

ESS101 Mod & Sim
SP1 2022
Assignment 2 - System Identification
Due date: 2022-10-06

In this assignment, we will explore various aspects of parameter estimation including least-squares estimation and system identification. Write a small report providing the answers to the questions. For compute exercise, provide the answers in the report and a file with the Matlab code, too.

Instructions

The assignments comprise an important part of the examination in this course. Hence, it is important to comply with the following rules and instructions:

- The assignment is pursued in groups of two students. For group formation procedures, see Canvas.
- The findings from each assignment are described in a short report, written by each group independently.
- The report should provide clear answers to the questions, including your motivations, explanations, observations from simulations etc. Figures included in the report should have legends, and axes should be labelled.
- Since the assignments are part of the examination in the course, plagiarism is of course not allowed. If we observe that this happens anyway, it will be reported.
- The report should be uploaded to Canvas *before the deadline*. Matlab code is uploaded as separate files. If the deadline is not met, we cannot guarantee that you will be able to complete the assignment during the course.
- The written report is graded PASS, REVISE or RESUBMIT. With a decision REVISE, you will get a new due date to provide a revised report with changes highlighted. With a decision RESUBMIT, you will need to submit a completely rewritten report.

First part - Estimators and Least Squares.

1. **Unbiased estimate.** Consider the data samples $\{x[0], \dots, x[N-1]\}$, where each sample is distributed as $\mathcal{U}[0, \theta]$ (uniform distribution in the interval $[0, \theta]$) and the samples are IID.

(a) Find an unbiased estimator for θ . The range of θ is $0 < \theta < \infty$.

Hint: Think about the relation between the mean value of the data samples and the parameter you want to estimate!

2. **Linear least squares.** Consider the linear system

$$y[k] = \theta u[k] + e[k], \quad k = 0, \dots, N-1 \quad (3)$$

with θ being a parameter to estimate and $e[k] \sim \mathcal{N}(1, \sigma^2)$.

- (a) Compute the least-squares estimate of θ .
- (b) Is the estimate biased? If yes, give the expression of the bias.
- (c) What happens to the estimation of θ if $u[k] = 0, \forall k$?
3. **Computer exercise. Curve fitting.** In Matlab, create a noise-corrupted 2 dimensional dataset $\{x(i), y(i)\}$, $i = 1, \dots, N$ following the relation denoted by $y = a_0 + b_0 * x + e$ where $e \sim \mathcal{N}(\mu, \sigma^2)$, with $\mu = 0, \sigma = 3.8$, true parameters $a_0 = -2.923, b_0 = 7.18, N = 100$, and limiting values of x up to 50. x can be chosen uniformly in range $[0, 50]$. (If you would like to generate the same figures you can use `rng(1)` command to fix the seed to generate the same random data.)

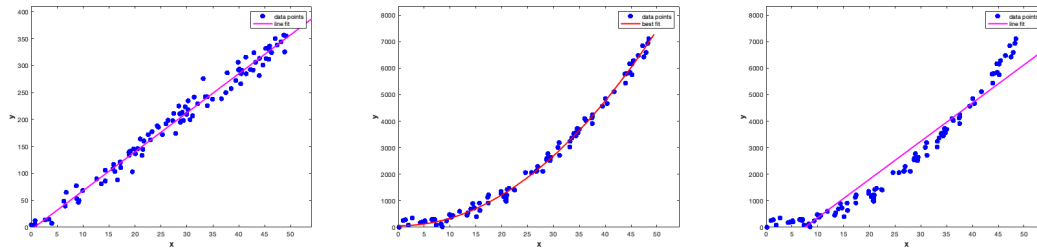


Figure 1: Linear regression based on least squares

- Report the parameters that correspond to the line that best fits this data using the least squares solution, and plot the line together with the data points as in the first plot in Fig. 1.
- Fit a 2nd order polynomial to a dataset created with $y = a_0 + b_0 * x + c_0 * x^2 + e$, $\mu = 0$, $\sigma = 12.8$, $a_0 = -2.923$, $b_0 = 7.18$, $c_0 = 2.8$ and x values up to 50, only altering θ in the previous line fitting exercise and plot the resulting model fit as in the second plot in Fig. 1, also fit a line to the same data in the third plot in Fig. 1 and compare the resulting residual of the both polynomial and line fitting values, $\varepsilon = 1/N \sum_{i=1}^N (y_{(i)} - \hat{y}_{(i)})^2$.

Second part - Identification of linear models for dynamical systems.

- One-step ahead predictor.** Consider the following model structure

$$y(t) + a_1 y(t-1) + a_2 y(t-2) = b_0 u(t) + e(t) + c_1 e(t-1) \quad (5)$$

- What kind of model structure is it (ARX, FIR, ARMAX, OE)?
- Which are the expressions of the plant model G and noise model H ?
- Find the 1-step-ahead predictor for Model (5).
- Is the 1-step-ahead predictor you found a linear function of the parameters?

- Prediction or simulation?** Consider the following model structure

$$y(t) + a_1 y(t-1) = b_0 u(t) + e(t) + a_1 e(t-1) \quad (9)$$

- What kind of model structure is it (ARX, FIR, ARMAX, OE)?
- Find the 1-step-ahead predictor for the model in (a). Something weird is happening! Can you explain what and why?

- Computer exercise. Identification of an ARX model.** In Matlab, load the data files input.mat and output.mat from Canvas. The data contained in the file are generated using an ARX model with additive white Gaussian noise ($\sigma^2 = 0.01$). Consider the following candidate ARX model structures,

$$y(t) + a_1 y(t-1) + a_2 y(t-2) = b_0 u(t) + e(t) \quad (12a)$$

$$y(t) + a_1 y(t-1) + a_2 y(t-2) = b_0 u(t) + b_1 u(t-1) + e(t) \quad (12b)$$

$$y(t) + a_1 y(t-1) + a_2 y(t-2) + a_3 y(t-3) = b_1 u(t-1) + e(t) \quad (12c)$$

- Split the data in estimation and validation data, and identify the parameters of each candidate ARX model, using the estimation data. *Hint: ARX models provide a 1-step-ahead predictor, which is a linear regression in the parameters. Hence, compute the predictor and use linear least-squares to find a closed-form expression for the parameter estimates.*
- For each estimated model, write a small script for computing the simulated and predicted (1-step) output, using the validation data. Compare those outputs with the true data (you can use Root Mean Square Error as a measure of the goodness of the comparison). Which model is the best in simulation? And which one in prediction?