Exercises for Course on Modeling and System Identification (MSI)
Albert-Ludwigs-Universität Freiburg – Winter Term 2021-2022

# Exercise 4: Weighted Linear Least-Squares

**(to be returned on November 29, 2021, 9:00)**

Prof. Dr. Moritz Diehl, Katrin Baumgärtner, Yizhen Wang, David Berrazueta, Mohammed Hababeh

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. ON PAPER: We would like to find the parameters $\hat{\theta}_{\text{LS}}$ of a linear model $y_k = \phi_k \cdot \theta + \epsilon_k$, where $\epsilon_k \sim \mathcal{N}(0, \sigma_\epsilon^2)$ is an additive i.i.d. zero-mean Gaussian noise that perturbed a series of $N$ scalar measurements $y = [y_1, \dots, y_N] \in \mathbb{R}^N$ . From the lecture we know that $\hat{\theta}_{\text{LS}}$ can be computed using least-squares:

$$\hat{\theta}_{\text{LS}} = \arg\min_{\theta} \frac{1}{2}\|y - \Phi\theta\|_2^2$$

where $\Phi \in \mathbb{R}^{N \times d}$. Assume that $\sigma_\epsilon^2$ is known.

   (a) Define the matrix $\Phi$ and state the closed form solution of least squares problem. $\hat{\theta}_{\text{LS}} = \dots$

   (b) Calculate the covariance of the least squares estimate. $\text{cov}\left(\hat{\theta}_{\text{LS}}\right) = \dots$

*Hint: Recall from Exercise 2 that the covariance matrix of a vector-valued variable $Y = AX + b$ for a constant $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$ is given by* $\text{cov}(Y) = A \, \text{cov}(X) \, A^\top$.

(2 points)

2. ON PAPER: Consider a series of $N$ scalar measurements $y = [y_1, \dots, y_N] \in \mathbb{R}^N$ and a linear model $y_k = \phi_k \cdot \theta + \epsilon_k$, where $\epsilon_k \sim \mathcal{N}(0, \sigma_k^2)$ and all $\sigma_k^2, \;\; k = 1, \dots, N$ can be different, e.g. time depending. The measurements thus are perturbed by additive independent zero-mean noise that is not identically distributed. In order to give a lower weight to the measurements with stronger noise, we introduce a weightning positive definite matrix $W \in \mathbb{R}^{N\times N}$. Consider the following weighted least-squares optimization problem (WLS)

$$\min_{\theta \in \mathbb{R}^d} \quad \frac{1}{2}\|r\|_W^2 = \frac{1}{2} r^\top W r \tag{1}$$

where $r = [r_1, \dots, r_N] \in \mathbb{R}^N$ is the vector of the prediction errors $r_k = y_k - \phi_k \cdot \theta$ , the regression vectors are denoted by $\phi_1, \dots, \phi_N \in \mathbb{R}^d$ and the unknown parameters by $\theta \in \mathbb{R}^d$ .

   (a) Please re-write the WLS optimization problem (1) as an unweighted LLS problem, i.e. specify $\tilde{y}$ and $\tilde{\Phi}$ such that

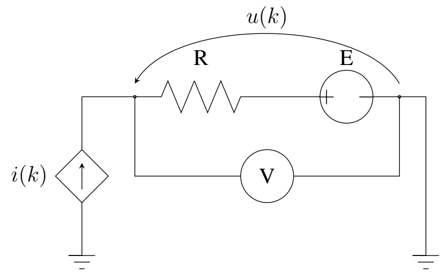
$$\min_{\theta \in \mathbb{R}^d} \quad \frac{1}{2}\|\tilde{y} - \tilde{\Phi}\theta\|_2^2 \; = \; \min_{\theta \in \mathbb{R}^d} \quad \frac{1}{2} r^\top W r$$

(1 point)

   *Hint: The reformulation found above and (1) are equivalent, use either one for (b) and (c)*

   (b) Is it a convex problem? Prove it. (1 point)

3. Recall the resistance estimation example from the last exercise sheet. Again, we consider the following experimental setup:




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

$$u(k) = R_0 i(k) + E_0 + n_u(k) \quad \sigma_k^2 = c \cdot k,\ k = 1, \dots, N_m$$

where $n_u(k) \sim \mathcal{N}(0, \sigma_k^2)$ 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 $i(k)$, $k = 1, \dots, N_m$.

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) ON PAPER: On Grader 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 $\sigma_k^2$ is proportional to the timestep $k$, i.e.

$$\sigma_k^2 = c \cdot k,\ k = 1, \dots, N_m,$$

for some fixed value of $c$ (you may start by chossing $c = 1$). How do you make use of this assumption when applying weighted linear least-squares? (1 point)

(b) MATLAB: For student 1, perform both linear least-squares (LLS) and weighted linear least-squares (WLS) to obtain estimates of the parameter $\theta_0 = [E_0, R_0]^\top$. Plot the data of student 1, as well as the fit obtained from LLS and WLS in a single figure.

ON PAPER: What do you observe? (1 point)

(c) MATLAB: For each student $d = 1, \dots, N_e$, compute $\theta_{\text{LLS}}^{(d)}$ and $\theta_{\text{WLS}}^{(d)}$. (1 point)

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

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

Here $*$LS refers to LLS and WLS. (1 point)

(e) MATLAB: Plot $\theta_{\text{LLS}}^{(d)}$ and $\theta_{\text{WLS}}^{(d)}$, $d = 1, \dots, N_e$, where the $x$-axis corresponds to the estimated $R$ values and the $y$-axis corresponds to the estimated $E$ values.

Plot the mean and $1\sigma$-confidence ellipsoids for both $\theta_{\text{LLS}}$ and $\theta_{\text{WLS}}$ in the same figure.

ON PAPER: What do you observe? (1 point)

(f) ON PAPER: In part (b) we assumed that the measurement noise is proportional to $k$. Does $\theta_{\text{WLS}}$ depend on the choice of the proportionality factor? Why (not)? (1 point)

*This sheet gives in total 10 points.*