LUND
UNIVERSITY

Department of
AUTOMATIC CONTROL

FRTN35 System Identification

Final Exam October 30, 2019, 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. From measured input and output data you are given the test to estimate a model that describes the dynamics of the system of the form:

$$A(z) \cdot y_k = B(z) \cdot u_{k-\tau} + w_k.$$

You have estimated three candidate models and evaluated the performance criteria shown in Table 1 below.

a. Describe the criteria in the table and explain how they can be used for model evaluation. (2 p)

b. Which of the three models in the table would you choose and why? (1 p)

c. Suggest an additional model evaluation method and describe how it would complement the three criteria in Table 1. (1 p)

Table 1 The performance criteria in Problem 1.

	Model 1	Model 2	Model 3
Fit	90.94 %	28.08 %	79.81 %
AIC	-9.92	-2.53	-6.45
FPE	4.92e-05	0.08	0.0016

Solution

a. **Fit:** Gives the percentage of the measured output that was explained by the model.
AIC: This index includes both the estimated variance and the model order complexity. The AIC decreases as the variance of the residual decreases and increases as the number of parameters in a model increases.
FPE: This criterion simulates the situation where the model is tested on a different data set. The criterion decreases as the residual variance decreses, it increases as the number of parameters in the model increases and decreases as the number of observations in the data increases. It tends to underestimate the correct model order.

b. For model 1, the Fit is best and the FPE lowest. Hence, with this evaluation, model 1 should be chosen.

c. For instance, residual analysis could be considered.

2. Consider the least-squares (LS) estimation problem:

$$\mathscr{Y}_N = \Phi_N \theta + \varepsilon,$$

with

$$\mathscr{Y}_N := \begin{bmatrix} y(0) \\ \vdots \\ y(N-1)) \end{bmatrix}, \qquad \Phi_N := \begin{bmatrix} \varphi^\top(0) \\ \vdots \\ \varphi^\top(N-1) \end{bmatrix}, \qquad \theta := \begin{bmatrix} a_1 \\ \vdots \\ a_n \\ b_1 \\ \vdots \\ b_m \end{bmatrix}.$$

Assume that the noise ε, is a zero-mean Gaussian with a given variance matrix $E\left\{\varepsilon\varepsilon^\top\right\} = R$. The matrix R is positive definite.

Our goal is to derive an unbiased linear estimator $\hat{\theta}$ of the form

$$\hat{\theta} = Z^\top \mathcal{Y}_N, \tag{1}$$

which minimizes its variance.

For a given regressor Φ, prove the following statements in **a**-**c**:

a. If a linear estimator of the form (1) is unbiased, then $Z^\top\Phi = I$. (1 p)

b. The covariance matrix of any linear unbiased estimator of the form (1) is given by $\mathrm{Cov}\left\{\hat{\theta}\right\} = Z^\top RZ$. (1 p)

c. For $Z = R^{-1}\Phi$, the estimator $\hat{\theta}_Z = \left(\Phi^\top R^{-1}\Phi\right)^{-1}\Phi^\top R^{-1}\mathcal{Y}_N$ with $\mathrm{Cov}\left\{\hat{\theta}_Z\right\} = \left(\Phi^\top R^{-1}\Phi\right)^{-1}$, exhibits the smallest variance in the class of all unbiased estimators:

$$\mathrm{Cov}\left\{\hat{\theta}_Z\right\} \leq \mathrm{Cov}\left\{\hat{\theta}\right\}.$$

$\hat{\theta}_Z$ is called the Best Linear Unbiased Estimator (BLUE). (2 p)

Hint: All covariance matrices are positive semi-definite. The inverse of a positive definite matrix is also positive definite.

d. Name one disadvantage of $\hat{\theta}_Z$, the BLUE estimator in **c**. (1 p)

Solution

a. For a linear estimator of the form (1) to be unbiased we require that

$$\theta = E\left\{\hat{\theta}\right\}.$$

Hence, for zero-mean Gaussian noise, ε, and fixed Φ we get

$$\theta = E\left\{Z^\top \mathcal{Y}_N\right\} = E\left\{Z^\top(\Phi\theta + \varepsilon)\right\} = Z^\top\Phi\theta,$$

which implies that

$$Z^\top\Phi = I.$$

b. The covariance matrix of any linear unbiased estimator of the form (1) is

$$\begin{aligned}
\mathrm{cov}\left\{\hat{\theta}\right\} &= E\left\{\left(\hat{\theta}-\theta\right)\left(\hat{\theta}-\theta\right)^\top\right\} && \left[\hat{\theta} = Z^\top\mathcal{Y}_N\right] \\
&= E\left\{\left(Z^\top\mathcal{Y}_N - \theta\right)\left(Z^\top\mathcal{Y}_N - \theta\right)^\top\right\} && \left[\mathcal{Y}_N = (\Phi\theta+\varepsilon)\right] \\
&= E\left\{\left(Z^\top(\Phi\theta+\varepsilon) - \theta\right)\left(Z^\top(\Phi\theta+\varepsilon) - \theta\right)^\top\right\} && \left[Z^\top\Phi = I\right] \\
&= Z^\top E\left\{\varepsilon\varepsilon^\top\right\}Z && \left[E\left\{\varepsilon\varepsilon^\top\right\} = R\right] \\
&= Z^\top RZ.
\end{aligned}$$

c. For $\hat{\theta}_Z$ to be the BLUE in the class of all unbiased estimators, we want to show that $\operatorname{cov}\{\hat{\theta}\} - \operatorname{cov}\{\hat{\theta}_Z\} \geq 0$,

$$\begin{aligned}
\operatorname{cov}\{\hat{\theta}\} - \operatorname{cov}\{\hat{\theta}_Z\} &= Z^\top RZ - \left(\Phi^\top R^{-1}\Phi\right)^{-1}\left[Z^\top\Phi = I\right] \\
&= Z^\top RZ - Z^\top\Phi\left(\Phi^\top R^{-1}\Phi\right)^{-1}\Phi^\top Z \\
&= Z^\top\left[R - \Phi\left(\Phi^\top R^{-1}\Phi\right)^{-1}\Phi^\top\right]Z \\
&= Z^\top FZ,
\end{aligned}$$

where we define $F = R - \Phi\left(\Phi^\top R^{-1}\Phi\right)^{-1}\Phi^\top$. We need to show that F is positive definite such that $Z^\top FZ$ is positive. We can show that

$$\begin{aligned}
F^\top R^{-1}F =& \left(R - \Phi\left(\Phi^\top R^{-1}\Phi\right)^{-1}\Phi^\top\right)^\top R^{-1}\left(R - \Phi\left(\Phi^\top R^{-1}\Phi\right)^{-1}\Phi^\top\right) \\
=& \left(R - \Phi\left(\Phi^\top R^{-1}\Phi\right)^{-1}\Phi^\top\right)\left(I - R^{-1}\Phi\left(\Phi^\top R^{-1}\Phi\right)^{-1}\Phi^\top\right) \\
=& R - \Phi\left(\Phi^\top R^{-1}\Phi\right)^{-1}\Phi^\top - \Phi\left(\Phi^\top R^{-1}\Phi\right)^{-1}\Phi^\top + \\
& \Phi\left(\Phi^\top R^{-1}\Phi\right)^{-1}\Phi^\top R^{-1}\Phi\left(\Phi^\top R^{-1}\Phi\right)^{-1}\Phi^\top \\
=& R - \Phi\left(\Phi^\top R^{-1}\Phi\right)^{-1}\Phi^\top \\
=& F.
\end{aligned}$$

Since R is positive definite, R^{-1} is positive definite, we can conclude that $F = F^\top R^{-1}F$ has to be positive and therefore $\operatorname{cov}\{\hat{\theta}_Z\} \leq \operatorname{cov}\{\hat{\theta}\}$. Hence, $\hat{\theta}_Z$ is the estimator with the smallest variance.

d. Notice that BLUE is a function of the assembly of Z which require knowledge of the error variance R. Since noise measurements are generally unavailable, this represents a major restriction for this estimation approach.

3. Consider the transfer function

$$H(z) = \frac{0.5}{z^2 - z + 0.5}$$

a. The following is a state-space realization of H:

$$\begin{aligned}
x_{k+1} &= \begin{pmatrix} 0.718 & -0.546 \\ 0.546 & 0.282 \end{pmatrix} x_k + \begin{pmatrix} 0.572 \\ -0.572 \end{pmatrix} u_k \\
y_k &= \begin{pmatrix} 0.572 & 0.572 \end{pmatrix} x_k
\end{aligned}$$

Actually, the state-space realization is a balanced realization. Determine the asymptotic reachability Gramian P and the asymptotic observability Gramian Q. (2 p)

b. Given the Gramians in **a**, determine whether it is suitable or not to do a model reduction. (1 p)

c. Relate your answer in **b** to the poles and zeros of $H(z)$, i.e., give an intuitive explanation to why it is advisable (or not advisable) to perform model reduction on this system. (1 p)

Solution

a. For a balanced realization, the asymptotic reachability Gramian P is equal to the asymptotic observability Gramian Q. The diagonal matrix $\Sigma = P = Q$ fulfills the discrete-time Lyapunov equations

$$\begin{aligned} \Phi\Sigma\Phi^T - \Sigma + \Gamma\Gamma^T &= 0 \\ \Phi^T\Sigma\Phi - \Sigma + C^T C &= 0 \end{aligned}$$

Solving the first equation gives

$$\Sigma = P = Q = \begin{pmatrix} 1.12 & 0 \\ 0 & 0.72 \end{pmatrix}$$

A check gives that also the second Lyapunov equation is fulfilled.

b. Since the elements in the Gramians do not vary with a factor of magnitude, it is not suitable to perform a state reduction.

c. H has complex poles, i.e., it is very difficult to reduce this to a system with one (real) pole that gives similar behavior.

4. Assume that you and your friend want to identify a two-tank process, consisting of two tanks in series and with the input to the upper tank. From a basic control course, you know that it may be described by following nonlinear state-space system:

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

a. Your friend wants to try to identify a nonlinear model of the two-tank by using a grey-box model. Is it a suitable approach? Motivate your answer! (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. Yes, as we have a model that has a known structure and parametrization but with unknown parameters, a grey-box model would be a good approach.

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 ω 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.

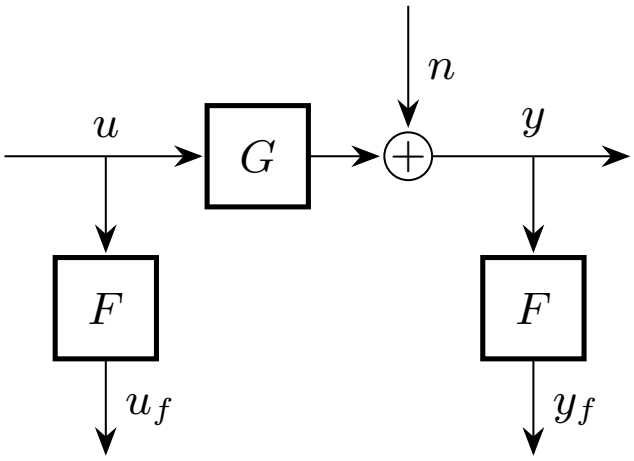

Figure 1 The identification experiment in Problem 5.

5. Suppose we are going to identify the system G in Figure 1. Assume that u and y are two measured signals which are possibly interrelated by

$$Y(s) = G(s)U(s) + N(s)$$

where N is a noise term and uncorrelated to U. The coherence function $\gamma(\omega)$ between signals u and y is defined as

$$\gamma(\omega) = \frac{|S_{uy}(i\omega)|}{\sqrt{S_{uu}(i\omega)S_{yy}(i\omega)}}.$$

a. Determine how $\gamma(\omega)$ is affected by prefiltering of the data sequences u and y.
Hint: Introduce $u_f = Fu$ and $y_f = Fy$. (2 p)

b. Express $\gamma(\omega)$ such that it can be used to judge if the excitation signal is good enough for identification. What conclusions can you draw? (1 p)

Solution

a. The cross spectra of the filtered data sequences $u_f = Fu$ and $y_f = Fy$ is

$$S_{u_f u_f} = |F|^2 S_{uu} \qquad S_{y_f y_f} = |F|^2 S_{yy}$$

and the cross spectrum

$$S_{u_f y_f} = S_{u_f u_f} G^* = |F|^2 S_{uu} G^* = |F|^2 S_{uy}$$

where the asterisk denotes complex conjugate. Then, the coherence function for the filtered sequences

$$\gamma_f(\omega) = \frac{|S_{u_f y_f}(i\omega)|}{\sqrt{S_{u_f u_f}(i\omega) S_{y_f y_f}(i\omega)}} = \frac{|F|^2 |S_{uy}|}{\sqrt{|F|^2 S_{uu} |F|^2 S_{yy}}} = \gamma(\omega)$$

Hence, γ is not affected by the prefiltering.

b. According to Equation (a)

$$\gamma^2(\omega) = \frac{|S_{uy(i\omega)}|^2}{S_{uu}(i\omega) S_{yy}(i\omega)}$$

that is

$$\begin{aligned} \gamma^2(\omega) &= \frac{|G(i\omega)|^2 S_{uu}^2(i\omega)}{S_u u(i\omega)(|G(i\omega)|^2 S_{uu}(i\omega) + S_{nn}(i\omega))} \\ &= \frac{1}{1 + \frac{S_{nn}(i\omega)}{|G(i\omega)|^2 S_{uu}(i\omega)}}. \end{aligned}$$

Consequently, $\gamma(\omega)$ is close to 1 when the noise is small compared to the input signal and close to 0 when the noise is large compared to the input signal.

6. Consider N independent samples, each with associated density

$$p(y_i|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-1}{2\sigma^2}(y_i - \mu)^2\right)$$

a. Evaluate the Fisher Information Matrix $\mathscr{I}_N(\mu, \sigma^2)$ for parameters μ and σ^2. (2 p)

Hint: The Fisher Information Matrix is defined as

$$\mathscr{I}_N(\mu, \sigma^2) = -\mathbf{E}\left(\begin{bmatrix} \frac{\partial^2 \log(L)}{\partial^2 \mu} & \frac{\partial^2 \log(L)}{\partial \mu \partial \sigma^2} \\ \frac{\partial^2 \log(L)}{\partial \sigma^2 \partial \mu} & \frac{\partial^2 \log(L)}{\partial^2 \sigma^2} \end{bmatrix}\right),$$

where L is the Maximum Likelihood function.

b. Show that the estimator for μ

$$\bar{\theta}(y_1 \dots y_N) = \frac{1}{N}\sum_{i=1}^{N} y_i$$

achieves the Cramer-Rao lower bound in terms of its variance. (1 p)

Hint: The Cramer-Rao lower bound is defined by $\text{Cov}[\bar{\theta}(\mathbf{y})] \geq \mathscr{I}_N^{-1}(\mu, \sigma^2)$ where $\mathscr{I}_N(\mu, \sigma^2)$ is the Fisher Information Matrix.

Solution

a. Let $\mathbf{y} = (y_1, \dots, y_N)$ be a random sample from $\mathscr{N}(\mu, \sigma^2)$ and let $f_N(\mathbf{y} \mid \mu, \sigma^2)$ denote the probability density function of the data. Then the Fisher information matrix for parameters μ and σ^2 is given by

$$\begin{aligned} \mathscr{I}_N(\mu, \sigma^2) &:= -\mathrm{E}\begin{pmatrix} \frac{\partial^2 \log(f_N(\mathbf{y}|\mu,\sigma^2))}{\partial^2 \mu} & \frac{\partial^2 \log(f_N(\mathbf{y}|\mu,\sigma^2))}{\partial \mu \partial \sigma^2} \\ \frac{\partial^2 \log(f_N(\mathbf{y}|\mu,\sigma^2))}{\partial \sigma^2 \partial \mu} & \frac{\partial^2 \log(f_N(\mathbf{y}|\mu,\sigma^2))}{\partial^2 \sigma^2} \end{pmatrix} \\ &= N \times \mathscr{I}_1(\mu, \sigma^2). \end{aligned}$$

The last equality holds as y_i's are independent and identically distributed. For $N = 1$ the log-likelihood function is

$$\log(f_1(y \mid \mu, \sigma^2)) = l(y \mid \mu, \sigma^2) = -\frac{1}{2}\log(2\pi\sigma^2) - \frac{(y-\mu)^2}{2\sigma^2}.$$

Therefore,

$$\left(\frac{\partial l}{\partial \mu}, \frac{\partial l}{\partial \sigma^2}\right) = \left(\frac{(y-\mu)}{\sigma^2}, \frac{(y-\mu)^2}{2\sigma^4} - \frac{1}{2\sigma^2}\right).$$

Hence the Fisher information matrix becomes

$$\begin{aligned} \mathscr{I}_N(\mu, \sigma^2) &= -N \times \mathrm{E}\begin{pmatrix} -\frac{1}{\sigma^2} & -\frac{(y-\mu)}{\sigma^4} \\ -\frac{(y-\mu)}{\sigma^4} & \frac{1}{2\sigma^4} - \frac{(y-\mu)^2}{\sigma^6} \end{pmatrix} \\ &= \begin{pmatrix} \frac{N}{\sigma^2} & 0 \\ 0 & \frac{N}{2\sigma^4} \end{pmatrix}. \end{aligned} \tag{2}$$

b. Recall that, if $\bar{\theta}(\mathbf{y})$ is an unbiased estimator of a parameter θ, then the Cramer-Rao inequality assures that

$$\text{Cov}[\bar{\theta}(\mathbf{y})] \geq \mathscr{I}_N^{-1}(\mu, \sigma^2).$$

Clearly,

$$\mathbb{E}[\bar{\theta}(\mathbf{y})] = \mu\,,$$

and

$$\text{Cov}[\bar{\theta}(\mathbf{y})]_{(1,1)} = \text{Var}(\bar{\theta}(\mathbf{y}))) = \frac{1}{N^2}\sum_{i=1}^{N}\text{Var}(y_i) = \frac{N\sigma^2}{N^2} = \frac{\sigma^2}{N} = \mathscr{I}_N^{-1}(\mu, \sigma^2)_{(1,1)}.$$

Thus, it can be seen from equation (2) that the variance of $\bar{\theta}(\mathbf{y})$ achieves the Cramer-Rao lower bound.

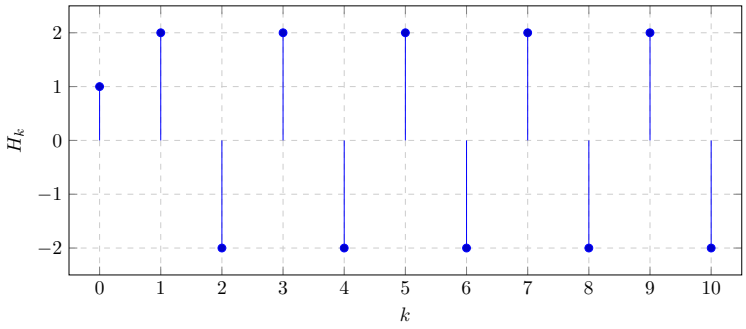

Figure 2 The impulse response H_k in Problem 7.

7. A linear, discrete-time, single-input single-output, first order system S is to be identified. Its impulse response $\{H_k\}_{k=0}^{\infty}$ is shown in Figure 2 for $k = 0$ to 10. Using Ho-Kalman-Kung's algorithm, find a state-space representation of S.
Hint: Choose the Hankel matrices as scalars.

(2 p)

Solution

a. Using Ho-Kalman-Kung's algoritm with (from the impulse response figure): $H_0 = 1$, $H_{1,1}^{(0)} = 2$, $H_{1,1}^{(1)} = -2$, we get

$$H_{1,1}^{(0)} = 2 = U\Sigma V^T$$

A possible solution is $U = V = 1$, $\Sigma = 2$. We thus get with $E_u = E_y = 1$:

$$\begin{array}{rclcrcl} A & = & \Sigma^{-1/2}U^T H_{1,1}^{(1)} V \Sigma^{-1/2} & = & \frac{1}{\sqrt{2}} \cdot 1 \cdot -2 \cdot 1 \cdot \frac{1}{\sqrt{2}} & = & -1 \\ B & = & \Sigma^{1/2}V^T E_u & = & \sqrt{2} \cdot 1 \cdot 1 & = & \sqrt{2} \\ C & = & E_y^T U \Sigma^{1/2} & = & 1 \cdot 1 \cdot \sqrt{2} & = & \sqrt{2} \\ D & = & H_0 & & & = & 1 \end{array}$$

One possible state-space realization of S is thus:

$$\begin{array}{rcl} x_{k+1} & = & -x_k + \sqrt{2}u_k \\ y_k & = & \sqrt{2}x_k + u_k \end{array}$$