Course 5SMB0: System Identification 2022
Exercise Set 5 (lectures 9-10)

Problem 1 (Frequency domain identification)
You have measured data available in (`datavecs.mat`) on some dynamical system that has to be identified. In this exercise this system will be identified with use of frequency domain parametric methods. The system has been excited with a periodic multisine signal on a particular set of frequencies, with a period of $N = 1000$ samples.

1. First the data has to be prepared for identification. In order to reduce the effect of initial transients in the data (the effect of starting the experiments), remove the data of the first period in input and output. Then take the Fourier transform of the time domain data (`fft()`). Note that matlab orders the fft such that it corresponds to the frequencies from 0 to 2π, and not $-\pi$ to π.

2. All frequencies that have not been excited by the input do not contain information on the system. Plot and inspect the absolute values of the input frequency data to check which frequencies have been excited (`loglog`). A vector that contains these excited frequencies is generated by the command `w_vec(9:8:1201)`. In a similar way, form data vectors of input and output that contain only the frequencies that were excited. You will have an input and output vector containing 150 frequencies each.

3. Now the identification procedure begins. You can load the frequency data in the GUI of the identification toolbox by choosing 'frequency domain data' at 'import data'.

   - Inspect the data set through 'Data spectra'.
   - Suppose that the lower-frequent behaviour in the data is system-behaviour, and the higher frequent behaviour is due to noise. Which model order would you choose to model the system behaviour?
   - What would you expect that would happen if you identify an output error model, on the basis of all frequencies in the data?
   - Verify your expectation by executing this. Use an output error model with the 'Focus' set to 'Simulation'.
   - In order to focus more on the system behaviour, you can make a selection of frequencies that need to be taken into account in the identification. Choose a passband `[0, x]` in the output error model, by selecting `'Focus' -> 'Filter'`, and perform the identification using this passband. Estimate an Output Error (OE) model while choosing some different passbands `[0 x]`. How does the choice of `x` affect your result?
   - What causes the difference between the estimated models?

Selecting a limited frequency band is one way to separate system dynamics from noise-induced effects. However this can only be done if noise and system dynamics are non overlapping in the frequency domain. When noise is present in the same frequency range as where system dynamics are, then it would not be desirable to exclude these frequencies from the identification criterion. Moreover these higher frequencies contain information, which is thrown away by filtering it out completely. In order to deal with this situation a (non-parametric) noise model will be estimated, which will be used as a weighting in the identification criterion (slide 9-26). Since there are no facilities for this in the GUI, the following steps need to be executed in the Matlab command window.

4. Make a non-parametric noise model of the data as done in the lecture slides. This noise model will consist of 1000 frequency points, as that is the length of the period. Inspect the non-parametric noise model, is it the shape you expected based on your observation of the output data? [Matlab commands `mean, repmat`]

5. The 150 excited frequencies in input and output match with the `2:151` elements from the non-parametric noise model. Pre-filter the excited frequencies of input and output with the noise model by pointwise division of the fft vectors of input and output through $\hat{\sigma}_V(k)$.

6. Load the filtered data into the toolbox GUI, inspect the content of the filtered data and repeat the identification steps from 3. Are the estimated models different? Why is that?

Problem 2 (Experiment design)

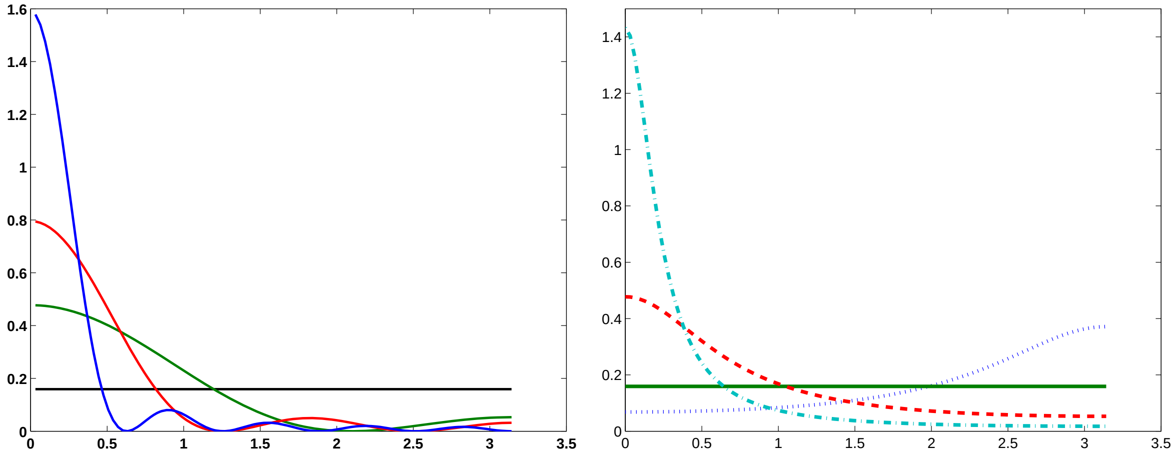

Figure 1: Left: Spectrum $\frac{1}{2\pi}\Phi_u(\omega)$ of RBS with basic clock period $N_c = 1$ (solid), $N_c = 3$ (dash-dotted), $N_c = 5$ (dotted), and $N_c = 10$ (dashed); Right: Spectrum $\frac{1}{2\pi}\Phi_u(\omega)$ for RBS with non-switching probabilities $p = 0.5$ (solid),$p = 0.75$ (dashed), $p = 0.90$ (dash-dotted) and $p = 0.3$ (dotted).

In Figure 1 the effect on the spectral density of an input signal is sketched for two different principles for influencing the spectral density of a (P)RBS. In the left Figure this is done through changing the clock period of the binary signal, and in the right Figure it is done by changing the switching probability of the binary signal.
What is the prime difference between the two results for the spectral densities, and discuss the importance of this for identification.

Problem 3 (Experiment design)

You perform a preparatory step experiment on a system, and notice that after measuring 300 samples the output response has converged to its steady state value. Your final objective is to estimate an accurate second order parametric model of the system, on the basis of a persistently exciting input, e.g. a random binary sequence (RBS). Would it be enough if you measure 1000 samples for estimating an accurate second order model (i.e. 5 parameters in G)?

Problem 4 (MIMO models)

Consider the multivariable model in state space form:

$$\begin{aligned} x(t+1) &= Ax(t) + Bu(t) + Ke(t) \\ y(t) &= Cx(t) + Du(t) + Le(t) \end{aligned}$$

(a) What is the motivation in an identification problem to restrict to the situation of $L = I$?

(b) Consider the situation $L = I$. Do you see any similarities with the polynomial model structures used in identification, like FIR/ARX/ARMAX/OE/BJ, and if so with which particular forms and why?

Problem 5 (Frequency and time-domain identification)

Consider a system $\mathcal{S}$:

$$y(t) = \frac{B^0(q^{-1})}{A^0(q^{-1})}u(t) + H^0(q)e(t) \tag{1}$$

that is excited by a multisine input signal, and disturbed by a filtered noise, with $e(t)$ a unit variance white Gaussian noise process. We consider three different situations:

(a) $H^0(q) = \frac{1}{A^0(q^{-1})}$

(b) $H^0(q) = 1$

(c) $H^0(q) = \frac{C^0(q^{-1})}{D^0(q^{-1})}$.

We apply two estimation methods to these three situations:

1. We estimate a time domain model, by choosing a parametric model structure that satisfies $\mathcal{S} \in \mathcal{M}$;
2. We estimate a model through a frequency domain method, estimating a parametric model for G^0, satisfying $\mathcal{G}^0 \in \mathcal{G}$ and a non-parametric noise model.

Each of the two methods we have applied to 20 different different data sets, genrated by using 20 different noise realizations of the process $e(t)$. Figure 2 shows the results of the estimated models in terms of the mean squared error of the frequency responses of the estimated models, i.e.

$$RMS = \sqrt{\frac{1}{20}\sum_{i=1}^{20} |G^0(e^{i\omega}) - \hat{G}^{(i)}(e^{i\omega})|^2}$$

with $\hat{G}^{(i)}(e^{i\omega})$ the estimate based on data realization i.
In the Figure it is illustrated that the results of the time domain and the frequency domain method are relatively close to each other.

- Can you provide an explanation for the differences between the time- and frequency domain method, for the cases (a), (b), and (c), respectively?

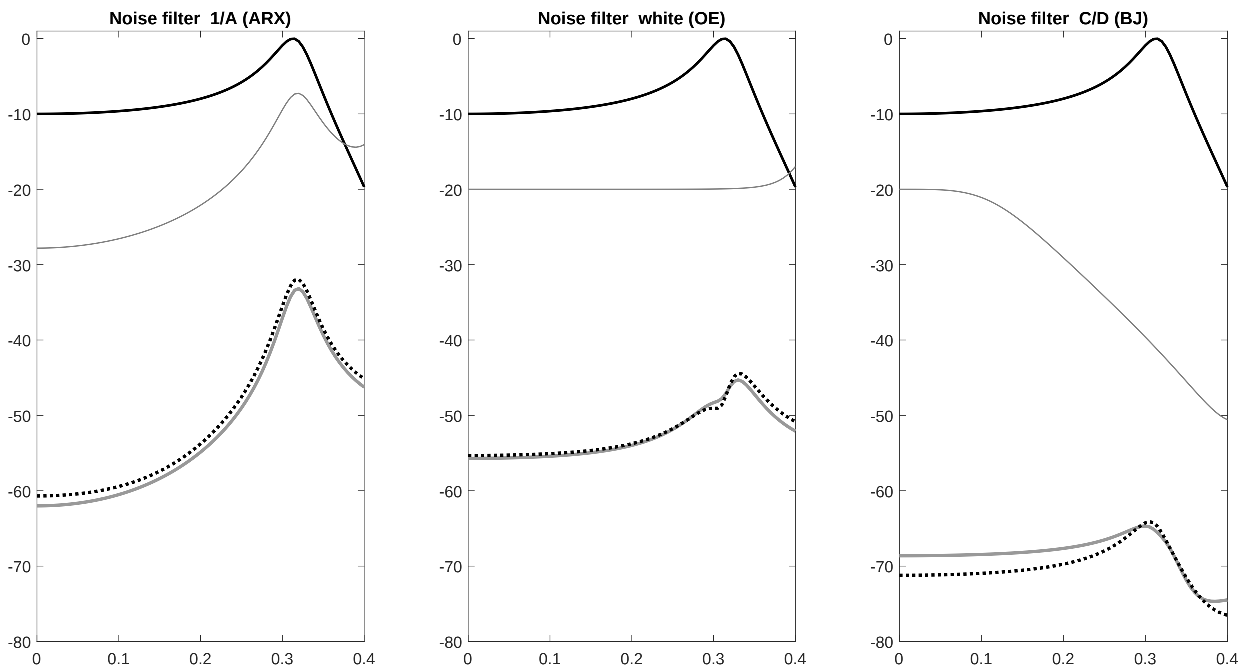

Figure 2: Comparison of time and frequency domain methods for three different noise types. Black line: $|G_0(e^{i\omega})|$; Thin gray line: $|H_0(e^{i\omega})|^2$, i.e. the power spectrum of disturbing noise; dotted line: RMS error for the frequency domain method; bold gray line: RMS error for the time domain method. Left: Situation (a); middle: Situation (b); right: Situation (c).