These exercises will not be graded and do not need to be returned.

1. Suppose $\rho_1 \sim \mathcal{N}(m_1, C_1)$ and $\rho_2 \sim \mathcal{N}(m_2, C_2)$. Prove that

$$d_{\mathrm{H}}(\rho_1, \rho_2)^2 = 1 - \frac{\det(C_1)^{1/4}\det(C_2)^{1/4}}{\det\left(\frac{C_1+C_2}{2}\right)^{1/2}}.$$

2. Consider a simple Bayesian inverse problem

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

where $x, \eta \in \mathbb{R}$ and $y \in \mathbb{R}$ is the measurement. Suppose that the unknown x has the prior distribution $x \sim \mathcal{N}(0, 1)$ and this is independent of the observational noise, which has the probability density

$$\pi_{\text{noise}}(\eta) = \begin{cases} 2\exp(-2\eta) & \text{if } \eta \geq 0, \\ 0 & \text{if } \eta < 0. \end{cases}$$

   (a) Derive the posterior density $\pi_{\text{post}}(x|y)$ up to a constant factor, where you can consider the marginal distribution $\pi(y)$ as part of the (non-explicit) normalization constant.

   (b) Solve the *maximum a posteriori estimate* when we observe $y = \frac{1}{2}$.

   (c) Use importance sampling to approximate the conditional mean estimate for this problem.

3. Let $x, y, \eta \in \mathbb{R}^2$. Consider the Bayesian inverse problem

$$y = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} x + \eta$$

with additive noise $\eta \sim \mathcal{N}(0, \gamma^2 I_2)$, where $I_2 \in \mathbb{R}^{2\times 2}$ is an identity matrix. Suppose that the prior distribution is given by $x \sim \mathcal{N}(0, I_2)$, and this is independent of the noise. What is the posterior distribution of $x|y$ if we observe $y = \begin{pmatrix} 1 & 2 \end{pmatrix}^{\mathrm{T}}$? What is the posterior covariance? What happens to the posterior distribution and posterior covariance under decreasing noise ($\gamma \downarrow 0$)?

4. Suppose $A = \operatorname{diag}(\lambda_1, \ldots, \lambda_n) \in \mathbb{R}^{n\times n}$, where $\lambda_1 \geq \cdots \geq \lambda_n > 0$, and let $y \in \mathbb{R}^n$.

   (a) Prove that the mapping $T\colon \mathbb{R}^n \to \mathbb{R}^n$ given by

$$T(x) = x + \beta(A^{\mathrm{T}}y - A^{\mathrm{T}}Ax)$$

   is a contraction when $\beta > 0$ is small enough.

(b) Define the Landweber–Fridman iteration. Describe briefly how it can be used to regularize an ill-posed system

$$Ax = y,$$

where $x \in \mathbb{R}^n$ is the unknown. Why it is necessary in the Landweber–Fridman iteration that the mapping T is a contraction?

5. Let us consider the following inverse problem: suppose that a particle with charge $q \in \mathbb{R}$ is located at some unknown location $x^* \in [0, 1]$ in the interval $[0, 1]$ and our goal is to locate it based on measurements of voltage at the interval end points $x = 0$ and $x = 1$. The voltage at any point $x \in [0, 1]$ is given by

$$y(x) = \frac{q}{|x^* - x|}.$$

Assume that each voltage measurement is corrupted by mutually independent additive normally distributed noise with zero mean and a known variance σ^2, which is the same for both sensor locations. Assume further that we know *a priori* that the particle is within the interval $[0, 1]$ and suppose that our prior information about the charge is that it is normally distributed around some fixed $q_0 \in \mathbb{R}$ with known variance γ^2.

(a) Write down the posterior density $\mathbb{P}(x, q \mid y)$ for the pair (x, q) given the measurement y.

(b) Our goal is to find the location of the particle, so we treat the charge q as a nuisance parameter and marginalize the posterior density with respect to it. Write explicitly the marginal density

$$\mathbb{P}(x \mid y) = \int_{-\infty}^{\infty} \mathbb{P}(x, q \mid y)\, \mathrm{d}q.$$

(c) Suppose that the true values are $(x^*, q) = (1/\pi, 1)$. Using MATLAB, simulate some measurement data using these true parameter values and add normally distributed mean-zero noise to the measurements with some reasonably chosen value for the standard deviation $\sigma > 0$. Assume that $q_0 = 1.1$ and visualize the posterior density, trying out a range of different values for the standard deviation $\gamma > 0$ corresponding to the charge. How does the level of uncertainty in the charge affect the posterior density?

6. 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

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

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:

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

As we saw in exercise 8, the posterior density for $(x, \alpha)|y$ is given by

$$\mathbb{P}(x, \alpha|y) \propto \alpha^{1/2} \exp\left(-\frac{1}{2}\left(y - \frac{1}{2}x\right)^2 - \frac{1}{2}\alpha x^2 - \frac{1}{2}\alpha^2 \right) \quad \text{for } \alpha > 0,$$

where the implied coefficient does not depend on x or α. Moreover, $(x, \alpha) = (1, 1/2)$ is the *maximum a posteriori* (MAP) estimate when we observe $y = 3/2$. Verify numerically that $(x, \alpha) = (1, 1/2)$ is the MAP estimate by trying out the following. Define the negative log-posterior

$$J(x, \alpha) := -\frac{1}{2}\log \alpha + \frac{1}{2}\left(\frac{3}{2} - \frac{1}{2}x\right)^2 + \frac{1}{2}\alpha x^2 + \frac{1}{2}\alpha^2$$

and consider the following alternating minimization algorithm:

- Set $k = 0$ and choose an initial guess for α, e.g., $\alpha_0 = 1$.

  repeat

  - Find the Tikhonov regularized solution

  $$x_k = \arg\min_{x \in \mathbb{R}} J(x, \alpha_k) = \arg\min_{x \in \mathbb{R}} \left\{ \left(\frac{3}{2} - \frac{1}{2}x\right)^2 + \alpha_k x^2 \right\}.$$

  - Solve $\alpha > 0$ from $\frac{\partial J(x_k, \alpha)}{\partial \alpha} = 0$ and set $\alpha_{k+1} = \alpha$.
  - Set $k \leftarrow k + 1$.

  until convergence

Does the sequence (x_k, α_k) approach $(1, 1/2)$?

7. Suppose our inverse problem is given by

$$y = Ax + \eta,$$

where y is our observation and $A \in \mathbb{R}^{N \times J}$ is our matrix modeling the measurement. Moreover, the noise distribution is given by $\eta \sim \mathcal{N}(0, I)$. Suppose we have that $x \sim \mathcal{N}(0, (\tau^2 I + L)^{-1})$, where $L \in \mathbb{R}^{N \times N}$ is a positive definite symmetric matrix. Moreover, the hierarchical parameter $\tau \in \mathbb{R}$ is unknown and is modelled with the density

$$\rho_{\text{hpr}}(\tau) = \begin{cases} C\exp(-\tau), & \text{when } \tau \geq 0 \\ 0, & \text{when } \tau < 0. \end{cases}$$

Write down the posterior distribution. How would you solve the *maximum a posteriori* estimator?

8. Let us consider priors for two-dimensional unknowns (pixel images). Below left is a picture with 4 rows and 5 columns, with pixel values numbered in the Matlab convention. We use zero boundary conditions, indicated in gray. The pixels containing the gray zeros are not part of the actual image.

	0	0	0	0	0	
0	f_1	f_5	f_9	f_{13}	f_{17}	0
0	f_2	f_6	f_{10}	f_{14}	f_{18}	0
0	f_3	f_7	f_{11}	f_{15}	f_{19}	0
0	f_4	f_8	f_{12}	f_{16}	f_{20}	0
	0	0	0	0	0	

			f_{10}			
		f_7	f_{11}	f_{15}		
			f_{12}			

For defining a smoothness prior for the vector $f \in \mathbb{R}^{20}$ we consider two-dimensional convolution with a discrete Laplace operator. This is done by moving a five-point mask over the image so that the location of the mask shown above on the right corresponds to the following element (with index 11) in the result of the convolution: $-4f_{11} + f_7 + f_{10} + f_{12} + f_{15}$. In the above case the matrix would be

$$L = \begin{bmatrix}
-4 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & -4 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & -4 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & -4 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & -4 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 1 & -4 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 1 & -4 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 1 & -4 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -4 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & -4 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & -4 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & -4 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -4 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & -4 & 1 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & -4 & 1 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & -4 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -4 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & -4 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & -4 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & -4
\end{bmatrix}.$$

Produce random samples from the Gaussian prior density

$$\pi_F(f) \propto \exp\left(-\frac{1}{2} f^T \Gamma^{-1} f \right)$$

with different choices of covariance matrix Γ.

(a) White noise prior. Take $\Gamma = I$.

(b) Smoothness prior. Take $\Gamma^{-1} = L^T L$ with L the discrete Laplace matrix.

9. The true state of a time-varying system is $v_k^* = 4 + \sin(0.03k)$, $k = 1, \ldots, 500$. The *observational model* is

$$y_k = v_k + \eta_k, \quad \eta_k \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 0.1^2).$$

Using the random walk *evolution model*

$$v_{k+1} = v_k + \xi_k, \quad \xi_k \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \gamma^2),$$

consider the following tasks:

(a) Simulate the measurements using v_k^* and implement the Kalman filter algorithm to compute the estimates $\mathbb{E}[v_k|y_1,\dots,y_k]$, $k=1,\dots,500$. Use the value $\gamma=0.1$ and an initial distribution $v_0 \sim \mathcal{N}(m_0,\sigma_0^2)$, where $m_0=4$ and $\sigma_0=1$.

(b) Run the Kalman filter with different combinations of m_0, σ_0, and γ and make inferences about their effects on the behavior of the estimated state.

Remark: The Kalman filter was considered in the 8$^{\text{th}}$ exercise sheet.

10. Consider the boundary value problem

$$-\frac{\mathrm{d}}{\mathrm{d}x}\left(a(x)\frac{\mathrm{d}}{\mathrm{d}x}u(x)\right)=1 \quad \text{for } x\in(0,1),$$
$$u(0)=0,$$
$$u'(1)=0.$$

It is well-known that this problem can be solved for u as

$$u(x)=\int_0^x \frac{1-y}{a(y)}\,\mathrm{d}y. \tag{1}$$

Let $x_k=hk$, $h=1/100$, $k=0,\dots,100$. The integral in (1) can be discretized using the trapezoidal rule as

$$\int_0^{x_k} F(y)\,\mathrm{d}y \approx h\sum_{i=1}^{k}\frac{F(x_i)+F(x_{i-1})}{2} \quad \text{for } k=1,\dots,n \quad \text{with } F(y):=\frac{1-y}{a(y)}.$$

This leads to the discrete measurement model

$$\boldsymbol{u}=G\frac{1}{\boldsymbol{a}}, \tag{2}$$

where $G\in\mathbb{R}^{100\times 101}$, $\boldsymbol{u}=[u(x_1),\dots,u(x_{100})]^{\mathrm{T}}$, $\boldsymbol{a}=[a(x_0),\dots,a(x_{100})]^{\mathrm{T}}$, and $\frac{1}{\boldsymbol{a}}=\mathtt{1./}\boldsymbol{a}=\left(\frac{1}{a(x_{i-1})}\right)_{i=1}^{101}$ denotes the elementwise reciprocal vector of $\boldsymbol{a}$.

Let us consider the inverse problem of recovering $\boldsymbol{a}$ based on noisy measurements $\boldsymbol{u}$ using the statistical inversion paradigm. Download the file `pde.mat` from the course website and run

```
load pde u
```

in MATLAB.[1] The vector `u` contains the values of u at the grid points $x_1,\dots,x_{100}$, contaminated with i.i.d. additive Gaussian noise with mean 0 and standard deviation 0.1 % relative to the maximal data component. Here, you can estimate the noise level of the measurements as

$$\sigma=10^{-3}\cdot\max_{\mathtt{i,j}=1,\dots,100}|\mathtt{u(i)}-\mathtt{u(j)}|.$$

[1] In Python: `import scipy.io` and run `u = scipy.io.loadmat('pde.mat')['u']`

In addition, suppose that we know *a priori* that the unknown coefficient $a(x)$ is very smooth. This suggests using a smoothness prior

$$\pi_{\mathrm{pr}}(\boldsymbol{a}) \propto \exp\left(-\frac{1}{2\omega^2}\|L\boldsymbol{a}\|^2 \right), \quad \omega > 0.$$

In order to avoid being overly committal with the boundary values, let us consider the so-called Aristotelian prior with

$$L = \begin{bmatrix} \delta & 0 & & & & \\ -1 & 2 & -1 & & & \\ & -1 & 2 & -1 & & \\ & & \ddots & \ddots & \ddots & \\ & & & -1 & 2 & -1 \\ & & & & 0 & \delta \end{bmatrix} \in \mathbb{R}^{101\times 101}, \quad \delta = 0.005.$$

The value $\delta = 0.005$ has been chosen here somewhat heuristically.

(a) Construct the system matrix $G \in \mathbb{R}^{100\times 101}$ in MATLAB (the commands `ones`, `eye`, `tril`, and `diag` may be useful) and write down the explicit formulae of the likelihood and posterior densities for this inverse problem.

(b) Explain why the maximum a posteriori (MAP) estimate for the problem (2) can be obtained by solving the minimization problem

$$\boldsymbol{a}_{\mathrm{MAP}} = \arg\min_{\boldsymbol{a}\in\mathbb{R}^{101}} \left\{ \left\| \boldsymbol{u} - G\frac{1}{\boldsymbol{a}} \right\|^2 + \lambda^2 \|L\boldsymbol{a}\|^2 \right\}, \quad \lambda = \frac{\sigma}{\omega}.$$

Define the objective function $S(\boldsymbol{a}) := \|\boldsymbol{u} - G\frac{1}{\boldsymbol{a}}\|^2 + \lambda^2\|L\boldsymbol{a}\|^2$ and consider the following algorithm for solving the minimization problem.

Gauss–Newton algorithm. Write the objective function as

$$S(\boldsymbol{a}) = \sum_{i=1}^{201} r_i(\boldsymbol{a})^2, \quad \text{where } r_i(\boldsymbol{a}) = \begin{cases} (\boldsymbol{u} - G\frac{1}{\boldsymbol{a}})_i & \text{if } 1 \le i \le 100, \\ \lambda(L\boldsymbol{a})_{i-100} & \text{if } 101 \le i \le 201. \end{cases}$$

Starting from an initial guess $\boldsymbol{a}^{(0)}$ for the minimum, iterate

$$\boldsymbol{a}^{(k+1)} = \boldsymbol{a}^{(k)} - J^{+} r(\boldsymbol{a}^{(k)}) \quad \text{for } k = 0, 1, 2, \ldots,$$

where $r(\boldsymbol{a}) = [r_1(\boldsymbol{a}), \ldots, r_{201}(\boldsymbol{a})]^{\mathrm{T}}$, $J = (J_{i,j})_{1\le i\le 201,\ 1\le j\le 101}$ is the Jacobi matrix of $r(\boldsymbol{a}^{(k)})$ defined elementwise as

$$J_{i,j} = \frac{\partial r_i}{\partial a_j}(\boldsymbol{a}^{(k)}) \quad \text{for } 1 \le i \le 201,\ 1 \le j \le 101, \tag{3}$$

and J^{+} denotes the Moore–Penrose pseudoinverse of the matrix J (see the command `pinv` in MATLAB / `numpy.linalg.pinv` in Python).
(See also: `https://en.wikipedia.org/wiki/Gauss%E2%80%93Newton_algorithm`)

For this problem, the Jacobi matrix (3) can be written (using MATLAB notation) as $\texttt{J} = [\texttt{G} * \texttt{diag}(\texttt{power}(\texttt{a}, -2)); \texttt{lambda} * \texttt{L}]$, where `a` denotes the k^{th} iterate $\boldsymbol{a}^{(k)}$ and `lambda` stands in for λ.

Implement the Gauss–Newton algorithm and compute the (approximate) MAP estimate $\boldsymbol{a}_{\text{MAP}}$. In order to find a good value for $\omega > 0$, use the Morozov discrepancy principle, i.e., ensure that the condition $\|\boldsymbol{u} - G\frac{1}{\boldsymbol{a}}\| \approx \sigma\sqrt{100}$ holds approximately. Here, it is generally a good idea to use $\boldsymbol{a}^{(0)} = \texttt{ones}(101, 1)$ as the initial guess for your experiments.

(c) Define what is meant by the conditional mean (CM) estimate of $\boldsymbol{a}$. Then, using your favorite MCMC method (for example, the random walk Metropolis–Hastings algorithm works well here), compute the (approximate) CM estimate of $\boldsymbol{a}$ using the value for ω that you obtained in part (b). Compare your CM reconstruction with the MAP estimate you obtained in part (b). Do they look alike? How did you assess the convergence and quality of your MCMC sampler?