Some Exercises in System Identification

Karl Henrik Johansson

Revised: Rolf Johansson
Maria Karlsson
Brad Schofield

Department of Automatic Control
Lund University
January 2007

Introduction to System Identification

1. This exercise is meant as an introduction to computer-based system identification. We run an identification demonstration in Matlab. Data from a laboratory scale hairdryer is used to estimate a model. The demonstration includes several steps that will be discussed thoroughly later on in the course, so do not expect to understand all steps taken in the demonstration.

Start

```
Matlab
```

Run the demonstration through the command

```
>> iddemo
```

and select demonstration `2)` in the menu. Follow the information on the screen.

When the demonstration is finished you end up in the menu. Select `1)` and follow the demonstration of the graphical user interface. About the same modeling is performed, but this time most Matlab commands are hidden by the interface.

Test Signals

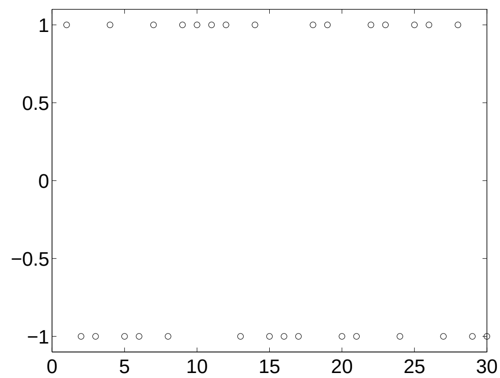


Figure 1 A pseudo-random binary sequence.

2. Figure 1 shows thirty samples from a pseudo-random binary sequence (PRBS). The PRBS is often derived from a shift-register, see Example 8.6 in the textbook.

a. Write a Matlab function `prbs(n)` that derives `n` samples from a PRBS sequence, for example, by using the following shift-register.

Let `x` denote the shift-register state and `u` the output. Then, a shift-register of length sixteen can be implemented in the following way. The output is derived as

```
u(k) = 2*x(16) - 1;
```

and then the states are updated as

```
d = x(16) + x(15) + x(13) + x(4);
x(2:16) = x(1:15);
x(1) = mod(d,2);
```

Draw a shift-register circuit corresponding to these equations, see Figure 8.5 in the textbook.

b. If the period time of the shift-register is sufficiently long, its output has approximately the same frequency content as white noise, that is, the frequency spectrum is constant. Often in control applications it is useful to have a test signal that has "white noise characteristics" only up to a certain frequency, for example, the bandwidth of the system. This can be conveniently implemented in the PRBS algorithm by introducing the parameter `period`.

The PRBS in **a** could shift value every time the shift-register was updated. Modify your Matlab function such that the output can only shift value at time instances that are multiples of `period`. The function should be called by the command `prbs(n,period)`. Figure 2 shows thirty samples from a PRBS with `period` = 3. Figure 1 corresponds to `period` = 1.

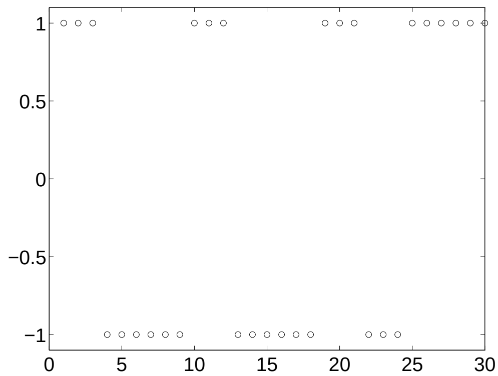


Figure 2 A pseudo-random binary sequence with `period` = 3.

c. Estimate and plot the frequency responses of the PRBS with various values on `period`. There are dips in the spectra. How does the number of dips relate to the parameter `period`?

Frequency Response Analysis

3. Consider a system

$$y_k = G(q)u_k + v_k$$

controlled by proportional feedback

$$u_k = -Ky_k$$

Derive a transfer function estimate $\widehat{G}$ via discrete Fourier transformation of y_k and u_k and draw some conclusions?

4. Non-parametric identification results in a model given as a frequency response, for example, a Bode plot. This type of identification it is often useful to start with to get an idea of the process dynamics. However, the parametric methods (for example, based on ARX or ARMAX models) we discuss later on are more reliable and better suited for control design.

Create input and output data via the Matlab commands

```
b = [0 1 0.5];
a = [1 -1.5 0.7];
th = poly2th(a,b);

u = randn(300,1);
e = randn(300,0.5);
y = idsim([u e],th0);
z = [y u];

idplot(z)
```

What is the relation between y, u, and e? Which are the poles and zeros of the process in discrete time, and what does that corresponds to in continuous time if the sampling time is 1 s?

Perform spectral analysis based on discrete Fourier transform of the covariance function and discrete Fourier transform of the data direct. The Matlab commands are `spa` and `etfe`. Note that default a window is used in `spa`, which smoothes the result.

State the equations for the spectral analysis above, when no extra filtering is done.

Linear Regression

5. Run the demonstrations in Exercise 1 again. This time go through each step in the demos carefully. Then, type

```
type iddemo1
```

and try to understand each command performed. Use `help` on Matlab commands you are not familiar with. Answer the following questions. Why do we have to remove trends from data sets, and what is the name of the Matlab command? Explain how the step response is estimated from correlation analysis. An ARX-model is estimated. What difference equation does the model correspond to?

Finally run the overall Matlab demonstration via `demo` and study some of the features. In particular, run some demos from the control system, robust control, and nonlinear control toolboxes.

Time-Series Analysis

6. Consider the system

$$\mathcal{S}: \qquad y_k = -ay_{k-1} + bu_{k-1} + v_k = \varphi_k^T \theta + v_k$$

where v is a white noise sequence (not necessarily Gaussian). We are interested in estimating the parameters a and b based on the data $y_1, y_2, \ldots, y_N$ and $u_1, u_2, \ldots, u_N$.

a. Determine the loss function that is minimized for the maximum-likelihood (ML) estimate, if v_k has the density function

$$f_v(v_k) = \frac{1}{\sqrt{2}\sigma} e^{-\sqrt{2}|v_k|/\sigma}$$

b. Derive the least-squares (LS) estimate. What loss function is minimized?

c. Determine the loss function that is minimized for the ML estimate, if v_k is normal distributed and, hence,

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

Also, derive the estimate that minimizes the loss function.

d. What conclusions can be drawn?

7. Consider the system

$$\mathcal{S}: \qquad y_k + ay_{k-1} = b_1 u_{k-1} + b_2 u_{k-2} + v_k$$

which is estimated with the instrument variable (IV) method using delayed inputs as instruments

$$z_k = \begin{pmatrix} u_{k-1} & u_{k-2} & u_{k-3} \end{pmatrix}^T$$

Assume u is white Gaussian noise with zero mean and unit variance, and that u and v are independent. Determine when the IV estimate is consistent.

8. Consider the system

$$\mathcal{S}: \qquad y_k + a_0 y_{k-1} = u_k$$

where u is zero mean white noise with variance λ^2. We will examine two ways of predicting y_k two steps ahead.

a. Let the model be

$$\mathcal{M}_1: \qquad y_k + ay_{k-1} = v_k$$

and estimate a. Then, the predictor is $\hat{y}_{k+2|k} = \hat{a}^2 y_k$. Derive the asymptotic least-squares estimate of a and its variance.

b. Another way to get a two-step predictor is to consider the model

$$\mathcal{M}_2 : \qquad y_k + \beta y_{k-2} = w_k$$

and estimate β. The predictor is $\hat{y}_{k+2|k} = -\hat{\beta} y_k$. Derive the asymptotic least-squares estimate of β. Show that w is not white noise by deriving its covariance function. (Hence, it is not possible to use the standard formula to calculate the variance of $\hat{\beta}$. You do not have to derive this variance.)

c. Assume you have to choose one of the methods in **a** and **b** above in order to predict y_k two steps ahead. Discuss what criteria you would consider to make your choice in a practical situation.

Model Validation and Reduction

9. Consider the system

$$\mathcal{S} : \qquad y_{k+2} - y_{k+1} + 0.5 y_k = u_{k+1} - 0.5 u_k + e_{k+2}$$

where $e_k \in \mathcal{N}(0, 1)$. Let the input u be a PRBS signal with unit amplitude, and perform an identification experiment collecting 200 data points. Estimate AR(n, n)-models, that is,

$$\mathcal{M} : \qquad y_{k+n} + a_1 y_{k+n-1} + \cdots + a_n y_k = b_0 u_{k+n-1} + \cdots + b_{n-1} u_k + e_{k+n}$$

where $n = 1, \ldots, 8$. Derive and compare the loss function (the average prediction error), Akaike's information criterion and Akaike's final prediction error criterion of the models.

10. Consider the ARMA-system

$$\mathcal{S} : \qquad y_k = -a y_{k-1} + e_k + c e_{k-1}$$

where $\{e_k\}$ is a white noise sequence with variance σ^2. Assume that we use the model

$$\mathcal{M} : \qquad y_k = -a y_{k-1} + e_k + c e_{k-1}$$

and estimate a and c based on a prediction error method. Then it can be shown that

$$\widehat{\theta} \in \mathrm{As}\mathcal{N}(\theta, \mathrm{Cov}(\widehat{\theta}))$$

where $\widehat{\theta} = \begin{pmatrix} \widehat{a} & \widehat{c} \end{pmatrix}^T$ is the prediction at time N and

$$\mathrm{Cov}(\widehat{\theta}) = \frac{\sigma^2}{N(c-a)^2} \times \begin{pmatrix} (1-a^2)(1-ac)^2 & (1-a^2)(1-ac)(1-c^2) \\ (1-a^2)(1-ac)(1-c^2) & (1-ac)^2(1-c^2) \end{pmatrix}$$

"As$\mathcal{N}$" denotes "asymptotically normal distributed," so the distribution of $\widehat{\theta}$ will tend to be normal as N increases.

In the following subproblems we study how close a and c are. We see from $\mathcal{S}$ that

$$y_k = \frac{q+c}{q+a} e_k$$

so if a and c are sufficiently close the model reduces to $y_k = e_k$.

a. Determine the asymptotic variance of $\widehat{a} - \widehat{c}$.

b. Test the hypothesis

$$\mathcal{H}_0 : \qquad a = c$$

on the significance level 0.05 based.

c. Do an F-test based on the two models

$$\begin{aligned} \mathcal{M}_1 : &\qquad y_k = e_k \\ \mathcal{M}_2 : &\qquad y_k + a y_{k-1} = e_k + c e_{k-1} \end{aligned}$$

and the hypotheses

$$\begin{aligned} \mathcal{H}_0 : &\qquad \mathcal{M}_1 \text{ is as good model for the system as } \mathcal{M}_2 \\ \mathcal{H}_1 : &\qquad \mathcal{M}_2 \text{ models the system better than } \mathcal{M}_1 \end{aligned}$$

that is,

$$\begin{aligned} \mathcal{H}_0 : &\qquad \sigma_1^2 = \sigma_2^2 \\ \mathcal{H}_1 : &\qquad \sigma_1^2 > \sigma_2^2 \end{aligned}$$

where σ_1^2 and σ_2^2 are the variances for the white noise e_k in the model $\mathcal{M}_1$ and $\mathcal{M}_2$, respectively. Can we reject $\mathcal{H}_0$ on significance level 0.05?

11. We are interested in reducing the order of the model

$$Y(z) = \frac{z - 0.55}{(z-0.5)(z-0.8)} U(z) \tag{1}$$

Therefore we derive a balanced state-space model

$$\begin{aligned} x_{k+1} &= \begin{pmatrix} 0.7790 & -0.0766 \\ -0.0766 & 0.5210 \end{pmatrix} x_k + \begin{pmatrix} 0.9884 \\ 0.1521 \end{pmatrix} u_k \\ y_k &= \begin{pmatrix} 0.9884 & 0.1521 \end{pmatrix} x_k \end{aligned}$$

The eigenvalues of the Gramians P and Q (see the textbook) are

$$\begin{pmatrix} \sigma_1 \\ \sigma_2 \end{pmatrix} = \begin{pmatrix} 2.4853 \\ 0.0517 \end{pmatrix}$$

Is it advisable to reduce the model order? If so, determine the reduced-order model.

12. Consider the transfer function

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

a. Show that

$$\begin{aligned} x_{k+1} &= \begin{pmatrix} 0.72 & -0.55 \\ 0.55 & 0.28 \end{pmatrix} x_k + \begin{pmatrix} 0.57 \\ -0.57 \end{pmatrix} u_k \\ y_k &= \begin{pmatrix} 0.57 & 0.57 \end{pmatrix} x_k \end{aligned}$$

is a state-space realization of H.

b. Actually, the state-space realization in **a** is a balanced realization. Determine the asymptotic reachability Gramian P and the asymptotic observability Gramian Q.

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

Real-Time Identification

13. Consider the system

$$\mathcal{S}: \qquad y_k = \varphi_k^T \theta + e_k$$

and the model

$$\mathcal{M}: \qquad y_k = \varphi_k^T \theta$$

The recursive least-squares (RLS) estimate

$$\begin{aligned} \widehat{\theta}_k^{RLS} &= \widehat{\theta}_{k-1}^{RLS} + K_k(y_k - \varphi_k^T \widehat{\theta}_{k-1}^{RLS}) \\ & K_k = P_k \varphi_k \\ & P_k = \frac{1}{\lambda}\left(P_{k-1} - \frac{P_{k-1}\varphi_k \varphi_k^T P_{k-1}}{\lambda + \varphi_k^T P_{k-1} \varphi_k} \right) \end{aligned} \tag{2}$$

asymptotically minimizes the loss function

$$V(\bar{\theta}, k) = \frac{1}{2} \sum_{i=1}^{k} \lambda^{k-i} (y_i - \varphi_i^T \bar{\theta})^2$$

Discuss how the forgetting factor λ influences the estimate.

14. Study the initial conditions influence on the RLS estimate $\widehat{\theta}_k^{RLS}$ in (2). In particular, show how $\widehat{\theta}_k^{RLS}$ depends on $\widehat{\theta}_0^{RLS}$, the initial covariance estimate P_0, and the forgetting factor λ, by stating and solving the difference equations for P_k^{-1} and $P_k^{-1}\widehat{\theta}_k^{RLS}$. Compare the results to the *batch* LS estimate (all data $i = 0, \ldots, k$ available)

$$\widehat{\theta}_k^{LS} = (\Phi_k^T \Phi_k)^{-1} \Phi_k^T Y_k = \left(\sum_{i=1}^{k} \varphi_i \varphi_i^T \right)^{-1} \sum_{i=1}^{k} \varphi_i y_i$$

Closed-Loop Identification

15. We will in this problem investigate estimation of the transfer function H in the control loop in Figure 3 by means of spectrum analysis. Assume that H, K, and the closed-loop system are stable. The reference signal r and the noise signal v are independent stochastic processes with mean value zero and with spectral densities $S_{rr}(i\omega)$ and $S_{vv}(i\omega)$, respectively. Let the estimate of H be given as

$$\widehat{H}(e^{i\omega}) = \frac{S_{yu}(i\omega)}{S_{uu}(i\omega)} \tag{3}$$

Derive an expression for the estimate in (3) in terms of $S_{rr}(i\omega)$, $S_{vv}(i\omega)$, $H(e^{i\omega})$ and $K(e^{i\omega})$. Investigate what happens as $S_{rr}(i\omega) \to 0$.

The following formula is useful. Let v be as above and let $x = H_1 v$ and $y = H_2 v$, where H_1 and H_2 are stable filters. Then $S_{xy}(i\omega) = H_1(e^{i\omega})\overline{H_2(e^{i\omega})}S_{vv}(i\omega)$

Note that we are investigating an asymptotic estimate. In other words we use $S_{yu}(i\omega)$ and $S_{uu}(i\omega)$ instead of estimates of them as in formula (4.24) in the text book.

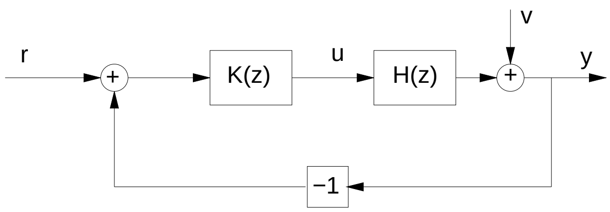


Figure 3 Feedback system

16. Consider the system

$$\mathcal{S}: \qquad y_k + a y_{k-1} = b u_{k-1} + e_k$$

where e_k is white noise with variance λ^2. Let the input signal to the system be

$$u_k = -K y_k + r_k$$

where the reference signal r_k is white noise with variance σ^2. Assume that we use the least-squares method based on the model

$$\mathcal{M}: \qquad y_k + a y_{k-1} = b u_{k-1}$$

to identify a and b.

a. Show that the estimates are consistent.

b. Determine the asymptotic parameter variances and show that they will tend to infinity as $\sigma^2 \to 0$.

17. Consider the unstable first-order process

$$\mathcal{S}: \qquad y_k + a y_{k-1} = b u_{k-1} + e_k, \qquad E e_k^2 = \lambda^2$$

Our model is

$$\mathcal{M}: \qquad y_k + \bar{a} y_{k-1} = \bar{b} u_{k-1}$$

The process is controlled by a proportional controller chosen such that the closed-loop system is stable. To the control signal an external signal is added to excite the system, hence,

$$u_k = -f y_k + v_k, \qquad E v_k^2 = \sigma^2$$

The stochastic processes e and v are independent white noises with means equal to zero. Which one of the following two methods gives best accuracy?

1. Measure u and y. Derive least-squares estimates of a and b.
2. Measure v and y. Derive a least-squares estimate of the closed-loop system. Out of this estimate, calculate the estimates of a and b.

Hint: Recall from the Computer-Controlled Systems course that for a first-order system

$$y_k = \frac{b_0 q + b_1}{a_0 q + a_1} u_k$$

the variance of y_k is

$$E y_k^2 = \frac{(b_0^2 + b_1^2) a_0 - 2 b_0 b_1 a_1}{a_0 (a_0^2 - a_1^2)} E u_k^2$$

Subspace Identification

18. In this exercise, we will investigate the use of state-space model identification.

In particular, subspace-based methods will be introduced. As preparation, read Chapter 13 and Appendix E in System Modeling and Identification Exercises We will use Matlab software throughout the exercise. Begin by installing the SMI-1.0 toolbox, available at:

- `http://www.control.lth.se/~FRT041/local/smi1_0d.tgz`

Download the SMI Toolbox Manual:

- `http://www.control.lth.se/~FRT041/local/smimanual-1.0.pdf`

The SMI-1.0 toolbox implements the Multivariable Output-Error State Space (MOESP) class of state space algorithms. Read through the manual, and run through the examples in Chapter 5. Try to identify some parametric models using the data from the examples. Compare their performance with that of the state space models given by the

MOESP algorithms using some common measure. The Matlab command n4sid uses subspace methods to identify state space models. Acquaint yourself with the use of the command, and try to identify some models using the data from the SMI toolbox manual examples. Compare the performance with the SMI toolbox algorithms and with other parametric models. Subspace algorithms may be directly applied to multivariable systems. Investigate this by using the rss command in Matlab to generate random state space systems of a given order and input/output dimensions, and trying to identify the system. What can be said about the use of parametric identification for multivariable systems?

Finally, download the identification data:

- `http://www.control.lth.se/~FRT041/data.zip`

and use state-space identification methods to obtain a model.

Solutions

Introduction to System Identification

1.

Test Signals

2.

a. See the solution in **b**.

b.

```
function u = prbs(n,period);
% PRBS  Pseudo random binary sequence
%
%       u = prbs(n,period) gives a PRBS row vector of
%       length n. The sequence periodiod has length
%       2^16-1. period is the same variable as in the
%       computer program logger (default 1).

u = zeros(n,1);
x = [zeros(1,15) 1];            % Initial state
if nargin==1,
  period = 1;
end

p = period;
for k = 1:n,
  if p == period,
    p  = 1;
    u(k) = 2*x(16) - 1;
    d = x(16) + x(15) + x(13) + x(4);
    x(2:16) = x(1:15);
    x(1) = mod(d,2);
  else
    u(k) = u(k-1);
    p = p + 1;
  end
end;
```

c. The Matlab function `fft` derives the discrete Fourier transform of a vector. The following commands gives the plots in Figure 4:

```
n=2^12;
p1=5;
p2=6;
plot(abs(fft(prbs(n,p1))));
plot(abs(fft(prbs(n,p2))));
```

Notice that the point 2048 on the x-axis corresponds to the Nyquist frequency (corresponding to the end of the spectrum). We see that

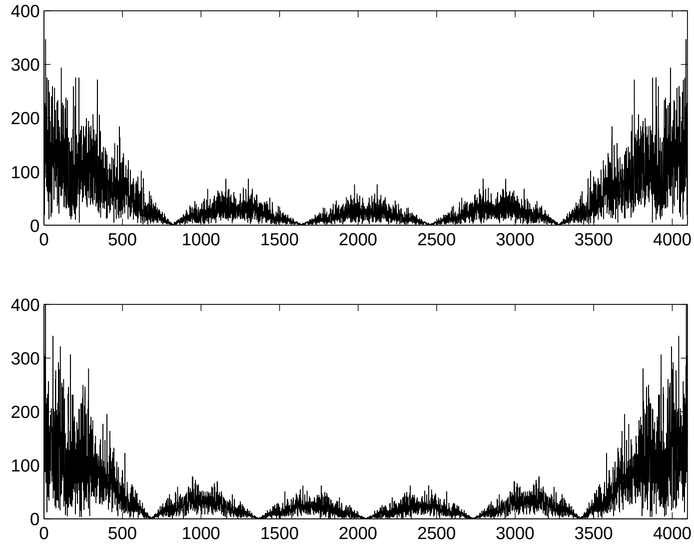


Figure 4 Discrete Fourier transforms of PRBS.

the spectrum has 2 dips if `period` $= 5$ and 2.5 if `period` $= 6$ (the dip at the Nyquist frequency is divided in two halves). Hence, a guess is that the number of dips equals $(\texttt{period} - 1)/2$. It is possible to show this formula analytically from the definition of the discrete Fourier transform.

Frequency Response Analysis

3. We have

$$\widehat{G}(e^{i\omega h}) = \frac{Y_N(i\omega)}{U_N(i\omega)} = \frac{h\sum_{m=0}^{N-1} y_{mh}e^{-i\omega hm}}{-Kh\sum_{m=0}^{N-1} y_{mh}e^{-i\omega hm}} = -\frac{1}{K}$$

Hence, the transfer function estimate is the inverse of the controller and not an estimate of G.

This simple example shows a common problem in identifying closed-loop systems; the correlation between the input signal u and the output y due to the control algorithm destroys the consistency property of the identification method. This is further discussed in Chapter 8 in the textbook.

4.

Linear Regression

5.

Time-Series Analysis

6.

a. The likelihood function to be maximized is

$$L(\bar{\theta}) = \mathcal{P}(\varepsilon|\bar{\theta}) = f_\varepsilon(\varepsilon_2, \ldots, \varepsilon_N|\bar{\theta}) = f_v(\varepsilon_2, \ldots, \varepsilon_N) = \prod_{k=2}^{N} f_v(\varepsilon_k)$$

where f_ε is the density function of the residuals $\varepsilon_k = y_k - \varphi_k^T \bar{\theta}$, and $\bar{\theta}$ is assumed to be the true estimate. Notice that we have used that since v is white noise,

$$f_v(v_2, \ldots, v_N) = \prod_{k=2}^{N} f_v(v_k)$$

Further,

$$\log L(\bar{\theta}) = -(N-1)\log(\sqrt{2}\sigma) - \frac{\sqrt{2}}{\sigma}\sum_{k=2}^{N} |\varepsilon_k|$$

Hence, the loss function to be minimized is

$$J(\bar{a}, \bar{b}) = \sum_{k=2}^{N} |y_k - (-\bar{a}y_{k-1} + \bar{b}u_{k-1})|$$

b. The LS estimate is

$$\begin{aligned} \widehat{\theta} &= (\Phi^T\Phi)^{-1}\Phi^T\mathcal{Y} \\ &= \begin{pmatrix} \sum_{k=2}^{N} y_{k-1}^2 & -\sum_{k=2}^{N} y_{k-1}u_{k-1} \\ -\sum_{k=2}^{N} y_{k-1}u_{k-1} & \sum_{k=2}^{N} u_{k-1}^2 \end{pmatrix}^{-1} \begin{pmatrix} -\sum_{k=2}^{N} y_k y_{k-1} \\ \sum_{k=2}^{N} y_k u_{k-1} \end{pmatrix} \end{aligned} \tag{4}$$

It minimizes the loss function

$$J(\bar{a}, \bar{b}) = \sum_{k=2}^{N} (y_k - (-\bar{a}y_{k-1} + \bar{b}u_{k-1}))^2 \tag{5}$$

c. In the case of normal distributed noise v, we redo the calculations in Section 6.3 in the textbook. Then, the loss function becomes identical to (5). Hence, the estimate is the same as in (4).

d. In the case of white Gaussian (normal distributed) noise, we conclude that the LS estimate is the optimal estimate. If the noise is not Gaussian, the LS estimate is usually not optimal. Further, if the white noise assumption is violated, then the ML estimate is different from the LS estimate even if the noise is normal distributed.

7. We have the system

$$\mathcal{S}: \qquad \mathcal{Y} = \Phi\theta + V$$

and the model

$$\mathcal{M}: \qquad \mathcal{Y} = \Phi\theta \tag{6}$$

where

$$\Phi = \begin{pmatrix} \varphi_4^T \\ \vdots \\ \varphi_N^T \end{pmatrix} = \begin{pmatrix} -y_3 & u_3 & u_2 \\ \vdots & & \vdots \\ -y_{N-1} & u_{N-1} & u_{N-2} \end{pmatrix}$$

and

$$\theta = \begin{pmatrix} a & b_1 & b_2 \end{pmatrix}^T \qquad V = \begin{pmatrix} v_4 & \dots & v_N \end{pmatrix}^T$$

By multiplying left- and right-hand side of (6) with

$$Z = \begin{pmatrix} z_4^T \\ \vdots \\ z_N^T \end{pmatrix}$$

we get the IV estimate

$$\widehat{\theta} = (Z^T\Phi)^{-1}Z^T\mathcal{Y} = \theta + (Z^T\Phi)^{-1}Z^TV$$

Hence, the estimate is consistent if $E\{z_k\varphi_k^T\}$ is non-singular and z is uncorrelated with v. We have

$$\begin{aligned} Z^T\Phi &= \begin{pmatrix} u_3 & \dots & u_{N-1} \\ u_2 & \dots & u_{N-2} \\ u_1 & \dots & u_{N-3} \end{pmatrix}^T \begin{pmatrix} -y_3 & u_3 & u_2 \\ \vdots & & \vdots \\ -y_{N-1} & u_{N-1} & u_{N-2} \end{pmatrix} \\ &= \begin{pmatrix} -\sum_3^{N-1} u_k y_k & \sum_3^{N-1} u_k^2 & \sum_3^{N-1} u_k u_{k-1} \\ -\sum_3^{N-1} u_{k-1} y_k & \sum_3^{N-1} u_{k-1} u_k & \sum_3^{N-1} u_{k-1}^2 \\ -\sum_3^{N-1} u_{k-2} y_k & \sum_3^{N-1} u_{k-2} y_k & \sum_3^{N-1} u_{k-2} u_{k-1} \end{pmatrix} \end{aligned}$$

Thus,

$$E\{z_k\varphi_k^T\} = \lim_{N\to\infty} \frac{1}{N-3} Z^T\Phi = \begin{pmatrix} 0 & 1 & 0 \\ -E\{u_k y_{k+1}\} & 0 & 1 \\ -E\{u_k y_{k+2}\} & 0 & 0 \end{pmatrix}$$

where

$$\begin{aligned} E\{u_k y_{k+1}\} &= E\{u_k(-ay_k + b_1u_k + b_2u_{k-1} + v_{k+1}) = b_1 \\ E\{u_k y_{k+2}\} &= E\{u_k(-a(y_k + b_1u_k + b_2u_{k-1} + v_{k+1}) + b_1u_{k+1} + b_2u_k + v_{k+2}) \\ &= -ab_1 + b_2 \end{aligned}$$

so that $E\{Z^T\Phi\}$ is non-singular if

$$\det E\{z_k\varphi_k^T\} = b_2 - ab_1 \neq 0$$

Notice that the transfer function from u to y given by $\mathcal{S}$ is

$$\frac{b_1 q + b_2}{q(q+a)}$$

The equality $b_2 - ab_1 = 0$ is equal to a pole-zero cancellation in $\mathcal{S}$, and is thus a degenerated case. Finally, since z and v are uncorrelated, we conclude that the IV estimate is consistent.

8.

a. We have

$$\begin{aligned}
\hat{a} &= \left(\Phi_N^T \Phi_N\right)^{-1} \Phi_N^T \mathcal{Y}_N \to \left(\mathrm{E} y_{k-1}^2\right)^{-1} (-\mathrm{E} y_k y_{k-1}) \\
&= \left(\frac{\lambda^2}{1-a_0^2}\right)^{-1} \frac{a_0 \lambda^2}{1-a_0^2} = a_0
\end{aligned}$$

and

$$\operatorname{Var} \hat{a} = \lambda^2 \left(\Phi_N^T \Phi_N\right)^{-1} \to \frac{\lambda^2}{N} \left(\mathrm{E} y_{k-1}^2\right)^{-1} = \frac{1-a_0^2}{N}$$

b. It holds that

$$\begin{aligned}
\hat{\beta} &= \left(\Phi_N^T \Phi_N\right)^{-1} \Phi_N^T \mathcal{Y}_N \to \left(\mathrm{E} y_{k-2}^2\right)^{-1} (-\mathrm{E} y_k y_{k-2}) \\
&= \left(\frac{\lambda^2}{1-a_0^2}\right)^{-1} \frac{-a_0^2 \lambda^2}{1-a_0^2} = -a_0^2
\end{aligned}$$

From $\mathcal{S}$, we have

$$y_k - a_o^2 y_{k-2} = u_k + a_0 u_{k-1}$$

and since $\hat{\beta}$ is an unbiased estimate, asymptotically

$$w_k = u_k + a_0 u_{k-1}$$

Hence, the covariance function

$$C_{ww}(\tau) = \begin{cases} (1+a_0^2)\lambda^2, & \tau = 0 \\ -a_0 \lambda^2, & |\tau| = 1 \\ 0, & |\tau| \geq 2 \end{cases}$$

c. It is reasonable to choose the predictor $\hat{y}_{k+2|k}$ minimizing $\|y_{k+2} - \hat{y}_{k+2|k}\|$. Depending on the choice of norm $\|\cdot\|$, this problem can be hard to solve analytically.

In our case, both estimates $\hat{a}^2$ and $\hat{\beta}$ are asymptotically unbiased. However, in a practical situation the data sequences have a limit length. Therefore, it is interesting to study how fast these estimates and their variances converge.

Model Validation and Reduction

9. Let $p = 2n$ denote the number of parameters in the models $\mathcal{M}$ and N the number of data points. The loss function

$$V(\widehat{\theta}) = \frac{1}{N}\sum_{k=1}^{N}\frac{1}{2}\varepsilon_k^2(\widehat{\theta}), \qquad \varepsilon_k(\widehat{\theta}) = y_k - \varphi_k^T\widehat{\theta}$$

the Akaike information criterion

$$\mathrm{AIC}(p) = \log V(\widehat{\theta}) + \frac{2p}{N}$$

and the final prediction error criterion

$$\mathrm{FPE}(p) = \frac{N+p}{N-p}V(\widehat{\theta})$$

are derived in Matlab by the following commands.[1]

```
% System ---------------------------------

a_true=poly([0.5+0.5i,0.5-0.5i]);
b_true=[0 1 -0.5];
% a_true =
%     1.0000   -1.0000    0.5000
% b_true =
%          0    1.0000   -0.5000
th_true=poly2th(a_true,b_true);

% Identification experiment ---------------

N=200;
u=prbs(N);
randn('seed',2)
e=randn(N,1);
y=idsim([u e],th_true);

% Model estimation -----------------------

NN=[ 1     1     1;    % Model structure [na nb nk]
     2     2     1;
     3     3     1;
     4     4     1;
     5     5     1;
     6     6     1;
     7     7     1;
     8     8     1;];
M=arxstruc([y(1:N) u(1:N)],[y(1:N) u(1:N)],NN);

% Criteria calculation --------------------
```

[1]The commands prbs and akaike are *not* included in standard Matlab packages.

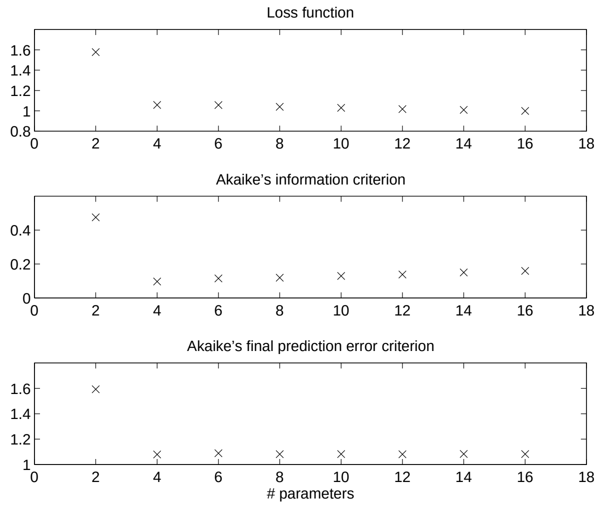


Figure 5 Loss function, AIC, and FPE in Exercise 4.

```
V=M(1,1:size(M,2)-1);            % The loss function
[aic,fpe]=akaike(V,NN,N);        % AIC and FPE
```

Figure 5 shows V, AIC, and FPE as functions of the number of estimated parameters p. We note that the loss function is a monotonously decreasing function, whereas both AIC and FPE has a minimum for $p = 4$ (the latter one is however hard to see). Hence, in this example, minimizing AIC or FPE gives a model with a number of parameters equal to the number in the system.

10.

a. The variance is

$$\begin{aligned}\operatorname{Var}(\widehat{a}-\widehat{c}) &= \operatorname{Var}(\begin{pmatrix} 1 & -1 \end{pmatrix}\widehat{\theta}) = \begin{pmatrix} 1 & -1 \end{pmatrix}\operatorname{Cov}(\widehat{\theta})\begin{pmatrix} 1 \\ -1 \end{pmatrix} \\ &= \frac{1-a^2c^2}{N}\end{aligned}$$

b. From **a** we know that under $\mathcal{H}_0$ it holds that

$$\widehat{a}-\widehat{c} \in \operatorname{As}\mathcal{N}(0,\mu^2), \qquad \mu^2 = \frac{1-a^2c^2}{N}$$

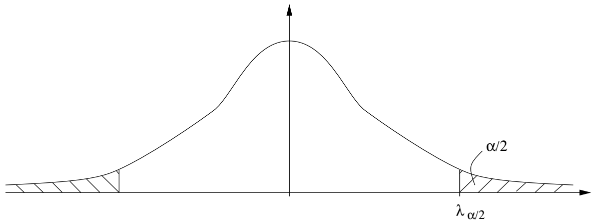
Introduce the normalized stochastic variable

$$X = \frac{\widehat{a}-\widehat{c}}{\mu} = \frac{\widehat{a}-\widehat{b}}{\sqrt{\frac{1-a^2c^2}{N}}} \in \operatorname{As}\mathcal{N}(0,1)$$

From the statistic course, we know that

$$\mathcal{P}(|X| > \lambda_{\alpha/2}) = \alpha$$

where α is the probability that the null hypothesis is rejected if it is true, and $\lambda_{\alpha/2}$ can be found in statistic tables.



Hence, reject $\mathcal{H}_0$ (on significance level α) if

$$|\widehat{a} - \widehat{c}| > \lambda_{\alpha/2}\sqrt{\frac{1-a^2c^2}{N}}$$

For $\alpha = 0.05$ we have $\lambda_{\alpha/2} = 1.96$.

c. F-tests are discussed in Section 9.3, B.3, and B.5 in the textbook. Notice that these tests are performed under the assumption that a correct model (under a certain hypothesis) gives optimal prediction errors ε_k, that is, $\varepsilon_k = e_k$.

Let V_1 and V_2 denote the sum of the squared residuals for the two models, and $p_1 = 0$ and $p_2 = 2$ the number of estimated parameters. Then, under the hypothesis $\mathcal{H}_0$,

$$\frac{V_i}{\sigma^2} = \sum_{k=p_i+1}^{N} \frac{\varepsilon_{ik}}{\sigma^2} \in \chi^2(N-p_i), \qquad i = 1,2$$

where $N - p_i$ is called *the degree of freedom*. Taking the difference between V_1 and V_2 (under the assumption that the prediction errors are optimal), it follows that the degree of freedom reduces to $p_1 - p_2$, so that

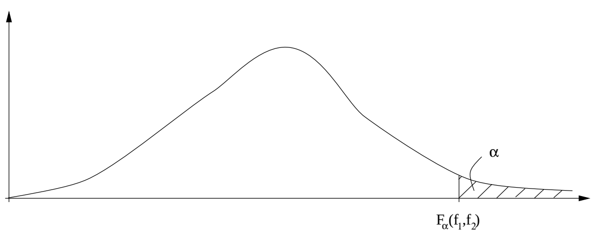
$$(V_1 - V_2)/\sigma^2 \in \chi^2(p_2 - p_1)$$

The quotient of two χ^2-distributed variables is F-distributed. We have under $\mathcal{H}_0$ that

$$\frac{V_1 - V_2}{V_2} \cdot \frac{N - p_2}{p_2 - p_1} \in F(p_2 - p_1, N - p_2)$$

Let F_α denote the α-percentile. Then, we reject $\mathcal{H}_0$ if

$$\frac{V_1 - V_2}{V_2} \cdot \frac{N-2}{2} > F_\alpha(2, N-2)$$



There is a probability α that $\mathcal{H}_0$ is rejected even if it is true. $F_\alpha(2, N-2)$ is given in tables. For instance, for $\alpha = 0.05$ we have

$$\lim_{N\to\infty} F_\alpha(2, N-2) = 2.60$$

11. A balanced realization has similar properties of reachability and observability. The magnitude of the elements of the Gramian expresses the relative importance of each state. Since $\sigma_1 \gg \sigma_2$, the second state has low influence on the input-output behavior. Therefore, it seems advisable to reduce the model to a first order model. We eliminate the second state, so that $x^{(2)}_{k+1} = x^{(2)}_k$. This gives the following equation for $x^{(2)}$

$$x^{(2)}_k = -0.0766x^{(1)}_k + 0.5210x^{(2)}_k + 0.1521u_k$$

Solving for $x^{(2)}$ and substituting in the state space model gives the reduced order state-space model

$$\begin{aligned} x_{k+1} &= 0.7912x_k + 0.9641u_k \\ y_k &= 0.9641x_k + 0.0483u_k \end{aligned}$$

The transfer function for this model is

$$H_{red}(z) = \frac{0.0483z + 0.8912}{z - 0.7912}$$

Since

$$H_{red}(z) \approx \frac{0.9}{z - 0.8}$$

the reduction is almost a pole-zero cancellation. However, we have to be careful; not all "almost pole-zero cancellation" should be done (compare exercise 10.2 in the textbook).

12.

a. For the given state-space realization $\{\Phi, \Gamma, C\}$, direct calculations give

$$C(zI - \Phi)^{-1}\Gamma = H(z)$$

b. 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.

c. Since the elements in the Gramians do not vary with a factor of magnitude, it is not suitable to perform a state reduction. (Notice that H has complex poles.)

Real-Time Identification

13. The forgetting factor $\lambda \in (0, 1]$ is typically chosen between 0.97 and 0.995. Its value should reflect the time-variation in the process. A forgetting factor $\lambda = 1$ gives that all data points are weighted equally, and is, hence, the theoretical choice for a time-invariant process. From the Taylor expansion around $\lambda = 1$

$$\ln(1 + (\lambda - 1)) = (\lambda - 1) + \dots$$

we get

$$\lambda^{k-i} = e^{(k-i)\ln\lambda} \approx e^{-(k-i)(1-\lambda)} = e^{-(k-i)/T}$$

where $T = 1/(1 - \lambda)$ is the time constant defining the approximate number of samples included in V. The suggested choice $\lambda \in [0.97, 0.995]$ corresponds to $T \in [33, 200]$. Roughly, there is no influences from data older than $2T$.

14. In the batch LS problem we consider the matrix

$$\bar{P}_k = (\Phi_k^T \Phi_k)^{-1} = \left(\sum_{i=1}^{k} \varphi_i \varphi_i^T\right)^{-1}$$

It satisfies the recursive equation

$$\bar{P}_k^{-1} = \bar{P}_{k-1}^{-1} + \varphi_k \varphi_k^T$$

We get the corresponding RLS covariance matrix by including the forgetting factor:

$$P_k^{-1} = \lambda P_{k-1}^{-1} + \varphi_k \varphi_k^T$$

with initial value P_0. The solution is

$$P_k^{-1} = \lambda^k P_0^{-1} + \sum_{i=1}^{k} \lambda^{k-i} \varphi_i \varphi_i^T \tag{7}$$

Further, $P_k^{-1}\widehat{\theta}_k^{RLS}$ satisfies

$$\begin{aligned} P_k^{-1}\widehat{\theta}_k^{RLS} &= P_k^{-1}\widehat{\theta}_{k-1}^{RLS} + P_k^{-1} K_k (y_k - \varphi_k^T \widehat{\theta}_{k-1}^{RLS}) \\ &= \lambda P_{k-1}^{-1}\widehat{\theta}_{k-1}^{RLS} + \varphi_k \varphi_k^T \widehat{\theta}_{k-1}^{RLS} + \varphi_k y_k - \varphi_k \varphi_k^T \widehat{\theta}_{k-1}^{RLS} \\ &= \lambda P_{k-1}^{-1}\widehat{\theta}_{k-1}^{RLS} + \varphi_k y_k \end{aligned}$$

so that

$$P_k^{-1}\widehat{\theta}_k^{RLS} = \lambda^k P_0^{-1}\widehat{\theta}_0^{RLS} + \sum_{i=1}^{k} \lambda^{k-i} \varphi_i y_i \tag{8}$$

From (7) and (8), we have

$$\widehat{\theta}_k^{RLS} = \left(\lambda^k P_0^{-1} + \sum_{i=1}^{k} \lambda^{k-i} \varphi_i \varphi_i^T\right)^{-1} \left(\lambda^k P_0^{-1}\widehat{\theta}_0^{RLS} + \sum_{i=1}^{k} \lambda^{k-i} \varphi_i y_i\right)$$

Hence,

- If $\lambda < 1$, the influences of initial conditions tend to zero as $k \to \infty$.
- If $\lambda = 1$ and $P_0 = \rho I$, then the influences of initial conditions tend to zero as $\rho \to \infty$.
- If $\lambda = 1$, then

$$
\begin{aligned}
\widehat{\theta}_k^{RLS} - \widehat{\theta}_k^{LS} &= P_k(\lambda^k P_0^{-1}\widehat{\theta}_0^{RLS} + \sum_{i=1}^{k} \lambda^{k-i}\varphi_i y_i) - P_k P_k^{-1}\widehat{\theta}_k^{LS} \\
&= P_k\left(P_0^{-1}\widehat{\theta}_0^{RLS} + \sum_{i=1}^{k}\varphi_i y_i - (P_0^{-1} + \sum_{i=1}^{k}\varphi_i\varphi_i^T)\widehat{\theta}_k^{LS}\right) \\
&= P_k(P_0^{-1}\widehat{\theta}_0^{RLS} - P_0^{-1}\widehat{\theta}_k^{LS}) = P_k P_0^{-1}(\widehat{\theta}_0^{RLS} - \widehat{\theta}_k^{LS})
\end{aligned}
$$

Thus,

$$\lim_{P_k \to 0} \widehat{\theta}_k^{RLS} - \widehat{\theta}_k^{LS} = 0$$

Further, by differentiating the loss function

$$V(\bar{\theta}) = (\bar{\theta} - \theta_0)^T P_0^{-1}(\bar{\theta} - \theta_0) + \varepsilon^T \varepsilon \tag{9}$$

where $\varepsilon = \mathcal{Y} - \Phi\bar{\theta}$, we get

$$\frac{dV}{d\bar{\theta}} = -\mathcal{Y}^T\Phi + \bar{\theta}^T\Phi^T\Phi + (\bar{\theta} - \theta_0)^T P_0^{-1} = 0$$

for $\bar{\theta} = \widehat{\theta}^{RLS}$. Hence, minimizing the loss function (9) gives a batch estimate equal to the RLS estimate with $\lambda = 1$.

Closed-Loop Identification

15. We have

$$
\begin{aligned}
u &= \frac{K}{1 + KH}(r - v) \\
y &= \frac{1}{1 + HK}v + \frac{HK}{1 + HK}r
\end{aligned}
$$

and since r and v are independent, we get

$$
\begin{aligned}
S_{uu}(i\omega) &= \frac{|K(e^{i\omega})|^2}{|1 + H(e^{i\omega})K(e^{i\omega})|^2}(S_{rr}(i\omega) + S_{vv}(i\omega)) \\
S_{yu}(i\omega) &= \frac{H(e^{i\omega})|K(e^{i\omega})|^2}{|1 + H(e^{i\omega})K(e^{i\omega})|^2}S_{rr}(i\omega) - \frac{\overline{K(e^{i\omega})}}{|1 + H(e^{i\omega})K(e^{i\omega})|^2}S_{vv}(i\omega)
\end{aligned}
$$

The estimate is

$$\widehat{H}(e^{i\omega}) = H(e^{i\omega}) - \frac{H(e^{i\omega})K(e^{i\omega}) + 1}{K(e^{i\omega})}\frac{S_{vv}(i\omega)}{S_{vv}(i\omega) + S_{rr}(i\omega)}$$

and as $S_{rr}(i\omega) \to 0$, we get

$$\widehat{H}(e^{i\omega}) \to -\frac{1}{K(e^{i\omega})}$$

16. The model is given by

$$\mathcal{M}: \qquad y_k = \begin{pmatrix} -y_{k-1} & u_{k-1} \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \phi_k^T \theta$$

so that we obtain the estimates as

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

where

$$\Phi = \begin{pmatrix} -y_1 & u_1 \\ \vdots & \vdots \\ -y_{N-1} & u_{N-1} \end{pmatrix} \qquad \mathcal{Y} = \begin{pmatrix} y_2 \\ \vdots \\ y_N \end{pmatrix}$$

a. We have that $\mathcal{Y} = \Phi\theta + \mathcal{E}$. This gives

$$\widehat{\theta} = \theta + (\frac{1}{N}\Phi^T\Phi)^{-1}(\frac{1}{N}\Phi^T\mathcal{E}) \to \theta, \quad N \to \infty$$

since interacting elements of Φ and $\mathcal{E}$ are uncorrelated.

b. Using the central limit theorem (eq. 6.89), we have that

$$\sqrt{N}(\widehat{\theta} - \theta) \in \mathrm{As}\mathcal{N}(0, \Sigma), \qquad \Sigma = \lambda^2 \mathrm{E}(\frac{1}{N}\Phi^T\Phi)^{-1}$$

which gives

$$\mathrm{Cov}(\widehat{\theta}) = \frac{1}{N}\Sigma = \frac{\lambda^2}{N} \lim_{N\to\infty} \left(\frac{1}{N} \begin{pmatrix} \sum y_k^2 & -\sum y_k u_k \\ -\sum y_k u_k & \sum u_k^2 \end{pmatrix} \right)^{-1}$$

Deriving the asymptotic diagonal elements gives

$$\mathrm{Var}(\widehat{a}) = \frac{\lambda^2}{N} \frac{r_u(0)}{r_y(0) r_u(0) - r_{yu}^2(0)} \qquad \mathrm{Var}(\widehat{b}) = \frac{\lambda^2}{N} \frac{r_y(0)}{r_y(0) r_u(0) - r_{yu}^2(0)}$$

Introduce $\alpha = a + Kb$ to simplify further calculations. Using the closed loop system

$$\mathcal{S}: \qquad y_k = -(a + Kb) y_{k-1} + b r_{k-1} + e_k$$

and some calculations we obtain

$$r_y(0) = \frac{b^2\sigma^2 + \lambda^2}{1 - \alpha^2} \qquad r_u(0) = \sigma^2 + K^2 r_y(0) \qquad r_{yu}(0) = -K r_y(0)$$

so that

$$\mathrm{Var}(\widehat{a}) = \frac{\lambda^2}{N} \left(\frac{K^2}{\sigma^2} + \frac{1 - \alpha^2}{b^2\sigma^2 + \lambda^2} \right) \qquad \mathrm{Var}(\widehat{b}) = \frac{\lambda^2}{N} \cdot \frac{1}{\sigma^2}$$

From these expressions we observe that the asymptotic parameter variances will tend to infinity if $\sigma \to 0$.

17. The models we are using in this problem is

$$\mathcal{M}_1: \qquad y_k + \bar{a} y_{k-1} = \bar{b} u_{k-1} + e_k$$

and

$$\mathcal{M}_2: \qquad y_k + \alpha y_{k-1} = \bar{b} v_{k-1} + e_k$$

where $\alpha = a + fb$.

Since the signals u_k and v_k are both persistently exciting and the identification set-ups are both on the form $y_k = \varphi_k \theta + e_k$, the least-squares estimates are consistent. It remains to compare the estimates variances.

From the textbook (p. 79), we know that the covariance matrix of a least-squares estimate $\widehat{\theta}$ is

$$\text{Cov } \widehat{\theta} \approx (\Phi_N^T \Phi_N)^{-1} E e_k^2 \approx \frac{1}{N} E \varphi_k \varphi_k^T \lambda^2$$

for large N. For the first method, we have

$$\varphi_k = \begin{pmatrix} -y_{k-1} & u_{k-1} \end{pmatrix}^T$$

Then,

$$\begin{aligned} E\varphi_k \varphi_k^T &= \begin{pmatrix} Ey_k^2 & Ey_k u_k \\ Ey_k u_k & Eu_k^2 \end{pmatrix} \\ &= \frac{1}{1 - a^2 - b^2 f^2} \begin{pmatrix} b^2\sigma^2 + \lambda^2 & -f(b^2\sigma^2 + \lambda^2) \\ -f(b^2\sigma^2 + \lambda^2) & f^2\lambda^2 + \sigma^2(1 - a^2) \end{pmatrix} \end{aligned}$$

Further,

$$(E\varphi_k \varphi_k^T)^{-1} = \frac{1}{\sigma^2} \begin{pmatrix} \frac{f^2\lambda^2 + \sigma^2(1-a^2)}{b^2\sigma^2 + \lambda^2} & * \\ * & 1 \end{pmatrix}$$

where $*$ denotes elements we are not interested in. Thus,

$$\begin{aligned} \text{Var } \hat{a} &= \frac{\lambda^2}{N} \cdot \frac{f^2\lambda^2 + \sigma^2(1 - a^2)}{\sigma^2(b^2\sigma^2 + \lambda^2)} \\ \text{Var } \hat{b} &= \frac{\lambda^2}{N} \cdot \frac{1}{\sigma^2} \end{aligned}$$

In the same way, for the second model we have

$$E\varphi_k \varphi_k^T = \begin{pmatrix} \frac{b^2\sigma^2 + \lambda^2}{1 - a^2 - b^2 f^2} & 0 \\ 0 & \sigma^2 \end{pmatrix}, \qquad \varphi_k = \begin{pmatrix} -y_{k-1} & v_{k-1} \end{pmatrix}^T$$

so that

$$\begin{aligned} \text{Var } \hat{\alpha} &= \frac{\lambda^2}{N} \cdot \frac{1 - a^2 - b^2 f^2}{b^2\sigma^2 + \lambda^2} \\ \text{Var } \hat{b} &= \frac{\lambda^2}{N} \cdot \frac{1}{\sigma^2} \end{aligned}$$

Hence, the estimate of b is as good in the first method as in the second. In the second method, $\hat{a}$ is derived from $\hat{\alpha}$. Then,

$$
\begin{aligned}
\operatorname{Var}\hat{a} &= \operatorname{Var}\hat{\alpha} + f^2 \operatorname{Var}\hat{b} - 2f\operatorname{Cov}\,(\hat{\alpha},\hat{b}) = \frac{\lambda^2}{N}\left(\frac{1-a^2-b^2f^2}{b^2\sigma^2+\lambda^2} + \frac{f^2}{\sigma^2}\right) \\
&= \frac{f^2\lambda^2+\sigma^2(1-a^2)}{\sigma^2(b^2\sigma^2+\lambda^2)}
\end{aligned}
$$

which is identical to the variance in the first method. To conclude, in this case it does not matter if the identification is done in open or closed loop.