Ecole Polytechnique de Louvain

[LINMA1731] - Stochastic processes : estimation and prediction

Kalman Filter: Theorical Questions

Professors :
Absil Pierre-Antoine
Vandendorpe Luc

Authors :
Debray Grégoire - 8196 17 00
Delcoigne Ben - 3877 17 00

Année académique 2020-2021

1 Alternative derivation of Kalman Filter

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

$$\begin{aligned} x_k &= Ax_{k-1} + Bu_k + w_k \\ y_k &= Cx_k + v_k \end{aligned} \tag{1}$$

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 $v_k \sim \mathcal{N}(0, R)$, with $R \in \mathbb{R}^{n_y \times n_y}$.

1.1 Forecast step

The distribution of $x_{k-1}|y_{1:k-1}$ is known. The goal is to find the distribution of $x_k|y_{1:k-1}$ Equation (1) give that

$$x_k|y_{1:k-1} = Ax_{k-1}|y_{1:k-1} + Bu_k + w_k$$

and with the following distribution

$$\begin{aligned} &x_{k-1}|y_{1:k-1} \sim \mathcal{N}(\mu_{k-1}, P_{k-1}) \\ &w_k \sim \mathcal{N}(0, Q) \end{aligned}$$

Using the proprieties of the normal distribution give the result :

$$x_k|y_{1:k-1} \sim \mathcal{N}(A\mu_{k-1} + Bu_k, AP_{k-1}A^T + Q)$$

By identification, it gives

$$\begin{aligned} \widetilde{\mu}_k &= A\mu_{k-1} + Bu_k \\ \widetilde{P}_k &= AP_{k-1}A^T + Q \end{aligned}$$

1.2 Update step

The Kalman Filter equations gives

$$\hat{x}_k|y_{1:k} = \hat{x}_k|y_{1:k-1} + K_k\widetilde{y}_k|y_{1:k-1}$$

with

$$\begin{aligned} \widetilde{y}_k|y_{1:k-1} &\triangleq y_k - \hat{y}_k|y_{1:k-1} \\ &= y_k - C\hat{x}_k|y_{1:k-1} \end{aligned}$$

These r.v. are distributed as

$$\begin{aligned} &\hat{x}_k|y_{1:k-1} \sim \mathcal{N}(\widetilde{\mu}_k, \widetilde{P}_k) \\ &\widetilde{y}_k|y_{1:k-1} \sim \mathcal{N}(y_k - C\widetilde{\mu}_k, C\widetilde{P}_k) \end{aligned}$$

Then the use of the proprieties of the normal distribution gives

$$\hat{x}_k|y_{1:k} \sim \mathcal{N}(\widetilde{\mu}_k + K_k(y_k - C\widetilde{\mu}_k), \widetilde{P}_k + K_kC\widetilde{P}_k)$$

1.3 Alternative expressions

Say

$$\mu_k = P_k(\widetilde{P}_k^{-1}\widetilde{\mu}_k + C^T R^{-1} y)$$

By substitution of P_k

$$\mu_k = (\widetilde{P}_k - K_k C\widetilde{P}_k) \cdot (\widetilde{P}_k^{-1}\widetilde{\mu}_k + C^T R^{-1} y_k)$$

Application of the double distributivity give

$$\mu_k = \widetilde{\mu}_k + (\widetilde{P}_k C^T R^{-1} - K_k C\widetilde{P}_k C^T R^{-1}) y_k - K_k C\widetilde{\mu}_k$$

Which must be the same as

$$\mu_k = \widetilde{\mu}_k + K_k(y_k - C\widetilde{\mu}_k$$

To proof that the following equation must be satisfied (Kalman gain)

$$K_k = \widetilde{P}_k C^T R^{-1} - K_k C\widetilde{P}_k C^T R^{-1}$$

By isolating K_k

$$K_k R + K_k C\widetilde{P}_k C^T = \widetilde{P}_k C^T R^{-1}$$
$$K_k = \widetilde{P}_k C^T R^{-1} (C\widetilde{P}_k C^T + R)^{-1}$$

Which is indeed the equation of the Kalman gain

1.4 Computationally expensive and memory-wise greedy for the KF

For the forecast step, the computation of $\widetilde{P}_k$ require the product of three square matrices of dimention $n_x \times n_x$ therefore the problem is computationally expensive. Furthermore, that computation require at least the storage of three $n_x \times n_x$ matrices which can take a huge amount of storage. For the update step the computation of K_k is the most computationally expensive and memory-wise greedy. Indeed, the computation require 4 matrix products and the inversion of the factor $(C\widetilde{P}_k C^T + R)$ also take a huge place in the memory when n_x or n_y become large.

2 Ensemble Kalman Filter (EnKF)

Suppose $\hat{x}_{k-1}^i$known for $i = 1 : N$ Applying the state equation give

$$\widetilde{x}_k^i = A\hat{x}_{k-1}^i + Bu_k^i + w^i$$

2.1 Forecast step

The goal is to proof that $\widetilde{x}_k^i | y_{1:k-1}$ follows the same distribution that $\widetilde{x}_k | y_{1:k-1}$.
For the expectation

$$\begin{aligned}
\mu_k^i &= \mathbb{E}[\hat{x}_k^i] \\
&= \mathbb{E}[Ax_{k-1}^i + Bu_k + w^i] \\
&= A\mu_{k-1} + Bu_k \\
&= \widetilde{\mu}_k
\end{aligned}$$

For the variance

$$
\begin{aligned}
P_k^i &= \mathbb{E}[(x_k^i - \mu_k^i) \cdot (x_k^i - \mu_k^i)^T] \\
&= \mathbb{E}[(Ax_{k-1}^i + Bu_k + w^i - A\mu_{k-1} - Bu_k) \cdot (Ax_{k-1}^i + Bu_k + w^i - A\mu_{k-1} - Bu_k)^T] \\
&= \mathbb{E}[(A(x_{k-1}^i - \mu_{k-1}) + w^i) \cdot (A(x_{k-1}^i - \mu_{k-1}) + w^i)^T] \\
&= A\mathbb{E}[(x_{k-1}^i - \mu_{k-1}) \cdot (x_{k-1}^i - \mu_{k-1})^T]A^T + \mathbb{E}[w^i] \\
&= AP_{k-1}A^T + Q \\
&= \widetilde{P}_k
\end{aligned}
$$

2.2 Update step

The goal is to proof that $\widetilde{x}_k^i|y_{1:k}$ follows the same distribution that $\widetilde{x}_k|y_{1:k}$.
For the expectation

$$
\begin{aligned}
\mu_k^i &= \mathbb{E}[\hat{x}_k^i] \\
&= \mathbb{E}[(I_{nx} - K_kC)x_{k-1}^i + K_ky_k] \\
&= (I_{nx} - K_kC)\mathbb{E}[x_{k-1}^i] + \mathbb{E}[K_ky_k] \\
&= (I_{nx} - K_kC)\widetilde{\mu}_k + K_ky_k \\
&= \mu_k
\end{aligned}
$$

For the variance

$$
\begin{aligned}
P_k^i &= \mathbb{E}[(x_k^i - \mu_k^i) \cdot (x_k^i - \mu_k^i)^T] \\
&= \mathbb{E}[(Ax_{k-1}^i + Bu_k + w^i - A\mu_{k-1} - Bu_k) \cdot (Ax_{k-1}^i + Bu_k + w^i - A\mu_{k-1} - Bu_k)^T] \\
&= \mathbb{E}[(A(x_{k-1}^i - \mu_{k-1}) + w^i) \cdot (A(x_{k-1}^i - \mu_{k-1}) + w^i)^T] \\
&= A\mathbb{E}[(x_{k-1}^i - \mu_{k-1}) \cdot (x_{k-1}^i - \mu_{k-1})^T]A^T + \mathbb{E}[w^i] \\
&= AP_{k-1}A^T + Q \\
&= \widetilde{P}_k
\end{aligned}
$$

2.3 Computationnality and memory storage

The EnKF reduces the computaion and the memory storage because the factor $\widetilde{P}_k$ is not explicitly computed. Indeed the reduction of the dimention give a little lost in precision but a huge gain in memory and computation.