

# Rob 501 Handout: Grizzle

## Discrete-Time Kalman Filter and its Derivation

**Model** Linear time-varying discrete-time system with “white<sup>1</sup>” Gaussian noise

$$\begin{aligned}x_{k+1} &= A_k x_k + G_k w_k, \quad x_0 \text{ initial condition} \\ y_k &= C_k x_k + v_k\end{aligned}$$

$x \in \mathbb{R}^n$ ,  $w \in \mathbb{R}^p$ ,  $y \in \mathbb{R}^m$ ,  $v \in \mathbb{R}^m$ . Moreover, the random vectors  $x_0$ , and, for  $k \geq 0$ ,  $w_k$ ,  $v_k$  are all independent<sup>2</sup> Gaussian (normal) random vectors.

**Notation:**  $\delta_{kl} = 1 \Leftrightarrow k = l$  (and  $\delta_{kl} = 0$ ,  $k \neq l$ )

### Precise Assumptions on the Random Variables

- For all  $k \geq 0$ ,  $l \geq 0$ ,  $x_0$ ,  $w_k$ ,  $v_l$  are jointly Gaussian.
- $w_k$  is a 0-mean white noise process:  $\mathcal{E}\{w_k\} = 0$ , and  $\text{cov}(w_k, w_l) = R_k \delta_{kl}$
- $v_k$  is a 0-mean white noise process:  $\mathcal{E}\{v_k\} = 0$ , and  $\text{cov}(v_k, v_l) = Q_k \delta_{kl}$
- Uncorrelated noise processes:  $\text{cov}(w_k, v_l) = 0$
- The initial condition  $x_0$  is uncorrelated with all other noise sequences.
- We denote the mean and covariance of  $x_0$  by

$$\bar{x}_0 = \mathcal{E}\{x_0\} \quad \text{and} \quad P_0 = \text{cov}(x_0) = \text{cov}(x_0, x_0) = \mathcal{E}\{(x_0 - \bar{x}_0)(x_0 - \bar{x}_0)^\top\}$$

---

<sup>1</sup>Recall that in white light, all frequencies are present. When only certain frequency components are present, you get “colored” light, such as blue light or red light. The term “white” noise means that if you compute the power spectral density of the noise random process, it is a constant, meaning that all frequency components are equally represented, just as in white light.

<sup>2</sup>Recall that for normal random variables, uncorrelated and independent are the same thing. This is one of several special properties of Gaussian random variables.

**Short-hand notation for the noise modeling assumptions:**

$$\text{COV} \left( \begin{bmatrix} w_k \\ v_k \\ x_0 \end{bmatrix}, \begin{bmatrix} w_l \\ v_l \\ x_0 \end{bmatrix} \right) = \begin{bmatrix} R_k \delta_{kl} & 0 & 0 \\ 0 & Q_k \delta_{kl} & 0 \\ 0 & 0 & P_0 \end{bmatrix}, \quad \delta_{kl} = \begin{bmatrix} 1, & k = l \\ 0, & k \neq l \end{bmatrix}$$

**Lemma (Propertis of  $x_k$  and  $y_k$  Coming from the Model)**

- For all  $k \geq 1$ ,  $x_k$  is a linear combination of  $x_0$  and  $w_0, \dots, w_{k-1}$ . In particular,  $x_k$  is uncorrelated with  $w_k$ .
- For all  $k \geq 1$ ,  $y_k$  is a linear combination of  $x_0, w_0, \dots, w_{k-1}$ , and  $v_0, \dots, v_k$ . In particular,  $y_k$  is uncorrelated with  $w_k$ .
- For all  $k \geq 0$ ,  $v_k$  is uncorrelated with  $x_k$ .

The proof is by induction using the recursive nature of the discrete-time model. We skip it. The reader can easily fill it in.

**Remark:** On the next page, we give (one form of) the discrete-time Kalman Filter. After that, we do the main elements of its derivation. There are many variations of the basic filter, all equivalent to the one we give, but some preferable over others for numerical reasons. Page 8 provides a version of the filter with the measurement update and prediction steps combined.

## The Kalman Filter

### Definition of Terms:

$$\hat{x}_{k|k} := \mathcal{E}\{x_k | y_0, \dots, y_k\}$$

$$P_{k|k} := \mathcal{E}\{(x_k - \hat{x}_{k|k})(x_k - \hat{x}_{k|k})^\top | y_0, \dots, y_k\}$$

$$\hat{x}_{k+1|k} := \mathcal{E}\{x_{k+1} | y_0, \dots, y_k\}$$

$$P_{k+1|k} := \mathcal{E}\{(x_{k+1} - \hat{x}_{k+1|k})(x_{k+1} - \hat{x}_{k+1|k})^\top | y_0, \dots, y_k\}$$

### Initial Conditions:

$$\hat{x}_{0|-1} := \bar{x}_0 = \mathcal{E}\{x_0\}, \quad \text{and} \quad P_{0|-1} := P_0 = \text{cov}(x_0)$$

For  $k \geq 0$

### Measurement Update Step:

$$K_k = P_{k|k-1} C_k^\top (C_k P_{k|k-1} C_k^\top + Q_k)^{-1}$$

(Kalman Gain)

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k (y_k - C_k \hat{x}_{k|k-1})$$

$$P_{k|k} = P_{k|k-1} - K_k C_k P_{k|k-1}$$

### Time Update or Prediction Step:

$$\hat{x}_{k+1|k} = A_k \hat{x}_{k|k}$$

$$P_{k+1|k} = A_k P_{k|k} A_k^\top + G_k R_k G_k^\top$$

**End of For Loop** (Just stated this way to emphasize the recursive nature of the filter)

## Preliminaries

**Measurements:** We collect all of the measurements at time  $k$  as

$$Y_k = (y_k, y_{k-1}, \dots, y_0).$$

Strictly speaking, we should be stacking them up into a column vector as we have done for all of our estimation problems, but notationally, it is more convenient to write them in a row. Also, it is more convenient to put the most recent measurement at the head of the list. We note that

$$Y_k = (y_k, Y_{k-1}).$$

Hence,

$$\hat{x}_{k|k} := \mathcal{E}\{x_k | Y_k\}$$

$$P_{k|k} := \mathcal{E}\{(x_k - \hat{x}_{k|k})(x_k - \hat{x}_{k|k})^\top | Y_k\}$$

mean and covariance of the conditional normal random vector  $x_k | Y_k$

$$\hat{x}_{k+1|k} := \mathcal{E}\{x_{k+1} | Y_k\}$$

$$P_{k+1|k} := \mathcal{E}\{(x_{k+1} - \hat{x}_{k+1|k})(x_{k+1} - \hat{x}_{k+1|k})^\top | Y_k\}$$

mean and covariance of the conditional normal random vector  $x_{k+1} | Y_k$

### Important Remarks:

- The conditional random vector  $x_k | Y_k$  is distributed  $N(\hat{x}_{k|k}, P_{k|k})$ .
- The conditional random vector  $x_{k+1} | Y_k$  is distributed  $N(\hat{x}_{k+1|k}, P_{k+1|k})$ .

## Filter Derivation Using Induction and Properties of Conditional Distributions of Gaussian Random Vectors

**Base step:** The initial conditions of the filter at time  $k = 0$ , namely

$$\hat{x}_{0|-1} := \bar{x}_0, \quad \text{and} \quad P_{0|-1} := P_0$$

**Induction step:** At time  $k \geq 0$ , we suppose that  $(\hat{x}_{k|k-1}, P_{k|k-1})$  are known, and we derive  $(\hat{x}_{k|k}, P_{k|k})$  and  $(\hat{x}_{k+1|k}, P_{k+1|k})$ .

**Key idea of the development:** We need to compute the distribution (or density) of the conditional random vector

$$x_k | Y_k = x_k | (y_k, Y_{k-1})$$

From HW 7, we learned **Fact 4**  $X | (Y, Z) = X | Z \mid Y | Z$ . From this we obtain

$$x_k | Y_k = x_k | (y_k, Y_{k-1}) = x_k | Y_{k-1} \mid y_k | Y_{k-1} \quad (*)$$

where we have identified

$$x_k \leftrightarrow X, \quad y_k \leftrightarrow Y, \quad \text{and} \quad Y_{k-1} \leftrightarrow Z.$$

Hence, if we can compute the distribution (or density) of

$$\begin{bmatrix} x_k \\ y_k \end{bmatrix} \mid Y_{k-1}$$

then we can apply **Fact 1** to obtain (\*).

The following calculations are aimed at doing just this.

**Measurement Update:** We seek to derive the filter equations on page 3. To begin, we have that  $y_k = C_k x_k + v_k$ . It follows by linearity that the conditional random variable  $y_k|Y_{k-1}$  is equal to

$$y_k|Y_{k-1} = C_k x_k|Y_{k-1} + v_k|Y_{k-1}.$$

From our assumptions on the noise,  $v_k$  is independent of both  $x_k$  and  $Y_{k-1}$ , and hence by **Fact 2**,  $v_k|Y_{k-1}$  is independent of the conditional random variable  $x_k|Y_{k-1}$ . Moreover, because  $v_k$  is independent of  $Y_{k-1}$ ,  $v_k|Y_{k-1} = v_k$ . Putting this together, we have that

$$y_k|Y_{k-1} = C_k x_k|Y_{k-1} + v_k,$$

and  $x_k|Y_{k-1}$  and  $v_k$  are independent. Hence

$$\begin{aligned} \hat{y}_{k|k-1} &:= \mathcal{E}\{y_k|Y_{k-1}\} \\ &= \mathcal{E}\{C_k x_k|Y_{k-1}\} + \mathcal{E}\{v_k\} \\ &= C_k \mathcal{E}\{x_k|Y_{k-1}\} + \mathcal{E}\{v_k\} \\ &= C_k \hat{x}_{k|k-1} + 0 \\ &= C_k \hat{x}_{k|k-1}. \end{aligned}$$

Moreover, the independence of  $x_k|Y_{k-1}$  and  $v_k$  with **Fact 3** yields

$$\text{cov}(y_k|Y_{k-1}, y_k|Y_{k-1}) = C_k P_{k|k-1} C_k^\top + Q_k.$$

We use independence again to obtain

$$\text{cov}(x_k|Y_{k-1}, y_k|Y_{k-1}) = \text{cov}(x_k|Y_{k-1}, C_k x_k|Y_{k-1}) = P_{k|k-1} C_k^\top.$$

With this information, we conclude that the vector

$$\begin{bmatrix} x_k \\ y_k \end{bmatrix} | Y_{k-1}$$

is jointly normally distributed, with mean and covariance

$$\begin{bmatrix} \hat{x}_{k|k-1} \\ C_k \hat{x}_{k|k-1} \end{bmatrix}, \begin{bmatrix} P_{k|k-1} & P_{k|k-1} C_k^\top \\ C_k P_{k|k-1} & C_k P_{k|k-1} C_k^\top + Q_k \end{bmatrix} \quad (*)$$

As discussed on the previous page, to compute the distribution of  $(x_k | Y_k)$ , we have from **Fact 4**

$$x_k | Y_k = x_k \Big| (y_k, Y_{k-1}) = x_k | Y_{k-1} \Big| y_k | Y_{k-1},$$

and thus applying **Fact 1** to  $(*)$  we compute the mean and covariance of  $x_k | Y_k = x_k | Y_{k-1} \Big| y_k | Y_{k-1}$  to be

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + P_{k|k-1} C_k^\top [C_k P_{k|k-1} C_k^\top + Q_k]^{-1} (y_k - C_k \hat{x}_{k|k-1})$$

$$P_{k|k} = P_{k|k-1} - P_{k|k-1} C_k^\top [C_k P_{k|k-1} C_k^\top + Q_k]^{-1} C_k P_{k|k-1}$$

**Remark:** We note that  $P_{k|k}$  is the Schur complement  $C_k P_{k|k-1} C_k^\top + Q_k$  in the covariance of

$$\begin{bmatrix} x_k \\ y_k \end{bmatrix} | Y_{k-1}$$

**Prediction or Time Update:** We seek to derive the filter equations on page 3. This time we use the state-variable model instead of the output model, namely

$$x_{k+1} = A_k x_k + G_k w_k,$$

and we are interested in the random vector

$$x_{k+1}|Y_k = A_k x_k|Y_k + G_k w_k|Y_k.$$

Because  $x_k$  and  $Y_k$  are both independent of  $w_k$ , by **Fact 2**,  $x_k|Y_k$  and  $w_k|Y_k$  are also independent. It follows that

$$\begin{aligned}\hat{x}_{k+1|k} &= \mathcal{E}\{x_{k+1}|Y_k\} \\ &= \mathcal{E}\{A_k x_k + G_k w_k|Y_k\} \\ &= A_k \mathcal{E}\{x_k|Y_k\} + G_k \mathcal{E}\{w_k|Y_k\} \\ &= A_k \hat{x}_{k|k} + G_k \mathcal{E}\{w_k\} \\ &= A_k \hat{x}_{k|k},\end{aligned}$$

where we have used  $w_k|Y_k = w_k$ , and  $\mathcal{E}\{w_k\} = 0$ .

Next, we use **Fact 3** and the conditional independence of the random vectors  $x_k|Y_k$  and  $w_k|Y_k$  to evaluate the covariance of  $x_{k+1}|Y_k$  as

$$P_{k+1|k} = A_k P_{k|k} A_k^\top + G_k R_k G_k^\top.$$

**That's the Proof Folks!**



## Combined Update Version of the Filter (often easier to implement)

Here, we will assume the model also has a *deterministic* input  $u_k$ , and thus

$$\begin{aligned}x_{k+1} &= A_k x_k + B_k u_k + G_k w_k \\ y_k &= C_k x_k + v_k,\end{aligned}$$

with the assumptions on the random vectors  $x_0$ ,  $w_k$  and  $v_k$  the same as before.

**Combined Filter:** The measurement-update step and time-update step of the Kalman Filter can be combined into a single step. The algorithm becomes:

**Initial Conditions:**

$$\hat{x}_{0|-1} := \bar{x}_0 = \mathcal{E}\{x_0\}, \quad \text{and} \quad P_{0|-1} := P_0 = \text{cov}(x_0)$$

**For**  $k \geq 0$

$$K_k = (P_{k|k-1} C_k^\top) [C_k P_{k|k-1} C_k^\top + Q_k]^{-1}$$

$$\hat{x}_{k+1|k} = A_k \hat{x}_{k|k-1} + B_k u_k + A_k K_k (y_k - C_k \hat{x}_{k|k-1})$$

$$P_{k+1|k} = A_k [P_{k|k-1} - K_k C_k P_{k|k-1}] A_k^\top + G_k R_k G_k^\top$$

**End of For Loop**

**Remark:** You do not have to start at  $k = 0$ . In MATLAB, it is often easier to begin with  $k = 1$ . In that case, the initial conditions are

$$\hat{x}_{1|0} := \bar{x}_0 = \mathcal{E}\{x_0\}, \quad \text{and} \quad P_{1|0} := P_0 = \text{cov}(x_0)$$

**Remark:**  $K_k C_k P_{k|k-1} = (P_{k|k-1} C_k^\top) [C_k P_{k|k-1} C_k^\top + Q_k]^{-1} C_k P_{k|k-1}$  is symmetric positive semi-definite and represents the value of the measurement in reducing the covariance of the state estimate, just as in the MVE.