
Problem for Day 5. *In this exercise we will create a smooth prior and use it to solve the hydraulic conductivity problem.*

To create a smooth prior we use a covariance kernel $f : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ and define a covariance matrix $\mathcal{C}$. We can the use this covariance matrix to define a Gaussian random variable.

*Use the code **`exercise_1.py`** in day 5 folder and follow the instructions below*

1. *create a discretization of the interval $[0, 1]$ as a vector $\mathbf{s}$ and store it in a Python variable `s`.*

2. *create a python function **`gaussian_cov_func`** to evaluate the Gaussian covariance kernel:*
$$f(\mathbf{x}, \mathbf{x}') = \exp(-\frac{|\mathbf{x} - \mathbf{x}'|^2}{2\ell^2})$$
with $\ell = 0.05$ the correlation length.

3. *create a covariance matrix $\mathcal{C}$ with elements:*
$$[\mathcal{C}]_{ij} = f(\mathbf{s}_i, \mathbf{s}_j).$$

4. *use **`np.random.multivariate_normal`** to draw 5 samples from $\mathcal{N}(0, \mathcal{C})$. Plot the samples.*

5. *repeat the experiment for $\ell = 0.2, 0.1$ and 0.05. And plot the samples. How does ℓ effects the samples?*

6. *now repeat the experiment above with the exponential kernel:*
$$f(\mathbf{x}, \mathbf{x}') = \exp(-\frac{|\mathbf{x} - \mathbf{x}'|}{\ell})$$
How does the behavior of the samples change?

7. *Often in inverse problems, it is preferred to work with independent samples. However, components of $\mathcal{N}(0, \mathcal{C})$ are inherently correlated. We can perform a change of variable and create a set of independent variables using the change of variables*
$$X = \mathcal{C}^{1/2} Z \tag{1}$$
where X is the original random variable, $\mathcal{C}^{1/2}$ is the Cholesky factor of $\mathcal{C}$, and now, Z is a new variable with distribution $\mathcal{N}(0, I_n)$. For the Gaussian kernel above draw 5 samples by first drawing $Z \sim \mathcal{N}(0, I_n)$ and then multiplying them by the $\mathcal{C}^{1/2}$.

we will now use this prior to formulate the posterior distribution of the hydraulic imaging problem.

The code **`hydraulic.py`** *contains the numerical approximation of the forward operator. You can follow the code* **`exercise_2.py`** *to see how you can pass a conductivity (the unknown* $\mathbf{x}$*) to create 5 pressure profiles for 5 number of injections. The code will return a matrix where the i-th row corresponds to the discrete pressure profile for i-th injection/measurement.*

For the Hydraulic inverse problem we consider the statistical inverse problem:

$$Y_i = F_i(X) + E_i, \quad i = 1, \dots 5 \tag{2}$$

where F_i *is the forward operator of the ith injection (in the code, the* **`hydraulic.py`** *will perform all* F_i *simultanuously), And* $E_i = \mathcal{N}(0, \sigma_i I_n)$*, for* $i = 1, \dots, 5$*. Remember that the noise standard deviations* σ_i *are different for each measurement.*

8. *For a prior distribution* $X \sim N(0, \mathcal{C})$ *write the posterior distribution, up to constant of proportionality, i.e.,*

$$\pi_{X|Y_1=\mathbf{y}_1,\dots,Y_5=\mathbf{y}_5}(\mathbf{x}) \propto \cdots \tag{3}$$

   Hint: Use the fact that the 5 measurements are independent, and therefore,

$$\pi_{Y_1,\dots,Y_5|X}(\mathbf{y}_1, \dots, \mathbf{y}_5) = \pi_{Y_1|X}(\mathbf{y}_1) \cdots \pi_{Y_5|X}(\mathbf{y}_5).$$

9. *Reformulate and rewrite the posterior using the change of variable* (1)*, i.e.,*

$$\pi_{Z|Y_1=\mathbf{y}_1,\dots,Y_5=\mathbf{y}_5} \propto \cdots$$

   Recall that in this case the prior distribution is simply

$$Z \sim \mathcal{N}(0, I_n)$$

10. *One measurement data is given to you in the file* **`obs.pickle`**. *The Python code* **`exercise_3.py`** *will read this file and create 2 variables* **`y_obs`** *and* **`sigma`** *which will hold the measurements* $\mathbf{y}_1, \dots, \mathbf{y}_5$ *and the noise standard deviations* $\sigma_1 \dots, \sigma_5$.

11. *complete the code in* **`exercise_3.py`** *to sample the posterior. Plot the posterior mean and point-wise posterior variance. Where in the solution do you suspect that there is a sudden increase in porosity?*

    These are steps to complete **`exercise_3.py`**

    (a) *create a discretization of the interval* $[0, 1]$ *in a vector* `s` *with* `N=100` *points.*

    (b) *initiate the hydraulic class with* `N` *with the command* `hydraulic = hydraulic_class(N)` .

    (c) *Create a covariance matrix of the prior using a Gaussian covariance kernel with length scale 0.1, i.e.,* `length_scale=0.1`.

    (d) *create the Cholesky factor of the covariance matrix.*

    For part (c) and (d) you can copy the code from the previous exercise.

(e) *create a Python function that evaluates the log of the un-normalized prior density $\pi_Z(\mathbf{z})$ (standard normal distribution).*

(f) *load the measurement file* `obs.pickle`.

(g) *create a Python function that evaluates the log of the un-normalized likelihood density.*

(h) *combine the two to create a Phython function that evaluates the log of posterior.*

(i) *perform the random walk Metropolis-Hastings sampling method to draw sample 10000 samples from the posterior. The step size $c = 0.003$ is appropriate for this problem.*

for (i) you can copy the code from the exercises in day 4.

(j) *compute the mean and the point-wise variance.*

12. ***(optional)*** *what happens to uncertainty in estimation when pressure sensors are broken in half of the domain? You can simulate faulty pressure sensors by throwing away half of the output of* `hydraulic.forward`, *i.e., the forward operator becomes*

```
hydraulic_broken = lambda x:  hydraulic.forwad(x)[:,N/2:]
```

similarly you should discard half of the measurements in `y_obs`, *i.e.,*

```
y_obs_broken = y_obs[:,N/2:]
```