

Kalman Filter

Control Theory, Lecture 10

by Sergei Savin

Spring 2025

- Random variables, mean, autocovariance
- Models with uncertainty, observer
 - ▶ Process noise, measurement noise
 - ▶ Open loop observer
 - ▶ Estimation error autocovariance propagation
 - ▶ Kalman filter
- Kalman filter gain

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

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

$$\bar{\mathbf{v}} = E[\mathbf{v}] \quad (1)$$

Mean has a number of properties:

$$E[\mathbf{a}] = \mathbf{a}, \quad \mathbf{a} = \text{const} \quad (2)$$

$$E[\mathbf{x} + \mathbf{y}] = E[\mathbf{x}] + E[\mathbf{y}] \quad (3)$$

$$E[\alpha \mathbf{x}] = \alpha E[\mathbf{x}] \quad \alpha = \text{const} \in \mathbb{R} \quad (4)$$

$$E[\mathbf{A}\mathbf{x}] = \mathbf{A}E[\mathbf{x}] \quad \mathbf{A} = \text{const} \quad (5)$$

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

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

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:

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

Autocovariance has a number of properties:

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

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

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

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

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

MEAN OF A LINEAR TRANSFORM

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} :

$$\mathbf{y} = \mathbf{M}\mathbf{x} \quad (12)$$

We can find mean of \mathbf{y} :

$$E[\mathbf{y}] = E[\mathbf{M}\mathbf{x}] \quad (13)$$

$$E[\mathbf{y}] = \mathbf{M}E[\mathbf{x}] \quad (14)$$

$$E[\mathbf{y}] = \mathbf{M}\bar{\mathbf{x}} \quad (15)$$

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

AUTO-COVARIANCE OVER LINEAR TRANSFORM

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:

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

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

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

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

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

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 $E[\tilde{\mathbf{x}}_{i+1}]$:

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

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:

$$\mathbf{y}_i = \mathbf{H}\mathbf{x}_i + \mathbf{v}_i \quad (20)$$

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 $\mathbf{v}_i \sim \mathcal{N}(0, \mathbf{R})$, where $\mathbf{R} \succ 0$.

We can propose the following modification to the observer:

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

where $\hat{\mathbf{x}}_{i+1}^-$ is an *a priori* estimate. We can re-write the last equation as $\hat{\mathbf{x}}_{i+1} = \hat{\mathbf{x}}_{i+1}^- + \mathbf{L}_i(\mathbf{H}\mathbf{x}_i - \mathbf{H}\hat{\mathbf{x}}_{i+1}^- + \mathbf{v}_i)$.

We can re-write all this in terms of state estimation error, defining $\tilde{\mathbf{x}}_{i+1}^- = \mathbf{x}_{i+1} - \hat{\mathbf{x}}_{i+1}^-$. For the last eq. in (21), we subtract \mathbf{x}_{i+1} from both sides:

$$\hat{\mathbf{x}}_{i+1} - \mathbf{x}_{i+1} = \hat{\mathbf{x}}_{i+1}^- - \mathbf{x}_{i+1} + \mathbf{L}_i(\mathbf{H}\mathbf{x}_i - \mathbf{H}\hat{\mathbf{x}}_{i+1}^- + \mathbf{v}_i) \quad (22)$$

and flip the sign:

$$\begin{cases} \tilde{\mathbf{x}}_{i+1}^- = \mathbf{A}\tilde{\mathbf{x}}_i + \mathbf{w}_i, \\ \tilde{\mathbf{x}}_{i+1} = (\mathbf{I} - \mathbf{L}_i\mathbf{H})\tilde{\mathbf{x}}_{i+1}^- + \mathbf{L}_i\mathbf{v}_i \end{cases} \quad (23)$$

We can compute estimation error mean dynamics (*propagation*):

$$E[\tilde{\mathbf{x}}_{i+1}^-] = E[\mathbf{A}\tilde{\mathbf{x}}_i + \mathbf{w}_i] = E[\mathbf{A}\tilde{\mathbf{x}}_i] + E[\mathbf{w}_i] = \mathbf{A}E[\tilde{\mathbf{x}}_i].$$

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

So, we obtain the following mean dynamics:

$$\begin{cases} E[\tilde{\mathbf{x}}_{i+1}^-] = \mathbf{A}E[\tilde{\mathbf{x}}_i], \\ E[\tilde{\mathbf{x}}_{i+1}] = (\mathbf{I} - \mathbf{L}_i\mathbf{H})E[\tilde{\mathbf{x}}_{i+1}^-] \end{cases} \quad (24)$$

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:

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

Reminder: $E[\mathbf{w}_i\mathbf{w}_i^\top] = \mathbf{Q}$ since $\mathbf{w} \sim \mathcal{N}(0, \mathbf{Q})$, $E[\tilde{\mathbf{x}}_i\mathbf{w}_i^\top] = 0$ since the two variables are independent, and $E[\tilde{\mathbf{x}}_i\tilde{\mathbf{x}}_i^\top] = \mathbf{P}_i$ by definition.

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

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

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:

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

Kalman filter gain

Before discussing how we can propose Kalman filter gain, we need two mathematical facts. First, inner and outer product:

$$\mathbf{x}^\top \mathbf{x} = \text{tr}(\mathbf{x}\mathbf{x}^\top) \quad (25)$$

where $\text{tr}(\cdot)$ is a trace operation.

Example:

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \quad \mathbf{x}^\top \mathbf{x} = x_1^2 + x_2^2 + x_3^2,$$
$$\mathbf{x}\mathbf{x}^\top = \begin{bmatrix} x_1^2 & x_1x_2 & x_1x_3 \\ x_2x_1 & x_2^2 & x_2x_3 \\ x_3x_1 & x_3x_2 & x_3^2 \end{bmatrix}, \quad \text{tr}(\mathbf{x}\mathbf{x}^\top) = x_1^2 + x_2^2 + x_3^2.$$

Second, derivatives of a trace:

$$\frac{\partial(\text{tr}(\mathbf{A}\mathbf{X}))}{\partial\mathbf{X}} = \frac{\partial(\text{tr}(\mathbf{X}\mathbf{A}))}{\partial\mathbf{X}} = \mathbf{A} \quad (26)$$

$$\frac{\partial(\text{tr}(\mathbf{A}\mathbf{X}^\top))}{\partial\mathbf{X}} = \frac{\partial(\text{tr}(\mathbf{X}^\top\mathbf{A}))}{\partial\mathbf{X}} = \mathbf{A}^\top \quad (27)$$

$$\frac{\partial(\text{tr}(\mathbf{A}\mathbf{X}))}{\partial\mathbf{X}^\top} = \frac{\partial(\text{tr}(\mathbf{X}\mathbf{A}))}{\partial\mathbf{X}^\top} = \mathbf{A}^\top \quad (28)$$

$$\frac{\partial(\text{tr}(\mathbf{X}^\top\mathbf{A}\mathbf{X}))}{\partial\mathbf{X}} = \mathbf{X}^\top(\mathbf{A} + \mathbf{A}^\top) \quad (29)$$

$$\frac{\partial(\text{tr}(\mathbf{X}\mathbf{A}\mathbf{X}^\top))}{\partial\mathbf{X}} = (\mathbf{A} + \mathbf{A}^\top)\mathbf{X}^\top \quad (30)$$

$$\frac{\partial(\text{tr}(\mathbf{X}^\top\mathbf{A}\mathbf{X}))}{\partial\mathbf{X}^\top} = (\mathbf{A} + \mathbf{A}^\top)\mathbf{X} \quad (31)$$

$$\frac{\partial(\text{tr}(\mathbf{X}\mathbf{A}\mathbf{X}^\top))}{\partial\mathbf{X}^\top} = \mathbf{X}(\mathbf{A} + \mathbf{A}^\top) \quad (32)$$

Here we will attempt to derive optimal Kalman gain \mathbf{L}_i for the i -th step, such that the following cost function is minimized:

$$J = E \left[\sum \tilde{x}_{i+1}^2 \right] \quad (33)$$

meaning that we minimize mean value of the square of the estimation error. We also know that as long as estimation error on the $i + 1$ -th step has zero mean (as a random variable), covariance takes the following form: $\mathbf{P}_{i+1} = E[\tilde{\mathbf{x}}_{i+1}\tilde{\mathbf{x}}_{i+1}^\top]$. Its trace gives us the cost function J :

$$\begin{aligned} J &= E \left[\text{tr}(\tilde{\mathbf{x}}_{i+1}\tilde{\mathbf{x}}_{i+1}^\top) \right] = \text{tr}(E \left[\tilde{\mathbf{x}}_{i+1}\tilde{\mathbf{x}}_{i+1}^\top \right]) = \text{tr}(\mathbf{P}_{i+1}) = \\ &= \text{tr}((\mathbf{I} - \mathbf{L}_i\mathbf{H})\mathbf{P}_{i+1}^-(\mathbf{I} - \mathbf{L}_i\mathbf{H})^\top + \mathbf{L}_i\mathbf{R}\mathbf{L}_i^\top) = \\ &= \text{tr}(\mathbf{P}_{i+1}^- - \mathbf{L}_i\mathbf{H}\mathbf{P}_{i+1}^- - \mathbf{P}_{i+1}^-\mathbf{H}^\top\mathbf{L}_i^\top + \mathbf{L}_i(\mathbf{H}\mathbf{P}_{i+1}^-\mathbf{H}^\top + \mathbf{R})\mathbf{L}_i^\top) \end{aligned}$$

Next, we find derivative of J with respect to \mathbf{L}_i and set it to zero:

$$\begin{aligned}\frac{\partial J}{\partial \mathbf{L}_i} &= -\mathbf{H}\mathbf{P}_{i+1}^- - (\mathbf{P}_{i+1}^- \mathbf{H}^\top)^\top + 2(\mathbf{H}\mathbf{P}_{i+1}^- \mathbf{H}^\top + \mathbf{R})\mathbf{L}_i^\top = 0 \\ &\quad -2\mathbf{H}\mathbf{P}_{i+1}^- + 2(\mathbf{H}\mathbf{P}_{i+1}^- \mathbf{H}^\top + \mathbf{R})\mathbf{L}_i^\top = 0 \\ &\quad \mathbf{L}_i(\mathbf{H}\mathbf{P}_{i+1}^- \mathbf{H}^\top + \mathbf{R}) = \mathbf{P}_{i+1}^- \mathbf{H}^\top \\ &\quad \mathbf{L}_i = \mathbf{P}_{i+1}^- \mathbf{H}^\top (\mathbf{H}\mathbf{P}_{i+1}^- \mathbf{H}^\top + \mathbf{R})^{-1}\end{aligned}$$

So, the Kalman gain can be optimally chosen as

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

There are alternative but equivalent ways to pick \mathbf{L}_i . We can do it "the same way" as we did with LQR:

$$\mathbf{L}_i = \mathbf{P}_{i+1} \mathbf{H}^\top \mathbf{R}^{-1} \quad (34)$$

The equivalence of this formula to the earlier one will be shown in the Appendix B.

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

Lecture slides are available via Github:

github.com/SergeiSa/Control-Theory-2025



Appendix A

MEAN OF AN AFFINE TRANSFORM

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

$$\mathbf{y} = \mathbf{M}\mathbf{x} + \mathbf{c} \quad (35)$$

We can find mean of \mathbf{y} :

$$E[\mathbf{y}] = E[\mathbf{M}\mathbf{x} + \mathbf{c}] \quad (36)$$

$$E[\mathbf{y}] = \mathbf{M}E[\mathbf{x}] + \mathbf{c} \quad (37)$$

$$E[\mathbf{y}] = \mathbf{M}\bar{\mathbf{x}} + \mathbf{c} \quad (38)$$

AUTOCOVARANCE WITH ZERO MEAN

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

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

AUTO-COVARIANCE OVER AFFINE TRANSFORM

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

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

Appendix B

Given observer gain $\mathbf{L}_i = \mathbf{P}_{i+1}\mathbf{H}^\top\mathbf{R}^{-1}$ and autocovariance propagation $\mathbf{P}_{i+1} = (\mathbf{I} - \mathbf{L}_i\mathbf{H})\mathbf{P}_{i+1}^-(\mathbf{I} - \mathbf{L}_i\mathbf{H})^\top + \mathbf{L}_i\mathbf{R}\mathbf{L}_i^\top$, we can derive expression for \mathbf{L}_i as a function of \mathbf{P}_{i+1}^- :

$$\mathbf{L}_i\mathbf{R} = \mathbf{P}_{i+1}\mathbf{H}^\top \quad (39)$$

$$\mathbf{L}_i\mathbf{R}\mathbf{L}_i^\top = \mathbf{P}_{i+1}\mathbf{H}^\top\mathbf{L}_i^\top \quad (40)$$

$$\mathbf{P}_{i+1} = (\mathbf{I} - \mathbf{L}_i\mathbf{H})\mathbf{P}_{i+1}^-(\mathbf{I} - \mathbf{L}_i\mathbf{H})^\top + \mathbf{P}_{i+1}\mathbf{H}^\top\mathbf{L}_i^\top \quad (41)$$

$$\mathbf{P}_{i+1}(\mathbf{I} - \mathbf{H}^\top\mathbf{L}_i^\top) = (\mathbf{I} - \mathbf{L}_i\mathbf{H})\mathbf{P}_{i+1}^-(\mathbf{I} - \mathbf{L}_i\mathbf{H})^\top \quad (42)$$

Assuming that $\det(\mathbf{I} - \mathbf{L}_i\mathbf{H})^\top \neq 0$, we can multiply on the right by $(\mathbf{I} - \mathbf{L}_i\mathbf{H})^{-\top}$:

$$\mathbf{P}_{i+1} = (\mathbf{I} - \mathbf{L}_i\mathbf{H})\mathbf{P}_{i+1}^- \quad (43)$$

$$\mathbf{P}_{i+1} = (\mathbf{I} - \mathbf{L}_i \mathbf{H}) \mathbf{P}_{i+1}^- \quad (44)$$

$$\mathbf{P}_{i+1} \mathbf{H}^\top \mathbf{R}^{-1} = (\mathbf{I} - \mathbf{L}_i \mathbf{H}) \mathbf{P}_{i+1}^- \mathbf{H}^\top \mathbf{R}^{-1} \quad (45)$$

$$\mathbf{L}_i = (\mathbf{I} - \mathbf{L}_i \mathbf{H}) \mathbf{P}_{i+1}^- \mathbf{H}^\top \mathbf{R}^{-1} \quad (46)$$

$$\mathbf{L}_i \mathbf{R} = (\mathbf{I} - \mathbf{L}_i \mathbf{H}) \mathbf{P}_{i+1}^- \mathbf{H}^\top \quad (47)$$

$$\mathbf{L}_i \mathbf{R} + \mathbf{L}_i \mathbf{H} \mathbf{P}_{i+1}^- \mathbf{H}^\top = \mathbf{P}_{i+1}^- \mathbf{H}^\top \quad (48)$$

$$\mathbf{L}_i (\mathbf{R} + \mathbf{H} \mathbf{P}_{i+1}^- \mathbf{H}^\top) = \mathbf{P}_{i+1}^- \mathbf{H}^\top \quad (49)$$

$$\mathbf{L}_i = \mathbf{P}_{i+1}^- \mathbf{H}^\top (\mathbf{R} + \mathbf{H} \mathbf{P}_{i+1}^- \mathbf{H}^\top)^{-1}. \quad \square \quad (50)$$