LUND INSTITUTE
OF TECHNOLOGY
Lund University

Department of
**AUTOMATIC CONTROL**

# FRT 041 System Identification

**Final Exam March 7, 2011, 2pm - 7pm**

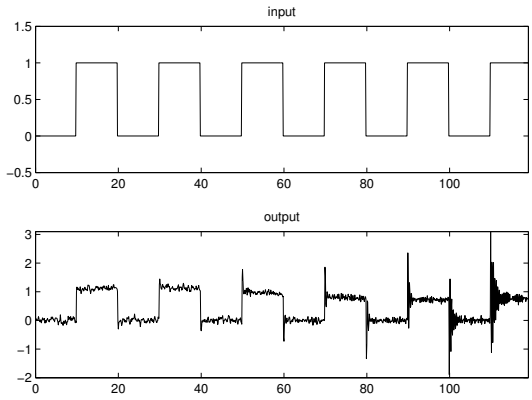
### General Instructions

This is an open book exam. You may use any book you want, but no note, 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 results of the exam will be posted at the latest March 21 on the note board on the first floor of the M-building.




**Figur 1** Input-output data in Problem 1.

**1.** Input-output data from an unknown system is given in Figure 1. One attempt to model the system is to use an ARX-model according to (1).

$$y(k) + a_1 y(k-1) + a_2 y(k-2) = b_1 u(k-1) + b_2 u(k-2) \tag{1}$$

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*

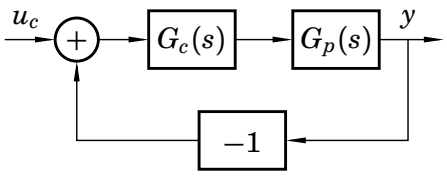
It is clearly seen in the figure that the system dynamics is changing, and a time-invariant model is therefore probably not the best choice of model.

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.

**2.** 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 2. The controller is given according to (2).

$$G_c(s) = \frac{10s + 20}{s} \tag{2}$$

You decide to use indirect identification, and the result after system identification is the transfer function from $u_c$ to $y$, given in (3).




**Figur 2** The control system in Problem 2.

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

**a.** Calculate the process transfer function. 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 (4) (calculated from the block diagram).

$$G_{yu_c}(s) = \frac{G_p(s)G_c(s)}{1 + G_p(s)G_c(s)} \tag{4}$$

The process transfer function is now given as (5)

$$G_p(s) = \frac{G_{yu_c}(s)}{G_c(s)\left(1 - G_{yu_c}(s)\right)} \tag{5}$$

By inserting the given transfer functions $G_p(s)$ can be calculated, according to (6).

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

$$= \frac{10s^2 + 70s + 100}{(10s + 20)(s^2 + 7s - 3)} = \frac{(10s + 20)(s + 5)}{(10s + 20)(s - 1)(s + 3)} = \frac{s + 5}{(s - 1)(s + 3)} \tag{6}$$

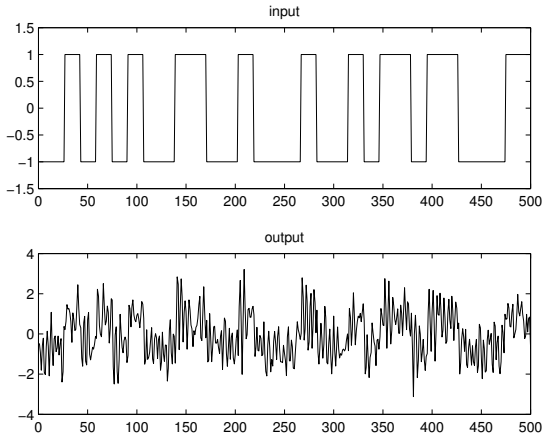
The transfer function clearly has an unstable pole.

**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.

**3.** An identification experiment is performed on a process, see Figure 3. According to physical knowledge of the process it makes sense to assume that a model of it will have the form (7), where $\{w_k\}$ is assumed to be white noise with unknown variance.

$$y_k + ay_{k-2} = bu_{k-1} + w_k \tag{7}$$




**Figur 3** Input-output data in Problem 3.

**a.** The input signal used in this experiment was a pseudorandom binary sequence. Discuss why this choice of input is a good choice in general. (1 p)

**b.** Calculate a least-squares estimate of the parameters $a$ and $b$, also calculate an estimate of the noise variance. The following partial results might help you in the calculations, the number of samples is $N = 500$. (3 p)

$$\sum_{k=1}^{N-2} y_k = -32.48 \qquad \sum_{k=1}^{N-2} y_k^2 = 691.88$$

$$\sum_{k=2}^{N-1} y_{k+1}^2 = 690.65 \qquad \sum_{k=2}^{N-1} y_{k+1}y_k = 325.16$$

$$\sum_{k=1}^{N-2} y_{k+2}y_k = -17.10 \qquad \sum_{k=1}^{N-2} y_k u_{k+1} = 341.98$$

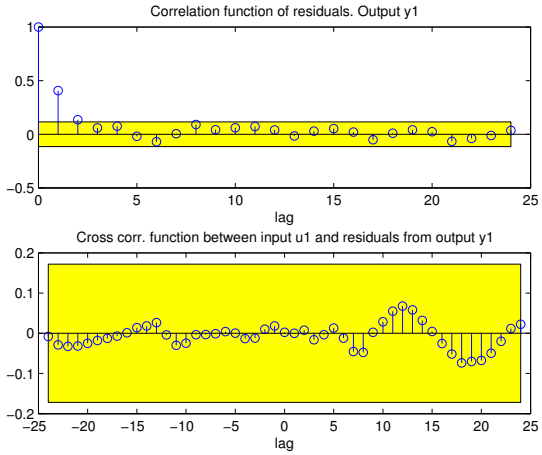
$$\sum_{k=1}^{N-2} y_k u_{k+2} = 310.81 \qquad \sum_{k=2}^{N-1} y_{k+1}u_k = 380.46$$

$$\sum_{k=2}^{N-1} u_k = -64.00 \qquad \sum_{k=2}^{N-1} u_k^2 = 498.00$$

**c.** After a model has been identified it has to be validated in some way, to assure that it fulfills the modelling requirements. There are several tests which can be used for model validation. Mention one limitation to each one of the following test methods.

*i*) Residual analysis. (1 p)

*ii*) Coherence spectrum. (1 p)




**Figur 4** Residual analysis performed in Problem 3. The highlighted area indicates indicates a 99 % confidence region.

**d.** Residual analysis is performed on the previously identified model, see Figure 4. What conclusions can you draw from this? What would be your next step in the identification process? (1 p)

*Solution*

**a.** A pseudorandom binary sequence is convenient to work with as it has a given amplitude, has low autocorrelation and the period of it can be choosen arbitrarily long.

**b.** The model can be written as (8) and all observations can be collected into (9).

$$y_k = \phi_k^T \theta + w_k \quad , \quad \phi_k^T = \begin{pmatrix} -y_{k-2} & u_{k-1} \end{pmatrix} \quad , \quad \theta^T = \begin{pmatrix} a & b \end{pmatrix} \tag{8}$$

$$\begin{aligned} Y_N &= \Phi_N \theta + W, & Y_N^T &= \begin{pmatrix} y_3 & \dots & y_N \end{pmatrix} \\ \Phi_N^T &= \begin{pmatrix} \phi_3^T & \dots & \phi_N^T \end{pmatrix}, & W^T &= \begin{pmatrix} w_3 & \dots & w_N \end{pmatrix} \end{aligned} \tag{9}$$

The least-squares estimate is now given by (10), where the components of this expression is given in (11) and (12).

$$\hat{\theta} = \left( \Phi_N^T \Phi_N \right)^{-1} \Phi_N^T Y_N \tag{10}$$

$$\Phi_N^T \Phi_N = \begin{pmatrix} \sum_{k=1}^{N-2} y_k^2 & -\sum_{k=1}^{N-2} y_k u_{k+1} \\ -\sum_{k=1}^{N-2} y_k u_{k+1} & \sum_{k=2}^{N-1} u_k^2 \end{pmatrix} \tag{11}$$

$$\Phi_N^T Y_N = \begin{pmatrix} -\sum_{k=1}^{N-2} y_{k+2} y_k \\ \sum_{k=2}^{N-1} y_{k+1} u_k \end{pmatrix} \tag{12}$$

Insertion of the values in the sums gives (13).

$$\hat{\theta} = \begin{pmatrix} 691.88 & -341.98 \\ -341.98 & 498.00 \end{pmatrix}^{-1} \begin{pmatrix} 17.10 \\ 380.46 \end{pmatrix} = \begin{pmatrix} 0.6091 \\ 1.1822 \end{pmatrix} \tag{13}$$

An unbiased estimate of the noise variance is given by (14), where $p$ is the number of estimated parameters and $V$ is the loss function.

$$\hat{\sigma}_w^2 = \frac{2}{N-p} V(\hat{\theta}) \tag{14}$$

The loss function can be calculated according to (15)

$$V(\hat{\theta}) = \frac{1}{2} \varepsilon_N^T \varepsilon_N = \frac{1}{2} \left( Y_N - \Phi_N \hat{\theta} \right)^T \left( Y_N - \Phi_N \hat{\theta} \right) =$$

$$= \frac{1}{2} \left( Y_N^T Y_N - 2\hat{\theta}^T \Phi_N^T Y_N + \hat{\theta}^T \Phi_N^T \Phi_N \hat{\theta} \right) \tag{15}$$

The first term inside the parenthesis has not been calculated before, but it will be given by (16). Now inserting numerical values and performing the calculation gives $V(\hat{\theta}) = 115.22$. Finally using this in (14), together with $p = 2$ as the number of estimated parameters is 2, gives $\hat{\sigma}_w^2 = 0.46$.

$$Y_N^T Y_N = \sum_{k=2}^{N-1} y_{k+1}^2 \tag{16}$$

**c.** *i*) 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 for predicting the behavior. But if we perform a cross validation with data that have not been previously used we might see that the model can not predict the behavior.

*ii*) It can only be used to verify if one can expect good result of an identified 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.

**d.** The second plot in Figure 4 shows that all input-output behaviour has been modelled within a 99 % confidence level. The upper plot, however, indicates that there is some correlation left in the residuals, as lag 2 and 3 are outside the confidence region. This is an indication of that the assumption of white noise is not correct, and the next step in the identification process should therefore be to add a noise model.

**4.** A system is given by (17), where $\{v_k\}$ is an independent Gaussian noise process with probability density function (18).

$$y_k = ay_{k-1} + bu_{k-1} + v_k \tag{17}$$

$$f_v(v_k) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{v_k^2}{2\sigma^2}} \tag{18}$$

Assume that input-output data from the process is available, i.e. $\{u_k\}_{k=1}^N$ and $\{y_k\}_{k=1}^N$. Derive the maximum-likelihood estimate of $a$, $b$ and $\sigma$. (3 p)

*Solution*

The likelihood function is given by (19), where the residual $\varepsilon_k$ is defined in (20).

$$L(\theta,\sigma) = P(\varepsilon|\theta) = f_\varepsilon(\varepsilon_2,...,\varepsilon_N) = \prod_{k=2}^{N} f_v(\varepsilon_k) = \prod_{k=2}^{N} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{\varepsilon_k^2}{2\sigma^2}} \tag{19}$$

$$\varepsilon_k = y_k - ay_{k-1} - bu_{k-1} \tag{20}$$

The logarithm of the likelihood function is given by (21).

$$\log L(\theta,\sigma) = -(N-1)\log(\sqrt{2\pi}\sigma) - \frac{1}{2\sigma^2}\sum_{k=2}^{N}\varepsilon_k^2 \tag{21}$$

The partial derivative of (21) with respect to $\sigma$ is (22).

$$\frac{\partial \log L(\theta,\sigma)}{\partial\sigma} = -\frac{N-1}{\sigma} + \frac{1}{4\sigma^3}\sum_{k=2}^{N}\varepsilon_k^2 \tag{22}$$

By putting this derivative to zero and solving for $\sigma^2$ suggests that the estimate $\hat{\sigma}^2$ should be (23).

$$\hat{\sigma}^2 = \frac{1}{4(N-1)}\sum_{k=2}^{N}\varepsilon_k^2 \tag{23}$$

Inserting this into (21) gives (24).

$$\log L(\theta,\hat{\sigma}) = -(N-1)\log\left(\sqrt{2\pi}\sqrt{\frac{1}{4(N-1)}\sum_{k=2}^{N}\varepsilon_k^2}\right) - 2(N-1) =$$

$$= \frac{N-1}{2}\log\left(\frac{\pi}{2(N-1)}\sum_{k=2}^{N}\varepsilon_k^2\right) - 2(N-1) =$$

$$= \frac{N-1}{2} \log \left( \frac{\pi}{2(N-1)} \sum_{k=2}^{N} (y_k - a y_{k-1} - b u_{k-1})^2 \right) - 2(N-1) =$$

$$= \frac{N-1}{2} \left( \log \pi - \log 2(N-1) + \log \left( \sum_{k=2}^{N} (y_k - a y_{k-1} - b u_{k-1})^2 \right) \right) - 2(N-1) \tag{24}$$

(24) is to be maximized to find the estimates of the parameters $a$ and $b$, this is equivalent to maximizing (25).

$$J(a,b) = \log \left( \sum_{k=2}^{N} (y_k - a y_{k-1} - b u_{k-1})^2 \right) \tag{25}$$

Once this has been done, the resulting $\hat{a}$ and $\hat{b}$ can be inserted into (23) to find $\hat{\sigma}$.

**5.** A two-tank process, with two tanks in series and with the input to the upper tank, may be described with the nonlinear state-space system (26).

$$\begin{cases} \dot{x}_1(t) = -\gamma_1 \sqrt{x_1(t)} + \delta u(t) \\ \dot{x}_2(t) = \gamma_1 \sqrt{x_1(t)} - \gamma_2 \sqrt{x_2(t)} \\ y(t) = x_2(t) \end{cases} \tag{26}$$

**a.** How can you use this knowledge to identify a nonlinear model? (1 p)

**b.** Now assume that you know nothing about the process. Usually one then tries to identify a linear model. What do you have to think about to get a good model when the process is nonlinear? (1 p)

**c.** You decide to identify a discrete-time model of the process. What can be said about the choice of sampling frequency for the experiment? Discuss the potential risks of choosing to high respectively to low sampling frequency when doing system identification! (2 p)

*Solution*

**a.** This identification problem can be solved by using a grey-box model, i.e., a model that has a known structure and parametrization but with unknown parameters.

**b.** When the process is nonlinear one can not expect to get a good linear approximation for all states of the process, instead one has to consider an approximation around some state, which hopefully is linear. This can be achieved by using a small enough input amplitude, such that one stays in a region of the process that can be described by a linear model. (At the same time the amplitude should be chosen to a large value, to get a good signal to noise ratio)

**c.** As a rule of thumb, a reasonable way of choosing the sampling interval, $h$, is to let

$$\omega h = 0.2 - 0.6$$

where $\omega$ represents important frequencies of the system, such as the cross over frequency or the natural frequency. To choose the sampling interval properly we must therefore have knowledge about the significant frequencies of the system.

In general, if the sampling interval is chosen very short relative to the significant frequencies of the system, this could lead to numerical precision problems. On the other hand, by choosing a too long sampling interval, there is a risk that important dynamics above the Nyquist frequency is not described by the resulting model.

**6.** A second order transfer function (27) has been identified. You believe that it might be approximated by a first-order model instead.

$$G(s) = \frac{s + 0.25}{s^2 + 3s + 2} \tag{27}$$

**a.** Compute the first order Padé approximation (28) of $G(s)$. Conclude why this approximation is not good in this particular case. (1 p)

$$\hat{G}_1(s) = \frac{b}{s + a} \tag{28}$$

**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 (29). Decide if it is advisable to perform a model reduction for this system. (2 p)

$$\begin{cases} x(k+1) = \begin{pmatrix} -0.6639 & -0.5242 \\ -0.5242 & 0.2639 \end{pmatrix} x(k) + \begin{pmatrix} 0.8345 \\ -0.5511 \end{pmatrix} u(k) \\ y(k) = \begin{pmatrix} 0.8345 & -0.5511 \end{pmatrix} x(k) \end{cases} \tag{29}$$

*Solution*

**a.** The Padé approximation is based on the Taylor series expansion of $G(s)$. This is calculated in (30).

$$G(s) = G(0) + \frac{dG}{ds}(0) + \frac{1}{2}\frac{d^2G}{ds^2}(0) + ... = \frac{1}{8} + \frac{5}{16}s + \frac{1}{2}\left(-\frac{17}{16}\right)s^2 + ... \tag{30}$$

For a first order approximation we only need to keep the two first terms, giving the truncated polynomial (32).

$$G_1(s) = \frac{1}{8} + \frac{5}{16}s \tag{31}$$

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 (32).

$$B_1(s) = G_1(s)A_1(s) \quad \Rightarrow \quad b_0 = \left(\tfrac{1}{8} + \tfrac{5}{16}s\right)(s+a) = \tfrac{a}{8} + \left(\tfrac{1}{8} + \tfrac{5a}{16}\right)s + \tfrac{5}{16}s^2 \tag{32}$$

Skip the $s^2$-coefficient and match the other two, this results in (33), giving the Padé approximation (34).

$$\begin{cases} a = -\frac{2}{5} \\ b = -\frac{1}{20} \end{cases} \tag{33}$$

$$\hat{G}_1(s) = \frac{1/20}{2/5 - s} \tag{34}$$

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 (35), where $\Phi$ and $\Gamma$ are system matrices from (29), given in (36).

$$\Phi P \Phi^T - P + \Gamma\Gamma^T = 0 \tag{35}$$

$$\Phi = \begin{pmatrix} -0.6639 & -0.5242 \\ -0.5242 & 0.2639 \end{pmatrix} \quad , \quad \Gamma = \begin{pmatrix} 0.8345 \\ -0.5511 \end{pmatrix} \tag{36}$$

$P$ is a diagonal matrix (37), and inserting this and the matrices from (36) gives (38).

$$P = \begin{pmatrix} P_1 & 0 \\ 0 & P_2 \end{pmatrix} \tag{37}$$

$$\begin{pmatrix} -0.5592P_1 + 0.2748P_2 + 0.6964 & 0.3480P_1 - 0.1383P_2 - 0.4599 \\ 0.3480P_1 - 0.1383P_2 - 0.4599 & 0.2748P_1 - 0.9304P_2 + 0.3037 \end{pmatrix} = 0 \tag{38}$$

Solving (38) for $P_1$ and $P_2$ gives (39). As the two values not are of different magnitude, a model reduction is not advisable.

$$\begin{cases} P_1 = 1.6443 \\ P_2 = 0.8121 \end{cases} \tag{39}$$

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

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

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

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

(1 p)

*Solution*

The factorization property is verified by direct substitution of Markov parameters $h_k = CA^{k-1}B$.