Exercises for Lecture Course on Modeling and System Identification (MSI)
Albert-Ludwigs-Universität Freiburg – Winter Term 2019

Exercise 3: Optimality Conditions and Introduction to Linear Least Squares

(to be returned on Nov 20, 2019, 08:30 in SR 00-036 (Schick-Saal), or before in building 102, 1st floor, 'Anbau')

Prof. Dr. Moritz Diehl, Tobias Schöls, Naya Baslan, Jakob Harzer, Bryan Ramos

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

Exercise Tasks

1. ON PAPER: Given the function $f(x) : \mathbb{R}^n \to \mathbb{R}$ 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.

   (d) Show that if Q is symmetric and positive definite, x^* will be global minimizer if and only if x^* satisfies FONC. (4 points)

   Hint: You can re-write a matrix-vector product $b = Qx$ as $b_i = \sum_{j=1}^n Q_{ij}\, x_j, \quad for \quad i = 1, \dots, m.$

2. ON PAPER: Consider the function $f(x) : \mathbb{R}^2 \to \mathbb{R}$ with $f(x) = x^\top \begin{bmatrix} 4 & 1 \\ 1 & 2 \end{bmatrix} x + x^\top \begin{bmatrix} 3 \\ 4 \end{bmatrix} + 7.$

   (a) Compute the gradient $\nabla f(x)$ and the Hessian $\nabla^2 f(x)$ of $f(x)$.

   (b) Find all points that satisfy the first order necessary conditions (FONC). Which of them is the global minimizer and why?

   (2 points)

3. ON PAPER: In the lecture notes, the sample variance S^2 is defined as

$$S^2 = \frac{1}{N-1} \sum_{n=1}^{N} (Y(n) - M(Y(N)))^2,$$

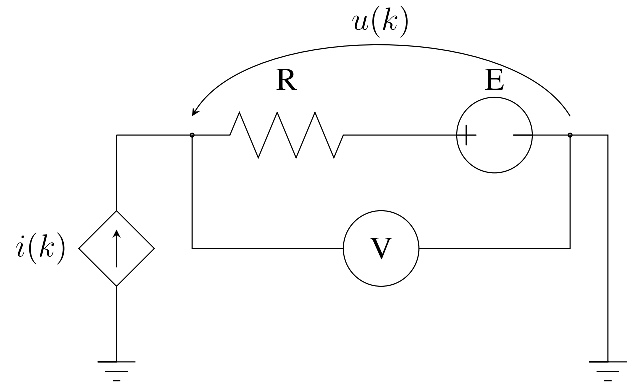
   where $M(Y(N))$ is the sample mean (see lecture notes ch. 2.4 p. 17). Explain, why the division by $N-1$ is preferable over N. (2 points)

   Hint: Calculate the expected value of the sample variance and compare it to the expected value of the mean squared deviations estimator.

4. MATLAB: Implement a function called `myVar(vec)` that computes the sample variance for a given vector `vec`. Use the formula from the previous exercise task to implement the variance. (1 point)

   Hint: You can use `MATLAB`'s `var` command to check your implementation.

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



You obtain two datasets each containing N measurements of the voltage $u(k)$ for different values of $i(k)$. The first dataset contains $\{u_1(k)\}_{k=1}^N$ and $\{i_1(k)\}_{k=1}^N$ and the second dataset contains $\{u_2(k)\}_{k=1}^N$ and $\{i_2(k)\}_{k=1}^N$. For cleaner and simpler notation, we omit the dataset indices, e.g. instead of $u_1(k)$ and $u_2(k)$ we write $u(k)$ but mean both.

We assume that the input measurement $i(k)$ is not affected by noise, but that the measurements $u(k)$ are affected by i.i.d. additive noise $n_{\mathrm{u}}(k)$. Under these assumptions the measurement model is given by:

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

Tasks: (4 points)

(a) MATLAB: Load the datasets provided on the website containing the measurements into MATLAB. Plot each dataset in a corresponding plot using the `subplot` command.

(b) ON PAPER: Formulate the problem as a least squares problem where $\theta = \begin{bmatrix} E \\ R \end{bmatrix}$ and define $\Phi \in \mathbb{R}^{N\times 2}$ and $y \in \mathbb{R}^N$ such that the optimizer is given by $\theta^* = \arg\min_\theta \|y - \Phi\theta\|_2^2$.

(c) MATLAB: Use the least squares estimator formulated in the previous subtask to find the experimental values of R and E for each of the two datasets individually.
Plot the linear fits through the respective measurement data.

(d) MATLAB: For each dataset plot a histogram of the residuals defined as $r(k) = m(k) - u(k)$, where $m(k) = E + R \cdot i(k)$ is the voltage determined by the model, and $u(k)$ are the obtained measurements.

(e) ON PAPER: Which dataset is noisier? Give an educated guess of the type of noise.

This sheet gives 13 points in total.