EE 60039: Discrete-time Kalman Filter **Autumn 2023-24**

Discrete-time Kalman Filter

Instructor: Prof. Ashish R. Hota

In this note, we will formally derive the discrete-time Kalman filter for estimating the state of a linear dynamical system.

Consider the discrete-time linear dynamical system:

$$x_{k+1} = A_k x_k + w_k, \tag{1.1a}$$
$$y_k = C_k x_k + v_k, \tag{1.1b}$$

where $x_k \in \mathbb{R}^n$ is the state of the system, $w_k \in \mathbb{R}^n$ is the *process noise* affecting the system dynamics, $v_k \in \mathbb{R}^p$ is the *measurement noise* affecting the observation process, and $y_k \in \mathbb{R}^p$ is the observed output at time k.

We assume that the system dynamics (A_k), the observation matrix (C_k) and the measurements or outputs (y_k) are known to us for $k \geq 0$.

Our goal is to estimate the state of the system from current and past observations under the following *probabilistic* assumptions on the initial state and the noise affecting the system.

Estimation in the absence of noise

In the absence of noise, we often design a class of observers, called **Luenberger Observer**, to estimate the states. Let the system be time-invariant. We start with an initial estimate $\widehat{x}_0$ and recursively update our estimate as

$$\begin{aligned} \widehat{x}_{k+1} &= A\widehat{x}_k + L[y_k - C\widehat{x}_k] \\ \implies \widehat{e}_{k+1} &= x_{k+1} - \widehat{x}_{k+1} \\ &= Ax_k - A\widehat{x}_k - L[y_k - C\widehat{x}_k] \\ &= (A - LC)e_k. \end{aligned}$$

If L is chosen such that all the eigenvalues of $(A-LC)$ lie within the unit circle, then the estimation error asymptotically converges to 0. The pair (C, A) for which such a matrix L can be found is called detectable.

Review of LMSE and Orthogonal Projection

Before introducing the Kalman filter, we first establish certain intermediate results. The first one is regarding certain properties of the linear mean square error (LMSE) estimators, and the second one shows the correlation between certain random variables.

Recall that the LMSE estimator of a random vector $X \in \mathbb{R}^n$ given another random vector $Y \in \mathbb{R}^m$ is

$$\Pi_{\mathcal{L}(Y)}(X) := \widehat{\mathbb{E}}[X|Y] := \mathbb{E}[X] + \mathtt{Cov}(X,Y)^{\mathsf{T}}\mathtt{Cov}(Y)^{-1}[Y - \mathbb{E}[Y]], \tag{1.2}$$

where $\mathcal{L}(Y)$ is the set of all random variables that are linear combinations of random variables in Y and $\Pi_{\mathcal{L}(Y)}(X)$ is the projection of X on the closed linear space $\mathcal{L}(Y)$.

Lemma 1 (Properties of projection)**.** *Let $\mathcal{V}$ be a closed linear subspace of $\mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$. The projection has the following properties.*

1. *Linearity:* $\quad \Pi_{\mathcal{V}}(a_1X_1 + a_2X_2) = a_1\Pi_{\mathcal{V}}(X_1) + a_2\Pi_{\mathcal{V}}(X_2), \quad \forall a_1, a_2 \in \mathbb{R}.$
2. *Orthogonal subspace projection: Let $\mathcal{V}_1$ and $\mathcal{V}_2$ be two closed linear subspaces such that $\mathbb{E}[Z_1Z_2] = 0$ for every $Z_1 \in \mathcal{V}_1$ and $Z_2 \in \mathcal{V}_2$. Let $\mathcal{V} := \{Z_1 + Z_2 : Z_i \in \mathcal{V}_i\}$ be the span of $\mathcal{V}_1$ and $\mathcal{V}_2$. Then,*
$$\Pi_{\mathcal{V}}(X) = \Pi_{\mathcal{V}_1}(X) + \Pi_{\mathcal{V}_2}(X).$$
3. *Uncorrelated random variables: Let X be a zero-mean random variable that is uncorrelated with Y. Then, $\Pi_{\mathcal{L}(Y)}(X) = \widehat{\mathbb{E}}[X|Y] = 0$.*

Proof. For the proof of the linearity property, see Proposition 3.3 in Hajek [2015]. For the proof of the second property, see Proposition 3.5 in Hajek [2015]. The third property is straightforward following equation (1.2) since $\mathtt{Cov}(X,Y) = 0$ when X and Y are uncorrelated. □

Estimation in presence of noise

Assumption 1. *The process and measurement noise are zero mean, and are uncorrelated in time, and with each other. Furthermore, the initial state is uncorrelated with the noise processes. In particular, we assume that*

$$\mathbb{E}[w_k] = \mathbb{E}[v_k] = 0, \quad \mathbb{E}[w_kw_k^{\mathsf{T}}] = \Sigma_w, \quad \mathbb{E}[v_kv_k^{\mathsf{T}}] = \Sigma_v, \quad \forall k, \tag{1.3a}$$
$$\mathbb{E}[w_kv_m^{\mathsf{T}}] = \mathbb{E}[x_0v_m^{\mathsf{T}}] = \mathbb{E}[x_0w_k^{\mathsf{T}}] = 0, \quad \forall k, m, \tag{1.3b}$$
$$\mathbb{E}[w_kw_m^{\mathsf{T}}] = \mathbb{E}[v_kv_m^{\mathsf{T}}] = 0, \quad \forall k \neq m. \tag{1.3c}$$

Remark 1. *Recall that if two random variables are independent, then they are uncorrelated. The converse only holds in case of Gaussian random variables. Therefore, the above assumptions are more general than assuming that the noise processes and the initial state are independent.*

We now fix the required notation. Let

- $Y_k := (y_0, y_1, \dots, y_k)$ be the set of all observations till time instant k,
- $\widehat{x}_{m|k}$ be the estimate of the state x_m based on observations Y_k,
- $\tilde{x}_{m|k} := x_m - \widehat{x}_{m|k}$ be the estimation error, and
- $\Sigma_{m|k} := \mathtt{Cov}(\tilde{x}_{m|k})$ be the covariance of the estimation error.

When $m > k$, $\widehat{x}_{m|k}$ is often called the **predicted** value of x_m given Y_k, and when $m = k$, $\widehat{x}_{m|k}$ is called the **estimate** of x_m given Y_k.

Kalman filter is a recursive method to compute an estimate of the state x_k as a linear function of the observations Y_k. The recursive algorithm is shown below.

Discrete-time Kalman Filter

Initial condition: We define $\widehat{x}_{0|-1} = \mathbb{E}[x_0]$ and $\Sigma_{0|-1} = \Sigma_0 = \texttt{Cov}(x_0)$. System model $(A_k, C_k)_{k\geq 0}$ is known.

At time k: We know the previous estimate $\widehat{x}_{k-1|k-1}$ and the error covariance $\Sigma_{k-1|k-1}$.

Prediction Steps: Before observing y_k, we compute:

1. Prediction of current state: $\widehat{x}_{k|k-1} = A_{k-1}\widehat{x}_{k-1|k-1}$
2. Prediction of current output: $\widehat{y}_{k|k-1} = C_k\widehat{x}_{k|k-1}$
3. Error covariance: $\Sigma_{k|k-1} = \texttt{Cov}(x_k - \widehat{x}_{k|k-1}) = A_{k-1}\Sigma_{k-1|k-1}A_{k-1}^{\mathtt{T}} + \Sigma_w$

Update Steps: After observing y_k, we compute:

1. Kalman gain: $L_k = \Sigma_{k|k-1}C_k^{\mathtt{T}}\left[C_k\Sigma_{k|k-1}C_k^{\mathtt{T}} + \Sigma_v\right]^{-1}$
2. Estimate of current state:

$$\begin{aligned}\widehat{x}_{k|k} &= \widehat{x}_{k|k-1} + L_k(y_k - \widehat{y}_{k|k-1}) \\ &= A_{k-1}\widehat{x}_{k-1|k-1} + L_k(y_k - C_k\widehat{x}_{k|k-1})\end{aligned}$$

3. Error covariance: $\Sigma_{k|k} = \texttt{Cov}(x_k - \widehat{x}_{k|k}) = \Sigma_{k|k-1} - L_k\left[C_k\Sigma_{k|k-1}C_k^{\mathtt{T}} + \Sigma_v\right]L_k^{\mathtt{T}}$

Repeat at $k+1$.

Lemma 2 (Correlation of noise with estimation error)**.** *Under Assumption 1, we have*

1. $x_k \in \mathcal{L}(x_0, w_0, w_1, \ldots, w_{k-1})$.
2. $y_k \in \mathcal{L}(x_0, w_0, w_1, \ldots, w_{k-1}, v_k)$.
3. $\mathcal{L}(Y_{k-1}) = \mathcal{L}(y_0, y_1, \ldots, y_{k-1}) = \mathcal{L}(x_0, w_0, w_1, \ldots, w_{k-2}, v_0, v_1, \ldots, v_{k-1})$.

As a consequence, we have

$$\widehat{\mathbb{E}}[w_{k-1}|Y_{k-1}] = \widehat{\mathbb{E}}[v_k|Y_{k-1}] = \mathbb{E}[\widehat{x}_{k|k-1}v_k^{\mathtt{T}}] = \mathbb{E}[\tilde{x}_{k|k-1}v_k^{\mathtt{T}}] = \mathbb{E}[\tilde{x}_{k-1|k-1}w_{k-1}^{\mathtt{T}}] = 0,$$

where $\widehat{x}_{k|k-1} = \widehat{\mathbb{E}}[x_k|Y_{k-1}]$, $\tilde{x}_{k|k-1} = x_k - \widehat{x}_{k|k-1}$, *and* $\tilde{x}_{k-1|k-1} = x_{k-1} - \widehat{x}_{k-1|k-1}$.

Proof. Note that $x_k = A_{k-1}x_{k-1} + w_{k-1}$. Therefore, x_k is a linear combination of x_{k-1} and w_{k-1}. Similarly, x_{k-1} is a linear combination of x_{k-2} and w_{k-2}, and finally x_1 is a linear combination of x_0

and w_0. Thus, $x_k \in \mathcal{L}(x_0, w_0, w_1, \ldots, w_{k-1})$. Now observe that $y_k = C_k x_k + v_k$, and it does not depend on y_{k-1} or any of the past measurement noise. Thus, we have $y_k \in \mathcal{L}(x_0, w_0, w_1, \ldots, w_{k-1}, v_k)$. The third identity follows from this argument as well.

Recall from (1.2) that $\widehat{\mathbb{E}}[w_{k-1}|Y_{k-1}] = \Pi_{\mathcal{L}(Y_{k-1})}(w_{k-1})$. From the above discussion, w_{k-1} is zero-mean and uncorrelated with the constituent random variables of $\mathcal{L}(Y_{k-1}) = \mathcal{L}(x_0, w_0, \ldots, w_{k-2}, v_0, v_1, \ldots, v_{k-1})$. Therefore, following part 3 of Lemma 1, we have $\widehat{\mathbb{E}}[w_{k-1}|Y_{k-1}] = 0$.

Following an analogous argument, note that v_k is zero-mean and uncorrelated with the constituent random variables of $\mathcal{L}(Y_{k-1}) = \mathcal{L}(x_0, w_0, \ldots, w_{k-2}, v_0, v_1, \ldots, v_{k-1})$. Thus, $\widehat{\mathbb{E}}[v_k|Y_{k-1}] = 0$.

Note that $\widehat{x}_{k|k-1} = \widehat{\mathbb{E}}[x_k|Y_{k-1}] = \Pi_{\mathcal{L}(Y_{k-1})}(x_k)$ is the projection of x_k on $\mathcal{L}(Y_{k-1})$. Therefore, $\widehat{x}_{k|k-1} \in \mathcal{L}(Y_{k-1})$, i.e., it is a linear combination of $\{x_0, w_0, \ldots, w_{k-2}, v_0, v_1, \ldots, v_{k-1}\}$. Thus, $\widehat{x}_{k|k-1}$ is uncorrelated with v_k, and we have $\mathbb{E}[\widehat{x}_{k|k-1} v_k^{\mathtt{T}}] = 0$.

Similarly, note that x_k is a linear combination of $\{x_0, w_0, \ldots, w_{k-1}\}$, and thus, $\tilde{x}_{k|k-1} = x_k - \widehat{x}_{k|k-1}$ is a linear combination of $\{x_0, w_0, \ldots, w_{k-1}, v_0, v_1, \ldots, v_{k-1}\}$. Therefore, $\mathbb{E}[\tilde{x}_{k|k-1} v_k^{\mathtt{T}}] = 0$.

Finally, $\tilde{x}_{k-1|k-1} = x_{k-1} - \Pi_{\mathcal{L}(Y_{k-1})}(x_{k-1})$ where x_{k-1} is a linear combination of $\{x_0, w_0, \ldots, w_{k-2}\}$ and $\Pi_{\mathcal{L}(Y_{k-1})}(x_{k-1})$ is a linear combination of $\{x_0, w_0, \ldots, w_{k-2}, v_0, v_1, \ldots, v_{k-1}\}$. Thus, $\tilde{x}_{k-1|k-1}$ is uncorrelated with w_{k-1} and $\mathbb{E}[\tilde{x}_{k-1|k-1} w_{k-1}^{\mathtt{T}}] = 0$. □

Lemma 3 (Conditional expectation of Gaussian random vectors)**.** *Let X and Y be jointly Gaussian random vectors. Then, the conditional expectation $\mathbb{E}[X|Y]$ is an affine function of Y, i.e.,*

$$\mathbb{E}[X|Y] = \widehat{\mathbb{E}}[X|Y] = \mathbb{E}[X] + \mathtt{Cov}(X,Y)^{\mathtt{T}} \mathtt{Cov}(Y)^{-1}[Y - \mathbb{E}[Y]]. \tag{1.4}$$

In other words, the LMSE and the minimum mean square estimators coincide.

Proof. Refer to Proposition 3.9 in Hajek [2015]. □

The following theorem states the optimality property of the Kalman filter. The prediction and update steps are derived in the proof.

Theorem 1. *Suppose Assumption 1 holds. The discrete-time Kalman filter shown above minimizes the mean square error $\mathbb{E}[||x_k - \widehat{x}_{k|k}||^2]$ among all* **linear** *estimators of x_k given Y_k. Furthermore, if x_0, w_k and v_k are Gaussian random vectors, then the Kalman filter minimizes the mean square error $\mathbb{E}[||x_k - \widehat{x}_{k|k}||^2]$ among all estimators of x_k given Y_k.*

Proof. We first consider the case where $\mathbb{E}[x_0] = 0$ to show that Kalman filter is the LMSE estimator, followed by the non-zero mean case. We then treat the case with Gaussian random vectors.

Case 1: $\mathbb{E}[x_0] = 0$.

From the linearity of expectation, we have $\mathbb{E}[x_k] = \mathbb{E}[y_k] = 0$. At time k, we assume that $\widehat{x}_{k-1|k-1}$ and the error covariance $\Sigma_{k-1|k-1}$ are known, and from this knowledge, we first derive the predicted values of the state and output at time k. We compute

$$\begin{aligned}\widehat{x}_{k|k-1} = \widehat{\mathbb{E}}[x_k|Y_{k-1}] &= \widehat{\mathbb{E}}[A_{k-1}x_{k-1} + w_{k-1}|Y_{k-1}] \\ &= A_{k-1}\widehat{\mathbb{E}}[x_{k-1}|Y_{k-1}] + \widehat{\mathbb{E}}[w_{k-1}|Y_{k-1}] \\ &= A_{k-1}\widehat{x}_{k-1|k-1},\end{aligned}$$

where the second equality holds because LMSE (being a projection) is a linear operator following Lemma 1, and the last equality holds because $\widehat{\mathbb{E}}[w_{k-1}|Y_{k-1}] = 0$ from Lemma 2.

Following analogous arguments, we compute

$$\begin{aligned}\widehat{y}_{k|k-1} = \widehat{\mathbb{E}}[y_k|Y_{k-1}] &= \widehat{\mathbb{E}}[C_k x_k + v_k|Y_{k-1}] \\ &= C_k\widehat{\mathbb{E}}[x_k|Y_{k-1}] + \widehat{\mathbb{E}}[v_k|Y_{k-1}] \qquad \text{(LMSE is a linear operator)} \\ &= C_k\widehat{x}_{k|k-1} \qquad \text{(from Lemma 2)}.\end{aligned}$$

Now, observe that

$$\tilde{x}_{k|k-1} = x_k - \widehat{x}_{k|k-1} = A_{k-1}x_{k-1} + w_{k-1} - A_{k-1}\widehat{x}_{k-1|k-1} = A_{k-1}\tilde{x}_{k-1|k-1} + w_{k-1}.$$

Therefore, we compute the covariance of the prediction error as

$$\begin{aligned}\Sigma_{k|k-1} &= \mathbb{E}[(A_{k-1}\tilde{x}_{k-1|k-1} + w_{k-1})(A_{k-1}\tilde{x}_{k-1|k-1} + w_{k-1})^{\mathtt{T}}] \\ &= A_{k-1}\mathbb{E}[\tilde{x}_{k-1|k-1}\tilde{x}_{k-1|k-1}^{\mathtt{T}}]A_{k-1}^{\mathtt{T}} + 2A_{k-1}\mathbb{E}[\tilde{x}_{k-1|k-1}w_{k-1}^{\mathtt{T}}] + \mathbb{E}[w_{k-1}w_{k-1}^{\mathtt{T}}] \\ &= A_{k-1}\Sigma_{k-1|k-1}A_{k-1}^{\mathtt{T}} + \Sigma_w,\end{aligned}$$

as $\mathbb{E}[\tilde{x}_{k-1|k-1}w_{k-1}^{\mathtt{T}}] = 0$ following Lemma 2. Since all quantities have zero-mean, covariance equals correlation.

Thus far, we have derived all three steps of the prediction stage. Once we observe y_k, we first compute the *innovation* contained in it. In particular, we have

$$\tilde{y}_{k|k-1} = y_k - \widehat{y}_{k|k-1} = y_k - C_k\widehat{x}_{k|k-1} = C_k[x_k - \widehat{x}_{k|k-1}] + v_k = C_k\tilde{x}_{k|k-1} + v_k. \tag{1.7}$$

Recall that $\widehat{y}_{k|k-1} = \Pi_{\mathcal{L}(Y_{k-1})}(y_k)$ is the projection of y_k on the linear space spanned by random variables Y_{k-1}. Therefore, by the orthogonality property of projection, $\tilde{y}_{k|k-1}$ is orthogonal to every random variable that is a linear combination of random variables in Y_{k-1}. In particular, $\mathcal{L}(Y_{k-1})$ is orthogonal to $\mathcal{L}(\tilde{y}_{k|k-1})$. Therefore, following part 2 of Lemma 1, we have

$$\begin{aligned}\widehat{x}_{k|k} = \widehat{\mathbb{E}}[x_k|Y_k] &= \Pi_{\mathcal{L}(Y_{k-1})}(x_k) + \Pi_{\mathcal{L}(\tilde{y}_{k|k-1})}(x_k) \\ &= \widehat{x}_{k|k-1} + \mathbb{E}[x_k\tilde{y}_{k|k-1}^{\mathtt{T}}]\mathbb{E}[\tilde{y}_{k|k-1}\tilde{y}_{k|k-1}^{\mathtt{T}}]^{-1}\tilde{y}_{k|k-1},\end{aligned} \tag{1.8}$$

since we have assumed that x_0 and therefore all random variables including x_k and $\tilde{y}_{k|k-1}$ are zero-mean. From (1.7), we compute

$$\begin{aligned}\mathbb{E}[x_k\tilde{y}_{k|k-1}^{\mathtt{T}}] &= \mathbb{E}[(\widehat{x}_{k|k-1} + \tilde{x}_{k|k-1})(C_k\tilde{x}_{k|k-1} + v_k)^{\mathtt{T}}] \\ &= \mathbb{E}[\widehat{x}_{k|k-1}(C_k\tilde{x}_{k|k-1} + v_k)^{\mathtt{T}}] + \mathbb{E}[\tilde{x}_{k|k-1}(C_k\tilde{x}_{k|k-1} + v_k)^{\mathtt{T}}] \\ &= 0 + \Sigma_{k|k-1}C_k^{\mathtt{T}} = \Sigma_{k|k-1}C_k^{\mathtt{T}},\end{aligned} \tag{1.9}$$

since $\widehat{x}_{k|k-1}$ is orthogonal with $\tilde{x}_{k|k-1}$ from the orthogonality property of projection, and both $\widehat{x}_{k|k-1}$ (being a function of random variables in Y_{k-1} and $\tilde{x}_{k|k-1}$ (being a function of x_k and $\widehat{x}_{k|k-1}$) are uncorrelated with v_k from Lemma 2. In an analogous manner, we compute

$$\begin{aligned}\mathbb{E}[\tilde{y}_{k|k-1}\tilde{y}_{k|k-1}^{\mathsf{T}}] &= \mathbb{E}[(C_k\tilde{x}_{k|k-1}+v_k)(C_k\tilde{x}_{k|k-1}+v_k)^{\mathsf{T}}] \\ &= C_k\mathbb{E}[\tilde{x}_{k|k-1}\tilde{x}_{k|k-1}^{\mathsf{T}}]C_k^{\mathsf{T}} + \mathbb{E}[v_kv_k^{\mathsf{T}}] \\ &= C_k\Sigma_{k|k-1}C_k^{\mathsf{T}} + \Sigma_v, \end{aligned} \tag{1.10}$$

where we have again used the fact that $\tilde{x}_{k|k-1}$ is uncorrelated with v_k.

We now substitute (1.9) and (1.10) in (1.8), and obtain

$$\widehat{x}_{k|k} = \widehat{x}_{k|k-1} + \Sigma_{k|k-1}C_k^{\mathsf{T}}\Big[C_k\Sigma_{k|k-1}C_k^{\mathsf{T}} + \Sigma_v\Big]^{-1}\tilde{y}_{k|k-1} \tag{1.11a}$$

$$=: \widehat{x}_{k|k-1} + L_k\tilde{y}_{k|k-1} \tag{1.11b}$$

$$= \widehat{x}_{k|k-1} + L_k(y_k - \widehat{y}_{k|k-1}), \tag{1.11c}$$

where $L_k := \Sigma_{k|k-1}C_k^{\mathsf{T}}\Big[C_k\Sigma_{k|k-1}C_k^{\mathsf{T}} + \Sigma_v\Big]^{-1}$ is the *Kalman gain.*

Thus, the estimate of x_k is a linear combination of the predicted value based on prior observations ($\widehat{x}_{k|k-1}$) and a feedback term where the error between the observed output (y_k) and the predicted value of the output ($\widehat{y}_{k|k-1}$) is multiplied by the Kalman gain L_k.

We now compute $\Sigma_{k|k} = \mathtt{Cov}(x_k - \widehat{x}_{k|k})$ to complete the derivation. Note that

$$\begin{aligned} x_k - \widehat{x}_{k|k} &= x_k - \widehat{x}_{k|k-1} - L_k(C_kx_k + v_k - C_k\widehat{x}_{k|k-1}) \\ &= (I - L_kC_k)\tilde{x}_{k|k-1} - L_kv_k, \end{aligned}$$

where I is the identity matrix of dimension n. Since $\tilde{x}_{k|k-1}$ is uncorrelated with v_k, we obtain

$$\begin{aligned}\Sigma_{k|k} &= \mathbb{E}[((I-L_kC_k)\tilde{x}_{k|k-1} - L_kv_k)((I-L_kC_k)\tilde{x}_{k|k-1} - L_kv_k)^{\mathsf{T}}] \\ &= (I-L_kC_k)\Sigma_{k|k-1}(I-L_kC_k)^{\mathsf{T}} + L_k\Sigma_vL_k^{\mathsf{T}} \\ &= \Sigma_{k|k-1} - L_kC_k\Sigma_{k|k-1} - \Sigma_{k|k-1}C_k^{\mathsf{T}}L_k^{\mathsf{T}} + L_kC_k\Sigma_{k|k-1}C_k^{\mathsf{T}}L_k^{\mathsf{T}} + L_k\Sigma_vL_k^{\mathsf{T}} \\ &= \Sigma_{k|k-1} - L_kC_k\Sigma_{k|k-1} - \Sigma_{k|k-1}C_k^{\mathsf{T}}L_k^{\mathsf{T}} + L_k\big[C_k\Sigma_{k|k-1}C_k^{\mathsf{T}} + \Sigma_v\big]L_k^{\mathsf{T}} \\ &= \Sigma_{k|k-1} - L_kC_k\Sigma_{k|k-1} - \Sigma_{k|k-1}C_k^{\mathsf{T}}L_k^{\mathsf{T}} + \Sigma_{k|k-1}C_k^{\mathsf{T}}L_k^{\mathsf{T}} \\ &= \Sigma_{k|k-1} - L_kC_k\Sigma_{k|k-1} \\ &= \Sigma_{k|k-1} - \Sigma_{k|k-1}C_k^{\mathsf{T}}\big[C_k\Sigma_{k|k-1}C_k^{\mathsf{T}} + \Sigma_v\big]^{-1}C_k\Sigma_{k|k-1} \\ &= \Sigma_{k|k-1} - L_k\big[C_k\Sigma_{k|k-1}C_k^{\mathsf{T}}L_k^{\mathsf{T}} + \Sigma_v\big]L_k^{\mathsf{T}}, \end{aligned}$$

where any of the last three equalities can be used depending on the context. This concludes the derivation for the zero-mean case. The second equality above is often used to compute $\Sigma_{k|k}$ as it is more robust to numerical errors.

Case 2: $\mathbb{E}[x_0] \neq 0$**.**

In this case, we express the overall dynamics (1.1) as the sum of a deterministic or nominal component and a zero-mean stochastic component. In particular, we define

$$x_k^C := x_k - \mathbb{E}[x_k], \quad y_k^C := y_k - \mathbb{E}[y_k].$$

Then,

$$x_{k+1}^C = x_{k+1} - \mathbb{E}[x_{k+1}] = A_k x_k + w_k - A_k \mathbb{E}[x_k] = A_k x_k^C + w_k, \tag{1.12a}$$

$$y_k^C = y_k - \mathbb{E}[y_k] = C_k x_k + v_k - C_k \mathbb{E}[x_k] = C_k x_k^C + v_k. \tag{1.12b}$$

Thus, x_k^C follows a similar dynamics as (1.1) and all the state and output quantities are zero-mean. From the previous case of the derivation, we have

$$\begin{aligned}
\widehat{x}_{k|k} &= \mathbb{E}[x_k] + \widehat{x}_{k|k}^C = \mathbb{E}[x_k] + \widehat{x}_{k|k-1}^C + L_k(y_k^C - \widehat{y}_{k|k-1}^C) \\
&= A_{k-1}\mathbb{E}[x_{k-1}] + A_{k-1}\widehat{x}_{k-1|k-1}^C + L_k(y_k - \mathbb{E}[y_k] - C_k \widehat{x}_{k|k-1}^C) \\
&= A_{k-1}\widehat{x}_{k-1|k-1} + L_k(y_k - \mathbb{E}[y_k] - C_k \widehat{x}_{k|k-1} + C_k \mathbb{E}[x_k]) \\
&= \widehat{x}_{k|k-1} + L_k(y_k - \widehat{y}_{k|k-1}).
\end{aligned}$$

Therefore, the same equations as in the zero-mean case holds.

Case 3: Gaussian random vectors.

When x_0, w_k and v_k are all Gaussian random vectors, then x_k, and y_k are also Gaussian random vectors for every k as we consider a linear dynamical system. Following Lemma 3, we note that for Gaussian random vectors X and Y, $\widehat{\mathbb{E}}[X|Y] = \mathbb{E}[X|Y]$, i.e., the LMSE and MMSE estimators coincide. Therefore, we can replace $\widehat{\mathbb{E}}$ by $\mathbb{E}$ in all the steps in the derivation in Case 1. As a result, the estimates minimize the mean square error. □

The above theorem shows that the Kalman filter minimizes the mean square error among all linear estimators.

Steady-State Behavior

First we establish steady-state properties of the following system:

$$x_{k+1} = A x_k + w_k.$$

It is easy to see that the mean $\bar{x}_k$ and the covariance Σ_k^x of the states evolve as

$$\begin{aligned}
\bar{x}_{k+1} &= A \bar{x}_k, \\
\Sigma_{k+1}^x &= A \Sigma_k^x A^\top + \Sigma_w.
\end{aligned}$$

The following theorem shows the asymptotic properties of the states.

Theorem 2. *Let Σ_w be positive definite. The followings are equivalent:*

- *For any* Σ_0^x, $\Sigma_{k+1}^x = A\Sigma_k^x A^\top + \Sigma_w$ *is such that* Σ_k^x *converges to* $\Sigma^\star$ *irrespective of choice of* Σ_0^x.
- *The Lyapunov equation* $\Sigma = A\Sigma A^\top + \Sigma_w$ *has a unique solution* $\Sigma^\star$.
- *The matrix* A *has all eigenvalues within the unit circle.*

We now establish steady-state properties of the estimation error under Kalman filter for the time-invariant system. First, we write the Kalman filter expressions in compact form as:

$$\begin{aligned}
L_k &= \Sigma_{k|k-1}C^\mathsf{T}\big[C\Sigma_{k|k-1}C^\mathsf{T} + \Sigma_v\big]^{-1},\\
\widehat{x}_{k+1|k} &= A\widehat{x}_{k|k} = A\widehat{x}_{k|k-1} + AL_k(y_k - \widehat{y}_{k|k-1})\\
&= (A - AL_kC)\widehat{x}_{k|k-1} + AL_ky_k,\\
\Sigma_{k+1|k} &= A\Sigma_{k|k}A^\top + \Sigma_w\\
&= A(\Sigma_{k|k-1} - \Sigma_{k|k-1}C^\mathsf{T}\big[C\Sigma_{k|k-1}C^\mathsf{T} + \Sigma_v\big]^{-1}C\Sigma_{k|k-1})A^\top + \Sigma_w\\
&= A\Sigma_{k|k-1}A^\top - A\Sigma_{k|k-1}C^\mathsf{T}\big[C\Sigma_{k|k-1}C^\mathsf{T} + \Sigma_v\big]^{-1}C\Sigma_{k|k-1}A^\top + \Sigma_w.
\end{aligned}$$

The following theorem shows the asymptotic properties of the error covariance.

Theorem 3. *Let* Σ_w *be positive semidefinite,* Σ_v *be positive definite,* (A, C) *be detectable and* (A, Σ_w) *be stabilizable. The followings are equivalent:*

- *The equation*
$$\Sigma = A\Sigma A^\top - A\Sigma C^\mathsf{T}\big[C\Sigma C^\mathsf{T} + \Sigma_v\big]^{-1}C\Sigma A^\top + \Sigma_w$$
has a unique solution $\Sigma^\star$ *with* $\Sigma^\star$ *being positive definite.*
- *For any* Σ_0^x, *under the Kalman filter update equation,* $\Sigma_{k+1|k}$ *converges to* $\Sigma^\star$ *irrespective of choice of* Σ_0.
- *The matrix* $A - AL^\star C$ *has all eigenvalues within the unit circle where*
$$L^\star = \Sigma^\star C^\mathsf{T}\big[C\Sigma^\star C^\mathsf{T} + \Sigma_v\big]^{-1}.$$
The matrix $L^\star$ *is called steady-state Kalman gain.*

References

Bruce Hajek. *Random Processes for Engineers.* Cambridge University Press, 2015.