Learning dynamical systems using system identification

(Course: SSY 230)

Exam 1st June 2021

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 COVID 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
Consider a bilinear model structure described by

$$\begin{aligned} x(t) + a_1 x(t-1) + a_2 x(t-2) &= b_1 u(t-1) + b_2 u(t-2) + c_1 x(t-1) u(t-1) \\ y(t) &= x(t) + v(t) \end{aligned}$$

where

$$\theta = [a_1,\ a_2,\ b_1,\ b_2,\ c_1]^T$$

(a) Assume $\{v(t)\}$ to be white noise and compute the predictor $\hat{y}(t,\theta)$ and give the expression for it in the pseudolinear regression form

$$\hat{y}(t,\theta) = \varphi^T(t,\theta)\theta$$

with a suitable vector $\varphi^T(t,\theta)$. (5p)

(b) Now, suppose that $\{v(t)\}$ is not white, but can be modeled as an (unknown) first order ARMA process. Then suggest a suitable predictor for the system. (5p)

Problem 2
Given N observations of $y(t)$ and $\varphi(t)$, the model

$$y(t) = \theta^T \varphi(t) + e(t).$$

and the assumption that the noise $e(t)$ is iid according to

$$f(x) = \frac{x}{\sigma^2} e^{-\frac{x^2}{2\sigma^2}}, \quad x > 0.$$

Formulate the Maximum-Likelihood estimate of σ^2 and θ.

Note: An explicit solution is not likely since the PDF of the noise is not Gaussian. Hence, your answer can be stated as the equations which need to be solved numerically, given the data.

(10p)

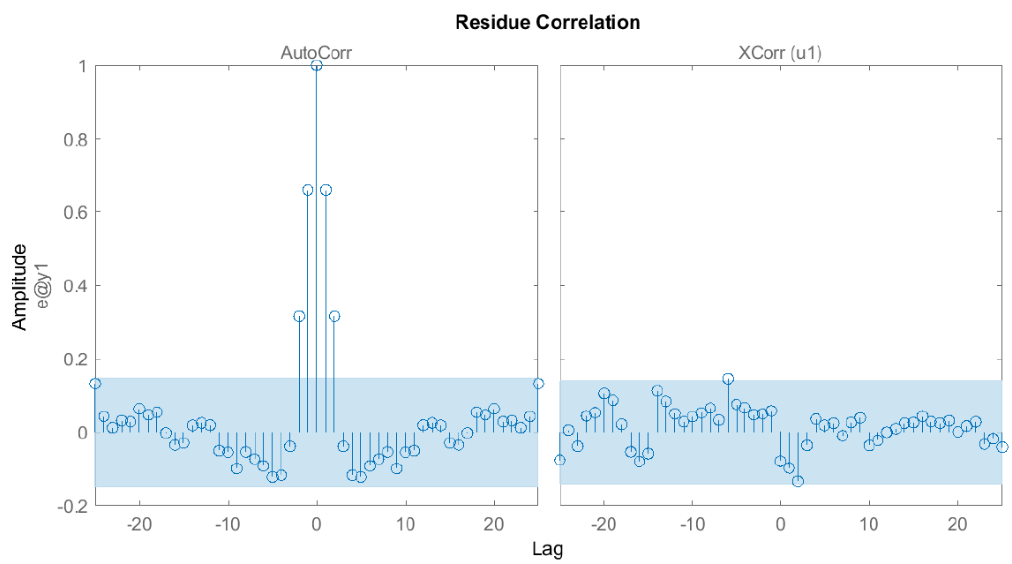
Problem 3

A general polynomial model can be described as

$$y(t) = \frac{B}{F}u(t) + \frac{C}{D}e(t)$$

where the first part is the process model and the second is the disturbance model. The degree of the polynomials F,B,C, and D indicate the model order. By setting some of the polynomials equal to 1, special named structures are obtained, eg, ARX, ARMAX, OE. If non-of them is set to 1 then the structure is called Box-Jenkins model.

A model with specific choice of the degree of the polynomials has been identified and the result of a residual test on validatin data is depicted in the figure. To the left is the auto-correlation of the residuals, and to the right is the cross-correlation between the input signal and the residuals.



(a) Any reason to suggest a change of the process model order?

(3p)

(b) Any reason to suggest a change of the disturbance model order?

(3p)

(c) Which of the polynomials F,B,C, and D, could be suspected to be fixed to 1 in the model? There is no 100% sure answer, but you can do some reasoning. Which of them do you suggest should be changed and to what order?

(4p)

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
Consider the model structure

$$\begin{aligned} \hat{x}(t+1,\theta) &= A(\theta)\hat{x}(t,\theta) + B(\theta)\,u(t) + K(\theta)e(t) \\ y(t) &= C(\theta)\hat{x}(t,\theta) + e(t) \end{aligned}$$

with

$$A(\theta) = \begin{bmatrix} -a_1 & 1 \\ -a_2 & 0 \end{bmatrix}, \qquad B(\theta) = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix},$$

$$C(\theta) = \begin{bmatrix} 1 & 0 \end{bmatrix}, \qquad K(\theta) = \begin{bmatrix} k_1 \\ k_2 \end{bmatrix},$$

$$\theta = \begin{bmatrix} a_1 & a_2 & b_1 & b_2 & k_1 & k_2 \end{bmatrix}^T, \quad \theta \in D_1 \subset \mathbb{R}^6$$

and another structure

$$A(\eta) = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}, \qquad B(\eta) = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix},$$

$$C(\eta) = \begin{bmatrix} \gamma_1 & \gamma_2 \end{bmatrix}, \qquad K(\eta) = \begin{bmatrix} \kappa_1 \\ \kappa_2 \end{bmatrix},$$

$$\eta = \begin{bmatrix} \lambda_1 & \lambda_2 & \mu_1 & \mu_2 & \gamma_1 & \gamma_2 & \kappa_1 & \kappa_2 \end{bmatrix}^T, \quad \eta \in D_2 \subset \mathbb{R}^8$$

Determin D_1 and D_2 so that the two model structures determine the same model set. What about identifiability properties?

Exam 1st June 2021: Solution schetches

1

(a) Given the system description and that $\{v(t)\}$ is white, the predictor becomes $\hat{y}(t, \theta) = x(t)$. It remains writing $x(t)$ as an inner product between $\theta = [a_1,\ a_2,\ b_1,\ b_2,\ c_1]^T$ and a pseudo-regressor. Looking at the expresion for $x(t)$ gives

$$\varphi^T(t, \theta) = [-x(t-1)\ -x(t-2)\ u(t-1)\ u(t-2)\ x(t-1)u(t-1)]$$

where $x(t-1)$ and $x(t-2)$ are generated with the system equation.

(b) In a), the best prediction of $\{v(t)\}$ was 0. Now a better prediction can be obtained using the knowlwedge that it is a first order ARMA process. That is,

$$v(t) = -\alpha\, v(t-1) + \beta\, e(t-1) + e(t)$$

This gives

$$v(t) = \frac{1 + \beta\, q^{-1}}{1 + \alpha\, q^{-1}} e(t) = H(q^{-1}) e(t)$$

and

$$\hat{v}(t|t-1) = (1 - H^{-1}(q^{-1}))v(t) = \frac{(\alpha - \beta)q^{-1}}{1 + \alpha\, q^{-1}} v(t)$$

Hence,

$$\hat{v}(t|t-1) = -\alpha\, \hat{v}(t-1|t-2) + (\alpha - \beta)v(t-1)$$

where $v(t-1) = y(t-1) - x(t)$ and the one step predictor can be put together as

$$\hat{y}(t, \theta) = x(t) + \hat{v}(t|t-1)$$

2

Model

$$y(t) = \theta^T \varphi(t) + e(t)$$

With θ and σ^2 as unknown parameters.

With

$$e(t) = y(t) - \theta^T \varphi(t)$$

The ML method then becomes

$$\max_{\sigma^2,\theta} \prod_{i=1}^{N} \frac{y(t)-\theta^T\varphi(t)}{\sigma^2} e^{-(y(t)-\theta^T\varphi(t))^2/2\sigma^2}$$

Taking the log this becomes:

$$\left[\max_{\sigma^2,\theta} -N\log\sigma^2 + \sum_{i=1}^{N}\log(y(t)-\theta^T\varphi(t)) + \sum_{i=1}^{N} -(y(t)-\theta^T\varphi(t))^2/2\sigma^2\right]$$

So if we want to determine both σ^2 and θ we take the derivative with respect to each separetly:

For σ^2: We have

$$\frac{-N}{\sigma^2} + \sum_{i=1}^{N}(y(t)-\theta^T\varphi(t))^2/2\sigma^4 = 0 \tag{1}$$

and

$$\hat{\sigma^2} = \frac{1}{2N}\sum_{i=1}^{N}(y(t)-\theta^T\varphi(t))^2$$

And for θ:

$$-\sum_{i=1}^{N}\frac{\varphi(t)}{y(t)-\theta^T\varphi(t)} + \sum_{i=1}^{N} 2\varphi(t)(y(t)-\theta^T\varphi(t))/\sigma^2 = 0 \tag{2}$$

Equations (1) and (2) is a set of equations to be solved under the condition

$$y(t)-\theta^T\varphi(t) > 0$$

To give the estimates.

Remark 1: (2) is one equation for each component of θ.

Remark 2: The solution is more complicated than what we have if we assume normal distributed disturbancies

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}}$$

where we have an explicit expression for the estimate.

3

(a) The process model seems to be very good, there is no indication on a correlation between the input signal and the residuals. Hence, nothing left to be modeled in the process model.

(b) The disturbance model is however poor, and there is clear evidense of correlation. The suggestion is therefore to change the disturbance model.

(c) The auto-correlation, $R_\varepsilon(\tau)$, seems to be non-zero for $\tau = \pm 1,\ \pm 2$ and then zero for larger τ. Hence, to explain this, a disturbance model of MA-type is necessary and the suggestion is to change C to a 2nd order polynomial.

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

We need to find the parameter sets so that the models have the same plant model, and also the same disturbance model. Hence, calculate the transfer functions for the two models.

The first model structure

$$
\begin{aligned}
\hat{x}(k+1) &= A\hat{x}(k) + B\hat{x}(k) + Ke(t) \\
y(t) &= C\hat{x} + e(t)
\end{aligned}
$$

gives transfer functions

$$G = C(qI - A)^{-1}B \text{ and } H = C(qI - A)^{-1}K + I$$

Inserting the matrices gives

$$G_1 = \frac{q\,b_1 + b_2}{q^2 + q\,a_1 + a_2} \text{ and } H_1 = \frac{q^2 + q(a_1 + k_1) + a_2 + k_2}{q^2 + q\,a_1 + a_2}$$

For the seconds system, similarly,

$$G_2 = \frac{q(\gamma_1\mu_1 + \gamma_2\mu_2) - \lambda_2\gamma_1\mu_1 - \gamma_2\mu_2\lambda_1}{(q - \lambda_1)(q - \lambda_2)}$$

and

$$H_2 = \frac{q^2 + q(\gamma_1\kappa_1 + \gamma_2\kappa_2 - \lambda_1 - \lambda_2) - \lambda_2\gamma_1\kappa_1 - \gamma_2\kappa_2\lambda_1 + \lambda_1\lambda_2}{(q - \lambda_1)(q - \lambda_2)}$$

Now we look at the transfer functions of the two systems and see when they can be the same.

For the first system:

$$G_1 = \frac{q\,b_1 + b_2}{q^2 + q\,a_1 + a_2} \text{ and } H_1 = \frac{q^2 + q(a_1 + k_1) + a_2 + k_2}{q^2 + qa_1 + a_2}$$

Two transere functions with any combinations of poles or zeros. There is no direct term in G_1 but otherwise the only limitation is that H and G must have the same poles.

However the second model,

$$G_2 = \frac{q(\gamma_1\mu_1 + \gamma_2\mu_2) - \lambda_2\gamma_1\mu_1 - \gamma_2\mu_2\lambda_1}{(q-\lambda_1)(q-\lambda_2)}$$

and

$$H_2 = \frac{q^2 + q(\gamma_1\kappa_1 + \gamma_2\kappa_2 - \lambda_1 - \lambda_2) - \lambda_2\gamma_1\kappa_1 - \gamma_2\kappa_2\lambda_1 + \lambda_1\lambda_2}{(q-\lambda_1)(q-\lambda_2)}$$

we have only real poles. To impose this restriction also for model 1 we need to restrict D_1 by including the constraint $a_1^2 \geq 4a_2$.

Further, if there is a double pole, ($\lambda_1 = \lambda_2$) then there is a zero for G_2 and H_2 at the same value. That is, the zero cancelles one of the poles. For model 1, double poles means $a_1^2 = 4a_2$ and the pole is at $\frac{-a_1}{2}$. As zero at the same place for both plant and disturbance transfere function gives the equations $-\frac{b_2}{b_1} = \frac{-a_1}{2}$ and $-\frac{k_2}{k_1} = \frac{-a_1}{2}$ which can be formulated in the following constraints $b_1 > 0$, $k_1 > 0$, $a_1b_1 = 2b_2$ and $a_1k_1 = 2k_2$ on D_1 which must hold in case there is a double pole.

So if we add these requirements on D_1 model 1 defines the same model set as model 2. On D_2 we need no conatraints.

We have identifiability for model 1 as the parameters correspond to the cofficients in the transfere functions. However for model 2 we have too many unkonowns for the elements in the transfer functions (8 parameters but only 6 coefficients in the transfer functions).