LUND UNIVERSITY

Department of
AUTOMATIC CONTROL

FRTN35 System Identification

Final Exam October 31, 2018, 8am - 13pm

General Instructions

This is an open book exam. You may use any book you want, including the slides from the lecture, but no exercises, exams, or solution manuals are allowed. Solutions and answers to the problems should be well motivated. The exam consists of 7 problems. The credit for each problem is indicated in the problem. The total number of credits is 25 points. Preliminary grade limits:

Grade 3:	12 – 16 points
Grade 4:	17 – 21 points
Grade 5:	22 – 25 points

Results

The result of the exam will become accessible through LADOK.

1. Consider the system

$$y_k = -ay_{k-1} + bu_{k-1} + e_k$$

where the $\{u_k\}$ and $\{e_k\}$ are noise process that are mutually independent and fulfill $u_k \sim N(0, \sigma_u^2)$ and $e_k \sim N(0, \sigma_e^2)$.

a. Assume that you want to estimate a and b using the model

$$y_k = -ay_{k-1} + bu_{k-1}$$

Derive the least-squares estimate of a and b based on N observations of the system. Show that the estimate will be unbiased. (2 p)

b. Your colleague has been collecting $N = 1000$ data points from the system and provided you with the results below. Use what you deem necessary to give a numerical estimate of the process parameters based on your findings in **a.**. Give also an estimate of the process noise variance σ_e^2.

$$\sum_{k=2}^N y_{k-1}^2 = 2629, \quad \sum_{k=2}^N y_{k-1}u_{k-1} = 18.1, \quad \sum_{k=2}^N u_{k-1}^2 = 979.9$$
$$\sum_{k=2}^N y_k u_{k-1} = 1892, \quad \sum_{k=2}^N y_k y_{k-1} = 482.8$$
$$V(\widehat{\theta}) = 1016$$

(1 p)

c. Assuming that the noise variance σ_e^2 is known, derive the loss function of the maximum-likelihood estimate of a and b. Show also that the maximum-likelihood estimate in this case coincides with the least-squares estimate. You may use the fact that the probability distribution function for a normally distributed variable v is given by

$$f_v(v) = \frac{1}{\sqrt{2\pi}\sigma_v} e^{-(v-\mu_v)^2/2\sigma_v^2}$$

(1 p)

Solution

a. Start by defining the following matrices

$$Y = \begin{pmatrix} y_2 \\ \vdots \\ y_N \end{pmatrix}, \quad \Phi = \begin{pmatrix} -y_1 & u_1 \\ \vdots & \vdots \\ -y_{N-1} & u_{N-1} \end{pmatrix}, \quad e = \begin{pmatrix} e_2 \\ \vdots \\ e_N \end{pmatrix}$$

For the equation $Y = \Phi\theta$, the least-squares estimate of θ is given by $\widehat{\theta} = (\Phi^\top\Phi)^{-1}\Phi^\top Y$, which is obtained through differentiation of the cost function

$$V(\theta) = (Y - \Phi\theta)^\top (Y - \Phi\theta)$$

Therefore we investigate

$$\begin{aligned}\left(\frac{1}{N-1}\Phi^T\Phi\right)^{-1} &= \left(\frac{1}{N-1}\begin{pmatrix} \sum_{k=2}^N y_{k-1}^2 & -\sum_{k=2}^N y_{k-1}u_{k-1} \\ -\sum_{k=2}^N y_{k-1}u_{k-1} & \sum_{k=2}^N u_{k-1}^2 \end{pmatrix}\right)^{-1} \\ &\longrightarrow \begin{pmatrix} E\{y_{k-1}^2\} & -E\{y_{k-1}u_{k-1}\} \\ -E\{y_{k-1}u_{k-1}\} & E\{u_{k-1}^2\} \end{pmatrix}^{-1} \text{ as } N \to \infty\end{aligned}$$

and

$$\frac{1}{N-1}\Phi^T Y = \frac{1}{N-1}\begin{pmatrix} -\sum_{k=2}^{N} y_k y_{k-1} \\ \sum_{k=2}^{N} y_k u_{k-1} \end{pmatrix} \longrightarrow \begin{pmatrix} -E\{y_k y_{k-1}\} \\ E\{y_k u_{k-1}\} \end{pmatrix} \text{ as } N \to \infty$$

Identifying the elements of the matrices we get

$$\begin{aligned} E\{y_{k-1}^2\} &= \frac{b^2\sigma_u^2 + \sigma_e^2}{1-a^2} \\ E\{y_{k-1}u_{k-1}\} &= 0 \\ E\{u_{k-1}^2\} &= \sigma_u^2 \\ E\{y_k y_{k-1}\} &= -aE\{y_{k-1}^2\} = -\frac{ab^2\sigma_u^2 + a\sigma_e^2}{1-a^2} \\ E\{y_k u_{k-1}\} &= b\sigma_u^2 \end{aligned}$$

Putting the pieces together results in

$$\widehat{\theta} \longrightarrow \begin{pmatrix} -\frac{E\{y_k y_{k-1}\}}{E\{y_{k-1}^2\}} \\ \\ \frac{E\{y_k u_{k-1}\}}{E\{u_{k-1}^2\}} \end{pmatrix} = \begin{pmatrix} a \\ b \end{pmatrix}$$

b. Inserting the given data in the estimator derived in **a** gives the following result

$$\widehat{\theta} = \begin{pmatrix} \widehat{a} \\ \widehat{b} \end{pmatrix} = \begin{pmatrix} -0.170 \\ 1.928 \end{pmatrix}$$

As for the noise variance we have $\widehat{\sigma}_e^2 = \frac{2V(\widehat{\theta})}{N-p} = 2.036$ (where $p = 2$).

c. The likelihood function is given by

$$L(\bar{\theta}) = f_e(\varepsilon_n, \ldots, \varepsilon_N) = \prod_{k=n}^{N} f_e(\varepsilon_k) = \left(\frac{1}{\sqrt{2\pi}\sigma_e}\right)^{N-n} e^{-\frac{1}{2\sigma_e^2}\sum_{k=n}^{N}\varepsilon_k^2}$$

where $\varepsilon_k = y_k - (-ay_{k-1} + bu_{k-1})$. Further on,

$$\log L(\bar{\theta}) = -(N-n)\log(\sqrt{2\pi}\sigma) - \frac{1}{2\sigma_e^2}\sum_{k=n}^{N}(y_k - (-ay_{k-1} + bu_{k-1}))^2$$

Hence, in order to maximize the log-likelihood with respect to a and b, it is sufficient to minimize the loss function

$$J(a,b) = \sum_{k=n}^{N}(y_k - (-ay_{k-1} + bu_{k-1}))^2$$

This loss function is the same that we minimize when solving the least-squares problem. We can therefore conclude that the maximum-likelihood and least-squares estimates coincide.

2. The purpose of model validation is to verify that the identified model fulfills the modeling requirements. There are several tests which can be used for model validation. Describe one limitation to each one of the following test methods. For full points, each of the described limitations should be *different*.

a) Cross-validation simulation. (1 p)

b) Residual analysis. (1 p)

c) Coherence spectrum. (1 p)

Solution

a) From only looking at a cross-validation simulation, it is difficult to know if the fit percentage value is sufficient or not, i.e. if the unexplained output variance is high or low.

b) Residual analysis can not be used to investigate if the estimation is consistent. If we have over-fitted the data the residual analysis still would give good result and draw the conclusion that the model could be used to predict the behavior. But if we perform a cross validation test with data that have not been previously used we might see that the model can not predict the behavior.

c) It can only be used to verify if can expect good result of an identified linear model by using the data. Even if the coherence is close to one it doesn't guarantee that we can find a good model. We can not separate if the low coherence value is due to high noise or nonlinearities.

3. As your employer's expert in system identification, you have been given the task of investigating if model-based controller design could be a solution for improving the performance of a process. The process can be assumed to be on the form

$$Y(z) = H(z)U(z) + V(z)$$

a. A colleague of yours, without your skills in system identification, suggests that you could estimate a process model by

$$\widehat{H}_A(e^{i\omega h}) = \frac{Y_N(i\omega)}{U_N(i\omega)}$$

where Y_N and U_N are the discrete Fourier transforms of a time-series of output and input data, respectively. Show your colleague why this might not be such a good idea from the viewpoints of consistency and noise sensitivity. (2 p)

b. How can the consistency of the estimator in **a)** be improved? (1 p)

c. Explain why the estimator

$$\widehat{H}_B(e^{i\omega h}) = \frac{\widehat{S}_{yu}(i\omega)}{\widehat{S}_{uu}(i\omega)}$$

can be expected to be less sensitive to noise than $\widehat{H}_A(e^{i\omega h})$. (1 p)

Solution

a. The suggested estimator is not statistically consistent. It is only defined for a fixed number of frequency points (which is a result of the discrete Fourier transform), and is asymptotically unbiased at these points. However, the variance of the estimator does not decrease as the number of data points increase.

Also, it has poor noise properties, since it does not include a noise model. Thus, any noise, white or colored, could potentially affect Y_N severely, and in length, the model estimate $\widehat{H}$. (2 p)

b. Divide the dataset into blocks of data, perform the estimation on each subset, and average the result. (1 p)

c. The expected contribution from the disturbance v in $\widehat{S}_{yu}$ is small in cases where the input and the disturbance are uncorrelated, whereas the disturbance contribution to $Y_N(i\omega)$ and thus $\widehat{H}_A(e^{i\omega h})$ might be considerable. (1 p)

4. A second order transfer function $G(s)$ has been identified:

$$G(s) = \frac{s + 0.25}{s^2 + 3s + 2}$$

You believe that it might be approximated by a first-order model instead.

a. Compute the first order Padé approximation of $G(s)$:

$$\hat{G}_1(s) = \frac{b}{s + a}$$

Conclude why this approximation is not good in this particular case. (1 p)

b. A balanced realization is another method that can be used for model reduction. Describe what is meant by a *balanced realization*. (1 p)

c. A balanced state-space realization of a stable discrete time linear system is shown below. Decide if it is advisable to perform a model reduction for this system. (2 p)

$$\begin{cases} x(k+1) = \begin{pmatrix} -0.6639 & -0.5242 \\ -0.5242 & 0.2639 \end{pmatrix} x(k) + \begin{pmatrix} 0.8345 \\ -0.5511 \end{pmatrix} u(k) \\ y(k) = \begin{pmatrix} 0.8345 & -0.5511 \end{pmatrix} x(k) \end{cases}$$

Solution

a. The Padé approximation is based on the Taylor series expansion of $G(s)$. This is calculated in (1).

$$G(s) = G(0) + \frac{dG}{ds}(0) + \frac{1}{2}\frac{d^2G}{ds^2}(0) + \ldots = \frac{1}{8} + \frac{5}{16}s + \frac{1}{2}\left(-\frac{17}{16}\right)s^2 + \ldots \tag{1}$$

For a first order approximation we only need to keep the two first terms, giving the truncated polynomial (3).

$$G_1(s) = \frac{1}{8} + \frac{5}{16}s \tag{2}$$

The resulting approximation should be a rational function B_1/A_1, which should match G_1. B_1 and A_1 is found by matching polynomial coefficients of (3).

$$B_1(s) = G_1(s)A_1(s) \quad \Rightarrow \quad b_0 = \left(\tfrac{1}{8} + \tfrac{5}{16}s\right)(s+a) = \tfrac{a}{8} + \left(\tfrac{1}{8} + \tfrac{5a}{16}\right)s + \tfrac{5}{16}s^2 \tag{3}$$

Skip the s^2-coefficient and match the other two, this results in (4), giving the Padé approximation (5).

$$\begin{cases} a = -\frac{2}{5} \\ b = -\frac{1}{20} \end{cases} \tag{4}$$

$$\hat{G}_1(s) = \frac{1/20}{2/5 - s} \tag{5}$$

This approximation is clearly unstable, which the original process was not. Therefore this is a bad approximation.

b. A balanced realisation has 'balanced' observability and reachability properties, that is to say the Gramians P and Q are equal.

c. To be able to determine if it advisable to perform a model reduction the observability or the reachability Gramian has to be calculated, which one does not matter as they are equal. Let us here consider the reachability Gramian, which is the solution P of (6), where Φ and Γ are system matrices from (c), given in (7).

$$\Phi P \Phi^T - P + \Gamma\Gamma^T = 0 \tag{6}$$

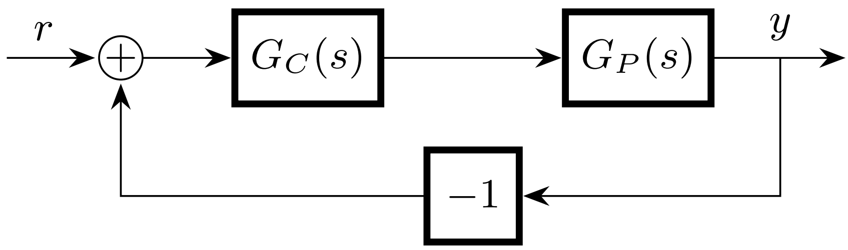

Figure 1 The control system in Problem 5.

$$\Phi = \begin{pmatrix} -0.6639 & -0.5242 \\ -0.5242 & 0.2639 \end{pmatrix} \quad , \quad \Gamma = \begin{pmatrix} 0.8345 \\ -0.5511 \end{pmatrix} \tag{7}$$

P is a diagonal matrix (8), and inserting this and the matrices from (7) gives (9).

$$P = \begin{pmatrix} P_1 & 0 \\ 0 & P_2 \end{pmatrix} \tag{8}$$

$$\begin{pmatrix} -0.5592P_1 + 0.2748P_2 + 0.6964 & 0.3480P_1 - 0.1383P_2 - 0.4599 \\ 0.3480P_1 - 0.1383P_2 - 0.4599 & 0.2748P_1 - 0.9304P_2 + 0.3037 \end{pmatrix} = 0 \tag{9}$$

Solving (9) for P_1 and P_2 gives (10). As the two values not are of different magnitude, a model reduction is not advisable.

$$\begin{cases} P_1 = 1.6443 \\ P_2 = 0.8121 \end{cases} \tag{10}$$

5. You are given the assignment to identify an unstable process. A stabilizing controller exists, but it is desired to increase the closed loop performance by using some kind of model-based based control scheme. The current control system is given in Figure 1. The controller is given as:

$$G_c(s) = \frac{10s+20}{s}$$

You decide to use indirect identification, and the result after system identification is the transfer function from r to y, given in:

$$G_{yr}(s) = \frac{10s^2+70s+100}{s^3+12s^2+67s+100}$$

a. Calculate the process transfer function $G_P(s)$. Confirm that the process is unstable. (2 p)

b. What kind of problems might occur when using indirect identification? (1 p)

c. Suggest an alternative identification strategy and conditions for that this strategy returns a correct estimate. (1 p)

Solution

a. The closed loop transfer function is given by (11) (calculated from the block diagram).

$$G_{yu_c}(s) = \frac{G_p(s)G_c(s)}{1+G_p(s)G_c(s)} \tag{11}$$

The process transfer function is now given as (12)

$$G_p(s) = \frac{G_{yu_c}(s)}{G_c(s)\left(1-G_{yu_c}(s)\right)} \tag{12}$$

By inserting the given transfer functions $G_p(s)$ can be calculated, according to (13).

$$G_p(s) = \frac{\frac{10s^2+70s+100}{s^3+12s^2+67s+100}}{\frac{10s+20}{s}\left(1-\frac{10s^2+70s+100}{s^3+12s^2+67s+100}\right)} = \frac{\frac{10s^2+70s+100}{s^3+12s^2+67s+100}}{\frac{(10s+20)(s^3+2s^2-3s)}{s(s^3+12s^2+67s+100)}} =$$

$$= \frac{10s^2+70s+100}{(10s+20)(s^2+7s-3)} = \frac{(10s+20)(s+5)}{(10s+20)(s-1)(s+3)} = \frac{s+5}{(s-1)(s+3)} \tag{13}$$

The transfer function clearly has an unstable pole.

b. Any nonlinearity in the controller, such as saturations and anti-windup schemes, directly degrade the result.

c. The alternative is to use direct identification. Here you must assure that you use an input signal, u_c that is exciting enough, otherwise you might end up with an inverse model of the controller.

6. The impulse response coefficients (or Markov parameters) $\{h_k\}_{k=1}^{\infty}$ form the transfer function

$$H(z) = \sum_{k=1}^{\infty} h_k z^{-k}, \quad h_k = CA^{k-1}B$$

a. Show that a Hankel matrix of these coefficients can be factorised as

$$\begin{aligned}\mathscr{H}_{r,s}^{(k)} &= \begin{pmatrix} h_{k+1} & h_{k+2} & \cdots & h_{k+s} \\ h_{k+2} & h_{k+3} & \cdots & h_{k+s+1} \\ \vdots & \vdots & \ddots & \vdots \\ h_{k+r} & h_{k+r+1} & \cdots & h_{k+r+s-1} \end{pmatrix} \\ &= \begin{pmatrix} C \\ CA \\ \vdots \\ CA^{r-1} \end{pmatrix} A^k \begin{pmatrix} B & AB & \dots & A^{s-1}B \end{pmatrix}\end{aligned}$$

(2 p)

b. How can this fact be exploited for system identification purposes? (1 p)

c. Explain the differences between realization-based and subspace-based methods for system identification. (1 p)

Solution

The factorization property is verified by direct substitution of Markov parameters $h_k = CA^{k-1}B$. Using a numerical factorization such as the singular value decomposition it is possible to find estimates of the extended observability and controllability matrices. In turn, this information can be used to determine a state-space realization $\{A, B, C\}$.

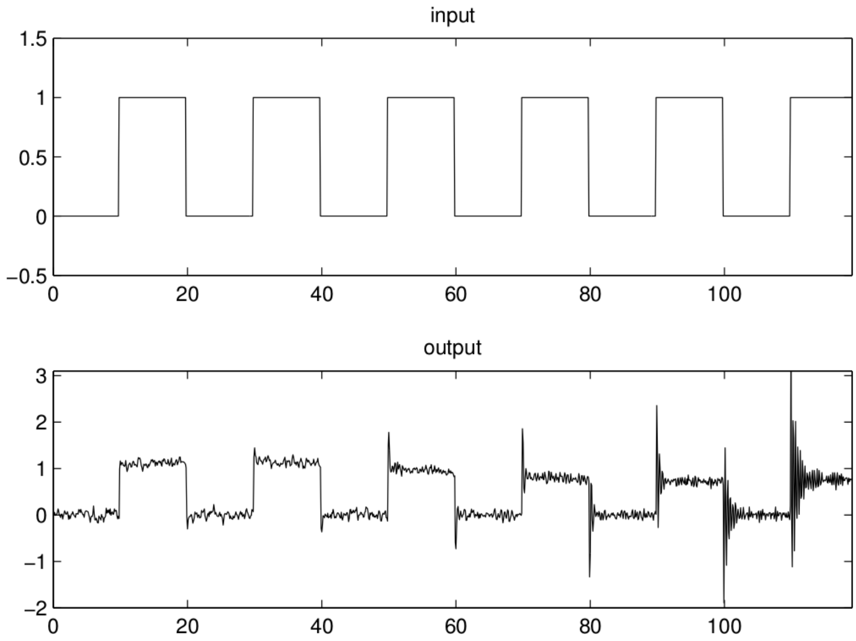

Figure 2 Input-output data in Problem 7.

7. Input-output data from an unknown system is given in Figure 2. One attempt to model the system is to use an ARX-model according to:

$$y(k)+a_1y(k-1)+a_2y(k-2)=b_1u(k-1)+b_2u(k-2)$$

Do you expect the model to be a good description of the data? Why? In case your answer is no to the first question, suggest a more appropriate model. (2 p)

Solution

It is clearly seen in the figure that the system dynamics is changing, and a time-invariant model is therefore probably not the best choice of model.

A more appropriate model would be time-varying, one could e.g. keep the ARX model, but estimate its parameters recursively with a Kalman filter or with exponential forgetting.