Exercise 10 – *Bayesian inference and Data assimilation*

Due by: Tuesday, 27 June 2023, 23:59 (CEST)

Problem 1 We perform very similar question to P3 of the last exercise. Consider the hidden state variable X with prior $X \sim N(0, 2)$. The observable

$$Y = h(X) + W, \qquad h(x) = x^4 + x^2$$

where $W \sim N(0, 1)$. Assume we observe $y = 4$.

1. Find the conditional density $\pi_{X|Y}(x \mid Y = y)$. You may omit the explicit formula for the normalization constant.

2. Numerically find a MAP estimator m using gradient descent method. The idea of gradient descent is as follows: To minimize a function $f(x)$, iterate

$$x_{n+1} = x_n - \lambda_n \nabla f(x_n)$$

   where $\lambda_n > 0$ is a small "learning rate". Iterate this until $|\nabla f|$ is small enough. There are many techniques to adjust the algorithm, but you do not have to apply the latest theory here.

3. Suppose $\pi(x) = \exp(-V(x))$ be a probability density. The Laplace approximation to π is defined as $\tilde{\pi} = N(m, V''(m)^{-1})$, where m is the MAP estimator you found in the previous step. Write down the Laplace approximation to the posterior of X.

4. Use the Langevin method we used in P2 of Exercise 8. Again use 10000 samples and run sufficiently long time until the distribution settles. Play with the terminal time and evaluate how long does it take to reach its invariant distribution. Overlay the histogram and the approximated density you obtained in step 3.

5. Discuss about the two method. You can discuss about, for instance, the accuracy, possible errors, computational effort, etc.

Problem 2 We consider a real-valued random variable X that has a probability density.

1. Show that $c = \mathbb{E}[X]$ minimizes the mean-squared error $\mathbb{E}\big[(X - c)^2\big]$.

2. Show that the median minimizes $\mathbb{E}\big[|X - c|\big]$. The median is defined as a number c such that $\mathbb{P}[X < c] = \mathbb{P}[X > c] = 0.5$.

Problem 3 Let X be a 2-dimensional Gaussian random variable given by $X = (X_1, X_2) \sim N(0, I)$. The observable Y is given by

$$Y = X_1 + X_2 + W$$

where W is an independent random noise given by $W \sim N(0, \sigma)$.

1. Given $Y = y$, what is the conditional distribution of X_1 and X_2? Find the mean and the variance in terms of y and σ.

2. How those two components behaves when $\sigma \to 0$ and $\sigma \to \infty$?