Learning dynamical systems using system identification

(Course: SSY 230)

Exam 31st May 2022

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 assignments on time can skip problem 1, and you recieve the maximum number of points on that problem.

This is an "open book" exam. Help material: 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: contact examiner.

Good Luck!

Department of Electrical Engineering Chalmers University

Problem 1
Consider the AR-process

$$y(t) = a_1 y(t-1) + e(t)$$

Calculate the k-step prediction for $k > 0$.

(10p)

Problem 2

Consider the output error system

$$y(t) = \frac{b_0\, q^{-1}}{1 + a_0\, q^{-1}} u(t) + e(t)$$

where

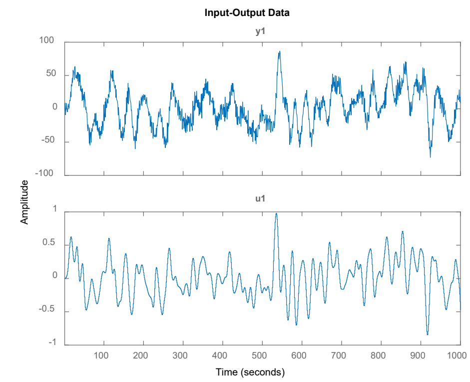
- $e(t)$ is a white stochastic measurement noise process of zero mean and variance σ_e^2.
- $u(t)$ is a white stochastic input process of zero mean and variance σ_u^2.
- the two processes $e(t)$ and $u(t)$ are independent.
- The actual system parameters are $a_0 = -0.8$ and $b_0 = 0.2$.

(a) Specify the one step ahead predictor which is necessary for estimating the a and b parameters using PEM. What do the PEM estimated parameter values asymptotically converge to as the number of data points N goes to infnity? Motivate. (4p)

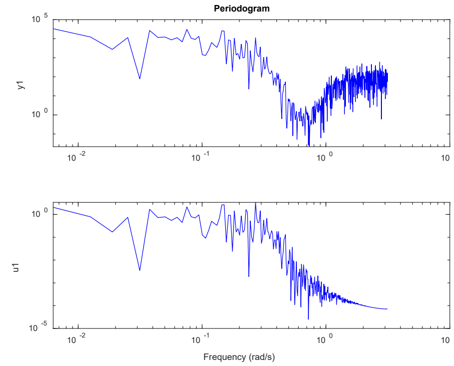
(b) Estimate the covariance of the parameter estimate for large N in the case that only b is estimated, and that a has been fixed to its correct value. (6p)

Problem 3

Input and output data has been collected from a dynamical system. The data is shown bellow, both in time domain and in frequency domain.

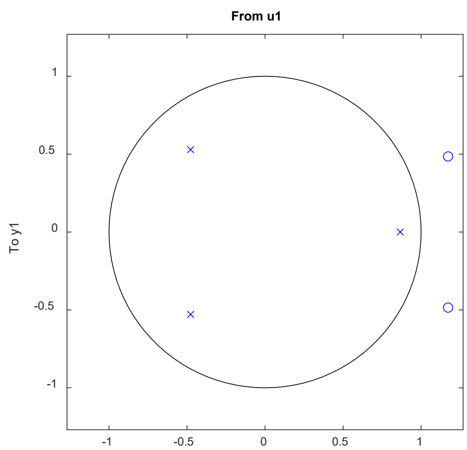


a)

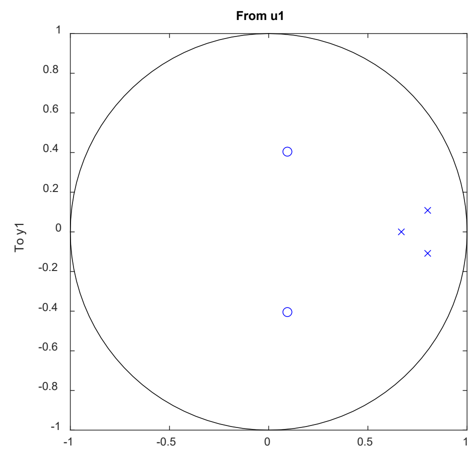


b)

Two third order models models are being identified one ARX model and one Output Error model. Bellow you see the pole-zero plots of the two models, and they are quite different.



a)



b)

(a) Which of the pole-zero plots is from the ARX model and which is from the Output Error model? (3p)

(b) Suggest some pre-processing of the data which likely improves the quality of the ARX model. (4p)

(c) How do the properties of output-error model change with the pre-processing? (3p)

Don't forget to motivate your answer.

Problem 4
For a time continuous system, assume that the ideal impulse cannot be implemented, and that it is replaced by the input

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

(a) How should the impulse response be determined from data obtained when using this input?

Assume that the impulse reponse is zero for $t < 0$, ie, the system is causal. Then, look for the soution in the intervals $0 < t < T$ and $t > T$ separately.

(b) Consider the case that T can be chosen very small. Give an approximate solution where the output does not need to be differentiated.

(10p)

Problem 5
In system identification, often you have the situation where you are interested of properties of a function depending on estimated parameters. Consider

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

the estimate of f depending on the estimated $\hat{\theta}_N$, where N is the number of data used and d is the dimension of the parameter vector. $\hat{\theta}_N$ is asymptotically normal distributed:

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

where θ_0 is the true parameter vector.

(a) Show that the random variable $P^{-1/2}\hat{\theta}_N$ is asymptotically normal distributed with variance I. This means that the components of $P^{-1/2}\hat{\theta}_N$ are independent of each other. (3p)

(b) A sum of the squares of d Gaussian distributed variables, with mean zero and unit variance is a new random variable with distribution $\chi^2(d)$. Show that $(\hat{\theta}_N - \theta_0)P_\theta^{-1}(\hat{\theta}_N - \theta_0)^T \in \chi^2(d)$ (3p)

(c) Discribe how an approximation of the confident region of degree α, of f, for any value of x can be obtained by sampling θ_k K times using the distribution described above.

(2p)

(d) Make scetchs for the cases of one and two dimensional parameter spaces (ie two scetches) where you indicate where the samples θ_k are taken.

(2p)

Exam 31st May 2022: Solution schetches

1 We have:

$$y(t) = a_1 y(t-1) + e(t)$$

Using the standardform

$$y(t) = G(q)u(t) + H(q)e(t)$$

we identify

$$G(q) = 0$$

and

$$H(q) = \frac{1}{(1 - a_1 q^{-1})} = \sum_{i=0}^{\infty} (a_1 q^{-1})^i$$

where we also Taylor expanded $H(q)$. This expansion can be divided for a k step predictor

$$H(q) = \sum_{i=0}^{k-1} (a_1 q^{-1})^i + \sum_{i=k}^{\infty} (a_1 q^{-1})^i$$

where first part is defined with

$$\bar{H}_k(q) = \sum_{i=0}^{k-1} (a_1 q^{-1})^i$$

We have for a k step predictor:

$$\hat{y}(t|t-k) = \bar{H}_k(q)H(q)^{-1}G(q)u(t) + (1 - \bar{H}_k(q)H(q)^{-1})y(t) =$$
$$(1 - (\sum_{i=0}^{k-1} (a_1 q^{-1})^i (1 - a_1 q^{-1})))y(t) = (1 - \sum_{i=0}^{k-1} (a_1 q^{-1})^i + a_1 q^{-1} \sum_{i=0}^{k-1} (a_1 q^{-1})^i)y(t) =$$
$$(\sum_{i=1}^{k-1} (a_1 q^{-1})^i + \sum_{i=1}^{k} (a_1 q^{-1})^i)y(t) = (a_1 q^{-1})^k y(t) = a_1^k y(t-k)$$

2 (a) This is an OE model. The predictor is

$$\hat{y}(t) = -a\, \hat{y}(t-1) + b\, u(t-1)$$

Since the plant is within the model set, and the input signal is percistent exiting, the parameters a and b converge towards their true values as $N \to \infty$.

(b) Assymptotic distribution of the parameter estimate is given by

$$\sqrt{N}(\hat{\theta} - \theta_0) \in AsN(0, \hat{\Sigma}_\theta)$$

where

$$\begin{aligned}
\hat{\sigma}_N^2 &= \frac{1}{N}\sum_{t=1}^{N} \varepsilon(t, \hat{\theta}_N)^2 \\
\hat{\Sigma}_\theta &= \hat{\sigma}_N^2 \left(\frac{1}{N}\sum_{t=1}^{N} \psi(t, \hat{\theta}_N)\psi^T(t, \hat{\theta}_N) \right)^{-1}
\end{aligned}$$

and

$$\psi(t, \hat{\theta}_N) = -\frac{d\,\hat{y}(t)}{d\,\theta}$$

Since the parameters converge towards the true parameters, $\lim_{N\to\infty} \hat{\sigma}_N^2 = \sigma_e^2$. In our case $\theta = [b]$ and

$$\psi(t, b) = -\frac{d\,\hat{y}(t)}{d\,b} = a\frac{d\,\hat{y}(t-1)}{d\,b} - u(t-1)$$

and

$$\psi(t, b)\psi^T(t, b) = a^2\psi(t-1, b)\psi^T(t-1, b) + u^2(t-1) - 2\,a\psi(t-1, b)u(t-1)$$

When $N \to \infty$ this is converge to the expectation value, and the last term vanish, hence

$$E\,\psi(t, b)\psi^T(t, b) = \frac{E\,u^2(t-1)}{1-a^2} = \frac{\sigma_u^2}{1-a^2}$$

The coveriance of the estimate of b then becomes

$$\frac{\sigma_e^2(1-a^2)}{N\,\sigma_u^2}$$

3 From the data plot one observe that the input signal is dominated by lower frequencies, but the output contains a lot of high-frequencies. The lack of input energy at higher frequencies indicate that there are high-frequency disturbancies on the output.

(a) Figure a) is from the ARX model. The poles are used to explain both plant and disturbance dynamic. Due to the high frequency disturbance, the poles are pulled toward higher frequencies. For the output-error model there is no such effect.

(b) LP-filter the data, both the input signal and the output signal so that frequencies above the bandwidth of the input signal are removed. This will remove the high frequency disturbancies and then the poles of the ARX are primarily used to model the plant.

(c) If the disturbance is removed by the preprocessing, then the uncertainty of the parameter estimate of the output-error model will decrease. The expectation value of the estimate does not change.

4 (a) We can express the input as two steps, a positive one at $t = 0$ and a negative one at $t = T$. The output becomes

$$y(t) = \frac{1}{T}(\int_0^t g(\tau)d\tau - \int_T^t g(\tau - T)d\tau) = \frac{1}{T}(\int_0^t g(\tau)d\tau - \int_0^{t-T} g(\tau)d\tau)$$

Taking the derivative

$$\dot{y}(t) = \frac{1}{T}(g(t) - g(t-T))$$

Now, for the interval $t = \{0, T\}$ the second term is zero since $g(\tau)$ is zero for τ negative. Hence, the impulse responce can be expressed as

$$g(t) = \begin{cases} T\,\dot{y}(t) & 0 < t < T \\ T\,\dot{y}(t) + g(t-T) & T < t \end{cases}$$

Note that to calculate $g(t)$ for $T < t$ requires that one first obtain $g(\tau)$ for $\tau \in \{0, t-T\}$.

(b) If T is small, then the integral in (a) can be approximated

$$y(t) = \frac{1}{T}(\int_0^t g(\tau)d\tau - \int_0^{t-T} g(\tau)d\tau) = \frac{1}{T}(\int_{t-T}^t g(\tau)d\tau \approx g(t)$$

This is the result we expect in the limit when T goes to zero. Note that what is meant by "T small" relates to the smoothness of g.

5 (a) Since $\hat{\theta}_N$ is asymptotically normal distributed, and $P_\theta^{-1/2}\hat{\theta}_N$ is a linear transform of it, so it is also asymptotically normal distributed.

$$\mathrm{E}\,[P^{-1/2}\hat{\theta}_N] = P_\theta^{-1/2}\mathrm{E}\,[\hat{\theta}_N] = P_\theta^{-1/2}\theta_0$$

$$\begin{aligned}\mathrm{Var}(P_\theta^{-1/2}\hat{\theta}_N) = \mathrm{E}\,[(P_\theta^{-1/2}\hat{\theta}_N - P_\theta^{-1/2}\theta_0)(P_\theta^{-1/2}\hat{\theta}_N - P_\theta^{-1/2}\theta_0)^T] =& \\ = P_\theta^{-1/2}\mathrm{Var}(\hat{\theta}_N)P_\theta^{-1/2} = P_\theta^{-1/2}P_\theta P_\theta^{-1/2} = I&\end{aligned}$$

(b) Introduce $\tilde{\theta} = P_\theta^{-1/2}(\hat{\theta}_N - \theta_0)$, a vector of d independent asymptotically normal distributed variables with mean 0 and variance 1 according to (a), ie assymptotically in N, $\tilde{\theta}^T\tilde{\theta} \in \chi^2(d)$

(c) If θ is scalar, this is straight forward, we can generate bounds on $\hat{f}(x)$ by simply adding the uncertainties of θ

$$f(x, \theta \pm \Delta\theta)$$

where

$$\Delta\theta = \sqrt{P_\theta}\sqrt{\chi^2_\alpha(1)}.$$

However with more dimensions this is no longer possible since it would require evaluating f on a hyper-surface of dimension $d-1$. Instead, we can explore f in a number of random directions, in the following way. $\Delta\theta_k$ is chosen randomly and scaled so that

$$\|\Delta\theta_k\|^2 = \chi^2(d)$$

So using this and the known covariance of θ we can generate a new parameter vector:

$$\theta_k = \theta + P^{1/2}\Delta\theta_k$$

From this, one obtains an estimate of the confidence interval with limits

$$\max_k f(x, \theta_k)$$

and

$$\min_k f(x, \theta_k)$$

(d) One dimensional case: there are only to possible samples to take, one sample at each side of the Gaussian. Hence, in the one dimensional case you do not need to take more than two samples.

Two dimensional case: Level curves of the Gaussian form elipsiods. Samples are taken on that elipsoid so that the Gaussian having in total α probability mass outside the elipsoid.