Projet LINMA1731

Georgios Birpoutsoukis, Jehum Cho

May 2021

1 Answers to Theoretical Questions

Consider a linear dynamical system described by the following Gaussian state-space model :

$$\begin{aligned} x_k &= A\ x_{k-1} + B\ u_k + w_k \qquad \text{(state equation)} \\ y_k &= C\ x_k + v_k \qquad \text{(measurement equation)} \end{aligned} \tag{1}$$

where $x \in \mathbb{R}^{n_x}$ denotes the state and $y \in \mathbb{R}^{n_y}$ the measurement vector. The noise representing the uncertainty of the state vector is given by $w_k \sim \mathcal{N}(0, Q)$, with $Q \in \mathbb{R}^{n_x \times n_x}$ the corresponding covariance matrix. Similarly, the measurement noise is given by $v_k \sim \mathcal{N}(0, R)$, $R \in \mathbb{R}^{n_y \times n_y}$.

a) Alternative derivation of Kalman Filter (KF)

Suppose that the following conditional distribution is available :

$$x_{k-1}|y_{1:k-1} \sim \mathcal{N}(\mu_{k-1}, P_{k-1}) \tag{2}$$

1. (Forecast step) It holds $x_k|y_{1:k-1} \sim \mathcal{N}(\tilde{\mu}_k, \tilde{P}_k)$. Define $\tilde{\mu}_k$ and $\tilde{P}_k$ (Hint : Use the state equation in (1)).

   Answer : Looking at the state equation, consider the expected value of the state vector. In the following, $|y_{1:k-1}$ is omitted to lighten the notation.

   $$\mathbb{E}[x_k] = \tilde{\mu}_k = \mathbb{E}[A\ x_{k-1} + B\ u_k + w_k] = A\ \mu_{k-1} + B\ u_k$$

   and for the covariance matrix for the state vector $\tilde{P}_k = \mathbb{E}[(x_k - \tilde{\mu}_k)(x_k - \tilde{\mu}_k)^T]$:

   $$\begin{aligned} \tilde{P}_k &= \mathbb{E}[(A(x_{k-1} - \mu_{k-1}) + w_k)(A(x_{k-1} - \mu_{k-1}) + w_k)^T] \\ &= A\mathbb{E}[(x_{k-1} - \mu_{k-1})(x_{k-1} - \mu_{k-1})^T]A^T + \mathbb{E}[w_k w_k^T] \\ &= AP_{k-1}A^T + Q \end{aligned}$$

2. (Update step) At step k the new measurement vector y_k is available. Show that $x_k|y_{1:k} \sim \mathcal{N}(\mu_k, P_k)$ where :

$$\begin{aligned} \mu_k &= \tilde{\mu}_k + K_k(y_k - C\tilde{\mu}_k) \\ P_k &= (I_{n_x} - K_k C)\tilde{P}_k \\ K_k &= \tilde{P}_k C^T (C\tilde{P}_k C^T + R)^{-1} \end{aligned} \tag{3}$$

where I_n denotes the identity matrix of size n (Hint : conditional distribution of jointly Gaussian distributed variables).

Answer : Remember that for two jointly Gaussian random variables distributed as :

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \tilde{\mu}_1 \\ \tilde{\mu}_2 \end{bmatrix}, \begin{bmatrix} \tilde{\Sigma}_{11} & \tilde{\Sigma}_{12} \\ \tilde{\Sigma}_{21} & \tilde{\Sigma}_{22} \end{bmatrix} \right)$$

the conditional distribution of x_1 on x_2 is also Gaussian, $x_1|x_2 \sim \mathcal{N}(\mu_1, \Sigma_1)$ where :

$$\begin{aligned} \mu_1 &= \tilde{\mu}_1 + \tilde{\Sigma}_{12}\tilde{\Sigma}_{22}^{-1}(x_2 - \tilde{\mu}_2) \\ \Sigma_1 &= \tilde{\Sigma}_{11} - \tilde{\Sigma}_{12}\tilde{\Sigma}_{22}^{-1}\tilde{\Sigma}_{21} \end{aligned}$$

In the case of the update step of the Kalman filter, we know that :

$$\begin{bmatrix} x_k|y_{1:k-1} \\ y_k|y_{1:k-1} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \tilde{\mu}_k \\ C\tilde{\mu}_k \end{bmatrix}, \begin{bmatrix} \tilde{P}_k & \tilde{P}_k C^T \\ C\tilde{P}_k & C\tilde{P}_k C^T + R \end{bmatrix} \right)$$

Given the joint Gaussian distribution above, the posterior distribution of x_k conditioned on the process data $y_{1:k}$ is given by $x_k|y_{1:k} \sim \mathcal{N}(\mu_k, P_k)$ where :

$$\begin{aligned} \mu_k &= \tilde{\mu}_k + K_k(y_k - C\tilde{\mu}_k) \\ P_k &= (I_{n_x} - K_k C)\tilde{P}_k \\ K_k &= \tilde{P}_k C^T (C\tilde{P}_k C^T + R)^{-1} \end{aligned}$$

3. (Alternative expressions) Show that the alternative expressions below are equivalent to (3).

$$\begin{aligned} \mu_k &= P_k(\tilde{P}_k^{-1}\tilde{\mu}_k + C^T R^{-1} y_k) \\ P_k^{-1} &= \tilde{P}_k^{-1} + C^T R^{-1} C \end{aligned} \tag{4}$$

Answer : In order to show that P_k^{-1} is characterized by the second equation in (4), it suffices to show that $P_k P_k^{-1} = I_{n_x}$.

$$\begin{aligned} P_k P_k^{-1} &= (I_{n_x} - K_k C)\tilde{P}_k(\tilde{P}_k^{-1} + C^T R^{-1} C) \\ &= I_{n_x} - K_k C + \tilde{P}_k C^T R^{-1} C - K_k C\tilde{P}_k C^T R^{-1} C \\ &= I_{n_x} + \tilde{P}_k C^T R^{-1} C - K_k(C\tilde{P}_k C^T + R)R^{-1} C \\ &= I_{n_x} \end{aligned}$$

For μ_k, the expression of μ_k in (4) can be rewritten successively as

$$
\begin{aligned}
\mu_k &= P_k(\tilde{P}_k^{-1}\tilde{\mu}_k + C^T R^{-1} y_k) \\
&= (I_{n_x} - K_k C)\tilde{P}_k(\tilde{P}_k^{-1}\tilde{\mu}_k + C^T R^{-1} y_k) \\
&= \tilde{\mu}_k - K_k C\tilde{\mu}_k + \tilde{P}_k C^T R^{-1} y_k - K_k C\tilde{P}_k C^T R^{-1} y_k \\
&= \tilde{\mu}_k - K_k C\tilde{\mu}_k + [\tilde{P}_k C^T R^{-1} - K_k(C\tilde{P}_k C^T + R)R^{-1} + K_k] y_k \\
&= \tilde{\mu}_k - K_k C\tilde{\mu}_k + K_k y_k \\
&= \tilde{\mu}_k + K_k(y_k - C\tilde{\mu}_k)
\end{aligned}
$$

which is the expression of μ_k in (3).

4. Why can the forecast and update step of KF become computationally expensive and memory-wise greedy when the size of the state n_x and/or the measurement vector n_y becomes large ?

   Answer : Computational complexity : $\mathcal{O}(n_x^3)$ or $\mathcal{O}(n_x^2 n_y)$ / Memory complexity : $\mathcal{O}(n_x^2)$.
   The cost of computing $\tilde{P}_k$ is $\mathcal{O}(n_x^3)$ if A is dense. If A is not dense, then the cost of computing $\tilde{P}_k$ depends on the sparsity of A, but in any case, $\tilde{P}_k$ is in general dense for all $k \geq 1$, and the cost of computing is $\mathcal{O}(n_x^2 n_y)$.

   It is clear that for the cases where n_x and/or n_y is large, the computational and memory needs grow significantly.

b) Ensemble Kalman Filter (EnKF)

Suppose that we have N samples (ensemble) following the distribution (2). Denote the samples (ensemble) as $\hat{x}_{k-1}^i,\ i = 1, \ldots, N$.

1. (Forecast step) Apply the state equation in (1) on each member of the ensemble to obtain $\tilde{x}_k^i,\ i = 1, \ldots, N$ as $\tilde{x}_k^i = A\ \hat{x}_{k-1}^i + B\ u_k + w^i$ where $w^i \sim \mathcal{N}(0, Q)$ is a noise realization added to the forecast of the ensemble member. Show that the forecast distribution $\tilde{x}_k^i | y_{1:k-1}$ is the same as the one you obtained in the forecast step of the KF (question a.1).

   Answer : In the following, $|y_{1:k-1}$ is omitted to lighten the notation.

$$
\mathbb{E}[\tilde{x}_k^i] = \mathbb{E}[A\ \hat{x}_{k-1}^i + B\ u_k + w_k] = A\ \mu_{k-1} + B\ u_k = \tilde{\mu}_k
$$

$$
\begin{aligned}
\mathrm{Cov}(\tilde{x}_k^i) &= \mathbb{E}[(\tilde{x}_k^i - \tilde{\mu}_k)(\tilde{x}_k^i - \tilde{\mu}_k)^T] \\
&= \mathbb{E}[(A(\hat{x}_{k-1} - \mu_{k-1}) + w_k)(A(\hat{x}_{k-1} - \mu_{k-1}) + w_k)^T] \\
&= A\mathbb{E}[(\hat{x}_{k-1} - \mu_{k-1})(\hat{x}_{k-1} - \mu_{k-1})^T]A^T + \mathbb{E}[w_k w_k^T] \\
&= AP_{k-1}A^T + Q \\
&= \tilde{P}_k
\end{aligned}
$$

2. (Update step) Apply the measurement equation in (1) on each member of the ensemble to obtain the forecast output vectors $\tilde{y}_k^i, \; i = 1, \ldots, N$ as $\tilde{y}_k^i = C \; \tilde{x}_k^i - v^i$, where $v^i \sim \mathcal{N}(0, R)$ is a measurement noise realization. Given the new measurement vector y_k, compute the update of each ensemble member as $\hat{x}_k^i = \tilde{x}_k^i + K_k(y_k - \tilde{y}_k^i)$. Show that $\hat{x}_k^i | y_{1:k} \sim \mathcal{N}(\mu_k, P_k)$, with μ_k and P_k defined in (3).

   Answer : In the following, $|y_{1:k}$ is omitted to lighten the notation.

$$
\begin{aligned}
\mathbb{E}[\hat{x}_k^i] &= \mathbb{E}[\tilde{x}_k^i + K_k(y_k - C\tilde{x}_k^i + v^i)] \\
&= \tilde{\mu}_k + K_k(y_k - C\tilde{\mu}_k) \\
&= \mu_k \\
\mathrm{Cov}(\hat{x}_k^i) &= \mathrm{Cov}(\tilde{x}_k^i) + \mathrm{Cov}(K_k\tilde{y}_k^i) - \mathrm{Cov}(\tilde{x}_k^i, K_k\tilde{y}_k^i) - \mathrm{Cov}(K_k\tilde{y}_k^i, \tilde{x}_k^i) \\
&= \tilde{P}_k + K_k(C\tilde{P}_kC^T + R)K_k^T - 2K_k(C\tilde{P}_kC^T + R)K_k^T \\
&= \tilde{P}_k - K_kC\tilde{P}_k \\
&= P_k
\end{aligned}
$$

3. Considering i) the forecast step of the EnKF, ii) update step of the EnKF as well as iii) the fact that the Kalman gain in (3) is often replaced by an estimate $\hat{K}_k = \hat{\tilde{P}}_k C^T (C\hat{\tilde{P}}_k C^T + R)^{-1}$ with $\hat{\tilde{P}}_k$ the covariance of the forecast ensemble $\tilde{x}_k^i, \; i = 1, \ldots, N$, how does the EnKF tackle the computational and memory storage issues that occur in the case of the KF (question a.4) ?

   Answer : Assume that $n_x \approx n_y$ and $n_x \gg N$. Computational complexity : $\mathcal{O}(n_x^2 N)$ / Memory complexity : $\mathcal{O}(n_x N)$.

   Compared to the complexities of the Kalman filter, it is clear that when the state of measurement vector becomes large, then considering only N samples from the initial state distribution restricts significantly the computational and memory needs.