Lecture 11 : Kalman Filter

H. METE SONER, ORFE, PRINCETON

ORF 418 OPTIMAL LEARNING

OCTOBER 8, 2025

version October 8, 2025

Context

- There is a stochastic process x_k like the position of an aircraft ;
- We know its dynamics ;
- We make noisy observations of it ;
- Based on these observations we want to estimate x_k.

It is important that these formula are tractable and can be implemented easily in real applications.

It was developed by Rudolf Emil Kálmán in 1960.

Outline

Derivation

We are given :

$$x_{k+1} = A_k \, x_k + \omega_{k+1}, \qquad \text{State Dynamics}$$
$$z_k = H_k \, x_k + \nu_k, \qquad \text{Observations}$$

where $\{\omega_k\}$, $\{\nu_k\}$ are independent Gaussian process and we assume that x_0 is Gaussian and independent of the noise.

We define

$$\widehat{x}_k := \mathbb{E}[\, x_k \mid z_1, \ldots, z_k \,], \qquad \text{Main Estimate;}$$
$$\widehat{x}_{k|k-1} := \mathbb{E}[\, x_k \mid z_1, \ldots, z_{k-1} \,], \qquad \text{Pre-observation Estimate;}$$
$$\iota_k := z_k - H_k \, \widehat{x}_{k|k-1}, \qquad \text{Innovation Process;}$$
$$P_k := \mathbb{C}(x_k - \widehat{x}_k) = \mathbb{E}[(\, x_k - \widehat{x}_k)(x_k - \widehat{x}_k)^\top \,], \qquad \text{Error Covariance;}$$
$$\overline{P}_k := \mathbb{C}(x_k - \widehat{x}_{k|k-1}) = \mathbb{E}[(x_k - \widehat{x}_{k|k-1})(x_k - \widehat{x}_{k|k-1})^\top], \qquad \text{Pre-observation Covariance.}$$

Useful Observations

- $$\widehat{x}_{k+1|k} = \mathbb{E}[\, x_{k+1} \mid z_1, \ldots, z_k \,] = \mathbb{E}[\, Ax_k + \omega_{k+1} \mid z_1, \ldots, z_k \,] = A\mathbb{E}[\, x_k \mid z_1, \ldots, z_k \,] = A\widehat{x}_k.$$

- $$\mathbb{E}[\iota_k] = \mathbb{E}[z_k - H_k \widehat{x}_{k|k-1}] = H_k \left(\mathbb{E}[x_k - \mathbb{E}[x_k \mid z_1, \ldots, z_{k-1}]\,]\right) = 0.$$

- Observation noise ν_k at time k is independent of previous processes prior to k.
- Therefore ν_k is independent of x_k and $\widehat{x}_{k|k-1}$. Hence,

$$\mathbb{C}(\iota_k) = \mathbb{E}[\iota_k \iota_k^\top] = \mathbb{E}[H_k(x_k - \widehat{x}_{k|k-1})(H_k(x_k - \widehat{x}_{k|k-1}))^\top] + \mathbb{E}[\nu_k \nu_k^\top] = H_k \overline{P}_k H_k^\top + R_k.$$

Theorem (4.2.3 in the Lecture Notes)

$$\widehat{x}_k = \widehat{x}_{k|k-1} + K_k \, \iota_k, \qquad \iota_k = z_k - H_k \, \widehat{x}_{k|k-1}, \qquad \widehat{x}_{k|k-1} = A_{k-1} \, \widehat{x}_{k-1},$$

Moreover, the gains matrix K_k *is given by,* $K_k = \overline{P}_k \, H_k^\top \, \left(H_k \, \overline{P}_k \, H_k^\top + R_k \right)^{-1}$.

Discussion of the proof :

- Formula for $\widehat{x}_{k|k-1}$ was derived before, and $\iota_k = z_k - H_k$ is the definition or the innovation process.
- We compute that $\mathbb{E}[(x_k - \widehat{x}_{k|k-1}) \, | \iota_1, \ldots, \iota_{k-1}] = 0$.
- Hence, $(x_k - \widehat{x}_{k|k-1})$ is independent of $\iota_1, \ldots, \iota_{k-1}$, and

$$\begin{aligned} \widehat{x}_k &= \widehat{x}_{k|k-1} + \mathbb{E}[(x_k - \widehat{x}_{k|k-1}) \mid z_1, \ldots, z_k] = \widehat{x}k|k-1 + \mathbb{E}[(x_k - \widehat{x}_{k|k-1}) \mid z_1, \ldots, z_k] \\ &= \widehat{x}_{k|k-1} + \mathbb{E}[(x_k - \widehat{x}_{k|k-1}) \mid \iota_1, \ldots, \iota_k] = \widehat{x}_{k|k-1} + \mathbb{E}[(x_k - \widehat{x}_{k|k-1}) \mid \iota_k]. \end{aligned}$$

- Above is the conditional expectation of a Gaussian given another Gaussian. By the general formula,

$$\widehat{x}_k = \widehat{x}_{k|k-1} + K \, \iota_k,$$

and $K = \mathbb{C}((x_k - \widehat{x}_{k|k-1}), \iota_k) / var(\iota_k)$.

From previous slight

$$\widehat{x}_k = \widehat{x}_{k|k-1} + K_k\, \iota_k, \qquad \iota_k = z_k - H_k\, \widehat{x}_{k|k-1}, \qquad \widehat{x}_{k|k-1} = A_{k-1}\, \widehat{x}_{k-1},$$

$$K_k = \overline{P}_k\, H_k^\top \left(H_k\, \overline{P}_k\, H_k^\top + R_k \right)^{-1}$$

- At time k, we have the estimate $\widehat{x}_{k-1}$ from the previous step, and we compute the pre-observation estimate $\widehat{x}_{k|k-1}$ by the formula $\widehat{x}_{k|k-1} = A_{k-1}\, \widehat{x}_{k-1}$.
- Given the observation z_k and computed $\widehat{x}_{k|k-1}$, we compute the innovation process ι_k by the formula $\iota_k = z_k - H_k\, \widehat{x}_{k|k-1}$.
- We update our estimate by the formula $\widehat{x}_k = \widehat{x}_{k|k-1} + K_k\, \iota_k$.

For this procedure we need to compute $P_k, \overline{P}_k$ and also the gains matrix K_k, prior to the calculations at time k. We discuss this in the next slight.

Theorem (4.2.3 in the Lecture Notes)

For $k = 1, 2, \dots$

$$\overline{P}_k = A_{k-1}\, P_{k-1}\, A_{k-1}^\top + Q_k, \qquad K_k = \overline{P}_k\, H_k^\top \left(H_k\, \overline{P}_k\, H_k^\top + R_k\right)^{-1}, \qquad P_k = (I - K_k\, H_k)\, \overline{P}_k.$$

Given these formulae we can compute them independent of the observations :

- We have $P_0 = \mathbb{C}(x_0)$ given.
- At time k, we have the P_{k-1} from the previous step, and compute $\overline{P}_k$ by the formula $\overline{P}_k = A_{k-1}\, P_{k-1}\, A_{k-1}^\top + Q_k$.
- We compute the gains matrix K_k by the formula $K_k = \overline{P}_k\, H_k^\top \left(H_k\, \overline{P}_k\, H_k^\top + R_k\right)^{-1}$.
- We update P_k by the formula $P_k = (I - K_k\, H_k)\, \overline{P}_k$.

Theorem (4.2.3 in the Lecture Notes)

For $k = 1, 2, \ldots$

$$\overline{P}_k = A_{k-1}\, P_{k-1}\, A_{k-1}^\top + Q_k, \qquad K_k = \overline{P}_k\, H_k^\top \left(H_k\, \overline{P}_k\, H_k^\top + R_k\right)^{-1}, \qquad P_k = (I - K_k\, H_k)\, \overline{P}_k.$$

Given these formulae we can compute them independent of the observations :

- We have $P_0 = \mathbb{C}(x_0)$ given.
- At time k, we have the P_{k-1} from the previous step, and compute $\overline{P}_k$ by the formula $\overline{P}_k = A_{k-1}\, P_{k-1}\, A_{k-1}^\top + Q_k$.
- We compute the gains matrix K_k by the formula $K_k = \overline{P}_k\, H_k^\top \left(H_k\, \overline{P}_k\, H_k^\top + R_k\right)^{-1}$.
- We update P_k by the formula $P_k = (I - K_k\, H_k)\, \overline{P}_k$.

Theorem (Innovation process, 4.2.2 in the Lecture Notes)

The innovation process has the following properties :

1. $\{\iota_k\}_{k=1,2,\dots}$ *are independent of each other and have mean zero.*
2. *For each* $k \geq 1$, $(x_{k+1} - \widehat{x}_{k+1|k})$ *is independent of all* $\iota_1, \dots, \iota_k$.
3. $\iota_1, \dots, \iota_k$ *is an affine function of* $z_1, \dots, z_k$. *Hence, it is Gaussian and conditioning on* $z_1, \dots, z_k$ *and* $\iota_1, \dots, \iota_k$ *are same.*

- Full proof is given in the notes.
- We fist show that $\mathbb{E}[\iota_{k+1}|z_1, \dots, z_k] = 0$.
- Since all r.v.'s are Gaussian, this implies that ι_{k+1} is independent of $z_1, \dots, z_k$.
- Also $\mathbb{E}[(x_{k+1} - \widehat{x}_{k+1|k})\,|\iota_1, \dots, \iota_k] = 0$. Therefore, the second claim follows.
- The last claim follows from the fact that there is one-to-one correspondence between the observations z and the innovations ι.

Theorem (Innovation process, 4.2.2 in the Lecture Notes)

The innovation process has the following properties :

1. $\{\iota_k\}_{k=1,2,\ldots}$ *are independent of each other and have mean zero.*
2. *For each* $k \geq 1$, $(x_{k+1} - \widehat{x}_{k+1|k})$ *is independent of all* $\iota_1, \ldots, \iota_k$.
3. $\iota_1, \ldots, \iota_k$ *is an affine function of* $z_1, \ldots, z_k$. *Hence, it is Gaussian and conditioning on* $z_1, \ldots, z_k$ *and* $\iota_1, \ldots, \iota_k$ *are same.*

- Full proof is given in the notes.
- We fist show that $\mathbb{E}[\iota_{k+1}|z_1, \ldots, z_k] = 0$.
- Since all r.v.'s are Gaussian, this implies that ι_{k+1} is independent of $z_1, \ldots, z_k$.
- Also $\mathbb{E}[(x_{k+1} - \widehat{x}_{k+1|k})\,|\iota_1, \ldots, \iota_k] = 0$. Therefore, the second claim follows.
- The last claim follows from the fact that there is one-to-one correspondence between the observations z and the innovations ι.

Gains Matrix

- $\widehat{x}_k = \mathbb{E}[x_k \mid z_1, \ldots, z_k] = \widehat{x}_{k|k-1} + \mathbb{E}[(x_k - \widehat{x}_{k|k-1}) \mid z_1, \ldots, z_k]$.
- From the second and third parts of the above theorem :

$$\mathbb{E}[(x_k - \widehat{x}_{k|k-1}) \mid z_1, \ldots, z_k] = \mathbb{E}[(x_k - \widehat{x}_{k|k-1}) \mid \iota_1, \ldots, \iota_k] = \mathbb{E}[(x_k - \widehat{x}_{k|k-1}) \mid \iota_k].$$

- $(x_k - \widehat{x}_{k|k-1}), \iota_k$ are jointly Gaussian, and we use the Bayesian estimate formula from Lecture 8 :

$$\mathbb{E}[(x_k - \widehat{x}_{k|k-1}) \mid \iota_k] = \mathbb{E}[x_k - \widehat{x}_{k|k-1}] + \mathbb{C}(x_k - \widehat{x}_{k|k-1}, \iota_k)\ \mathbb{C}(\iota_k)^{-1}\ \iota_k =: K_k\ \iota_k.$$

- Combining them all :

$$\widehat{x}_k = \widehat{x}_{k|k-1} + K_k\ \iota_k,$$

where using the formula derived earlier for $\mathbb{C}(\iota_k)$:

$$\mathbb{C}(\iota_k) = H_k \overline{P}_k H_k^\top + R_k;$$

$$\mathbb{C}(x_k - \widehat{x}_{k|k-1}, \iota_k) = \mathbb{C}(x_k - \widehat{x}_{k|k-1}, H_k(x_k - \widehat{x}_{k|k-1}) + \nu_k) = \overline{P}_k H_k^\top$$

$$K_k = \mathbb{C}(x_k - \widehat{x}_{k|k-1}, \iota_k)\ \mathbb{C}(\iota_k)^{-1} = \overline{P}_k H_k^\top (H_k \overline{P}_k H_k^\top + R_k)^{-1}.$$

Position Estimation : Section 4.1

This is discussed in Section 4.1 of the Lecture Notes.

- We would like to estimate the position of a vehicle trying follow a given path.
- To simplify we assume that it moves on a straight line and has unit mass.
- A target trajectory $p_d : \mathbb{R}_+ \mapsto \mathbb{R}$ is given and force is applied to the particle so as to keep its trajectory equal to the target.
- We first obtain the *dynamics without noise*. Let $p_a(t)$ be the actual position of the particle. Without noise $p_a(t) = p_d(t)$. Then, the force is given by,

$$\text{Force} = \text{mass} \times \text{acceleration} = p_d''(t).$$

- Therefore, *with noise*, the actual position satisfies $p_a''(t) = p_d''(t) + \text{noise}$.

- Set the continuous-time state vector be $x(t) = [\, p_a(t) - p_d(t)\,,\ p_a'(t)\ -\ p_d'(t))\,]^\top$.
- Then,

$$x'(t) = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} x(t)\ + \begin{bmatrix} 0 \\ 1 \end{bmatrix} \text{noise.}$$

- We need to discretize time. So we fix a small time step h and set $x_k := x(ht)$ and ξ_{k+1} be the unknown noise that influenced the particle during time interval $[kh, (k+1)h]$.
- Then, the following difference equation is a good approximation and we use it as our state equation :

$$\boxed{x_{k+1} = A\,x_k + \omega_{k+1}, \qquad \text{where} \qquad A = \begin{bmatrix} 1 & h \\ 0 & 1 \end{bmatrix},\ \ \omega_{k+1} := \begin{bmatrix} 0 \\ h \end{bmatrix} \xi_{k+1}.}$$

- The position also observed with noise ν_k. So the observations are given by,

$$\boxed{z_k = H\,x_k + \nu_k, \qquad \text{where} \qquad H = [1 \;\; 0].}$$

- We want to estimate

$$\mathbb{E}[\; x_k \;| z_1, \ldots z_k \;].$$

- This is exactly the setting of the Kalman filter.
- In the above calculations, we fixed the force to be equal to $p_d''(t)$, as this is the optimal force in the deterministic problem. But with random perturbations we have apply additional force to keep the trajectory closer to the given one based on the noisy observations. This problem of the autopilot. The solution will be the topic of next lecture.

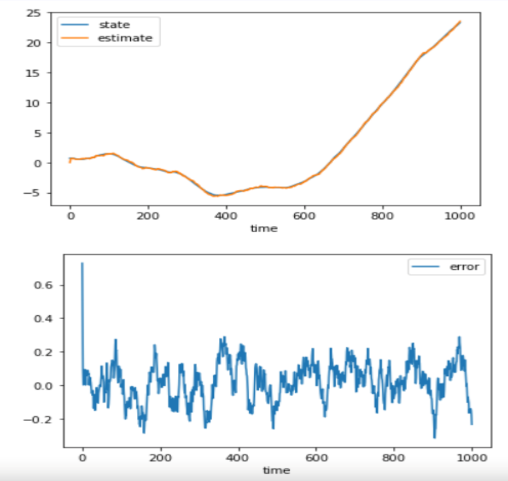
- We have $d = 2$, $\ell = 1$ and

$$A = \begin{bmatrix} 1 & h \\ 0 & 1 \end{bmatrix}, \quad H = [1 \;\; 0], \quad \omega_{k+1} := \begin{bmatrix} 0 \\ h \end{bmatrix} \xi_{k+1}, \quad \nu_k \in \mathbb{R}.$$

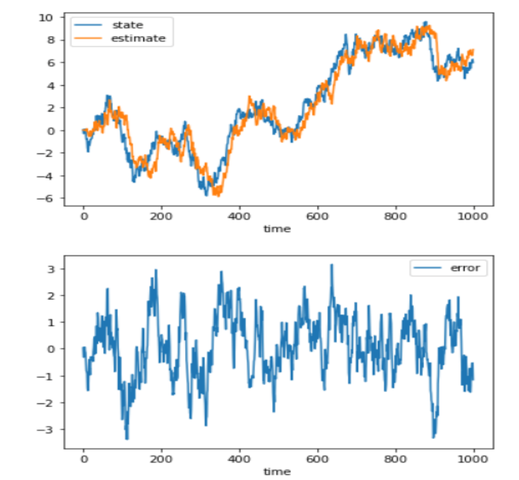
- Hence, $Q = \begin{bmatrix} 0 & 0 \\ 0 & q \end{bmatrix}$, $R = [\, r \,]$.
- We solve this system numerically for the values $h = 0.01$, $r = 0.1$, $q = 10\, h$, and

$$P_0 = \begin{bmatrix} 0.1 & 0 \\ 0 & 0.1 \end{bmatrix}.$$

- The numerical solution is easily obtained in the computer.



Position



Velocity

Long Time behavior : Section 4.2.4

Assumptions

- Suppose that all coefficients are time independent, i.e., $A = A_k$, $H = H_k$, $Q = Q_k$, $R = R_k$.
- Set

$$P_* := \lim_{k\to\infty} P_k.$$

- Under these assumptions, we have

$$\overline{P}_* := \lim_{k\to\infty} \overline{P}_k = A P_* A^\top + Q,$$

$$K_* := \lim_{k\to\infty} K_k = \overline{P}_* H^\top (H \overline{P}_* H^\top + R)^{-1},$$

$$P_* = (I - K_* H) \overline{P}_*.$$

- Combining all we obtain then following equation for P_*,

$$P_* = Q + A P_* A^\top - P_* H^\top (H P_* H^\top + R)^{-1} (A P_* A^\top + Q).$$

Additionally, $\overline{P}_*$ satisfies,

$$\overline{P}_* = Q + A \overline{P}_* A^\top - A \overline{P}_* H^\top (H \overline{P}_* H^\top + R)^{-1} H \overline{P}_* A^\top.$$

- Suppose that we initially we start with $P_0 = P_*$.
- Then, $P_k = P_*$ for every k. Similarly, $\overline{P}_k = \overline{P}_*$ and $K_k = K_*$ for all k. Moreover,

$$\widehat{x}_{k+1} = A\,\widehat{x}_k + K_*\,\iota_k.$$

- We know that the innovation process has mean zero, independent of each other and

$$\mathbb{C}(\iota_k) = H\,\overline{P}_k\,H^\top + R = H\,\overline{P}_*\,H^\top + R.$$

  Hence, the innovation process is also identical making it an i.i.d. process.
- This observation will be used in the next lecture.

Theorem (4.2.4 in the Lecture Notes)

The matrices $(P_*, \overline{P}_*, K_*) := \lim_{k\to\infty} (P_k, \overline{P}_k, K_k)$ *solve*

$$\overline{P}_* = Q + A\overline{P}_* A^\top - A\overline{P}_* H^\top (H\overline{P}_* H^\top + R)^{-1} H\overline{P}_* A^\top$$
$$K_* = \overline{P}_* H^\top (H\overline{P}_* H^\top + R)^{-1}$$
$$P_* = (I_d - K_* H)\overline{P}_*.$$

Moreover, if $P_0 = P_*$*, then the Kalman covariances* $(P_k, \overline{P}_k, K_k)$ *are all equal to* $(P_*, \overline{P}_*, K_*)$*.*
Additionally, the innovation process is i.i.d. with mean zero and

$$\mathbb{C}(\iota_k) = H\overline{P}_* H^\top + R.$$