

**Exercise 3: Optimality Conditions and Linear Least Squares**  
(to be returned on November 11th, 8:15)

Prof. Dr. Moritz Diehl, Katrin Baumgärtner, Jakob Harzer, Dan Wang,  
Ashwin Karichannavar, Premnath Srinivasan

---

The aim of this sheet is to strengthen your knowledge in optimality conditions and least squares estimation.

**Exercise Tasks**

1. ON PAPER: Given the function  $f(x) : \mathbb{R}^n \rightarrow \mathbb{R}$  with  $f(x) = x^\top Q x + c^\top x$  and fixed  $c \in \mathbb{R}^n$ .
  - (a) Consider the not necessarily symmetric matrix  $Q \in \mathbb{R}^{n \times n}$  and compute the gradient  $\nabla f(x) \in \mathbb{R}^n$  and the Hessian  $\nabla^2 f(x) \in \mathbb{R}^{n \times n}$  of this function for any  $x$ .
  - (b) If  $Q$  is symmetric, what properties does it have to fulfil such that the unique minimizer  $x^*$  can be computed?
  - (c) Compute the unique minimizer and the minimum function value  $f(x^*)$  under the correct assumptions. (3 points)
  
2. ON 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  $y_N = [y(1), \dots, y(N)] \in \mathbb{R}^N$ . From the lecture we know that  $\hat{\theta}_{\text{LS}}$  can be computed using least-squares:

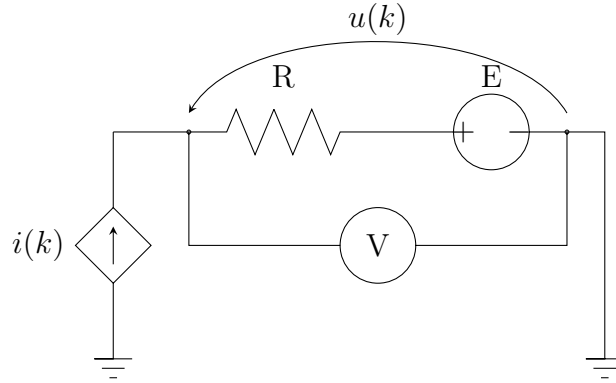
$$\hat{\theta}_{\text{LS}} = \arg \min_{\theta} \frac{1}{2} \|y_N - \Phi_N \theta\|_2^2$$

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

- (a) State the matrix  $\Phi_N$  and the solution of least squares problem  $\hat{\theta}_{\text{LS}}$ .
- (b) Calculate the covariance of the least squares estimate  $\text{cov}(\hat{\theta}_{\text{LS}})$ .

*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)*

3. Consider the following experimental set up to estimate the values of  $E$  and  $R$ .



You obtain a dataset containing  $N$  measurements of the voltage  $u(k)$  for different values of  $i(k)$ . We assume that the input measurements  $i(k)$  is not affected by noise, but that the measurements  $u(k)$  are affected by i.i.d. additive noise  $n_u(k)$ . Under these assumptions the measurement model is given by:

$$u(k) = m(k) + n_u(k) \quad \text{where} \quad m(k) = E + R \cdot i(k).$$

- CODE: Load the dataset containing the measurements. Plot the data with 'x' markers and add labels to both axis. (0.5 points)
- ON PAPER: Formulate the problem of estimating the parameters  $E$  and  $R$  as a least squares problem where  $\theta = [E \ R]^\top \in \mathbb{R}^2$  by defining  $\Phi \in \mathbb{R}^{N \times 2}$  and  $y \in \mathbb{R}^N$  such that the optimizer is given by  $\hat{\theta} = \arg \min_{\theta} \|y - \Phi\theta\|_2^2$ . (1 points)
- CODE: Use the least squares estimator formulated in the previous subtask to find the experimental values of  $R$  and  $E$  for the dataset. Plot the linear fit through the measurement data. (1 points)
- CODE: Plot a histogram of the residuals defined as  $r(k) = u(k) - m(k)$ , where  $m(k) = E + R \cdot i(k)$  is the voltage determined by the model, and  $u(k)$  are the obtained measurements. (0.5 points)
- ON PAPER: Give an educated guess for the type of noise distribution. (0.5 points)

As we will learn later in the lecture, for i.i.d. measurement noise, the covariance matrix of the  $d$  estimated parameters  $\hat{\theta}$  from a single experiment can be approximated by

$$\hat{\Sigma}_{\hat{\theta}} = \frac{\|y - \Phi\hat{\theta}\|_2^2}{N - d} (\Phi^\top \Phi)^{-1}$$

From this estimated covariance matrix, the estimated variances for the single components of the parameter vector can be extracted as the respective diagonal entry.

- CODE: Calculate the estimated variance of both the  $R$  and  $E$  estimate and print their values along with a  $1\text{-}\sigma$  confidence estimate in the format of

$$(\text{quantity}) = (\text{value}) \pm (1 \text{ standard deviation})$$

(1.5 points)

*This sheet gives 10 points in total.*