Learning dynamical systems using system identification

(Course: SSY 230)

Exam 2th June 2020

Time: 14:00-18:00

Examiner: Jonas Sjöberg, phone 772 1855 (073-0346321)

Total number of credits is 50. Preliminary grade limits are 23 for grade 3, and 32 for grade 4, and 40 for grade 5.

Solutions should be clearly formulated so that it is easy to follow each step.

Those who passed the hand-in assignment can skip problem 1, and you recieve the maximum number of points on that problem.

Due to the changed examination conditions, all type of support is admitted, except communicating with other people. This includes

- The course book
- Mathematical handbook such as BETA, handbook with physical formulas such as Physics Handbook
- Standard Calculator
- Lecture slides

Solutions will be published on the course web after the exam.

Result of the exam is delivered in mail by the LADOK system.

Evaluation of grading to be decided.

Good Luck!

Department of Electrical Engineering Chalmers University

Problem 1
The P-matrix in the recursive least-squares (RLS) algorithm, except for the factor representing the noise variance, asymptotically represents the parameter covariance in the case of constant parameters and an uncorrelated noise sequence.

Modify/complement the RLS algorithm so that also an estimate of the noise variance is obtained.

Note that the noise variance estimate should be formulated so that the need of memory does not increase with time.

(10p)

Problem 2
Assume that the ideal discrete time impulse cannot be implemented, and that it is replaced by the input

$$u(t) = \begin{cases} 1/T, & \text{if } 0 \leq t \leq T \\ 0, & \text{if } t > T \end{cases}$$

How should the discrete time impulse response be determined from data obtained when using this input? Assume that the system is linear and time invariant.

(10p)

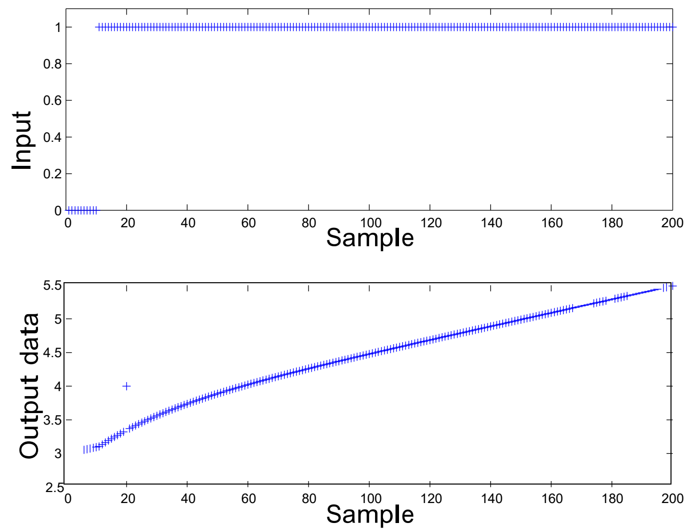
Problem 3

(a) When deciding upon the input signal to be used in an experiment to obtain data for system identification then the frequency contents is important. Also the amplitude of the signal is important, give aspects important when deciding the amplitude. used.

(3p)

(b) Cross-validation is an important part of the model validation process. Discuss how, in general, the data should be arranged for cross-validation (2p)

(c) You have performed a step response experiment and obtained both input and output data as shown in the figure bellow. Each observation is indicated with a "+sign.



Discuss if the sampling rate and observation interval have been well-chosen. (3p)

(d) What shall be done with the data before the identification procedure starts? Consider the case in (c). (2p)

Problem 4

In system identification, often you have the situation where you are interested of properties of a function depending on estimated parameters. Consider

$$\hat{y}(x,\theta) = f(x,\theta)$$

which express a model of some "true" relation $y(x)$. The model is parameterized by the parameter vector θ of dimension d. $\hat{\theta}_N$ is the estimate of θ based on N data. Assume that also P_N^θ, the variance of $\hat{\theta}_N$ has been estimated. We are now interested of the distribution of f as a function $\hat{\theta}_N$ and in properties like $E[\hat{y}(x,\theta)]$ and $var(\hat{y}(x,\theta))$. Note that these are functions of x.

(a) $\hat{\theta}_N$ is asymptotically normal distributed. What can you say about the distribution of $f(x,\hat{\theta}_N)$ (for a fixed x)? Consider different cases, eg, if f is linear (in x, or in θ), if f is smooth, if P_N^θ is large, or small etc. (3p)

(b) Give an algorithm how estimate the distribution of $f(x,\hat{\theta}_N)$ (for a fixed x) with a Monte Carlo method? (4p)

(c) Formulate estimates of $E_\theta[\hat{y}(x,\theta)]$ and var$(\hat{y}(x,\theta))$ using the estimated the distribution of $f(x,\hat{\theta}_N)$. (3p)

Problem 5

Consider the system

$$\mathcal{S}: \; y(t+1) = a\,(u(t) + w(t)), \;\; w(t) \in N(0, \sigma^2), \;\; \mathrm{E}[w(t)w(s)] = \sigma^2 \delta(t-s)$$

where $\{w(t)\}$ is a zero-mean white noise sequence and where $\{u(t)\}$ and $\{y(t)\}$ are measured variables. Formulate the Maximum Likelihood estimate of a?

(10p)

Exam 2th June 2020: Solutions

1

Let p be the number of estimated parameters. We formulate the estimate of the noise variance up to time t and from that we separate the last contribution so that we can form it as an update of the estimate from time $t-1$, that is

$$\hat{\sigma}_t^2 = \frac{1}{t-p}\sum_{k=1}^{t}\varepsilon_k^2 = \frac{1}{t-p}(\sum_{k=1}^{t-1}\varepsilon_k^2 + \varepsilon_t^2) = \frac{t-1-p}{t-p}\frac{1}{t-1-p}\sum_{k=1}^{t-1}\varepsilon_k^2 + \frac{1}{t-p}\varepsilon_t^2 = \frac{t-1-p}{t-p}\hat{\sigma}_{t-1}^2 + \frac{1}{t-p}\varepsilon_t^2$$

2

The input can be desribed as composed by two steps. The output of each step is a convolution of the unknown impulse responce and the step. Formulate the output at each t, which gives an equation. By observing $y(t)$ equations are obtained to determine the impulse responce.

3

(a) The amplitude of the input signal should be chosen as large as possible in order to achieve a good **signal-to-noise ratio** and to overcome problems with, for example, **static friction**. However, the amplitude may not be chosen larger than **the range in which the linearity assumption holds**. Typically saturations give an upper bound on the amplitude of the input signal. The mean value is in many cases non-zero in order to reduce friction problems or to give a linearized model around some, non-zero, stationary point.

(b) Data can be divided into three parts

i data with transients. These data sequencies can be used, but implies aditional difficulties. For example, the data cannot be considered as stationary realizations.

ii identification data

iii validation data

A standard decision is do divede data so that the identification data and the validation data become equal in size. There is, howeve, no strict rule, and depending on the particular problem other divisions may apply. Also, depending on the features of the data, for example varying set-poit, or vary-ing frequency contents, can motivate the division of the data.

(c) The closed loop system step response should be at least sampled 5-10 times during it's rise time. In our case we have around 100 samples during the rise time. It is unnecessary many samples which can be a problem if it extreem (100 is not extreem, it is only 10 times more than necessary). Hence, **sampling time is fine**. However, only parts of the step response is recorded. A **longer obesrvation window could be of interest** to obtain a better estimate of the end value of the step response, although no more input is injected during that time.

(d) Start by visually inspecting the data in a plot.

i Are there outliers to be removed? (yes, there seems to be at least one in this data set)

ii Remove linear trends and non-zero mean.

This should be done before the identification procedure starts.

4

(a) If f is linear with respect to θ then $f(x, \hat{\theta}_N)$ will also be normal distrubuted. On the other hand, if f is NOT linear with respect to θ then we cannot say anything. However, if f is smooth in θ and if the uncertainty in θ is small, it will be close to normal distributed.

(b) An approximation of the distribution of $f(x, \hat{\theta}_N)$ can be obtained by generating samples of $\hat{\theta}_N$. $\hat{\theta}_N$ is asymptotically normal distributed:

$$\hat{\theta}_N \in AsN(\theta_0, P_\theta) = \frac{1}{\sqrt{(2\pi)^2 \det P_\theta}} \exp(-\frac{1}{2}(\hat{\theta}_N - \theta_0)P_\theta^{-1}(\hat{\theta}_N - \theta_0)^T)$$

The true parameter value θ_0 and the true varance P_θ are unknown and must be replaced by there estimates. Then K samples can be generated

$$\{\hat{\theta}_N^k\}_{k=1}^K$$

and the set $\{f(x, \hat{\theta}_N^k)\}_{k=1}^K$ approximate the distribution of $f(x, \hat{\theta}_N)$.

Note: Approximating a distribution by taking samples of it, and propagating the samples ”through” a function and using the result as an estimate of the distribution of the function, is the main idea used in ”particle filters”. The samples are the particles.

(c) Estimate of expectation:

$$E_\theta[\hat{y}(x,\theta)] \approx \bar{y}(x,\theta) = \frac{1}{K}\sum_{k=1}^{K} f(x, \hat{\theta}_N^k)$$

which is close to $f(x, \hat{\theta}_N)$ if f linear.

Variance estimate:

$$E_\theta[\hat{y}(x,\theta)] \approx \frac{1}{K}\sum_{k=1}^{K}(f(x, \hat{\theta}_N^k) - \bar{y}(x,\theta))^2$$

5 Data is generated according to

$$y_{k+1} = a(u_k + w_k)$$

which we re-write so we get the noise term

$$w_k = \frac{1}{a}y_{k+1} - u_k$$

Inserting this in the distribution function gives

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(\frac{1}{a}y_{k+1}-u_k)^2}{2\sigma^2}}$$

This is not well-parameterized. Introduce $\theta = 1/a$ and formulating the MLE leeds to

$$\max_\theta \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^N e^{\sum_{i=1}^N -\frac{(\theta y_{i+1}-u_i)^2}{2\sigma^2}}$$

Logarithm of this:

$$\max_{\theta} \log\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^N + \sum_{i=1}^{N} -\frac{(\theta y_{i+1} - u_i)^2}{2\sigma^2}$$

Take the derivative with respect to $\theta$and set to zero,

$$\max_{\theta} \sum_{i=1}^{N} -y_{i+1}\frac{(\theta y_{i+1} - u_i)}{\sigma^2} = 0 \Leftrightarrow$$

$$\max_{\theta} \sum_{i=1}^{N} -\theta y_{i+1}^2 = -\sum_{i=1}^{N} y_{i+1}u_i \Rightarrow$$

$$\theta = \left(\sum_{i=1}^{N} y_{i+1}^2\right)^{-1} \sum_{i=1}^{N} y_{i+1}u_i = (Y^TY)^{-1}Y^TU$$

and

$$\hat{a} = 1/\theta = 1/(Y^TY)^{-1}Y^TU$$