Prof. Dr. Jens Honore Walther
Dr. Georgios Arampatzis
ETH Zentrum, CLT
CH-8092 Zürich

Set 2

Issued: 12.03.2021

In this exercise, you will learn about least squares fitting of data points and the sensitivity of the fit to the data quality and the amount of noise. In **Notebook 2.1** you will perform least squares fitting on 2D and 3D data and look into the sensitivity of LSQ to noise and outliers. In **Notebook 2.2** you will look into different ways to solve the linear system of equations present in least squares fitting and gauge the effect of the condition number.

Question 1: Linear least squares on 3D data

We will now perform a least squares fit using 3-D data $\{x_i, y_i, z_i\}_{i=1}^N$ and consider a function $z(x, y) = \alpha + \beta x + \gamma y$ with unknown coefficients α, β and γ.

a) Write the problem in the matrix formulation. Using the same procedure as in the 2D case given in the lecture notes in section (1.4.3), derive the least squares solution for the 3x3 matrix.

An applied counterpart to this question can be found in Notebook 2.1.

Question 2: [Advanced] Dependency of the LSQ fit on the noise

In this question you will understand how the LSQ fit behaves as a function of the number of data points (N) in the presence of noisy data. Let x_i, $i = 1, \ldots, N$ be points from a bounded domain (i.e., from an interval). Consider the data points $y_i = \alpha_0 + \beta_0 x_i$, $i = 1, \ldots, N$, and the noisy data $y_i^* = y_i + \epsilon_i$ where $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$. Proceed as follows:

1. Consider the model $f(x)$ and $f^*(x)$ with parameters α, β and α^*, β^*, that correspond to the data (x_i, y_i) and (x_i, y_i^*), respectively.
2. Express the difference between the two models in terms of differences in the parameters, $\alpha - \alpha^*$ and $\beta - \beta^*$.
3. Express $\alpha - \alpha^*$ in terms of elements of the matrix $H = A^T A$. Do the same for $\beta - \beta^*$.
4. Find out how the elements of matrix H^{-1} behave as a function of N

Hint: Assume x_i is bounded. Use the Central Limit Theorem.