Inverse Problems — Exercise 8

Sommersemester 2023

Return your written solutions either in person or by email to vesa.kaarnioja@fu-berlin.de by Tuesday 27 June, 2023, 10:15

Please note that there are a total of 3 tasks in this exercise sheet.

1. Let $x, y, \eta \in \mathbb{R}$ and consider a simple Bayesian inverse problem

$$y = \frac{1}{2}x + \eta$$

with additive noise $\eta \sim \mathcal{N}(0, 1)$. Assume that the prior model for the unknown is also Gaussian $x \sim \mathcal{N}(0, \frac{1}{\alpha})$, where $\alpha > 0$ is poorly known. It is possible to write the conditional prior for x, given α, as

$$\mathbb{P}(x|\alpha) = \frac{\alpha^{1/2}}{\sqrt{2\pi}} \exp\bigg(-\frac{1}{2}\alpha x^2\bigg).$$

Since the parameter α is not known, it is part of the inference problem. Assume that we set the following hyperprior density for the parameter α:

$$\mathbb{P}(\alpha) = \begin{cases} \sqrt{\frac{2}{\pi}} \exp(-\frac{1}{2}\alpha^2) & \text{if } \alpha > 0, \\ 0 & \text{if } \alpha \leq 0. \end{cases}$$

(a) Show that the posterior density for $(x, \alpha)|y$ is given by

$$\mathbb{P}(x, \alpha|y) \propto \alpha^{1/2} \exp\bigg(-\frac{1}{2}\bigg(y - \frac{1}{2}x\bigg)^2 - \frac{1}{2}\alpha x^2 - \frac{1}{2}\alpha^2\bigg),$$

where the implied coefficient does not depend on x or α.

(b) Show that $(x, \alpha) = (1, 1/2)$ is the *maximum a posteriori* (MAP) estimate when we observe $y = 3/2$.

You may assume in parts (a) and (b) that η and (x, α) are independent.

2. Let $A \in \mathbb{R}^{k \times d}$, $x \in \mathbb{R}^d$, $y, \eta \in \mathbb{R}^k$, and consider the linear measurement model with additive noise:

$$y = Ax + \eta.$$

During the lecture, we proved that if x is endowed with a Gaussian prior distribution $\mathcal{N}(x_0, \Gamma_{\rm pr})$, the noise η is assumed to have the Gaussian distribution $\mathcal{N}(\eta_0, \Gamma_{\rm n})$, and x and η are mutually independent, then the posterior distribution is Gaussian with posterior covariance

$$\Gamma_{\rm post} = (\Gamma_{\rm pr}^{-1} + A^{\rm T}\Gamma_{\rm n}^{-1}A)^{-1} \tag{1}$$

and posterior mean

$$\mu_{\rm post} = \Gamma_{\rm post}(A^{\rm T}\Gamma_{\rm n}^{-1}(y - \eta_0) + \Gamma_{\rm pr}^{-1}x_0). \tag{2}$$

Prove that these can alternatively be written as

$$\Gamma_{\text{post}} = \Gamma_{\text{pr}} - \Gamma_{\text{pr}} A^{\mathrm{T}} (A\Gamma_{\text{pr}} A^{\mathrm{T}} + \Gamma_{\text{n}})^{-1} A\Gamma_{\text{pr}} \tag{3}$$

and

$$\mu_{\text{post}} = x_0 + \Gamma_{\text{pr}} A^{\mathrm{T}} (A\Gamma_{\text{pr}} A^{\mathrm{T}} + \Gamma_{\text{n}})^{-1} (y - Ax_0 - \eta_0). \tag{4}$$

Hint: Use the Sherman–Morrison–Woodbury formula: for *any* conformable matrices A, B, C, and D such that A and C are invertible (square) matrices, there holds

$$(A + BCD)^{-1} = A^{-1} - A^{-1}B(C^{-1} + DA^{-1}B)^{-1}DA^{-1},$$

if $A + BCD$ is invertible (or, equivalently, if $C^{-1} + DA^{-1}B$ is invertible).

Begin by applying the Sherman–Morrison–Woodbury formula on (1); this should yield the formula (3). The formula (4) can then be proved by plugging the formula (3) into (2) and simplifying the resulting expression.

3. 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*

   $$x_{j+1} = Mx_j + \xi_{j+1}, \quad \xi_{j+1} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \Sigma), \tag{5}$$

   where $\Sigma \in \mathbb{R}^{d\times d}$ is a symmetric and positive 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 definite, what is the distribution of x_{j+1}?

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

   $$y_{j+1} = Hx_{j+1} + \eta_{j+1}, \quad \eta_{j+1} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \Gamma), \tag{6}$$

   where $\Gamma \in \mathbb{R}^{k\times k}$ is a symmetric and positive 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* (5)–(6) 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 definite covariance $C_0 \in \mathbb{R}^{d\times d}$.

      (ii) (*Prediction*) Set $x_j = x_0$ if $j = 0$ and $x_j = x_j|y_1, \ldots, y_j$ otherwise. Define x_{j+1} using the evolution model (5). 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).

(iii) (*Correction*) Define $x_{j+1}|y_{j+1}$ via the observation model (6), using $x_{j+1} \sim \mathcal{N}(\widehat{m}_j, \widehat{C}_j)$ from step (ii) as the prior. Then $x_{j+1} \sim \mathcal{N}(m_{j+1}, C_{j+1})$, 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 $\Delta = 0.01$, 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).