LUND UNIVERSITY

Department of
AUTOMATIC CONTROL

FRT 041 System Identification

Final Exam October 26, 2015, 2pm - 7pm

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 8 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

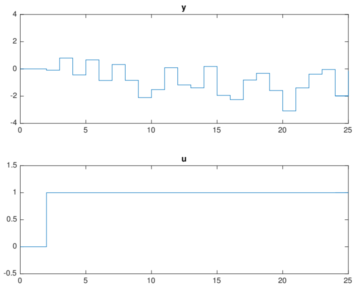
Results

The results of the exam will be posted at the latest November 2, 2015 on the note board on the first floor of the M-building.

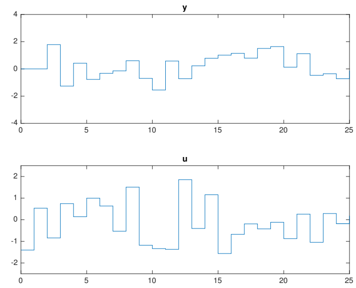
1. In Figure 1 is shown an excerpt of the input and output signals from two different experiments on the same unknown system, which could be modeled as

$$y_k = b_1 u_{k-1} + b_2 u_{k-2} + e_k, \quad k \geq 3 \tag{1}$$

where $e_k \sim \mathrm{N}(0,1)$. Both experiments were run for a length of $N = 1000$.



(a) Excerpt from experiment 1



(b) Excerpt from experiment 2

Figure 1: Excerpts from the two experiments in problem 1

a. Use the data below and calculate the least-squares parameter estimates. Explain the different results in the two experiments. Which results should you trust? (2 p)

		Experiment 1	Experiment 2
$\sum_{i=2}^{N-1} u_k^2$	:	997	1075.9
$\sum_{i=2}^{N-1} u_k u_{k-1}$	:	996	-1.9
$\sum_{i=2}^{N-1} u_{k-1} u_{k-1}$	:	996	1079.8
$\sum_{i=2}^{N-1} u_k y_{k+1}$	:	-221.7	339.0
$\sum_{i=2}^{N-1} u_{k-1} y_{k+1}$	:	-222.5	-529.5

b. Motivate why using a maximum-likelihood estimator will give, or not give, the same results as using least-squares in this case. (2 p)

Solution

a. The system has an obvious regressor form $y_k = \varphi_{k-1}\theta + e_k$, which enables us to define the matrices

$$Y = \Phi\theta + E = \begin{pmatrix} y_3 \\ \vdots \\ y_N \end{pmatrix} = \begin{pmatrix} u_2 & u_1 \\ \vdots & \vdots \\ u_{N-1} & u_{N-2} \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} + \begin{pmatrix} e_3 \\ \vdots \\ e_N \end{pmatrix}$$

Now we get

$$\Phi^T\Phi = \begin{pmatrix} \sum_{i=2}^{N-1} u_k^2 & \sum_{i=2}^{N-1} u_k u_{k-1} \\ \sum_{i=2}^{N-1} u_k u_{k-1} & \sum_{i=2}^{N-1} u_{k-1} u_{k-1} \end{pmatrix}$$

$$\Phi^T Y = \begin{pmatrix} \sum_{i=2}^{N-1} u_k y_{k+1} \\ \sum_{i=2}^{N-1} u_{k-1} y_{k+1} \end{pmatrix}$$

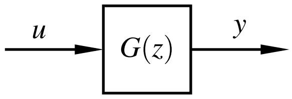


Figure 2: The input signal u is fed to the linear system G.

The least-squares estimate is then given by

$$\hat{\theta} = (\Phi^T\Phi)^{-1}\Phi^T Y$$

Inserting the values given we get the parameter estimates $\hat{\theta}_1 = (0.79, -1.02)$ and $\hat{\theta}_2 = (0.31, -0.49)$, which are clearly different although the experiments were done on the same system. This is because the input signal in experiment 1 is a step function, which is only persistently exciting of order 1, meaning that we can't identify the two parameters. By looking at the figure we can see that there clearly is more variations in the input signal for experiment 2, meaning that we can excite the system enough to identify both parameters (as a note, the input is white noise, which is persistently exciting up to arbitrary high order).

b. Since the noise is Gaussian with known mean and covariance, we get the following maximum-likelihood estimator:

$$\begin{aligned}\hat{\theta} &= \underset{\bar{\theta}}{\operatorname{argmax}}\, L(\bar{\theta}) = \underset{\bar{\theta}}{\operatorname{argmax}} \prod_{i=3}^{N} \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}(y_i - b_1u_{i-1} - b_2u_{i-2})^2) \\ &= \underset{\bar{\theta}}{\operatorname{argmax}}\ (\sqrt{2\pi})^{-N+3} \exp(-\frac{1}{2}\sum_{i=3}^{N}(y_i - b_1u_{i-1} - b_2u_{i-2})^2)\end{aligned}$$

where we see that the only part that depend on the parameters is the sum in the exponential, meaning that the ML estimator is given by the solution to

$$\underset{\bar{\theta}}{\operatorname{argmin}} \sum_{i=3}^{N}(y_i - b_1u_{i-1} - b_2u_{i-2})^2$$

which is the same cost function that is minimized by the least-squares estimator. Hence, the two methods produce identical results in this case.

2.

a. Consider a series of identification problems performed in the setting depicted in figure 2. It is known that the variance of the measurement noise is *quite* small compared to the variance of the input signal. Three different systems $G1, G2, G3$ have been fed a PRBS signal u, and estimates of the respective output spectra is shown in figure 3.

The three systems are to be modeled using ARX/ARMAX models. Which kind of model, and which order of the polynomials, would you use for the systems? To assist you in your choice, the autocorrelation functions of the output are shown in figure 4.

Provide an analysis of the information provided by the plotted estimates, what information do they convey about the order of the systems? If you deem that the information provided by the spectral estimates and autocorrelation functions are insufficient for determining your choice of model, describe in a short and *concise* way how you would proceed to further analyze the obtained identification data.

(3 p)

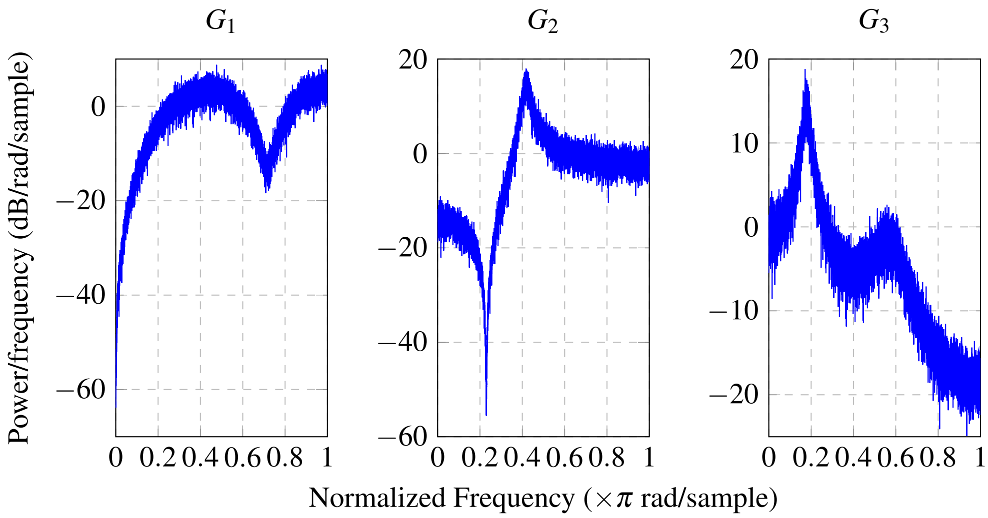


Figure 3: Spectral estimates of y_1, y_2, y_3

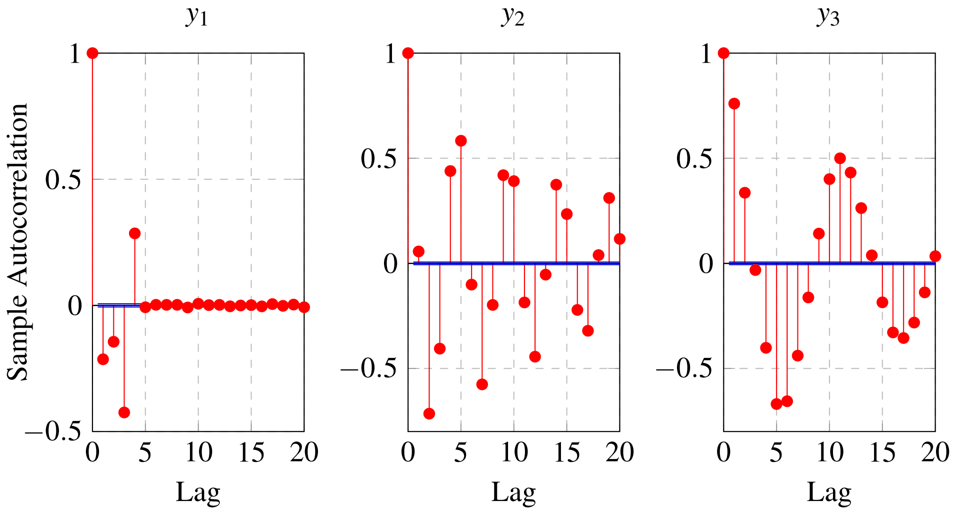


Figure 4: Autocorrelation function estimates for y_1, y_2, y_3

b. Consider now a setting where the system to be identified has a nonlinear function acting on the input, see figure 5. Describe how you would proceed before drawing conclusions from the estimated output spectra. (1 p)

c. The two spectra S_1, S_2 depicted in figure 6 represent the same signal, but are estimated using two different methods, of which one is the regular periodogram. Describe which spectra is likely to be estimated using the periodogram and why. What do you think has been done differently during in the other spectral estimate? Also describe the problems related to spectral estimation using the periodogram, what problems may occur and why?

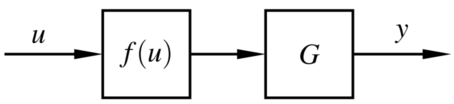


Figure 5: A nonlinear function $f(u)$ acts on the input signal to the linear system G

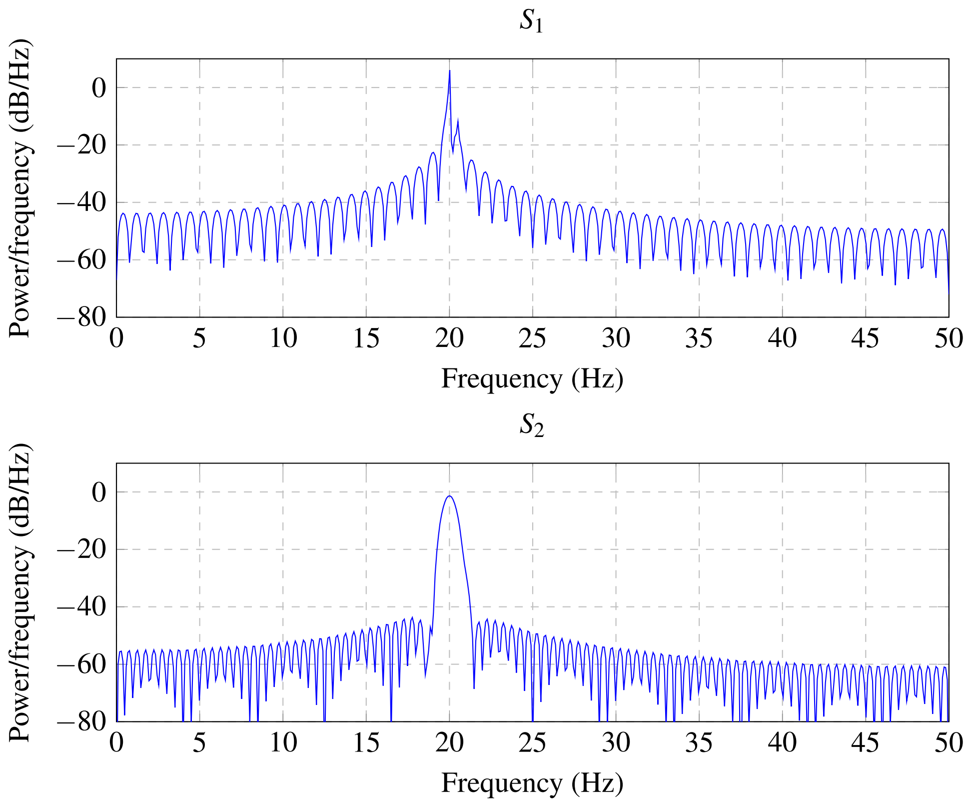


Figure 6: Spectral estimates for problem c

(2 p)

Solution

a.

- System G_1 seems to have two blocked frequencies, the blocked zero frequency indicates at least one zero and the blocked frequency at 0.75 indicates a pair of zeroes in the transfer function from u to y.
- System G_2 seems to have one blocked frequency, which indicates one pair of zeroes in the transfer function from u to y. G_2 further seems to have one resonant frequency which indicates a pair of poles in the transfer function from u to y.
- System G_3 seems to have two resonant frequencies, which indicates two pairs of poles in the transfer function from u to y.

To further analyze the data, the autocorrelation function is useful, y_1 seems to have no significant autocorrelation in lags higher than four, which indicates that there is no feedback in the system and supports the previous hypothesis of two pairs of zeroes.

Both y_2 and y_3 seems to have periodic ACFs, indicating the presence of feedback (poles/infinite impulse response) in the system.

The information provided is insufficient to say much about the characteristics of the noise. To analyze the data further, estimates of the cross correlation function, impulse response estimates and estimates of the partial autocorrelation function could be performed. One can also calculate some information criterion, such as AIC or FPE, and use this to support ones choice of model order.

To summarize

- G_1 comes from the process $y_k = u_k - u_{k-1} - 0.5u_{k-2} + u_{k-3}$, as supported by spectrum and autocorrelation function.
- G_2 comes from the process $y_k - 0.5y_{k-1} + 0.9y_{k-1} = u_k - 1.5u_{k-1} - u_{k-2}$, as supported by spectrum and partly by the autocorrelation function.
- G_3 comes from the process $y_k - 1.2y_{k-1} + 0.93y_{k-2} - 0.73y_{k-3} + 0.61y_{k-4} = u_k$, as supported by spectrum and partly by the autocorrelation function.

All systems are thus ARX models (it could be argued that the first system is an MA system where the moving average is over the input). If unsure, information criteria should be used to determine the proper model order.

b. In the presence of a nonlinearity, one must ensure oneself that the linear cross coherence spectrum is strong enough to allow identification of linear models. If this is not the case, nonlinear techniques must be adopted.

c. The spectrum S_1 is estimated using the periodogram. This can be seen from the higher level of spectral leakage which arises due to the finite measurement sequence and rectangular window. The second spectrum is estimated using Welch's method where a window (Hamming) has been applied to the data. This reduces the effect of the leakage, but does not remove it completely. It further reduces the frequency resolution in the estimate, as evidenced by the complete disappearance of the small peak at 20.5Hz. This peak is barely visible in the periodogram, but completely gone in the Welch estimate. Windowing also masks spectral content below the sidelobe attenuation. Choosing different windows will enable you to make tradeoffs between resolution (e.g., using a rectangular window) and sidelobe attenuation.

3. You are trying to estimate the parameters from the moving average process

$$y(k) = au_{k-1} + bu_{k-3} + e_k. \tag{2}$$

where $\{e_k\}$ is a zero-mean white noise process with variance σ^2 and $\{u_k\}$ is a zero-mean weakly stationary process with autocovariance function $C_{uu}(\tau) = (1/2)^{|\tau|}$ that is uncorrelated with $\{e_k\}$.

We are interested in finding the least-squares estimate for $\hat{\theta} = (\hat{a} \quad \hat{b})^T$. Does the parameter estimate have an asymptotic distribution? If so, what is the distribution and its parameters? (4 p)

Solution

The model (2) can be written as $y_k = \varphi_k^T \theta + e_k$, where $\theta = (a \ b)$ and

$$\varphi_k^T = (u_{k-1} \quad u_{k-3}).$$

Given N samples of observed input data the regression matrix is

$$\Phi_N = \begin{pmatrix} u_3 & u_1 \\ u_4 & u_2 \\ \vdots & \vdots \\ u_{N+2} & u_N \end{pmatrix}.$$

Now

$$\frac{1}{N}\Phi_N^T \Phi_N = \frac{1}{N}\begin{pmatrix} \sum_{k=3}^{N+2} u_k^2 & \sum_{k=3}^{N+2} u_k u_{k-2} \\ \sum_{k=3}^{N+2} u_k u_{k-2} & \sum_{k=1}^{N} u_k^2 \end{pmatrix}$$

so, due to ergodicity we have asymptotically,

$$\lim_{N\to\infty} \frac{1}{N}\,\Phi_N^T\,\Phi_N = E(\frac{1}{N}\Phi_N^T\,\Phi_N) = \begin{pmatrix} C_{uu}(0) & C_{uu}(2) \\ C_{uu}(2) & C_{uu}(0) \end{pmatrix} = \begin{pmatrix} 1 & 0.25 \\ 0.25 & 1 \end{pmatrix}.$$

Thus the regression matrix is invertible when the number of samples goes to infinity. This fact together with the fact that the input signal is uncorrelated with the noise signal ensures a consistent estimate. Therefore $E(\hat{\theta}) = \theta$ and the central limit theorem (6.98) on page 121 in the book gives the asymptotic distribution

$$\hat{\theta} \sim \mathrm{AsN}(\theta, \frac{\sigma^2}{N}\Sigma)$$

where

$$\Sigma = \begin{pmatrix} 1 & 0.25 \\ 0.25 & 1 \end{pmatrix}^{-1} = \frac{16}{15}\begin{pmatrix} 1 & -0.25 \\ -0.25 & 1 \end{pmatrix}$$

4. We want to estimate the parameters a and b in the system

$$y_k + ay_{k-1} = bu_{k-1} + v_k. \tag{3}$$

Due to peculiarities of the system the identification must be performed under proportional feedback. Determine if the parameters are identifiable under the follow feedback laws:

a. $u_k = -k_p y_k$ (1 p)

b. $u_k = -k_p y_{k-1}$ (1 p)

Solution

a. The closed loop system is $y_k = -(a + bk_p)y_{k-1} + v_k$. The parameters a and b can thus not be identified since they only affect the system as a sum.

b. Now the closed loop system is $y_k = -ay_{k-1} - bk_p y_{k-2} + v_k$ and the parameters are identifiable, no problem.

5. Consider the differential equation of a mechanical system

$$m\ddot{q} = -kq + f \tag{4}$$

with position coordinate q, force f, mass m, and stiffness k. Assume that q and f are available to measurement.

Formulate a regression model by means of a change of variables to the operator $\lambda = 1/(1+s\tau)$, thereby showing that it is possible to identify the parameters m and k from measurements of q and f. (3 p)

Solution

From $\lambda = 1/(1+s\tau)$ we find that $s = (1-\lambda)/\tau\lambda$. Application of the Laplace transform to Eq. (4) gives

$$ms^2 Q(s) = -kQ(s) + F(s) \tag{5}$$

A change of variables using λ gives

$$m(\frac{1-\lambda}{\tau\lambda})^2 Q = -kQ + F \tag{6}$$

Algebraic simplification of Eq. (6) gives

$$m(1-2\lambda+\lambda^2)Q = -k\tau^2\lambda^2 Q + \tau^2\lambda^2 F \tag{7}$$

from which we could formulate the time-domain regression model $y(t) = \phi(t)\theta$ where

$$y = \lambda^2\{f\}, \ \ \phi = (\,\tfrac{1}{\tau^2}(q - 2\lambda\{q\} + \lambda^2\{q\}) \quad \lambda^2\{q\}\,), \ \ \theta = \begin{pmatrix} m \\ k \end{pmatrix} \tag{8}$$

where $\lambda\{f\}$, $\lambda\{q\}$ and $\lambda^2\{q\}$ are low-pass filtered f and q.

6. Prediction error methods such as least-squares estimation for identification of ARX models minimize the square of the prediction errors $\frac{1}{N}\sum_{k=1}^{N}\varepsilon(\theta)^2 = \frac{1}{N}\sum_{k=1}^{N}(y_k - \hat{y}_k(\theta))^2$ to find an estimate $\hat{\theta}_N$ of the model parameters θ. Considering linear and Gaussian systems one could handwavingly argue that this sum is a sum of independent variables and thus converges to an expectation as $N \to \infty$. The residuals are not really independent though, but it turns out that it is indeed true for linear systems with Gaussian noise that

$$\frac{1}{N}\sum_{k=1}^{N}\varepsilon(\theta)^2 \to E\varepsilon(k,\theta)^2 = V(k,\theta),\ N \to \infty \tag{9}$$

and also that

$$\hat{\theta}_N \to \arg\min_{\theta} V(k,\theta),\ N \to \infty. \tag{10}$$

Now assume that we have a discrete-time model, where z^{-1} denotes the backwards shift operator,

$$y_k = G(z^{-1},\theta)u_k + e_k. \tag{11}$$

Assume further that the true system is given by

$$y_k = G_0(z^{-1})u_k + w_k. \tag{12}$$

Use Parseval's formula to state the optimization problem 10 in the frequency domain and argue that the input spectrum $S_{uu}(i\omega)$ can be used to choose for what frequency bands we get the best model fit. All reasonable assumptions about the signals are OK to use, if motivated well. (4 p)

Solution

We have

$$\varepsilon(k,\theta) = y_k - \hat{y}_k(\theta) = y_k - G(z^{-1},\theta)u_k = \left(G_0(z^{-1}) - G(z^{-1},\theta)\right)u_k + w_k. \tag{13}$$

Assuming independence between u_k and w_k the spectra is given by

$$S_{\varepsilon\varepsilon} = |G_0(i\omega) - G(i\omega,\theta)|^2 S_{uu}(i\omega) + S_{ww}(i\omega) \tag{14}$$

Parseval's relation gives

$$V(k,\theta) = \frac{1}{2\pi}\int_{-\pi}^{\pi} S_{\varepsilon\varepsilon}(i\omega,\theta)\,d\omega \tag{15}$$

The optimization problem is thus given by

$$\hat{\theta} = \arg\min_{\theta} V(k,\theta) = \frac{1}{2\pi}\int_{-\pi}^{\pi} S_{\varepsilon\varepsilon}(i\omega,\theta)\,d\omega \tag{16}$$

$$= \arg\min_{\theta} \frac{1}{2\pi}\int_{-\pi}^{\pi} |G_0(i\omega) - G(i\omega,\theta)|^2 S_{uu}(i\omega)\,d\omega \tag{17}$$

and we can see that S_{uu} acts like a weight function. Where $S_{uu}(i\omega)$ is big, a deviation between $G_0(i\omega)$ and $G(i\omega,\theta)$ is costly and we can thus choose for what frequencies the model fit will be good.

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

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

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

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

(1 p)

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

Solution

a. One way is to verify the factorization property by direct substitution of Markov parameters $h_k = CA^{k-1}B$ into the Hankel matrix.

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

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 $\{A, B, C\}$. In the factorization above, the matrix

$$\mathscr{O}_r = \begin{pmatrix} C \\ CA \\ \vdots \\ CA^{r-1} \end{pmatrix}$$

is the extended observability matrix and

$$\mathscr{C}_s = \begin{pmatrix} B & AB & \dots & A^{s-1}B \end{pmatrix}$$

the extended controllability matrix.

For $k = 0$, the factorization is then:

$$\begin{aligned}
H_{r,s}^{(0)} &= \mathscr{O}_r \cdot \mathscr{C}_s \\
&= U\Sigma V^T \\
&= U\Sigma^{1/2}\Sigma^{1/2}V^T
\end{aligned}$$

Where the second inequality is obtained through singular value decomposition. We then have:

$$\begin{aligned}
\mathscr{O}_r &= U \cdot \Sigma^{1/2} \\
\Rightarrow \mathscr{O}_r^{\dagger} &= \Sigma^{-1/2}U^T
\end{aligned}$$

$$\begin{aligned}
\mathscr{C}_s &= \Sigma^{1/2} \cdot V^T \\
\Rightarrow \mathscr{C}_s^{\dagger} &= V^T\Sigma^{-1/2}
\end{aligned}$$

The dagger sign on e.g. $\mathscr{O}_r^\dagger$ denotes the pseudoinverse.

The state space matrices are then, inserting the expressions for the pseudoinverse for the extended observability and controllability in for example $A = \mathscr{O}_r^\dagger H_{r,s}^{(1)} \mathscr{C}_s^\dagger$ and using the expressions above:

$$\begin{aligned}
\hat{A}_n &= \mathscr{O}_r^\dagger H_{r,s}^{(1)} \mathscr{C}_s^\dagger \\
&= \Sigma_n^{-1/2} U_n^T H_{r,s}^{(1)} V_n \Sigma_n^{-1/2} \\
\hat{B}_n &= \mathscr{C}_s \cdot \left[I_{m\times m} \; 0_{m\times(s-1)m} \right]^T \\
\hat{C}_n &= \left[\begin{pmatrix} I_{p\times p} \\ 0^T_{p\times(s-1)p} \end{pmatrix} \right]^T \mathscr{O}_r
\end{aligned}$$