Exercises for Course on Modeling and System Identification (MSI)
Albert-Ludwigs-Universität Freiburg – Winter Term 2022-2023

Exercise 4: Weighted Linear Least-Squares

(to be returned on Nov 21st, 8:00)

Prof. Dr. Moritz Diehl, Katrin Baumgärtner, Jakob Harzer, Yizhen Wang, Rashmi Dabir

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 find the parameters $\hat{\theta}_{\text{LS}}$ of a linear model $y(k) = \phi(k)^\top \theta + \epsilon(k)$, 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 $y_N = [y(1), \ldots, y(N)] \in \mathbb{R}^N$. From the lecture we know that $\hat{\theta}_{\text{LS}}$ can be computed using least-squares:
$$\hat{\theta}_{\text{LS}} = \arg\min_{\theta} \frac{1}{2} \|y_N - \Phi_N \theta\|_2^2$$
where $\Phi_N \in \mathbb{R}^{N \times d}$. Assume that σ_ϵ^2 is known.

   (a) State the matrix Φ_N and the solution of least squares problem $\hat{\theta}_{\text{LS}}$.

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

   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 $\text{cov}(Y) = A \,\text{cov}(X)\, A^\top$.

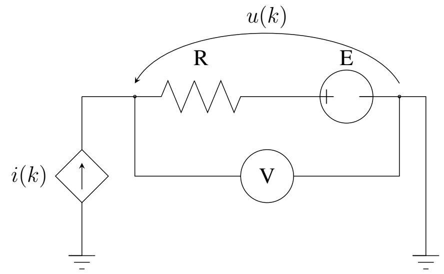
   (2 points)

2. PAPER: Consider a series of N scalar measurements $y_N = [y(1), \ldots, y(N)] \in \mathbb{R}^N$ and a linear model $y(k) = \phi(k)^\top \theta + \epsilon(k)$, where $\phi(k) \in \mathbb{R}^d$, $\theta \in \mathbb{R}^d$, $\epsilon(k) \sim \mathcal{N}(0, \sigma_\epsilon^2(k))$. 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 weighting matrix $W \in \mathbb{R}^{N \times N}$ which is positive definite. 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 $r = [r(1), \ldots, r(N)] \in \mathbb{R}^N$ is the vector of the prediction errors $r(k) = y(k) - \phi(k)^\top \theta$.

   (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)

   (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:

$$u(k) = R_0 i(k) + E_0 + n_u(k)$$

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 $i(k)$, $k = 1, \ldots, N_m$.

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.

$$\sigma_u^2(k) = c \cdot k, \ k = 1, \ldots, N_m,$$

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 $\theta = [R_0, E_0]^\top$. 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* $c = 1$.

PAPER: Which estimator fits better and why? (1 point)

(c) CODE: For each student $d = 1, \ldots, N_e$, compute $\theta_{\text{LLS}}^{(d)}$ and $\theta_{\text{WLS}}^{(d)}$. (0.5 point)

(d) CODE: Estimate the mean and covariance matrix of the random variables θ_{LLS} and θ_{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. (0.5 point)

(e) CODE: Plot $\theta_{\text{LLS}}^{(d)}$ and $\theta_{\text{WLS}}^{(d)}$, $d = 1, \ldots, N_e$, 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.

PAPER: What do you observe? (1.5 point)

(f) PAPER: In part (a) we assumed that the measurement noise is proportional to k. Does θ_{WLS} depend on the choice of the proportionality factor c? Why (not)? (1.5 point)

This sheet gives in total 10 points.