbf{y}\mathbf{y}^\top + E[\mathbf{y}]E[\mathbf{y}]^\top - \mathbf{y}E[\mathbf{y}]^\top - E[\mathbf{y}]\mathbf{y}^\top] = \\
= E[\mathbf{y}\mathbf{y}^\top + \bar{\mathbf{y}}\bar{\mathbf{y}}^\top - \mathbf{y}\bar{\mathbf{y}}^\top - \bar{\mathbf{y}}\mathbf{y}]^\top] = \\
= E[\mathbf{y}\mathbf{y}^\top] + \bar{\mathbf{y}}\bar{\mathbf{y}}^\top - E[\mathbf{y}]\bar{\mathbf{y}}^\top - \bar{\mathbf{y}}E[\mathbf{y}]^\top = \\
= E[\mathbf{y}\mathbf{y}^\top] + \bar{\mathbf{y}}\bar{\mathbf{y}}^\top - \bar{\mathbf{y}}\bar{\mathbf{y}}^\top - \bar{\mathbf{y}}\bar{\mathbf{y}}^\top \\
= E[(\mathbf{M}\mathbf{x} + \mathbf{c})(\mathbf{M}\mathbf{x} + \mathbf{c})^\top] - (\mathbf{M}\bar{\mathbf{x}} + \mathbf{c})(\mathbf{M}\bar{\mathbf{x}} + \mathbf{c})^\top \\
= E[\mathbf{M}\mathbf{x}\mathbf{x}^\top\mathbf{M}^\top + \mathbf{c}\mathbf{c}^\top + \mathbf{M}\mathbf{x}\mathbf{c}^\top + \mathbf{c}\mathbf{x}^\top\mathbf{M}^\top] - \\
-(\mathbf{M}\bar{\mathbf{x}}\bar{\mathbf{x}}^\top\mathbf{M}^\top + \mathbf{M}\bar{\mathbf{x}}\mathbf{c}^\top + \mathbf{c}\bar{\mathbf{x}}^\top\mathbf{M}^\top + \mathbf{c}\mathbf{c}^\top) = \\
= \mathbf{M}\mathbf{X}\mathbf{M}^\top + \mathbf{c}\mathbf{c}^\top + \mathbf{M}\bar{\mathbf{x}}\mathbf{c}^\top + \mathbf{c}\bar{\mathbf{x}}^\top\mathbf{M}^\top - \\
-(\mathbf{M}\bar{\mathbf{x}}\bar{\mathbf{x}}^\top\mathbf{M}^\top + \mathbf{M}\bar{\mathbf{x}}\mathbf{c}^\top + \mathbf{c}\bar{\mathbf{x}}^\top\mathbf{M}^\top + \mathbf{c}\mathbf{c}^\top) = \\
= \mathbf{M}\mathbf{X}\mathbf{M}^\top - \mathbf{M}\bar{\mathbf{x}}\bar{\mathbf{x}}^\top\mathbf{M}^\top
\end{aligned}$$