Projet LINMA1731

Georgios Birpoutsoukis, Jehum Cho

24/02/2021

1 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)).
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).
3. (Alternative expressions) Show that the alternative expressions below are equivalent to (3).

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

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 ?

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).
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).
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) ?

Algorithm 1: Stochastic EnKF

Start with an initial ensemble $\hat{x}_0^i,\ i = 1, \ldots, N$;
Repeat the following two steps;
1. Forecast Step : Calculate $\tilde{x}_k^i$ from $\hat{x}_{k-1}^i$ using state equation;
2. Update Step : $\hat{x}_k^i = \tilde{x}_k^i + \hat{K}_k(y_k - \tilde{y}_k^i)$

---

2 Practical Models Implementation

2.1 KF and EnKF for linear Double-tank system

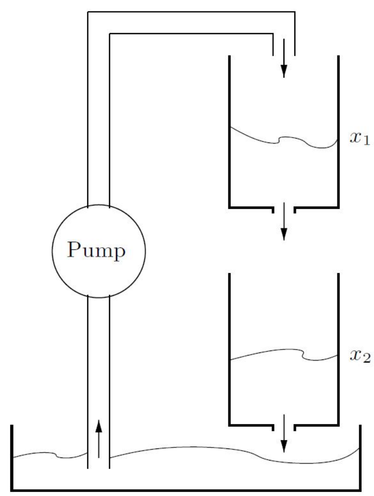

FIGURE 1 – A schematic view of the double-tank process.

Consider the double-tank process depicted in Figure 1. This process is characterized by the two states x_1 and x_2, which are the height of the water in the upper tank and that in the lower tank, respectively. The input signal u is the voltage to the pump that regulates the flow and the output signal $y = x_2$ is the level in the lower tank. A nonlinear continuous-time state-space description derived from Bernoulli's energy equation of the process is :

$$\frac{d}{dt}\begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} = \begin{bmatrix} -\alpha_1\sqrt{x_1(t)} + \beta u(t) \\ \alpha_1\sqrt{x_1(t)} - \alpha_2\sqrt{x_2(t)} \end{bmatrix}, \tag{5}$$

where $x = [x_1 \ x_2]^T$, u is the input signal, and α_1 and α_2 are the areas of the output flows of both tanks.

This continuous-time nonlinear model can be discretized and linearized into the following model :

$$x(k+1) = \begin{bmatrix} 0.9512 & 0 \\ 0.0476 & 0.9512 \end{bmatrix} x(k) + \begin{bmatrix} 0.0975 \\ 0.0024 \end{bmatrix} u(k) + w(k) \tag{6}$$

$$y(k) = \begin{bmatrix} 0 & 1 \end{bmatrix} x(k) + v(k), \tag{7}$$

where $w(k), v(k)$ are process noise following $\mathcal{N}(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, 10^{-3}\begin{bmatrix} 9.506 & 0.234 \\ 0.234 & 9.512 \end{bmatrix})$ and measurement noise following $\mathcal{N}(0, 0.0125)$, respectively. The sampling period is

$h = 0.1s$ and the prior distribution of initial conditions $[x_1(0)\ x_2(0)]^T$ follows $\mathcal{N}(\begin{bmatrix}5\\5\end{bmatrix}, \begin{bmatrix}1 & 0\\0 & 1\end{bmatrix})$. The goal is to estimate $x_1(k)$ and $x_2(k)$ on the basis of the noisy measurements of $y(k)$. Implement the KF and EnKF for the linearized version of the double-tank process using the given data (Section 2.3). In the case of EnKF, try different choices for the ensemble size. Report graphs to illustrate :

1. the real state trajectories (provided to you in the data)
2. the filtered states (result of the update step)
3. the 95% confidence intervals ($\pm 1.96\sigma$) around the filtered states (use of diagonal of updated covariance matrix)
4. Root-Mean-Square Deviation (RMSD) as a function of k.

All results must be properly analyzed and commented.

2.2 EnKF for nonlinear Double-tank system

The continuous-time state-space equations (9) can be discretized using the Runge-Kutta method to obtain the following discrete-time nonlinear state-space system of equations :

$$\begin{aligned}\begin{bmatrix}x_1(k+1)\\x_2(k+1)\end{bmatrix} &= \begin{bmatrix}x_1(k)\\x_2(k)\end{bmatrix} + \frac{h}{6}(f_1 + 2f_2 + 2f_3 + f_4) + \begin{bmatrix}w_1(k)\\w_2(k)\end{bmatrix}\\ y(k) &= x_2(k) + v(k)\end{aligned} \tag{8}$$

where

$$f(x(k), u(k)) = \begin{bmatrix}-\alpha_1\sqrt{x_1(k)} + \beta u(k)\\ \alpha_1\sqrt{x_1(k)} - \alpha_2\sqrt{x_2(k)}\end{bmatrix}, \tag{9}$$

$f_1 = f(x(k), u(k))$, $f_2 = f(x(k) + h\frac{f_1}{2}, u(k))$, $f_3 = f(x(k) + h\frac{f_2}{2}, u(k))$, $f_4 = f(x(k) + hf_3, u(k))$ and $h = 0.1s$ is the sampling period. The terms $w(k)$ and $v(k)$ denote the disturbance/noise added to the states and output, respectively, where $w(k)$ and $v(k)$ are distributed same as in the previous question. As such, given the set of parameters $\alpha_1 = \alpha_2 = \beta = 1$, as well as a pair of initial conditions $[x_1(0)\ x_2(0)]^T$, we can use (8) to simulate the nonlinear dynamic system.

Similar to the previous question, the goal is also here to use the measured data of the state x_2, i.e. y, and the EnKF in order to obtain the trajectories of the states x_1 and x_2. The EnKF allows the use of the nonlinear state-space (8) without the need to linearize around an operation point. To be precise, you can apply the forecast step of the EnKF by applying directly the nonlinear equation of the state evolution (8) to each ensemble member (do not omit to add the noise realization w^i, as in the linear case). Implement the EnKF for the nonlinear case using the given data (Section 2.3). For the initialization of the state vector $[x_1(0)\ x_2(0)]^T$, consider the same distribution as in the previous question.

Report graphs to illustrate :

1. the real state trajectories (provided to you in the data)
2. the filtered states (result of the update step)
3. the 95% confidence intervals ($\pm 1.96\sigma$) around the filtered states (use of diagonal of updated covariance matrix)
4. Root-Mean-Square Deviation (RMSD) as a function of k.

All results must be properly analyzed and commented.

2.3 Detailed Description of Given Data

Regarding question 2.1 :

1. the file "True_state_x1_linear_case.txt" contains the real state trajectory of x_1
2. the file "True_state_x2_linear_case.txt" contains the real state trajectory of x_2
3. the file "Input_linear_case.txt" contains the input signal u
4. the file "Measured_output_linear_case.txt" contains the measured output y

Regarding question 2.2 :

1. the file "True_state_x1_nonlinear_case.txt" contains the real state trajectory of x_1
2. the file "True_state_x2_nonlinear_case.txt" contains the real state trajectory of x_2
3. the file "Input_nonlinear_case.txt" contains the input signal u
4. the file "Measured_output_nonlinear_case.txt" contains the measured output y

Practical details

Modality	The project is carried out by groups of two students. If you need to work alone or do not know anybody to work with, please contact us to find an arrangement. Each group must register on Moodle by Friday 12 March 2021, 23.59 pm.
Supervision	Office hours on Teams from week 7 to week 13 : one hour of permanence on Monday 5 pm and Thursday 2 pm.
Report	English is strongly recommended. But the course is French friendly, hence French is allowed without penalty. However, reports in French, if any, will have to go through a different grading procedure. The goal is not to evaluate your English skills and we will therefore not pay attention to the quality of the language, but to the scientific quality of your report instead. Maximum 10 pages for both of the parts (e.g. 5 pages each).
Deadline	Part I : Friday 23 April 2021 at 18.15 pm on Moodle. For the mid-term deadline, you just need to submit the report of the first theoretical part — and not the code — (in pdf) with the filename `LINMA1731_2021_Project_Part1_NAME1_NAME2.pdf`. Part II : Friday 14 May 2021 at 18.15 pm on Moodle. The report (in pdf format) and the code (in Python .py or .ipynb) will be submitted together in a zip file named `LINMA1731_2021_Project_Part2_NAME1_NAME2.zip`.