Exercises for Lecture Course on Modeling and System Identification (MSI)
Albert-Ludwigs-Universität Freiburg – Winter Term 2016

Exercise 1: Brief Introduction to Statistics

(to be returned on Oct 31, 2016, 8:15 in SR 00-010/014, or before in building 102, 1st floor, 'Anbau')

Prof. Dr. Moritz Diehl, Robin Verschueren, Rachel Leuthold, Tobias Schöls, Mara Vaihinger

In this exercise you get to know how to fit a curve to a dataset. In addition, you investigate some important facts from statistics in numerical experiments.

For the the MATLAB exercises, create a MATLAB script called `main.m` with your code, possibly calling other functions/scripts. From running this script, all the necessary results and plots should be clearly visible. Compress all the files/functions/scripts necessary to run your code in a `.zip` file and send it to `msi.syscop@gmail.com`. Please state your name and the names of your team members in the e-mail.

Exercise Tasks

1. Consider the following experimental setup, where we measure the temperature-dependent expansion of a steel bar. Here L_0 [cm] is the length of the bar at the beginning of the experiment and $L(T)$ [cm] represents the length of the bar at temperature T [K]. The following relationship holds, between the length of the bar at temperature T_0 [K]: $L_0 = L(T_0)$. We define $\Delta T := T - T_0$ as the independent variable. Furthermore, we define $A := \alpha \cdot L_0$ [cm/K], where α [1/K] is the specific expansion coefficient. Then

$$L(\Delta T(k)) = A \cdot \Delta T(k) + L_0. \tag{1}$$

Below, you find the datapoints. Using the data, you will compute estimates for the parameters A and L_0. The following MATLAB commands might be helpful: `help`, `plot`, `hold on`.

ΔT [K]	5	15	35	60
L [cm]	6.55	9.63	17.24	29.64

   (a) Plot the $\Delta T(k)$, $L(k)$ relation using 'x' markers in MATLAB. (1 point)

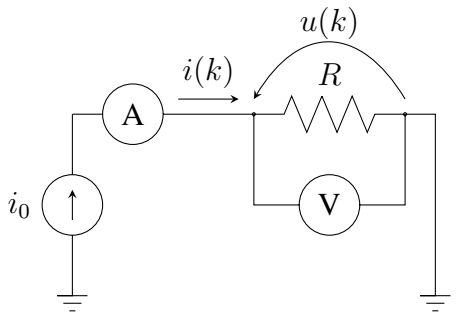
   (b) Compute the experimental values for the parameters A and L_0 using the model from above (Eq. 1). Minimize the sum of squared distance $d_k^2 = (L(k) - L)^2$ to find the solution. (Hint: Compute the solution by setting the gradient of the objective function with respect to the parameters (A,L_0) to zero. This will give you a 2×2 linear system. Check if the objective function is convex!) Plot the fit ($L(\Delta T(k)) = A \cdot \Delta T(k) + L_0$) through the $\Delta T(k)$, $L(k)$ data. (2 points)

   (c) Now, use a third order polynomial and fit it to the data. Again minimize the sum of squared distances to find optimal values for the coefficients of your model equation. Plot the fit in the same figure as before. (2 points)

   (d) You take another measurement: at $\Delta T = 70$K you measure a length of $L = 32.89$ cm. You can use this additional datapoint to validate your fit. Therefore plot it in the existing plot. Which fit looks more reasonable to you? The phenomenon of fitting a model to a data set which then does not pass validation is called 'overfitting'. (1 point)

2. Show that for any $A \in \mathbb{R}^{n \times m}$ holds that $A^\top A$ is symmetric and positive semi definite (PSD). (Hint: matrix B is PSD, if for any $x \in \mathbb{R}^n$ holds that $x^\top B x \geq 0$) (2 points)

3. Computer exercise with MATLAB. We consider the following experimental setup:



Imagine you are sitting in a class of 200 electrical engineering students and you want to estimate the value of R using Ohm's law. Since the value of the current i_0 flowing through the resistor is not known exactly, an ammeter is used to measure the current $i(k)$ and a voltmeter to measure $u(k)$. Every student is taking 1000 measurements. The measurement number is represented by k. We assume that the measurements are noisy:

$$i(k) = i_0 + n_i(k) \qquad \text{and} \qquad u(k) = u_0 + n_u(k)$$

where $u_0 = 10\,\mathrm{V}$ is the true values of the voltage across the resistor, $i_0 = 5\,\mathrm{A}$ is the true value of the current flowing through the resistor and $n_i(k)$ and $n_u(k)$ are the values of the noise.
Please download the data-set with all measurements of all students [`Dataset 1`] from the course website.

Let us now investigate the behaviour of the three different estimators which were introduced in the lecture:

$$\hat{R}_{\mathrm{SA}}(N) = \frac{1}{N}\sum_{k=1}^{N}\frac{u(k)}{i(k)} \qquad \hat{R}_{\mathrm{LS}}(N) = \frac{\frac{1}{N}\sum_{k=1}^{N}u(k)i(k)}{\frac{1}{N}\sum_{k=1}^{n}i(k)^2} \qquad \hat{R}_{\mathrm{EV}}(N) = \frac{\frac{1}{N}\sum_{k=1}^{N}u(k)}{\frac{1}{N}\sum_{k=1}^{N}i(k)}$$

We will use MATLAB to simulate the behavior of these estimators. For each of the three estimators, carry out the following tasks (useful MATLAB commands are `help`, `plot`, `for`, `mean`, `std`).

(a) First, compute the result of the function $\hat{R}_*(N)$, for $N = 1, \ldots, N_{\max}$ using your personal measurements (student 1 or experiment 1). Do this for each estimator ($*$ can be either SA, LS or EV). Plot the three curves in one plot. Do the estimators converge for $N \to \infty$?

(b) It is good practice to analyze the results of several experiments to cancel noise. Luckily, you get the datasets of all other students. Plot the function $\hat{R}_*(N)$, $N = 1, \ldots, N_{\max}$ for each estimator ($*$ can be either SA, LS or EV). To see the stochastic variations, plot all these functions in one graph per estimator using `hold on`. Do you see any difference to the plot from task (b)?

(c) Compute the mean of $\hat{R}_*(N)$ over all experiments (all 200 students) and plot it for N from 1 to $N_{\max}$.

(d) Plot a histogram containing all values of $\hat{R}_*(N_{\max})$.

(4 points)

This sheet gives in total 13 points (plus 4 bonus points from the previous sheet)