Kalman Filter Control Theory, Lecture 10

by Sergei Savin

Spring 2025

CONTENT

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

RANDOM VARIABLE, 1

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:

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

$$E[\mathbf{x} + \mathbf{y}] = E[\mathbf{x}] + E[\mathbf{y}] \tag{3}$$

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

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

RANDOM VARIABLE, 2

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

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

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

$$cov(x + a) = cov(x),$$
 $a = const$ (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} :

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

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

We can find mean of 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 $\bar{\mathbf{y}} = E[\mathbf{y}] = 0$.

AUTOCOVARIANCE 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} :

$$\mathbf{cov}(\mathbf{y}) = E[(\mathbf{y} - E[\mathbf{y}])(\mathbf{y} - E[\mathbf{y}])^{\top}] =$$

$$= E[\mathbf{y}\mathbf{y}^{\top}] =$$

$$= E[(\mathbf{M}\mathbf{x})(\mathbf{M}\mathbf{x})^{\top}] =$$

$$= E[\mathbf{M}\mathbf{x}\mathbf{x}^{\top}\mathbf{M}^{\top}] =$$

$$= \mathbf{M}\mathbf{X}\mathbf{M}^{\top}$$

STATE ESTIMATION ERROR - DYNAMICS

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 and $\mathbf{Q} \succeq 0$ (meaning that it is positive semidefinite). We can propose an open-loop observer:

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

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

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

STATE ESTIMATION ERROR - MEAN

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

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

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

$$= E[\mathbf{A}\tilde{\mathbf{x}}_i\tilde{\mathbf{x}}_i^{\top}\mathbf{A}^{\top} + \mathbf{A}\tilde{\mathbf{x}}_i\mathbf{w}_i^{\top} + \mathbf{w}_i\tilde{\mathbf{x}}_i^{\top}\mathbf{A}^{\top} + \mathbf{w}_i\mathbf{w}_i^{\top}]$$

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

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

CLOSED-LOOP OBSERVER, 1

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 **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 $\mathbf{R} \succ 0$.

CLOSED-LOOP OBSERVER, 2

We can propose the following modification to the observer:

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

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

and flip the sign:

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

CLOSED-LOOP OBSERVER - MEAN DYNAMICS

We can compute estimation error mean dynamics (propagation):

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

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

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

So, we obtain the following mean dynamics:

$$\begin{cases}
E[\tilde{\mathbf{x}}_{i+1}^{-}] = \mathbf{A}E[\tilde{\mathbf{x}}_{i}], \\
E[\tilde{\mathbf{x}}_{i+1}] = (\mathbf{I} - \mathbf{L}_{i}\mathbf{H})E[\tilde{\mathbf{x}}_{i+1}^{-}]
\end{cases}$$
(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 $E[\tilde{\mathbf{x}}_i] = 0$, $E[\tilde{\mathbf{x}}_i^-] = 0$.

CLOSED-LOOP OBSERVER - COVARIANCE DYNAMICS

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

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

$$= E[(\mathbf{A}\tilde{\mathbf{x}}_{i} + \mathbf{w}_{i})(\mathbf{A}\tilde{\mathbf{x}}_{i} + \mathbf{w}_{i})^{\top}] =$$

$$= \mathbf{A}\mathbf{P}_{i}\mathbf{A}^{\top} + \mathbf{Q}.$$

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.

CLOSED-LOOP OBSERVER - COVARIANCE DYNAMICS

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

PRELIMINARIES, 1

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

$$\mathbf{x}^{\top}\mathbf{x} = \operatorname{tr}(\mathbf{x}\mathbf{x}^{\top}) \tag{25}$$

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

Example:

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \quad \mathbf{x}^{\top} \mathbf{x} = x_1^2 + x_2^2 + x_3^2,$$

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

PRELIMINARIES, 2

Second, derivatives of a trace:

$$\frac{\partial(\operatorname{tr}(\mathbf{AX}))}{\partial\mathbf{X}} = \frac{\partial(\operatorname{tr}(\mathbf{XA}))}{\partial\mathbf{X}} = \mathbf{A}$$
 (26)

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

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

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

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

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

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

KALMAN GAIN, 1

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:

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

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:

$$J = E\left[\operatorname{tr}(\tilde{\mathbf{x}}_{i+1}\tilde{\mathbf{x}}_{i+1}^{\top})\right] = \operatorname{tr}(E\left[\tilde{\mathbf{x}}_{i+1}\tilde{\mathbf{x}}_{i+1}^{\top}\right]) = \operatorname{tr}(\mathbf{P}_{i+1}) =$$

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

$$\operatorname{tr}(\mathbf{P}_{i+1}^{-} - \mathbf{L}_{i}\mathbf{H}\mathbf{P}_{i+1}^{-} - \mathbf{P}_{i+1}^{-}\mathbf{H}^{\top}\mathbf{L}_{i}^{\top} + \mathbf{L}_{i}(\mathbf{H}\mathbf{P}_{i+1}^{-}\mathbf{H}^{\top} + \mathbf{R})\mathbf{L}_{i}^{\top})$$

KALMAN GAIN, 2

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

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

$$-2\mathbf{H}\mathbf{P}_{i+1}^{-} + 2(\mathbf{H}\mathbf{P}_{i+1}^{-}\mathbf{H}^{\top} + \mathbf{R})\mathbf{L}_i^{\top} = 0$$

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

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

So, the Kalman gain can be optimally chosen as $\mathbf{L}_i = \mathbf{P}_{i+1}^{-} \mathbf{H}^{\top} (\mathbf{H} \mathbf{P}_{i+1}^{-} \mathbf{H}^{\top} + \mathbf{R})^{-1}$.

KALMAN GAIN, 3

There are alternative but equivalent ways to pick L_i . We can do it "the same way" as we did with LQR:

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

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

FURTHER READING

■ 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} :

$$y = Mx + c (35)$$

We can find mean of y:

$$E[\mathbf{y}] = E[\mathbf{M}\mathbf{x} + \mathbf{c}] \tag{36}$$

$$E[\mathbf{y}] = \mathbf{M}E[\mathbf{x}] + \mathbf{c} \tag{37}$$

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

AUTOCOVARIANCE WITH ZERO MEAN

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

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

AUTOCOVARIANCE OVER AFFINE TRANSFORM

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

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

Appendix B

OBSERVER GAIN, 1

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}^{-} :

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

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

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

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

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

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

OBSERVER GAIN, 2

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

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

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

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

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

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

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