System Identification

227-0689-00L

Midterm examination: solution

Due: 10:00 on Wednesday, November 9^{th}, 2022

Overview

The midterm exam is composed of TWO problems. To solve these problems you need to write two separate MATLAB functions: one for Problem 1 and one for Problem 2. Each function will analyze the data associated with the corresponding problem.

This work must be done individually. You may not discuss these problems with anyone else.

Your functions will be submitted through Moodle. In this submission you must also include a scan of the signed declaration of originality form. In the form, the title of work should be specified as `System Identification midterm exam` and the **scanned** document should be named as `HS2022_SysID_midterm_DO_LegiNumber.pdf`, where "LegiNumber" must be your legi-number, without any dash, slash, backslash, etc., e.g.,

`HS2022_SysID_midterm_DO_12345678.pdf`

If you do not have a legi-number, before the due date of exam, you should send an email to `sysid@ee.ethz.ch` and ask for a temporary number to be assigned.

The Moodle forums are closed during the midterm examination. If you encounter any problems with the examination, please send an email to `sysid@ee.ethz.ch` or Prof. Roy Smith.

Your submission must be submitted to Moodle through the link

`https://moodle-app2.let.ethz.ch/mod/assign/view.php?id=787739`

by the due date and time given above. The submission should contain exactly 3 separate files (2 MATLAB m-files and a PDF).

Your grade for the midterm will be evaluated based on your performance in the two problems weighted in the following way: 50% for Problem 1, 50% for Problem 2.

Downloadable data

The required files for the individual parts will be provided through the Moodle platform under the "Midterm Matlab project" section. Here is the list of uploaded files:

File	Problem	Summary description
`HS2022_SysID_midterm_p1_GenerateData.p`	Problem 1	Generate the problem data
`HS2022_SysID_midterm_p1_12345678.m`	Problem 1	Solution template
`HS2022_SysID_midterm_p1_check.p`	Problem 1	Code compliance check function
`HS2022_SysID_midterm_p2_GenerateData.p`	Problem 2	Generate the problem data
`HS2022_SysID_midterm_p2_12345678.m`	Problem 2	Solution template
`HS2022_SysID_midterm_p2_check.p`	Problem 2	Code compliance check function
`HS2022_SysID_midterm_DO_12345678.pdf`	-	Declaration of originality

MATLAB function format

Each problem has a solution template (named `HS2022_SysID_midterm_p?_12345678.m`) which can be downloaded from Moodle. Replace the string `12345678` in the solution template with your legi-number (using only the digits) and use it as the starting point for your solution. You will also have to make the same change on the first line of the solution function. It is important that you do not change the name or the order of the output variables on the first line.

You will see that the solution template calls a function called

`HS2022_SysID_midterm_p?_GenerateData.p`

to generate the data for your analyses. There is a separate data generation function for each question and you will need to download both from Moodle.

Each problem also has a compliance check function (named `HS2022_SysID_midterm_p?_check.p`) which can be used to check if your functions meet the minimal requirements for submission. Running this function, with your legi-number as the input argument, will perform this check on the script matching the solution naming format in your current directory. **Do NOT submit any code that has not been successfully checked by this function.**

The description of the individual problems (given on the following pages) will specify the variables that must be returned by your functions. You will also be asked to answer certain questions and explain your results and the choices you are making. These explanations must be given in the command window of MATLAB, using the `disp` function. Explanations written in the comment of your functions will NOT be graded. Please keep your explanation concise (around 2-4 sentences). You will be graded on the quality of your explanation.

If you write any custom functions, include them in the solution file after the main function. This means that MATLAB treats them as subfunctions and calls them instead of any other functions in the path that have the same name.

The following are generic considerations for the behaviour of your code. They apply to each of the problems.

- You may only use functions from MATLAB and the Control Systems Toolbox. Functions from other toolboxes are not permitted. This is checked by the compliance check function.
- Your function must perform the calculations requested on the data generated by the data generation function. If the solution of the problem requires that you make a choice of parameter then this may be hard-coded into your function. You should explain in the command window why and how you made any such choice.
- The script should run unattended. Do not use the commands `pause`, `clear`, `clc`, `clf`, or `close` within your solution function. Do not define any global variables in your functions or subfunctions. If you skip one part of a problem, please keep the dummy variable as in the solution template.
- For any figures requested in the problem use the command `figure(xxx)`, where `xxx` is specified in the problem, to ensure that all figures remain visible on the screen after your code terminates.
- Each MATLAB figure can contain a single plot, except for Bode plots which can use the subplot function to display gain and phase separately.
- Specify all axis limits via the MATLAB command `axis`. Auto-scaling must be avoided as it works differently on different machines. If your plot uses a legend, use the `'Location'` flag in the `legend` command to make sure that relevant data are not covered.

Before submitting your functions, restart MATLAB and test them by running the compliance check function for each problem in a folder/directory with no other files.

Problem 1 (Weight: 50%)

Consider the problem of sensor calibration with a linear measurement model as follows:

$$y(k) = Cx(k) + e(k),$$

where $x(k) \in \mathbb{R}^{n_x}$, $y(k) \in \mathbb{R}$, $e(k) \in \mathbb{R}$, and $C \in \mathbb{R}^{1\times n_x}$ are the states, the sensor measurement, the noise, and the output matrix respectively, and n_x is the state dimension.

Your task is to estimate the output matrix C from state-measurement data. For this purpose, you are given the following experiment data.

- `p1_x` [dimension $N \times n_x$]: The state data, x. The k-th row corresponds to the data of $x^\top(k)$.
- `p1_y` [dimension $N \times 1$]: The measurement data, y. The k-th row corresponds to the data of $y(k)$.

Assume that the noise $e(k)$ is zero-mean normally distributed with

$$\mathbb{E}\left\{e(k)e(l)\right\} = \begin{cases} \alpha, & k = l \\ \beta, & |k-l| = 1 \\ \gamma, & |k-l| = 2 \\ 0, & \text{otherwise} \end{cases}.$$

The values of α, β, and γ will be displayed in the command window when you run the data generation function.

Using the provided experiment data, you should do the following:

a) Using the maximum likelihood method, find the maximum likelihood estimate $\hat{C}_{\text{ML}}$.

   Please explain the following in the command window:

   i) The optimization problem you need to solve, with symbols clearly defined.

   ii) Is the maximum likelihood estimate $\hat{C}_{\text{ML}}$ the best linear unbiased estimator? Why?

[**Weight: 15%**]

It is observed that the sensor measurement is subject to bias and drift as well. The model is thus augmented as

$$y(k) = Cx(k) + e(k) + b + d \cdot t,$$

where b, d, and t are the bias, the drift, and the time respectively. The identification data are collected at a sampling rate of 200Hz. You can assume that the first data point is at $t = 0$ in each variable.

b) To estimate the bias and the drift of the sensor, you are provided with additional measurement data with zero state values, i.e., $x(k) = \mathbf{0}$, $\forall\, k$.

- `p1_yf` [dimension $N_f \times 1$]: The measurement data, y, with zero state values. The k-th row corresponds to the data of $y(k)$.

Using the maximum likelihood method and only `p1_yf`, find the maximum likelihood estimate $\hat{\theta}_{bd} = \left[\hat{b}_{\text{ML}}\ \hat{d}_{\text{ML}}\right]^\top$, and the covariance matrix of the estimate Σ_{bd}.

Please explain the following in the command window:

i) The optimization problem you need to solve, with symbols clearly defined.

ii) The equation for computing Σ_{bd}, with symbols clearly defined. **[Weight: 15%]**

c) Consider the results in part b) as the prior distribution of b and d: $[b\ d]^\top \sim \mathcal{N}\left(\hat{\theta}_{bd}, \Sigma_{bd}\right)$. If you skipped part b), use the dummy variables provided in the solution template to continue.

Using the maximum a posteriori method, find the maximum a posteriori estimate $\hat{C}_{\text{MAP}}$ and $\hat{\theta}_{bd,\text{MAP}}$.

Please explain the following in the command window:

i) The equation for calculating $\hat{C}_{\text{MAP}}$ and $\hat{\theta}_{bd,\text{MAP}}$, with symbols clearly defined.

ii) If you were to find the maximum likelihood estimate of C, b, and d using both (`p1_x`, `p1_y`) and `p1_yf`, would the results be the same as the maximum a posteriori estimate? Why?

[Weight: 20%]

Your MATLAB function should return the following outputs of correct dimensions and sequences in the order listed:

- `p1_C_ML` (vector of dimension $1 \times n_x$): the maximum likelihood estimate $\hat{C}_{\text{ML}}$.
- `p1_theta_bd` (vector of dimension 2×1): the maximum likelihood estimate $\hat{\theta}_{bd}$.
- `p1_Sigma_bd` (matrix of dimension 2×2): the covariance matrix of the estimate $\hat{\theta}_{bd}$, Σ_{bd}.
- `p1_C_MAP` (vector of dimension $1 \times n_x$): the maximum a posteriori estimate $\hat{C}_{\text{MAP}}$.
- `p1_theta_bdMAP` (vector of dimension 2×1): the maximum a posteriori estimate $\hat{\theta}_{bd,\text{MAP}}$.

Your grade for this problem will depend on:

- The accuracy of the returned outputs. We will compare your outputs with the correct outputs obtained for your individual experiment data using the methods specified in the task description.
- The quality, clarity and accuracy of your explanations and answers in all parts.

Problem 2 (Weight: 50%)

The goal of this problem is to identify an unknown single-input, single-output plant G. For this purpose, you are given experiment data stored in `p2_data`.

- `p2_data` is a struct array with fields
  - `ID`: an integer in the range $\{1, \ldots, 15\}$ that uniquely identifies the experiment. The `ID` corresponds to the index in the struct array.
  - `type`: the type of input signal used in each experiment:
    `ID` $\in \{1, \ldots, 5\}$ use Chirp Signals,
    `ID` $\in \{6, \ldots, 10\}$ use Random Binary Signals (RBS),
    `ID` $\in \{11, \ldots, 15\}$ use Pseudo Random Binary Signals (PRBS).
  - `p2_u`: the input signal.
  - `p2_y`: the simulated output of the plant given the inputs `p2_u`. The plant starts from different and unknown initial conditions for each experiment. The output is corrupted by zero-mean Gaussian noise.
- The sample time is $T_s = 0.1\,\mathrm{s}$.
- The experiments are independent from each other.

For each of the following exercises you should pick the most appropriate experiment, each containing an input/output pair. Your choice should be returned, for each exercise, with the variable `p2<X>_ID`, where `<X>` is the exercise number $\in \{1, 2, 3\}$.

1) Estimate the plant using the ETFE method for the frequencies $f_k = 2\pi k/(MT_s)$, with $k \in \{1, \ldots, 127\}$, and $M = 255$.
   (a) What is the most appropriate input signal? Justify your choice.
   (b) Explain the procedure you used to compute the estimates.
   (c) Is your estimated $\hat{G}_1(e^{jw})$ an unbiased estimate?
   (d) Return:
      i. `p21_ID`, the index of the experiment you picked,
      ii. `p21_omega`, the vector of frequencies in $\mathrm{rad\,s^{-1}}$, and
      iii. `p21_G_hat`, the vector of plant estimates $\hat{G}_1(e^{j\omega})$ for the frequencies specified in `p21_omega`.
   (e) Plot both the phase in deg and the magnitude in dB of `p21_G_hat`, with `p21_omega` in $\mathrm{rad\,s^{-1}}$ as the x-axis in a logarithmic scale in `figure(211)`. **[Weight: 17.5%]**

2) Use a spectral method to estimate the plant for the frequencies $f_k = 2\pi k/(MT_s)$, with $k \in \{1, \ldots, 127\}$, and $M = 255$. As before, choose an appropriate input/output signal pair and complete the following questions. For this task, you can choose only from the 5 RBS signals.
   (a) What is the most appropriate input signal? Justify your choice.

(b) Explain the spectral method you used to compute the estimate.

(c) Is your estimated $\hat{G}_2(e^{jw})$ an unbiased estimate?

(d) Why are spectral methods preferred over ETFE for random signals?

(e) Return:

i. `p22_ID`, the index of the experiment you picked,

ii. `p22_omega`, the vector of frequencies in $\mathrm{rad\,s^{-1}}$,

iii. `p22_phi_u`, the power spectral density of the input,

iv. `p22_phi_yu`, the cross power spectral density, and

v. `p22_G_hat`, the vector of plant estimates $\hat{G}_2(e^{j\omega})$ for the frequencies specified in `p22_omega`.

(f) Plot the following with the frequencies `p22_omega` in $\mathrm{rad\,s^{-1}}$ as the x-axis, magnitudes in dB and phases in deg:

i. The magnitude of the input power spectral density `p22_phi_u`, with a linear x-scale in `figure(221)`.

ii. The magnitude of the cross power spectral density `p22_phi_yu`, with a linear x-scale in `figure(222)`.

iii. Both the phase and the magnitude of the plant estimate `p22_G_hat`, with a logarithmic x-scale in `figure(223)`.

[**Weight: 20%**]

3) Conduct another system identification experiment to generate a better estimate only for frequencies within $[1.12, 5.37]\,\mathrm{rad\,s^{-1}}$. Choose the most appropriate input/output signal pair and use ETFE to generate a new estimate $\hat{G}_3(e^{jw})$, for all $f_k = 2\pi k/(MT_s)$ within the specified frequency range, with $k \in \{k_1, \ldots, k_2\}$, and $M = 2047$.

(a) What is the most appropriate input signal? Justify your choice.

(b) Return:

i. `p23_ID`, the index of the experiment you picked,

ii. `p23_omega`, the vector of specified frequencies in $\mathrm{rad\,s^{-1}}$, and

iii. `p23_G_hat`, the vector of plant estimates $\hat{G}_3(e^{j\omega})$ for the frequencies specified in `p23_omega`. [**Weight: 12.5%**]