Statistics for Data Science | Exercise 13
Wintersemester 2023/24
Please complete these problems before the exercise session on Tuesday 30 January, 2024, 8:30. Please be prepared to present your solutions to any problems that you completed successfully.

1. Let $y \in \mathbb{R}^2$ the measurement, $x \in \mathbb{R}$ the unknown, and

$$y = \begin{pmatrix} 2 \\ 1 \end{pmatrix} x + \eta, \quad \eta \sim \mathcal{N}(0, \gamma^2 I_2),$$

where $I_2 \in \mathbb{R}^{2\times 2}$ is an identity matrix and $\gamma > 0$. Suppose that the prior distribution is given by $x \sim \mathcal{N}(0, 2)$, with x and η assumed to be independent. What is the posterior distribution if we observe $y = (1, 2)^{\mathrm{T}}$? What is the posterior variance? What happens to the posterior distribution and variance under decreasing noise ($\gamma \downarrow 0$)?

2–3. (This task is worth 2 points.) Let $M \in \mathbb{R}^{d\times d}$ and $H \in \mathbb{R}^{k\times d}$. Suppose that we have a sequence of measurements $\{y_j\}_{j\geq 1} \subset \mathbb{R}^k$ which correspond to a sequence of unknown states $\{x_j\}_{j\geq 1} \subset \mathbb{R}^d$.

(a) Suppose that the states obey an *evolution model*

$$x_{j+1} = Mx_j + \xi_{j+1}, \quad \xi_{j+1} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \Sigma), \tag{1}$$

where $\Sigma \in \mathbb{R}^{d\times d}$ is a symmetric and positive (semi)definite covariance matrix. If $x_j \sim \mathcal{N}(m_j, C_j)$, where $m_j \in \mathbb{R}^d$ and $C_j \in \mathbb{R}^{d\times d}$ is symmetric and positive (semi)definite, what is the distribution of x_{j+1}?

(b) Suppose that we have an *observation model*

$$y_{j+1} = Hx_{j+1} + \eta_{j+1}, \quad \eta_{j+1} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \Gamma), \tag{2}$$

where $\Gamma \in \mathbb{R}^{k\times k}$ is a symmetric and positive (semi)definite covariance matrix. The measurement y_{j+1} is given, with x_{j+1} and η_{j+1} assumed to be independent. Using the distribution you obtained in part (a) as the prior for x_{j+1}, what is the posterior distribution of $x_{j+1}|y_{j+1}$?

(c) Consider the *evolution-observation model* (1)–(2) and suppose that we are interested in finding the probability distribution of $x_{j+1}|y_1, \ldots, y_{j+1}$ (i.e., we wish to estimate the state at some future time step $j + 1$ given measurements at all previous time steps $1, 2, \ldots, j + 1$). Consider the following updating scheme:

(i) Set $j = 0$ and initialize $x_0 \sim \mathcal{N}(m_0, C_0)$ using some known mean $m_0 \in \mathbb{R}^d$ and symmetric, positive (semi)definite covariance $C_0 \in \mathbb{R}^{d\times d}$.

(ii) (*Prediction*) Define x_{j+1} using the evolution model (1). Then $x_{j+1} \sim \mathcal{N}(\widehat{m}_j, \widehat{C}_j)$, where $\widehat{m}_j$ and $\widehat{C}_j$ are the mean and covariance you derived in part (a).

The exercises continue on the next page!

(iii) (*Correction*) Using $x_{j+1} \sim \mathcal{N}(\widehat{m}_j, \widehat{C}_j)$ from step (ii) as the prior, we can obtain the posterior $x_{j+1}|y_1, \ldots, y_{j+1} \sim \mathcal{N}(m_{j+1}, C_{j+1})$ from the observation model (2), where m_{j+1} and C_{j+1} are the mean and covariance you derived in part (b).

(iv) Set $j = j + 1$ and return to step (ii).

This algorithm is known as the *Kalman filter*. It produces the so-called *filtering distributions* $x_{j+1}|y_1, \ldots, y_{j+1} \sim \mathcal{N}(m_{j+1}, C_{j+1})$ for $j = 0, 1, 2, \ldots$

Your task is to implement this algorithm numerically for the following model problem:

We wish to track the state $x_k := \begin{bmatrix} p_k \\ v_k \end{bmatrix} \in \mathbb{R}^2$ of a moving particle. The first component p_k corresponds to the position of the particle while the second component v_k is its velocity at time $k = 0, 1, 2, \ldots$. You may assume that you know the initial state of the particle perfectly: $x_0 = \mathbb{E}[x_0] = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \in \mathbb{R}^2$ and $C_0 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \in \mathbb{R}^{2\times 2}$. The evolution model for the particle is given by $M = \begin{bmatrix} 1 & \Delta t \\ 0 & 1 \end{bmatrix} \in \mathbb{R}^{2\times 2}$, with time step $\Delta t = 0.01$, and the innovation term is given by $\Sigma = \begin{bmatrix} \frac{1}{4}(\Delta t)^4 & \frac{1}{2}(\Delta t)^3 \\ \frac{1}{2}(\Delta t)^3 & (\Delta t)^2 \end{bmatrix} \in \mathbb{R}^{2\times 2}$. Meanwhile, we only measure the location of the particle so the observation model is given by $H = \begin{bmatrix} 1 & 0 \end{bmatrix} \in \mathbb{R}^{1\times 2}$ and the observational noise variance is assumed to be $\Gamma = [1] \in \mathbb{R}^{1\times 1}$.

Implement the Kalman filter for this model problem and plot the filtered positions $(t_k, \mathbb{E}[p_k|y_1, \ldots, y_k])_{k=1}^{2000}$ and velocities $(t_k, \mathbb{E}[v_k|y_1, \ldots, y_k])_{k=1}^{2000}$ as a function of time $t_k = k\Delta t$, $k = 1, \ldots, 2000$. To simulate the noisy measurements, you may assume that the true trajectory of the particle is given by the function $x(t) = 0.1(t^2 - t)$ for $t \in [0, 20]$, and the measurements are given by $y_k = x(t_k) + \eta_k$, where $\eta_k \sim \mathcal{N}(0, \Gamma)$ is additive i.i.d. noise for $k = 1, \ldots, 2000$.

Hint: Since all intermediate distributions in the Kalman filter algorithm are Gaussian (as long as the initial distribution for x_0 is Gaussian), from a computational point of view, we only need to keep track of the means and covariances using the update formulae you derived in parts (a) and (b).

4. Solve problem 3(c) using the ensemble Kalman filter with ensemble size $N = 100$. The covariance Σ of the innovation term in problem 3(c) is positive semidefinite. In order to avoid problems sampling from the distribution $\mathcal{N}(0, \Sigma)$, you can instead use, e.g., $\Sigma = \begin{bmatrix} \frac{1}{4}(\Delta t)^4 + 10^{-10} & \frac{1}{2}(\Delta t)^3 \\ \frac{1}{2}(\Delta t)^3 & (\Delta t)^2 + 10^{-10} \end{bmatrix}$ as a quick and dirty workaround. Note also that the prior distribution $x_0 \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \right)$ is degenerate, so you should use the initial ensemble $x_0^{(j)} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ for $j = 1, \ldots, N$.