Learning dynamical systems using system identification

(Course: SSY 230)

Exam 30th May 2024

Time: 8:00-12:00

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.

This is an "open book" exam. Help material: This includes

- The course book, or any other book on system identification, also as E-book.
- Mathematical handbook such as BETA, handbook with physical formulas such as Physics Handbook
- Standard Calculator
- Lecture slides, either on paper or on E-reader platform.

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

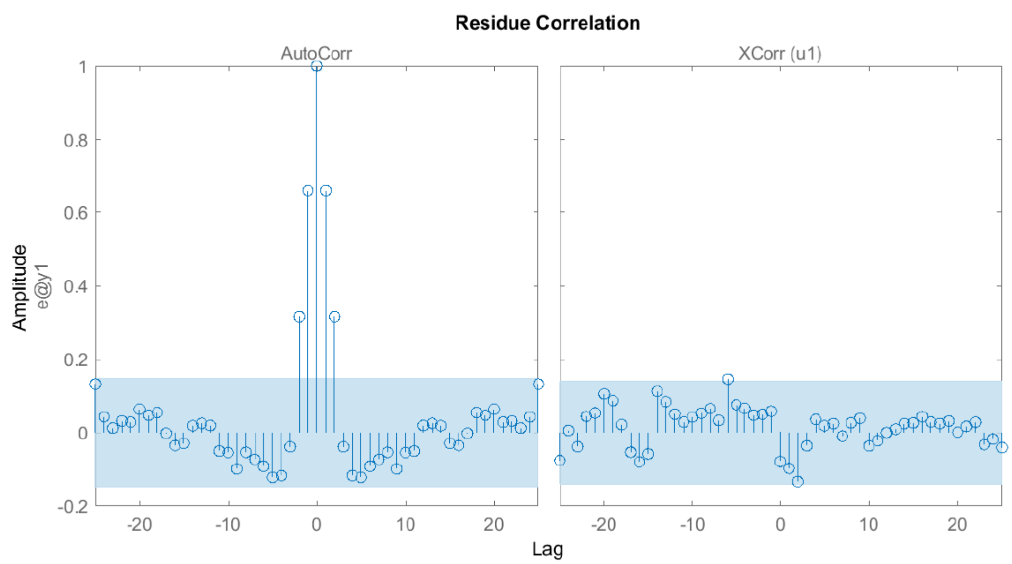
Problem 1

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 2

Closed-loop identification is sometimes necessary, for example in cases where the plant is unstable or when using operational data records from plants in production. Assume the following closed-loop setting

$$\begin{aligned} Y(s) &= G_P(s)U(s) \\ U(s) &= R(s) - G_C(s)Y(s) \end{aligned}$$

and that we have performed indirect identification with the following closed-loop model

$$G_{cl}(s) = \frac{(s+1)(s+3)}{(s+2)(s+3)(s+4)+s+1}$$

The used controller is given by

$$G_C(s) = \frac{1}{s+3}$$

(a) Find a model for the process $G_P(s)$. (6p)

(b) Discuss the main possible transfer function estimation problem associated with indirect identification. (4p)

Problem 3

Assume that input-output data $\{u(t)\}$ and $\{y(t)\}$ are observed from the system

$$\mathcal{S}: \; y(t) + a\,y(t-1) = b\,u(t-1) + w(t) + c\,w(t-1)$$

where $\{u(t)\}$ and $\{w(t)\}$ are independent white-noise sequences with variances $\sigma_w^2 = \sigma_u^2 = \sigma^2$. Assume that the model

$$\mathcal{M}: \; y(t) + a\,y(t-1) = b\,u(t-1)$$

is fitted to data by means of least-squares identification.

(a) Show that the expected squared prediction error is minimized for

$$\hat{\theta} = \begin{pmatrix} \hat{a} \\ \hat{b} \end{pmatrix} = \begin{pmatrix} a - c\,\sigma^2/\mathrm{E}\left[y^2(t)\right] \\ b \end{pmatrix}$$

(6p)

(b) Show also that the prediction error variance is smaller than that for $\hat{a} = a$ and $\hat{b} = b$. (2p)

(c) Calculate also the expression for $\mathrm{E}\left[y^2(t)\right]$. (2p)

Problem 4

Consider the system

$$y(t) = a_0 + b(t)e(t)$$

where

- $e(t)$ is a white Gaussian measurement noise process of zero mean and with unknown variance σ_e^2.
- a_0 is an unknown model parameter.
- $b(t)$ is a known and deterministic time series.

(a) Formulate the Maximum Likelihood estimate for a_0 and σ_e^2 given N measurements of $y(t)$. (5p)

(b) Give an expression for the assymptotic estimate of the variance of the ML estimate.

Hint: One way to solve this problem is to use the information provided on p.210 in S&S.

(5p)

Problem 5

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

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

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 $\mathrm{Var}(\hat{y}(x, \theta))$. 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 $\mathrm{Var}(\hat{y}(x, \theta))$ using the estimated the distribution of $f(x, \hat{\theta}_N)$. (3p)

Exam 30th May 2024: Solution schetches

1 (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.

2 (a) The closed-loop system is defined by

$$G_{cl}(s) = \frac{G_P(s)}{1 + G_C(s)G_P(s)}$$

gives

$$G_P(s) = \frac{G_{cl}(s)}{1 - G_{cl}(s)G_C(s)}$$

Now, use this equation and use the estimated closed-loop system to form an estimate of the plant

$$\hat{G}_P(s) = \frac{(s+1)(s+3)}{(s+2)(s+3)(s+4) + s + 1 - (s+1)} = \frac{s+1}{(s+2)(s+4)}$$

(b) The main disadvantage with indirect identification is that any error in $G_C(s)$ (including deviation from a linear regulator, due to input saturations or anti-windup measurements) will be incorporated in the estimate of $G_P(s)$.

3 (a) We have:

$$y_k + a\, y_{k-1} = b\, u_{k-1} + w_k + c\, w_{k-1}$$

which can re-writen as

$$y_k = -a\, y_{k-1} + b\, u_{k-1} + w_k + c\, w_{k-1}$$

and use the model

$$y_k + \hat{a}\, y_{k-1} = \hat{b}\, u_k$$

We want to find the $\hat{a}_k$ and $\hat{b}_k$ that minimize:

$$\begin{aligned} &\mathrm{E}\,(y_k - \hat{y}_k)^2 = \mathrm{E}\,(y_k + \hat{a}\,y_{k-1} - \hat{b}\,u_{k-1})^2 = \\ &\quad = \mathrm{E}\,(y_k^2 + 2\hat{a}\,y_k y_{k-1} - 2\hat{b}\,y_k u_{k-1} + \hat{a}^2\,y_{k-1}^2 - 2\hat{a}\,\hat{b}\,y_{k-1}u_{k-1} + \hat{b}^2\,u_{k-1}^2) \end{aligned} \quad (1)$$

Introduce

$$\mathrm{E}\,(y_k^2) = r_0$$

We need to express (1) in terms of σ^2 and r_0. Start with

$$\begin{aligned} &\mathrm{E}\,(y_k y_{k-1}) = \mathrm{E}\,(-a\,y_{k-1}y_{k-1} + b\,u_k y_{k-1} + w_k y_{k-1} + c\,w_{k-1}y_{k-1}) = \\ &\quad \mathrm{E}\,(-a\,y_{k-1}y_{k-1}) + \mathrm{E}\,(c\,w_{k-1}y_{k-1}) = -a\,r_0 + \mathrm{E}\,(c\,w_{k-1}y_{k-1}) \end{aligned} \quad (2)$$

and

$$\mathrm{E}\,(w_{k-1}y_{k-1}) = \mathrm{E}\,(w_{k-1}(-a\,y_{k-2} + b\,u_{k-2} + w_{k-1} + c\,w_{k-2})) = \sigma^2$$

Inserting in (2) gives

$$\mathrm{E}\,(y_k y_{k-1}) = -a\,r_0 + c\,\sigma^2$$

Next term

$$\mathrm{E}\,(y_k u_{k-1}) = \mathrm{E}\,(-a\,y_{k-1}u_{k-1} + b\,u_{k-1}u_{k-1} + w_k u_{k-1} + c\,w_{k-1}u_{k-1}) = b\sigma^2 \quad (3)$$

Next

$$\mathrm{E}\,(y_{k-1}u_{k-1}) = 0$$

Now, inserting in (1)

$$\mathrm{E}\,(y_k - \hat{y}_k)^2 = r_0(1 - 2\hat{a}\,a + \hat{a}^2) + 2\hat{a}\,c\,\sigma^2 - 2\hat{b}\,b\,\sigma^2 + \hat{b}^2\sigma^2 \quad (4)$$

Minimize for $\hat{b}$,

$$0 = -2b\,\sigma^2 + 2\hat{b}\,\sigma^2 \;\Rightarrow\; b = \hat{b}$$

and for $\hat{a}$,

$$\begin{aligned} &r_0(1 - 2\hat{a}\,a + \hat{a}^2) + 2\hat{a}\,c\,\sigma^2 - 2\hat{b}\,b\,\sigma^2 + \hat{b}^2\sigma^2 = \\ &\qquad - 2r_0 a + 2r_0\hat{a} + 2c\,\sigma^2 = 0 \;\Rightarrow\; \hat{a} = a - \frac{c\,\sigma^2}{r_0} \end{aligned}$$

(b) Take (4) for $\hat{a} = a$ and $\hat{b} = b$ minus (4) for the terms minimizing (4),

$$\begin{aligned}&r_0(1-2a\,a+a^2)+2a\,c\,\sigma^2-2b\,b\,\sigma^2+b^2\sigma^2\\&-r_0(1-2(a-\frac{c\,\sigma^2}{r_0})\,a+(a-\frac{c\,\sigma^2}{r_0})^2)-2(a-\frac{c\,\sigma^2}{r_0})\,c\,\sigma^2+2b\,b\,\sigma^2-b^2\sigma^2=\\&\qquad-r_0(\frac{2a\,c\,\sigma^2}{r_0}+\frac{c^2\,\sigma^4}{r_0^2}-\frac{2a\,c\,\sigma^2}{r_0})+\frac{2c^2\sigma^4}{r_0}=\frac{c^2\sigma^4}{r_0}>0\end{aligned}$$

(c)

$$\begin{aligned}r_0=\mathrm{E}\,y_k y_k=\mathrm{E}\,(-a\,y_{k-1}+b\,u_{k-1}+w_k+c\,w_{k-1})=\\a^2\,\mathrm{E}\,y_{k-1}^2+b^2\,\mathrm{E}\,u_{k-1}^2+\mathrm{E}\,w_k^2+c^2\,\mathrm{E}\,w_{k-1}^2-2a\,c\mathrm{E}\,y_k w_{k-1}=\\a^2\,r_0+b^2\,\sigma^2+\sigma^2+c^2\,\sigma^2-2a\,c\,\sigma^2\quad\Rightarrow\end{aligned}$$

$$r_0=\sigma^2\frac{b^2+c^2+1-2a\,c}{1-a^2}$$

4 (a) We have

$$y(t)=a_0+b(t)e(t)$$

and

$$e(t)=\frac{y(t)-a_0}{b(t)}$$

Now, $e(t)$ is normally distributed (Gaussian) so it follows the distribution,

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

Put in our values for e(t)

$$f(\frac{y(t)-a_0}{b(t)})=\frac{1}{\sqrt{2\pi\sigma_e^2}}e^{-\frac{(\frac{y(t)-a_0}{b(t)})^2}{2\sigma_e^2}}$$

We want the maximum probability for all N measurements of $y(t)$:

$$\max_{\theta}p(y|\theta)=\max_{a_0,\sigma_e}\prod_{t=1}^{N}\frac{1}{\sqrt{2\pi\sigma_e^2}}e^{-\frac{(\frac{y(t)-a_0}{b(t)})^2}{2\sigma_e^2}}=\max_{a_0,\sigma_e}(\frac{1}{\sqrt{2\pi\sigma_e^2}})^N e^{\sum_{t=1}^{N}-\frac{(y(t)-a_o)^2}{2b^2(t)\sigma_e^2}}$$

We take the logarithm of this

$$\begin{aligned}\max_{\theta}\log p(y|\theta)=\max_{a_0,\sigma_e}\log(\frac{1}{\sqrt{2\pi\sigma_e^2}})^N+\sum_{t=1}^{N}-\frac{(y(t)-a_o)^2}{2b^2(t)\sigma_e^2}=\\\max_{a_0,\sigma_e}N\log(\frac{1}{\sqrt{2\pi}})-\frac{N}{2}\log(\sigma_e^2)-\sum_{i=1}^{N}\frac{(y(t)-a_o)^2}{2b^2(t)\sigma_e^2}\end{aligned}$$

Take the derivative with respect to $a_{0,}\sigma_e^2$: For a_0:

$$\frac{d\,\log p(y|\theta)}{da_0} = 2\sum_{t=1}^{N}\frac{(y(t)-a_o)}{b^2(t)\sigma_e^2} = 0 \quad\Leftrightarrow$$

$$\sum_{t=1}^{N}\frac{y(t)}{b^2(t)} = \sum_{t=1}^{N}\frac{a_o}{b^2(t)} \quad\Leftrightarrow$$

$$a_0 = \frac{\sum_{t=1}^{N}\frac{y(t)}{b^2(t)}}{\sum_{t=1}^{N}\frac{1}{b^2(t)}}$$

For σ_e^2 (derivative with respect to σ_e^2)

$$\frac{d\,\log p(y|\theta)}{d\sigma_e^2} = -\frac{N}{2\sigma_e^2} + \sum_{t=1}^{N}\frac{(y(t)-a_o)^2}{2b^2(t)\sigma_e^4} = 0 \quad\Leftrightarrow$$

$$N\sigma_e^2 = \sum_{t=1}^{N}\frac{(y(t)-a_o)^2}{b^2(t)} \quad\Leftrightarrow$$

$$\sigma_e^2 = \frac{1}{N}\sum_{t=1}^{N}\frac{(y(t)-a_o)^2}{b^2(t)}$$

(b) The ML estimate is efficient (S& S p 210), so the asymptotic estimate of the variance converges to the Cramér-Rao lower bound. That is,

$$\mathrm{Cov}\,\hat{\theta} = M^{-1}$$

where M, is the fisher information matrix, so for our case:

$$\theta = \begin{bmatrix} a_0 \\ \sigma_e^2 \end{bmatrix}$$

and

$$M = -E\begin{bmatrix} \frac{d^2\log p(y|\theta)}{da_0^2} & \frac{d^2\log p(y|\theta)}{da_0 d\sigma_e^2} \\ \frac{d^2\log p(y|\theta)}{d\sigma_e^2 da_0} & \frac{d^2\log p(y|\theta)}{d(\sigma_e^2)^2} \end{bmatrix}$$

From before

$$\frac{d\,\log p(y|\theta)}{da_0} = 2\sum_{t=1}^{N}\frac{(y(t)-a_o)}{2b^2(t)\sigma_e^2}$$

Taking the derivative once more gives

$$\frac{d^2\log p(y|\theta)}{da_0^2} = -2\sum_{t=1}^{N}\frac{1}{2b^2(t)\sigma_e^2} = -\sum_{t=1}^{N}\frac{1}{b^2(t)\sigma_e^2}$$

This is a deterministic expression, so taking the expectation does not change anything, ie,

$$E\left[\frac{d^2 \log p(y|\theta)}{da_0^2}\right] = -\sum_{t=1}^{N} \frac{1}{b^2(t)\sigma_e^2}$$

Now to obtain $\frac{d^2 \log p(y|\theta)}{da_0 d\sigma_e^2}$, we use $\frac{d \log p(y|\theta)}{da_0}$ and take the derivative with respect to σ_e^2

$$\frac{d^2 \log p(y|\theta)}{da_0 d\sigma_e^2} = -2\sum_{t=1}^{N} \frac{(y(t)-a_o)}{2b^2(t)\sigma_e^4}$$

Taking the expectation gives

$$E\left[\frac{d^2 \log p(y|\theta)}{d\sigma_e^2}\right] = -E\left[2\sum_{t=1}^{N} \frac{(y(t)-a_o)}{2b^2(t)\sigma_e^4}\right] - \frac{1}{\sigma_e^4}\sum_{t=1}^{N} \frac{E\left[e(t)\right]}{b(t)} = 0$$

And for $\frac{d^2 \log p(y|\theta)}{d(\sigma_e^2)^2}$, first derivative

$$\frac{d\ \log p(y|\theta)}{d\sigma_e^2} = -\frac{N}{2\sigma_e^2} + \sum_{t=1}^{N} \frac{(y(t)-a_o)^2}{2b^2(t)(\sigma_e^2)^2}$$

Taking the derivative once more gives

$$\frac{d^2\ \log p(y|\theta)}{d(\sigma_e^2)^2} = \frac{N}{2\sigma_e^4} - 2\sum_{t=1}^{N} \frac{(y(t)-a_o)^2}{2b^2(t)\sigma_e^6}$$

And then apply the expectation to this

$$E\left[\frac{d^2\ \log p(y|\theta)}{d(\sigma_e^2)^2}\right] = \frac{N}{2\sigma_e^4} - \frac{1}{\sigma_e^6}\sum_{t=1}^{N} E\left[\frac{(y(t)-a_o)^2}{b^2(t)}\right] =$$

$$\frac{N}{2\sigma_e^4} - \frac{1}{\sigma_e^6}\sum_{t=1}^{N} E\left[e(t)^2\right] = \frac{N}{2\sigma_e^4} - \frac{N\sigma_e^2}{\sigma_e^6} = -\frac{N}{2\sigma_e^4}$$

If we put this together we obtain

$$M = \begin{bmatrix} \sum_{t=1}^{N} \frac{1}{b^2(t)\sigma_e^2} & 0 \\ 0 & \frac{N}{2\sigma_e^4} \end{bmatrix}$$

So asymptotically:

$$\mathrm{Cov}\,\hat{\theta} = M^{-1} = \begin{bmatrix} \frac{1}{\sum_{t=1}^{N} \frac{1}{b^2(t)\sigma_e^2}} & 0 \\ 0 & \frac{2\sigma_e^4}{N} \end{bmatrix}$$

5 (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. It is easy to show this in equations, and also numerically in a small Matlab example.

(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:

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

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

Variance estimate:

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