Exercises for Course on Modeling and System Identification (MSI)
Albert-Ludwigs-Universität Freiburg – Winter Term 2023-2024

Exercise 4: Regularized, Ill-Posed and Weighted Linear Least-Squares

(to be returned on Nov 27st, 8:30)

Prof. Dr. Moritz Diehl, Katrin Baumgärtner, Jakob Harzer, Yizhen Wang,
Adithya Anoop Thoniparambil, Premnath Srinivasan

The aim of this sheet is to strengthen your knowledge in least squares estimation and introduce some basic properties about quadratic functions and how they relate to weighted linear least-squares.

Exercise Tasks

1. PAPER: We would like to estimate a constant $\theta_0 \in \mathbb{R}$ that is corrupted by additive zero-mean Gaussian noise, i.e. we assume the following model

$$y = \theta_0 + \epsilon$$

where $\epsilon \sim \mathcal{N}(0, \sigma_\epsilon^2)$. To this end, we use *regularized* linear least-squares, i.e. we compute the estimate $\hat{\theta}_\mathrm{R}$ given by

$$\hat{\theta}_\mathrm{R} = \arg\min_{\theta \in \mathbb{R}} \frac{1}{2}\|y - \Phi\theta\|_2^2 + \frac{\alpha}{2}\|\theta\|_2^2$$

where $\theta \in \mathbb{R}$, $\Phi = (1, \dots, 1)^\top \in \mathbb{R}^{N \times 1}$ and $\alpha > 0$. From the lecture, we know that the solution to this optimization problem is given by

$$\hat{\theta}_\mathrm{R} = \left(\Phi^\top \Phi + \alpha \mathbb{I}\right)^{-1} \Phi^\top y$$

   (a) Calculate the expected value $\mathbb{E}\{\hat{\theta}_\mathrm{R}\}$ of $\hat{\theta}_\mathrm{R}$. Is the estimator unbiased and/or asymptotically unbiased?
   Hint: Check Section 4.5.1. of the lecture notes. (1 points)

   (b) Calculate the variance $\mathrm{var}(\hat{\theta}_\mathrm{R})$ of $\hat{\theta}_\mathrm{R}$. Compare with the unregularized case, i.e. $\alpha = 0$.
   Hint: Check Section 4.5.2. of the lecture notes. (1 points)

2. You are given the following ill-posed Linear Least-Squares problem:

$$\hat{\theta} = \arg\min_{\theta \in \mathbb{R}} \frac{1}{2}\|y - \Phi\theta\|_2^2 \qquad y = \begin{bmatrix} -3 \\ \vdots \\ 5 \end{bmatrix} \in \mathbb{R}^9 \qquad \Phi = \begin{bmatrix} \frac{1}{6} & \frac{1}{6} \\ \vdots & \vdots \\ \frac{1}{6} & \frac{1}{6} \end{bmatrix} \in \mathbb{R}^{9 \times 2} \qquad \theta \in \mathbb{R}^2$$

In the PYTHON script you will find code that visualizes the minimization problem in 3D.

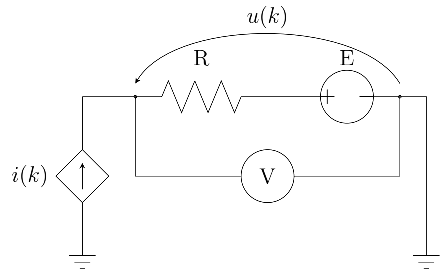
   (a) PAPER: Why is this an ill-posed problem? What issue do you run into when following the usual LLS approach of $\hat{\theta} = (\Phi^\top \Phi)^{-1} \Phi^\top y$? (0.5 points)

   (b) PAPER: Which two approaches do you know to solve this issue? (0.5 points)

   (c) CODE: Find a $\hat{\theta}$ using both methods from (b). Use $\alpha = 0.2$. (1 point)

   (d) PAPER: The original minimization problem is visualized in a figure with the two solutions (your $\hat{\theta}$ from the previous task) as red x. Why do the solutions end up where they are? Give a reason for each solution! (1 point)

3. Recall the resistance estimation example from the last exercise sheet. Again, we consider the following experimental setup, to find the values of the parameters R and E:



We assume that only our measurements of the voltage are corrupted by noise, i.e. we make the following model assumption:

$$u(k) = R \cdot i(k) + E + n_u(k)$$

where $n_u(k) \sim \mathcal{N}(0, \sigma_u^2(k))$ follows a zero-mean Gaussian distribution. You are given the data of N_e students, each of them performed the same experiment where they measured the voltage $u(k)$ for increasing values of $i(k)$, $k = 1, \ldots, N_\mathrm{m}$.

Unfortunately, the fan of your measuring device is broken. Thus, it starts heating up over the course of the experiment which decreases the accuracy of your measurements such that later measurements are much noisier than earlier ones.

(a) Paper: In the template we already provided a plot showing the measurements from all students. What do you observe?
To account for the decreasing accuracy of your measuring device, you decide to assume that the noise variance $\sigma_u^2(k)$ is proportional to the timestep k, i.e.

$$\sigma_u^2(k) = c \cdot k, \;\; k = 1, \ldots, N_\mathrm{m},$$

where c is a constant. How do you make use of this assumption to modify the LLS estimator? (1 point)

(b) Code: For student 1, perform both linear least-squares (LLS) and weighted linear least-squares (WLS) to obtain estimates of the parameter $\theta = [R, E]^\top$. Plot the data of student 1, as well as the fit obtained from LLS and WLS in a single figure. *Hint: for coding purpose, you can compute the weighting matrix assuming that $c = 1$.* (0.5 points)

(c) Paper: Which estimator fits better and why? (0.5 points)

(d) Code: For each student $d = 1, \ldots, N_\mathrm{e}$, compute $\theta_{\mathrm{LLS}}^{(d)}$ and $\theta_{\mathrm{WLS}}^{(d)}$. (0.5 points)

(e) Code: Estimate the mean and covariance matrix of the random variables θ_{LLS} and θ_{WLS} by calculating the sample mean $\bar{\theta}_{*\mathrm{LS}} = \frac{1}{N_\mathrm{e}} \sum_{d=1}^{N_\mathrm{e}} \theta_{*\mathrm{LS}}^{(d)}$ and the sample covariance matrix $\Sigma_{*\mathrm{LS}}$ that is given by

$$\Sigma_{*\mathrm{LS}} = \frac{1}{N_\mathrm{e} - 1} \sum_{d=1}^{N_\mathrm{e}} \left(\theta_{*\mathrm{LS}}^{(d)} - \bar{\theta}_{*LS}\right) \left(\theta_{*\mathrm{LS}}^{(d)} - \bar{\theta}_{*LS}\right)^\top .$$

Here $*$LS refers to LLS and WLS. (0.5 points)

(f) Code: Plot $\theta_{\mathrm{LLS}}^{(d)}$ and $\theta_{\mathrm{WLS}}^{(d)}$, $d = 1, \ldots, N_\mathrm{e}$, where the x-axis corresponds to the estimated R_0 values and the y-axis corresponds to the estimated E_0 values. Plot the mean and 1σ-confidence ellipsoids for both θ_{LLS} and θ_{WLS} in the same figure. (1 point)

(g) Paper: What do you observe? (1 point)

(h) Paper: In part (a) we assumed that the noise is proportional to k. Does θ_{WLS} depend on the choice of the proportionality factor c? Why (not)? (2 bonus points)

This sheet gives in total 10 points and 2 bonus points.