Learning dynamical systems using system identification, Problems

(Course: SSY 230)

Jonas Sjöberg

Department of Electrical Engineering Chalmers University

1 Model Uncertainty

Problem 1
Consider the situation where the following system is to be identified

$$y(t) = -\theta_0\, y(t-1) + e(t)$$

In an experiment $N = 100$ data are collected, $\{y(t)\}_{t=1}^{100}$.

(a) Give expression for Least Square (LS) stimate, $\hat{\theta}$, of θ_0.

(b) Give expression for the estimate of $e(t) = \varepsilon(t)$, called the prediction error. What is the estime of $\mathrm{Var}(e(t))$?

(c) Give the expression for $P_\theta = \mathrm{Var}(\hat{\theta})$

(d) Assymptotically (in N),

$$\hat{\theta}_N \in AsN(\theta_0, P_\theta) = \frac{1}{\sqrt{(2\pi)^2 \det P_\theta}} \exp(-\frac{1}{2}(\hat{\theta}_N - \theta_0) P_\theta^{-1} (\hat{\theta}_N - \theta_0)^T)$$

Assuming you can perform multiple experiments to collect data from the system, give an algorithm for validating P_θ and the distribution for $\hat{\theta}_N$ in the case $N = 100$.

Use (4.12) and (4.13) in S&S. Strictly to hold, φ should be determinsitic (Remark 1 p. 66), however, if $N \to \infty$ then the expresion also hold for stochastically generated φ. This is described in Example 7.6 p. 207.

Solution

Problem 2

In system identification, often you have the situation where you are interested of properties of a function depending on estimated parameters. Consider

$$\hat{f}(x) = f(x, \hat{\theta}_N)$$

the estimate of f depending on the estimated $\hat{\theta}_N$, where N is the number of data used and d is the dimension of the parameter vector. $\hat{\theta}_N$ is asymptotically normal distributed:

$$\hat{\theta}_N \in AsN(\theta_0, P_\theta) = \frac{1}{\sqrt{(2\pi)^2 \det P_\theta}} \exp(-\frac{1}{2}(\hat{\theta}_N - \theta_0)P_\theta^{-1}(\hat{\theta}_N - \theta_0)^T)$$

where θ_0 is the true parameter vector.

(a) Show that the random variable $P^{-1/2}\hat{\theta}_N$ is asymptotically normal distributed with variance I. This means that the components of $P^{-1/2}\hat{\theta}_N$ are independent of each other.

(b) A sum of the squares of d Gaussian distributed variables, with mean zero and unit variance is a new random variable with distribution $\chi^2(d)$. Show that $(\hat{\theta}_N - \theta_0)P_\theta^{-1}(\hat{\theta}_N - \theta_0)^T \in \chi^2(d)$

(c) Discribe how an approximation of the confident region of degree α, of f, for any value of x can be obtained by sampling θ_k K times using the distribution described above.

(d) Make scetchs for the cases of one and two dimensional parameter spaces (ie two scetches) where you indicate where the samples θ_k are taken.

Solution

2 Nonparametric Identification

Problem 3
Consider the system

$$y(t) = G_0(q)u(t) + v(t)$$

controlled by the regulator

$$u(t) = -F(q)y(t) + r(t)$$

where $r(t)$ is an external reference signal. r and v are independent and their spectra are $\Phi_r(\omega)$ and $\Phi_v(\omega)$, respectively. The usual spectral analysis estimate of $G_0(q)$ is given by S & S (3.33) or R.J (4.28). Show that in the limit, $N \to \infty$, and if a smoothing window is used for which $\gamma \to \infty$, then $\hat{G}_N(e^{i\omega})$ converge to

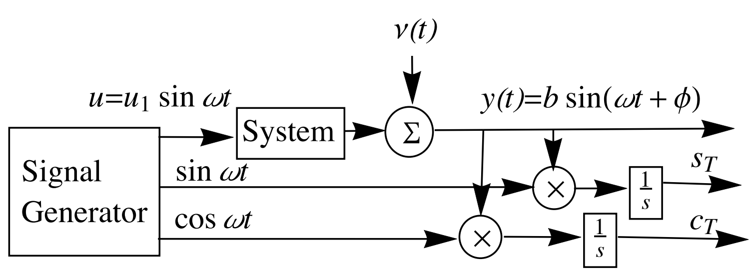
$$\hat{G}_*(e^{i\omega}) = \frac{G_0(e^{i\omega})\Phi_r(\omega) - F(e^{-i\omega})\Phi_v(\omega)}{\Phi_r(\omega) + |F(e^{i\omega})|^2\Phi_v(\omega)}$$

What happens in the two special cases $\Phi_r \equiv 0$ and $F \equiv 0$?

Is spectral analysis useful when there is feedback?

Solution

Problem 4
Consider the setup for frequency analysis in the following figure.



Assume that the disturbance ν affects the output of the system so that the resultant identification will be compromised. Also assume that the measurement time is n periods for each test signal with frequency ω_i.

(a) How will the Bode plot be affected if the disturbance is a sinusoid of frequency ω_ν, i.e., $\nu(t) = A_\nu \sin(\omega_\nu t)$? Consider both the case $\omega_\nu = \omega_i$ and $\omega_\nu \neq \omega_i$

(b) Assume that frequency response analysis has been performed with the measurement duration $T = 2\pi\ n/\omega_i$ and assume that ν is high bandwidth noise with mean 0 and variance σ^2. How much is it necessary to increase the measurement duration in order to reduce the variance of $\hat{G}$ by a factor of two?

(c) Assume that the frequency analysis and the Bode plots will be used as a basis for control design. Discuss how to choose a finite measurement time at each frequency to obtain optimal closed-loop control. Is it possible to choose a measurement strategy for this purpose?

Solution

Problem 5

Assume that the ideal discrete time impulse with amplitude 1 cannot be implemented, and that it is replaced by a pulse spread out over T samples, ie, the input

$$u(t) = \begin{cases} 1/T, & \text{if } 0 \leq t < T \\ 0, & \text{if } t \geq T \end{cases}$$

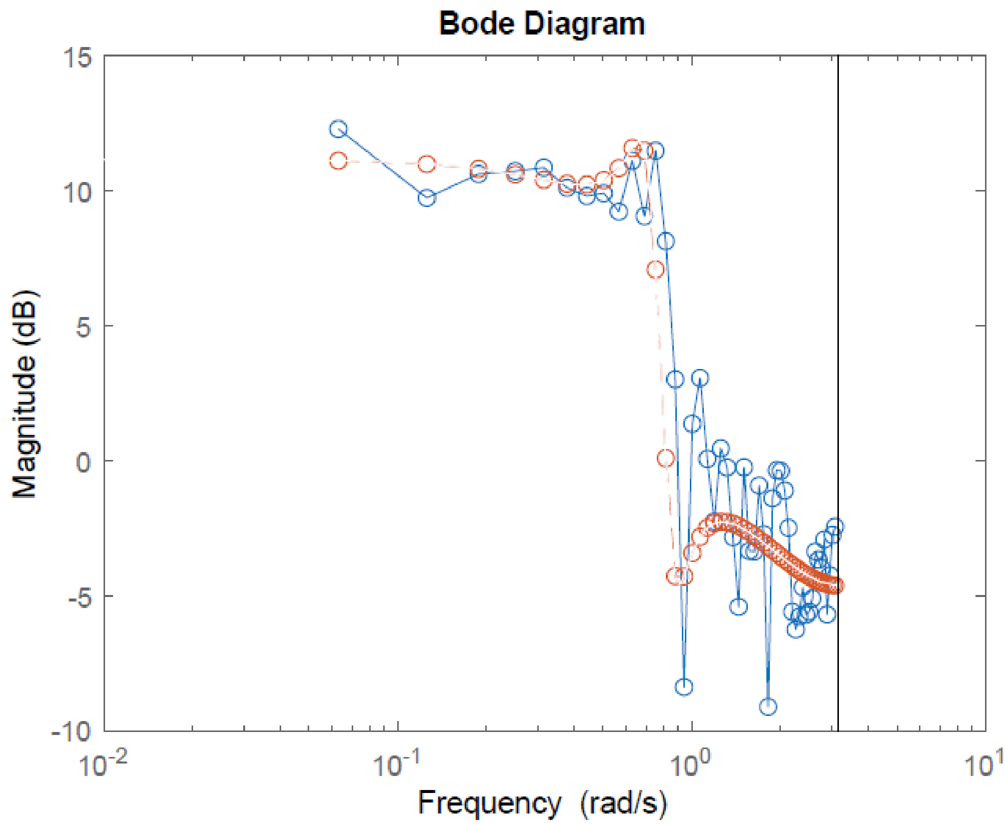
How should the discrete time impulse response be determined from data, $y(0),\ y(1), \ldots,$ obtained when using this input? Give the equations for $h(t)$. Assume that the system is linear and time invariant.

Solution

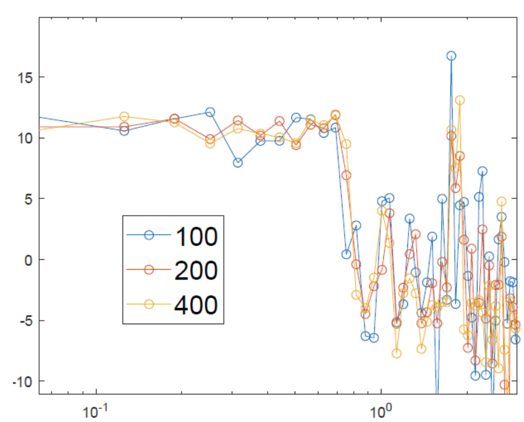
Problem 6

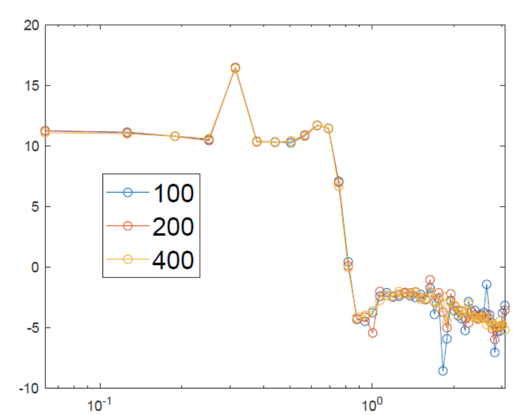
(a) Two periods of a periodic signal is used as input signal to a linear time-invariant system. The output from the system was recorded at almost ideal conditions so that it can be considered as disturbance-free. The data, the input and the input, was divided into two equally long data sets, and ETFE was computed on each of them. The results is shown in the figure.

Why are the two ETFE different? Which of the to ETFE is fom the fisrt data set and which is from the second? Motivate your answer.



(b) A periodic input is used to generate data for identification. Data from increasing number of periods are used for computing the ETFE. The same output from the system is measured in two ways, in one case, the disturbance is mainly a pure sinusoid, in the other case, the output is corrupted by almost white disturbancies. The result from the two outputs are shown in the following figures. The number of data are also indicated. Which is which? Motivate.





Solution

3 Linear Black-Box Models

Problem 7
Assume that input-output data $\{u(t)\}$ and $\{y(t)\}$ are observed from the system

$$\mathcal{S}: \; y(t) + a\,y(t-1) = b\,u(t-1) + w(t) + c\,w(t-1)$$

where $\{u(t)\}$ and $\{w(t)\}$ are independent white-noise sequences with variances $\sigma_w^2 = \sigma_u^2 = \sigma^2$. Assume that the model

$$\mathcal{M}: \; y(t) + a\,y(t-1) = b\,u(t-1)$$

is fitted to data by means of least-squares identification.

(a) Show that the expected squared prediction error is minimized for

$$\hat{\theta} = \begin{pmatrix} \hat{a} \\ \hat{b} \end{pmatrix} = \begin{pmatrix} a - c\,\sigma^2/\mathrm{E}\,[y^2(t)] \\ b \end{pmatrix}$$

(b) Show also that the prediction error variance is smaller than that for $\hat{a} = a$ and $\hat{b} = b$.

(c) Calculate also the expression for $\mathrm{E}\,[y^2(t)]$.

Solution

Problem 8
An impulse response test has given the result

Time t	0_+	0.2	0.4	0.6	0.8
Impulse response $h(t)$	3.4	2.3	1.7	1.2	0.9

(a) Assume that the impulse response can modeled as $h(t) = K\,e^{-t/\tau}$. Determine a least-squares estimate of the parameters K and τ.

(b) Determine an estimate of the parameters K and τ by means of a least-squares fit of the transformed model

$$\log h(t) = \log K - \frac{t}{\tau} = [1 \ -t] \begin{bmatrix} \log K \\ 1/\tau \end{bmatrix}$$

That is, $\log K$ ans $1/\tau$ are being estimated instead of K and τ.

(c) Reflect on the estimate in (a) and (b). What are the advantages and dissadvantages of each of them?

Solution

Problem 9

With a data set of N data, let Φ_N be the matrix of the regressor vector, and E_N the vector of the (unknown) noise.

(a) Show that the estimation error, ie, $\hat{\theta} - \theta_0$, of the least-squares estimate is $(\Phi_N^T \Phi_N)^{-1} \Phi_N^T E_N$.

(b) This is the total error, divide it into the bias error, and the variance error.

Solution

Problem 10

Assume that the noise sequence $\{e(t)\}_{t=1}^N$ consists of independent normally distributed components $e(t) \in N(0, \sigma^2)$. Show that the least-squares estimate $\hat{\theta}$ of the parameters of the model $y(t) = \varphi^T(t)\theta + e(t)$ is assymptotically normally distributed. Make neccessary assumptions on $\varphi(t)$.

Note: assume that mean values converge to expectation, as described in S& S Lemma B.1 and Lemma B.3.

Solution

Problem 11

(a) Consider a moving average (MA) process $y(t) = b_1 u(t-1) + \cdots + b_m u(t-m) + v(t)$, and formulate the least-squares estimate of the process parameters $b_1, \ldots, b_m$ given N measurements $\{u(t), y(t)\}_{t=1}^{N}$.

(b) Consider the two possibilities, $v(t)$ white and colored, respectively. Can you guarantee a consistent estimate if you make some assumption on the cross-correlation of $v(t)$ and $u(t)$ for the two cases?

Solution

Problem 12
Given the ARX model

$$y(t) = \frac{B(q)}{A(q)} u(t) + \frac{1}{A(q)} e(t)$$

where $A(q) = 1 - a_1 q^{-1} - a_2 q^{-2}$ and $B(q) = b_2 q^{-2}$.

(a) Write expressions for the 1-step prediction using transfer function expression, and as well as equation.

(b) The same thing as in (a) but the 2-step prediction.

(c) Write an expression for the k-step prediction as an equation using the m-step prediction for $m = k-1,\ k-2, \ldots,\ 1$. This equation is basically what you need in one of the tasks in Project 2.

Solution

Problem 13
Assume the true system is given by a ARMAX type system

$$y(t) = \frac{B(q)}{A(q)} u(t) + \frac{C(q)}{A(q)} e(t)$$

where all polynomials are of order 1. The noise $e(t)$ is white and the system operates in open loop.

Assume that an ARX model is fitted to data from the system and that the order of the ARX model, n, is much higher than the original system, eg, $n \leq 8$.

(a) Express the mean square error as a sum of two terms, one for the plant modelling error and one for the disturbance modelling error.

(b) Show that a good model of both the plant and the disturbance transfer functions can be obtained although the model is of incorrect type. Show that the quality improves for higher n.

(c) Are there any drawbacks with the obtained model compared to, eg, if a 1-order OE-model, if the model is intended to be used for control design?

Solution

Problem 14
Consider the AR-process

$$y(t) = a_1 y(t-1) + e(t)$$

Calculate the k-step prediction for $k > 0$.

Solution

Problem 15
Consider the MA-process

$$y(t) = c_1 e(t-1) + e(t)$$

Calculate the k-step prediction for $k > 0$.

Solution

4 Prediction Error Method

Problem 16
Given N observations of $y(t)$ and $\varphi(t)$, the model

$$y(t) = \theta^T \varphi(t) + e(t).$$

and the assumption that the noise $e(t)$ is iid according to

$$f(x) = \frac{x}{\sigma^2} e^{-\frac{x^2}{2\sigma^2}}, \quad x > 0.$$

Formulate the Maximum-Likelihood estimate of σ^2 and θ.

Note: An explicit solution is not likely since the PDF of the noise is not Gaussian. Hence, your answer can be stated as the equations which need to be solved numerically, given the data.

Solution

Problem 17

Consider the system

$$\mathcal{S}: \; y(t) = -a\, y(t-1) + b\, u(t-1) + v(t)$$

where $\{v(t)\}$ is a sequence of independent identically distributed stochastic variables, each with the probability density function $f(x) = \mu e^{-\mu x}, \;\; x \geq 0$. Design the Maximum-Likelihood method that permits estimation of a and b. You do this by formulating the optimization problem, with criterion to be minimized and constraints. The solution of the minimization problem does not need to be given.

What is the difference in algorithmic complexity for an unknown μ and for a known value of μ?

Solution

Problem 18

Consider the system

$$y(t) = a_0 + b(t)e(t)$$

where

- $e(t)$ is a white Gaussian measurement noise process of zero mean and with unknown variance σ_e^2.

- a_0 is an unknown model parameter.
- $b(t)$ is a known and deterministic time series.

(a) Formulate the Maximum Likelihood estimate for a_0 and σ_e^2 given N measurements of $y(t)$.

(b) Give an expression for the assymptotic estimate of the variance of the ML estimate.

Hint: One way to solve this problem is to use the information provided on p.210 in S&S.

Solution

5 Model Validation

Problem 19
Consider the ARMAX model

$$A(q)y(t) = B(q)u(t) + C(q)e(t)$$

Discuss how to test the hypothesis that $A(q) = C(q)$, that is, we have white measurement noise.

Solution

Problem 20
Let $\varepsilon(t)$ be a process independent on $u(t)$ and such that

$$e(t) = H^{-1}(q)\varepsilon(t)$$

is white noise with variance λ. Let

$$\zeta(t) = \begin{bmatrix} u(t-s-1) \\ \vdots \\ u(t-s-M) \end{bmatrix}$$

Let

$$\hat{R}_u = \frac{1}{N}\sum_{t=1}^{N}\zeta(t)\zeta^T(t), \quad \hat{r} = \frac{1}{\sqrt{N}}\sum_{t=1}^{N}[H^{-1}(q)\varepsilon(t)]\zeta(t)$$

Show that

$$\frac{1}{\lambda}\hat{r}^T\hat{R}_u^{-1}\hat{r} \in As\chi^2(M)$$

Solution

Problem 21

Consider the ARMA-system

$$y(t) = -a\,y(t-1) + e(t) + c\,e(t-1)$$

where $\{e(t)\}$ is a white noise sequence with variance σ^2. Assume that we use a model that is of the same structure,

$$y(t) = -a\,y(t-1) + e(t) + c\,e(t-1)$$

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

$$\hat{\theta}_N \to N(\theta_0, \text{Cov}(\hat{\theta}_N)),$$

where $\hat{\theta}_N = [\hat{a}\ \hat{c}]^T$ is the estimate based on N samples and

$$\text{Cov}(\hat{\theta}_N) = \frac{\sigma^2}{N(c-a)^2}\begin{pmatrix} (1-a^2)(1-a\,c)^2 & (1-a^2)(1-a\,c)(1-c^2) \\ (1-a^2)(1-a\,c)(1-c^2) & (1-a\,c)^2(1-c^2) \end{pmatrix}$$

Thus, the distribution of $\hat{\theta}_N$ will tend to the normal distribution as N increases. As for the validation of model complexity vs. statistical model accuracy, it makes sense to ask whether the estimates of a and c are sufficiently different in magnitude and accuracy to justify the model

$$Y(z) = \frac{z+c}{z+a}E(z)$$

If a and c are sufficiently close, the model would reduce to $y(t) = e(t)$.

(a) Determine the asymptotic mean and variance of $\hat{a} - \hat{c}$.

(b) How would you test the hypothesis

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

on the significance level $p = 0.05$?

(c) Make an F-test based on the two models

$$\begin{aligned} \mathcal{M}_1: &\quad y(t) = e(t) \\ \mathcal{M}_2: &\quad y(t) + a\, y(t-1) = e(t) + c\, e(t-1) \end{aligned}$$

and the hypotheses

$$\begin{aligned} \mathcal{H}_0: &\quad \mathcal{M}_1 \text{ is as good model for the system as } \mathcal{M}_2 \Rightarrow \sigma_1^2 = \sigma_2^2 \\ \mathcal{H}_1: &\quad \mathcal{M}_2 \text{ models the system better than } \mathcal{M}_1 \Rightarrow \sigma_1^2 > \sigma_2^2 \end{aligned}$$

where σ_1^2 and σ_2^2 are the variances for the white noise $e(t)$ in the model $\mathcal{M}_1$ and $\mathcal{M}_2$, respectively. If we have 102 measurements and the true values are $a = 0.1$, $c = 0.2$ would we expect to reject H_0 with a significant level of $p = 0.05$. What if $a = 0.7$ and $c = 0.2$?

Solution

Problem 22

Consider an unknown linear system

$$\mathcal{S}: \; y(t+n) + a_1 y(t+n-1) + \cdots + a_n y(t) = b_0 u(t+n) + \cdots + b_n u(t) + e(t+n)$$

where $e(t) \in N(0, \sigma^2)$ and n is unknown. We have performed an experiment collecting 200 data points where the input $u(t)$ was a PRBS signal with unit amplitude. We want to estimate ARX models, that is,

$$\mathcal{M}: \; y(t+n) + a_1 y(t+n-1) + \cdots + a_n y(t) = b_0 u(t+n) + \cdots + b_n u(t)$$

where $n = 1, \ldots, 5$. Determine the suitable number of parameters the ARX-model should have,

(a) according to the Akaike Information Criterion

(b) and the Final Prediction Error Criterion

if the loss function is given by

$$V(\theta) = \frac{1}{N} \sum_{t=1}^{N} \varepsilon^2(t, \theta), \qquad \varepsilon(t, \theta) = y(t) - \hat{y}(t, \theta)$$

and the obained values of the loss function are

$$V(\theta) = [1.6\ 1.0\ 0.97\ 0.95\ \ 0.94]$$

for $n = 1, \ldots, 5$, respectively.

Solution

6 Input Signal and Experimental Design & Closed Loop Identification

Problem 23

Pseudo random binary sequence (PRBS) is an easily generated signal with almost "white" properties. A PRBS is generated using feedback around a shift register. The length of the shift register determines the period of the sequence. A length N register results in a sequence with period $2N - 1$.

(a) The sequences generated by shift registers of length 2 and 3, respectively, look as follows

$$\text{length 2:} \qquad \ldots \underbrace{1, -1, 1,} \underbrace{1, -1, 1,} \ldots$$

$$\text{length 3:} \qquad \ldots \underbrace{1, -1, -1, 1, 1, 1, -1,} \underbrace{1, -1, -1, 1, 1, 1, -1,} \ldots$$

Determine the autocovariance functions $C_2(\tau)$ and $C_3(\tau)$ for the two sequences. What do you think $C_N(\tau)$ looks like? Remark: Although the PRBS is a deterministic signal it is still possible to define the covariance function as

$$\begin{aligned} \mu_u &= \lim_{N \to \infty} \frac{1}{N} \sum_{t=1}^{N} u(t) \\ C_{uu}(\tau) &= \lim_{N \to \infty} \frac{1}{N} \sum_{t=1}^{N} (u(t+\tau) - \mu_u)(u(t) - \mu_u)^T \end{aligned}$$

where μ_u is the mean value. It is also possible to change the PRBS to a stochastic signal by introducing a stochastic phase (i.e., the signal form is deterministic but the starting point in the sequence is stochastic realization).

(b) Determine the autospectra $S_2(i\,\omega)$ and $S_3(i\,\omega)$ of the two sequences in (a). What do you think $S_N(i\,\omega)$ looks like?

Discuss when a PRBS could be a good choise of inut signal. Try to give examples when it might be a good choise, and when it is a bad choice.

Solution

Problem 24
The coherence function γ_{yu} between signals u and y is defined as

$$\gamma_{yu}(\omega) = \frac{|S_{yu}(i\,\omega)|}{\sqrt{S_{uu}(i\,\omega)S_{yy}(i\,\omega)}}.$$

Suppose we are going to identify the system with transfer function G

$$y(t) = G\,u(t) + v(t)$$

where u is the input signal, y is the output signal, and v is a noise signal. Express $|S_{yu}(i\,\omega)|^2$ in terms of G, $S_{uu}(i\,\omega)$, and S_{vv}, so that it becomes obvious that $S_{yu}(i\,\omega)$ can be used to judge if the excitation in the input signal is suffucient for identification.

Solution

Problem 25
Consider the system

$$\mathcal{S}:\; y(t) + a\,y(t-1) = b\,u(t-1) + w(t) + c\,w(t-1)$$

where $\{w(t)\}$ is a zero-mean white-noise process with the variance σ^2. Assume that the system is operating in closed loop with the input $u(t) = -K\,y(t)$.

(a) Show that identifiability of a and b is lost if the system is identified with the direct approach.

(b) What is the asymptotic parameter estimate for large N when fitting the model

$$\mathcal{M}:\; y(t) = \alpha\,y(t-1)$$

with data from $\mathcal{S}$?

(c) Explain how a and b can be obtained if data is collected in several experiements where different values of K is used. How many experiemnts are needed to obtain estimates of the parameters?

Solution

Problem 26
Consider the system

$$\mathcal{S}: \; y(t) + a\, y(t-1) = b\, u(t-1) + w(t)$$

where $\{w(t)\}$ is a zero-mean white-noise process with the variance σ^2 and $u(t) = 1$ for all $t \geq 0$. Show that the input is persistently exciting of order 1 but not of order 2 and determine under what circumstances the least-squares estimate of a, b might give consistent estimates.

Solution

Problem 27
One problem with using normal distributed noise as excitation signal is its amplitude. There is no guarantee that the signal level will be within fixed limits, and it is easy to get saturation of the process input signal. Suppose the input signal to a certain process is limited to the amplitude range -1.0 to 1.0. The input signal is chosen as a normally distributed white-noise sequence with zero mean and variance σ^2. What is the largest σ^2 one can tolerate if the risk for input signal saturation should be less than 0.01?

Solution

Problem 28
Consider the system

$$y(t) = b_1 u(t-1) + b_2 u(t-2) + b_3 u(t-3)$$

Determine if it is possible to identify all three parameters in the model by means of the input $u(t) = \sin(t)$.

Solution

Problem 29

Closed-loop identification is sometimes necessary, for example in cases where the plant is unstable or when using operational data records from plants in production. Assume the following closed-loop setting

$$\begin{aligned} Y(s) &= G_P(s)U(s) \\ U(s) &= R(s) - G_C(s)Y(s) \end{aligned}$$

and that we have performed indirect identification with the following closed-loop model

$$G_{cl}(s) = \frac{(s+1)(s+3)}{(s+2)(s+3)(s+4)+s+1}$$

The used controller is given by

$$G_C(s) = \frac{1}{s+3}$$

(a) Find a model for the process $G_P(s)$.

(b) Discuss the main possible transfer function estimation problem associated with indirect identification.

Solution

7 State Space Identification

Problem 30

Uniqueness properties of a state-space model

Consider the state-space model

$$\begin{aligned} x(t+1) &= \begin{pmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{pmatrix} x(t) + \begin{pmatrix} b \\ 0 \end{pmatrix} u(t) \\ \theta &= (a_{11}\ a_{12}\ a_{22}\ b)^T \end{aligned}$$

Examine the identifiability for the following cases

(a) The first state variable is measured.

(b) The second state variable is measured.

(c) Both state variables are measured.

Solution

Problem 31
Consider the system description

$$\begin{aligned} x(t+1) &= a\,x(t) + b\,u(t) + \xi(t) \\ y(t) &= x(t) + e(t) \end{aligned}$$

where $e(t)$ is white Gaussian noise with variance σ^2 and $\xi(t)$ has the distribution

$$\begin{aligned} \xi(t) \quad &= 0, \quad w.p.\ 1-\lambda \\ \xi(t) \quad &= +1. \quad w.p.\ \lambda/2 \\ \xi(t) \quad &= -1. \quad w.p.\ \lambda/2 \end{aligned}$$

The coefficients $a,\ b$ and λ are adjustable parameters. Can this description be cast into the form

$$y(t) = G(q,\theta)u(t) + H(q,\theta)e_2(t)$$

where $\theta = [a\ b]$, and $f_e(x,\theta)$, the PDF of $e_2(t)$ and $e_2(t)$ is white noise. If so, at the expense of what approximation?

Solution

Problem 32
A state-space model of ship-steering dynamics is given as

$$\frac{d}{dt}\begin{bmatrix} v(t) \\ r(t) \\ h(t) \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & 0 \\ a_{21} & a_{22} & 0 \\ 0 & 1 & 0 \end{bmatrix}\begin{bmatrix} v(t) \\ r(t) \\ h(t) \end{bmatrix} + \begin{bmatrix} b_{11} \\ b_{21} \\ 0 \end{bmatrix} u(t)$$

where $u(t)$ is the rudder angle, $v(t)$ the sway velocity, $r(t)$ the turning rate, and $h(t)$ the heading angle.

Suppose only $u(t)$ and $y(t) = h(t)$ are measured. Show that the six parameters $a_{ij},\ b_{ij}$ are not identifiable. This is called *nonuniqueness* in S & S 6.3.

Solution

Problem 33
Sampling a simple continuous time system

Consider the following continuous time system:

$$
\begin{aligned}
\dot{x} &= \begin{pmatrix} 0 & K \\ 0 & 0 \end{pmatrix} x + \begin{pmatrix} K \\ 0 \end{pmatrix} u + \begin{pmatrix} 0 \\ 1 \end{pmatrix} v, \quad Ev(t)v(s) = r\,\delta(t-s) \\
y &= (1\ 0)x
\end{aligned}
$$

(Which can be written as

$$Y(s) = \frac{K}{s}U(s) + \frac{K}{s^2}V(s).)$$

Assume that the gain K is unknown and is to be estimated from *discrete time* measurements

$$y(t) = (1\ 0)x(t), \quad t = 1, \ldots, N$$

Find the discrete time description of the system, eg of the form (6.28) p. 157, S & S. Assum the input is constant over the sampling intervals. See also Continuous time models, p. 158, S & S.

Solution

8 Nonlinear System Identification

Problem 34

Consider the following nonlinear model structure

$$y(t) = \frac{b_1 u(t-1) + b_2 u(t-2)}{1 + a_1 y^2(t-1)} + v(t)$$

(a) Assuming $v(t)$ is white, give an expression for the predictor $\hat{y}(t, \theta|t-1)$, where θ contains the unknown parameters.

(b) Suppose $v(t)$ is not white but can be modelled by a first order (unknown) ARMA process. Suggest a predictor $\hat{y}(t, \theta|t-1)$ for this case.

(c) To estimate the unknown parameters using a prediction error method, expressions for

$$\psi(t, \theta) = -\frac{d\hat{y}(t, \theta|t-1)}{d\theta}$$

is needed. Give these for the two cases in (a) and (b).

Solution

Problem 35
Consider a Hammerstein model, ie, a linear transfere function with a static nonlinearity at the input,

$$G(f(u(t))$$

The static nonlinearity is a saturation parameterized as

$$f(u(t)) = \begin{cases} \theta_1 \cdot \theta_2, & \text{if } u(t) > \theta_2 \\ \theta_1 u(t), & \text{if } |u(t)| \leq \theta_2 \\ -\theta_1 \cdot \theta_2, & \text{if } u(t) < -\theta_2 \end{cases}$$

Suppose that the linear system can be described by a second-order ARX model. Write down, explicitly, the predictor for this model, parameterized in θ_1, θ_2 and the ARX parameters.

Solution

Problem 36
Consider a bilinear model structure described by

$$\begin{aligned} x(t) + a_1 x(t-1) + a_2 x(t-2) &= b_1 u(t-1) + b_2 u(t-2) + c_1 x(t-1) u(t-1) \\ y(t) &= x(t) + v(t) \end{aligned}$$

where

$$\theta = [a_1,\ a_2,\ b_1,\ b_2,\ c_1]^T$$

(a) Assume $\{v(t)\}$ to be white noise and compute the predictor $\hat{y}(t,\theta)$ and give the expression for it in the pseudolinear regression form

$$\hat{y}(t,\theta) = \varphi^T(t,\theta)\theta$$

with a suitable vector $\varphi^T(t,\theta)$.

(b) Now, suppose that $\{v(t)\}$ is not white, but can be modeled as an (unknown) first order ARMA process. Then suggest a suitable predictor for the system.

Solution

Problem 37

Consider the following nonlinear system with the unknown parameters a, b to be estimated from observations of u, y,

$$\ddot{y}(t) + a\,y(t)\,\dot{y}(t) + b\,y(t) = u(t)$$

This is a Liénard-type equation that typically arises when modeling mechanical systems with variable friction and damping. The Liénard equation often exhibits limit-cycle behavior that depends on the initial conditions (Birkhoff and Rota, 1974).

We are considering a situation where observations of $\dot{y}(t)$ and $y(t)$ (and $u(t)$) are available, possibly from position and velocity sensors. The parameters a and b are unknown and need to be estimated.

(a) If $\ddot{y}(t)$ was available, one could form the estimation as a least-squares problem. Hence, a straight-forward approach would be to numerically take the derivative of $\dot{y}(t)$ to obtain $\ddot{y}(t)$. However, taking numericl derivatives is often a bad solution due to that it amplifies high frequency noise. If the sampling time is short, the differenciation implies subtracting two similar values and the result might have lost all significance due to the numerical precision in the computer. Instead, it is better to integrate the whole equation and in that way transform it to an integral equation. Also, by integrating, the effect of noise decrcrease. In this system we do not have any explcit noise, but of course, if we have data then we have noise to some extent, at lest we have numerical noise due to the finite precision of the computer. This we also have if we generate the data in simulation.

Make a parametric model that allows a least squares estimate of the parameters a and b from observations of $\dot{y}$ and y by integrating $\ddot{y}$ "away".

(b) The initial conditions may have a strong influence on the system trajectories and partly determine the limit cycle behavior. Make a modification of the model so as to compensate for the harmful effect of initial conditions on the estimates. Do this by introducing a *modulating* function $p_k(t)$ and multiply the system equation with it.

Solution

9 Recurcive Identification

Problem 38

Consider a system with time-varying parameters. Identification of the model is performed using recursive least-squares (RLS) estimation with forgetting factor λ

$$\begin{aligned}
\hat{\theta}(t) &= \hat{\theta}(t-1) + K(t)(y(t) - \varphi(t)^T\hat{\theta}(t-1)) \\
K(t) &= P(t)\varphi(t) \\
P(t) &= \frac{1}{\lambda}\left(P(t-1) - \frac{P(t-1)\varphi(t)\varphi^T(t)P(t-1)}{\lambda + \varphi^T(t)P(t-1)\varphi(t)}\right)
\end{aligned}$$

which asymptotically minimizes the loss function

$$V(\theta) = \frac{1}{N}\sum_{\tau=1}^{t} \lambda^{t-\tau}(y(\tau) - \varphi(\tau)^T\hat{\theta}(\tau-1))^2$$

Remark: we have θ on the left side, but $\hat{\theta}(\tau)$ in the sum to the right. If the parameters are time-invariant then $\hat{\theta}(\tau)$ conerge w.p 1 when t goes to infinity. In that case $\hat{\theta}(\tau)$ can be replaced by θ.

(a) What value should the forgetting factor have if data older than 100 samples should be weighted by less than 0.15 in the loss function?

(b) Assume now that the true parameters are time-invariant. Why does the algorithm only minimize the loss function *assymptotically*?

Solution

Problem 39

Consider an RLS algorithm with a sliding rectangular window in the following way. The parameter estimate is defined by

$$\begin{aligned}
\hat{\theta}(t) &= \arg\min_{\theta} \sum_{s=t-m+1}^{t} \varepsilon^2(s,\theta) \\
\varepsilon(s,\theta) &= y(s) - \varphi^T(s)\theta
\end{aligned}$$

The number m of prediction errors used in the criterion remains constant.

(a) Show that $\hat{\theta}(t)$ can be computed recursively as

$$\begin{aligned}
\hat{\theta}(t) &= \hat{\theta}(t-1) + K_1(t)\varepsilon(t,\hat{\theta}(t-1)) - K_2(t)\varepsilon(t-m,\hat{\theta}(t-1)) \\
(K_1(t)\ \ K_2(t)) &= P(t-1)(\varphi(t)\ \ \varphi(t-m)) \\
&\quad \cdot \left\{ I + \begin{pmatrix} \varphi^T(t) \\ -\varphi^T(t-m) \end{pmatrix} P(t-1)(\varphi(t)\ \ \varphi(t-m)) \right\}^{-1} \\
P(t) &= P(t-1) - (K_1(t)\ \ K_2(t)) \begin{pmatrix} \varphi^T(t) \\ -\varphi^T(t-m) \end{pmatrix} P(t-1)
\end{aligned}$$

(b) The identification algorithm can be used for real-time applications, since by its construction it has a finite memory. How much memory does it require? Compare with the memory necessary for the exponential window.

(c) When can one expect a rectangular window be a good choice? Use the properties of the estimators to indicate properties of a problem to which the estimator would be a good choice.

Remark: See Young (1984) for an alternative treatment of this problem.

Solution

10 Adaptive Control

Problem 40
A process has the following transfer function

$$G(s) = \frac{b}{s(s+1)}$$

where b is a time varying parameter. The system is controlled by a propotional regulator

$$u(t) = k(r(t) - y(t))$$

It is desirable to choose the feedback gain so that the closed-loop system has the transfer function

$$G_c(s) = \frac{1}{s^2 + s + 1}$$

Design a model-reference adaptive system that gives the desired result and investigate the system by simulation. Use the MIT-rule.

Solution

Problem 41
Consider an MRAS where the plant is

$$G(s) = \frac{1}{s}$$

and the reference model

$$G_r(s) = \frac{\theta^0}{s}$$

The control is given by

$$u(t) = \theta\, r(t)$$

and the parameter adaption by

$$\theta(t) = -\gamma_1 r(t)e(t) - \gamma_2 \int^t r(\tau)e(\tau)d\tau$$

Determine the differential equation for $e(t)$ and discuss how γ_1 and γ_2 influence the convergence rate. Assume $r(t)$ to be constant.

Except γ_1 and γ_2, the differential equation can depend on $r(t)$, and θ_0. Dependence on $\theta(t)$ should be eliminated. To "obtain more equations" you take the derivative of the expressions of $e(t)$ and $\theta(t)$.

Solution

Problem 42

Consider the system

$$G(s) = G_1(s)G_2(s)$$

where

$$\begin{aligned} G_1(s) &= \frac{b}{s+a} \\ G_2(s) &= \frac{q}{s+p} \end{aligned}$$

where a and b are unknown parameters and q and p are known. Discuss how to make an MRAS based on the gradient approach. Let the desired model be described by

$$G_m(s) = \frac{\omega^2}{s^2 + 2\zeta\,\omega\,s + \omega^2}$$

Assume a general control structure described by three polynomials $J(s)$, $K(s)$, and $R(s)$,

$$J\,U(s) = -K\,Y(s) + T\,R(s)$$

Decide necessary degree of the polynomials, and find adaption law using the MIT-rule.

Solution

Problem 43

Consider the process

$$G(s) = \frac{1}{s(s+a)}$$

where a is an unknown parameter. Determine a controller that can give the closed-loop system

$$G_m(s) = \frac{\omega^2}{s^2 + 2\zeta\omega\,s + \omega^2}$$

Determine model-reference adaptive controllers based on gradient (and stability theory, respectively.)

Solution

Solutions

1 Back to problem

(a) We have:

$$y(t) = -\theta_0 y(t-1) + e(t)$$

We will disregard the minus sign as it can be included in θ_0.

Now with

$$\varphi(t) = y(t-1)$$

this can be written as linear regression

$$y(t) = \varphi(t)^T \theta_0 + e(t)$$

and we can apply a least square estimate. Define

$$\Phi = \begin{bmatrix} y(1) \\ y(2) \\ \vdots \\ y(N-1) \end{bmatrix}$$

and

$$Y = \begin{bmatrix} y(2) \\ y(3) \\ \vdots \\ y(N) \end{bmatrix}$$

We can calculate the estimate of θ_0 using the formula for least squares regression, (4.7 in S&S)*

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

(b) The prediction error is the difference between the predicted and true value,

$$e(t) = y(t) - \varphi(t)^T \theta_0$$

So we estimate it as,

$$\varepsilon(t) = y(t) - \varphi(t)^T \hat{\theta}$$

An estimator for the variance of $e(t)$ is given by (4.13) in S&S*

$$\mathrm{Var}(\varepsilon(t)) = s^2 = 2V(\hat{\theta})/(N-n)$$

Here N is the number of data, n is the number of parameters, and $V(\theta)$ is the least square loss function, ie,

$$V(\theta) = \frac{1}{2}\sum_{t=1}^{N}(y(t) - \varphi(t)^T\theta)^2$$

So what we use is,

$$V(\hat{\theta}) = \frac{1}{2}\sum_{t=1}^{N}(y(t) - \varphi(t)^T\hat{\theta})^2$$

To reflect on, why do we have $N-n$ in the denominator and not N? For larger N the difference between these alternatives decrease, hence, for large N, does it matter if we divide by $N-n$ or by N?

(c) The parameter uncertainty can be expressed using (4.12) in S&S*,

$$\mathrm{Cov}(\hat{\theta}) = \lambda^2(\Phi^T\Phi)^{-1}$$

Where $\lambda^2 = \mathrm{Var}(e(t))$, so we can use our estimate of it from b).

(d) Approach

(1) Generate 1000 sets of 100 data points.

(2) Use each set to estimate $\left\{\hat{\theta}_i\right\}_{i=1}^{1000}$.

(3) From this you can plot a histogram of $\hat{\theta}_i$ values and confirm that it follows the normal distribution.

Alternatively, you can use the 1000 estimates of $\hat{\theta}$ to estimate the mean and covariance of $\hat{\theta}$,

(a) $\bar{\theta} = \mathrm{E}\,(\hat{\theta}) = \frac{1}{1000}\sum_{k=1}^{1000}\hat{\theta}_i$

(b) $\mathrm{Var}(\hat{\theta}) = \frac{1}{1000}\sum_{k=1}^{1000}(\hat{\theta}_i - \bar{\theta})^2$ and verify that this variance is the same as the one in c). It will not be exactly the same although, since you estimate the variance using 1000 samples.

* For the expresions (4.12) and (4.13) in S& S are to hold, φ should be determinsitic (Remark 1 p. 66), however, if $N \to \infty$ then the expresion also hold for stochastically generated φ. This is described in Example 7.6 p. 207.

2 Back to problem

(a) Since $\hat{\theta}_N$ is asymptotically normal distributed, and $P_\theta^{-1/2}\hat{\theta}_N$ is a linear transform of it, so it is also asymptotically normal distributed.

$$\mathrm{E}\,[P^{-1/2}\hat{\theta}_N] = P_\theta^{-1/2}\mathrm{E}\,[\hat{\theta}_N] = P_\theta^{-1/2}\theta_0$$

$$\begin{aligned}\mathrm{Var}(P_\theta^{-1/2}\hat{\theta}_N) = \mathrm{E}\,[(P_\theta^{-1/2}\hat{\theta}_N - P_\theta^{-1/2}\theta_0)(P_\theta^{-1/2}\hat{\theta}_N - P_\theta^{-1/2}\theta_0)^T] = \\ = P_\theta^{-1/2}\mathrm{Var}(\hat{\theta}_N)P_\theta^{-1/2} = P_\theta^{-1/2}P_\theta P_\theta^{-1/2} = I\end{aligned}$$

(b) Introduce $\tilde{\theta} = P_\theta^{-1/2}(\hat{\theta}_N - \theta_0)$, a vector of d independent asymptotically normal distributed variables with mean 0 and variance 1 according to (a), ie assymptotically in N, $\tilde{\theta}^T\tilde{\theta} \in \chi^2(d)$

(c) If θ is scalar, this is straight forward, we can generate bounds on $\hat{f}(x)$ by simply adding the uncertainties of θ

$$f(x, \theta \pm \Delta\theta)$$

where

$$\Delta\theta = \sqrt{P_\theta}\sqrt{\chi^2_\alpha(1)}.$$

However with more dimensions this is no longer possible since it would require evaluating f on a hyper-surface of dimension $d-1$. Instead, we can explore f in a number of random directions, in the following way. $\Delta\theta_k$ is chosen randomly and scaled so that

$$\|\Delta\theta_k\|^2 = \chi^2(d)$$

So using this and the known covariance of θ we can generate a new parameter vector:

$$\theta_k = \theta + P^{1/2}\Delta\theta_k$$

From this, one obtains an estimate of the confidence interval with limits

$$\max_k\, f(x, \theta_k)$$

and

$$\min_k\, f(x, \theta_k)$$

(d) One dimensional case: there are only to possible samples to take, one sample at each side of the Gaussian. Hence, in the one dimensional case you do not need to take more than two samples.

Two dimensional case: Level curves of the Gaussian form elipsiods. Samples are taken on that elipsoid so that the Gaussian having in total α probability mass outside the elipsoid.

3 Back to problem

We have

$$\hat{G}_N(e^{i\omega}) = \frac{\hat{\phi}_{yu}(\omega)}{\hat{\phi}_{uu}(\omega)}$$

Consider the assymtotic case when the number of data $N \to \infty$ and the width of the frequency window goes to zero, $\gamma \to \infty$.

This means that $\hat{\phi}_{yu}(\omega) \to \phi_{yu}(\omega)$ and $\hat{\phi}_{uu}(\omega) \to \phi_{uu}(\omega)$. So we have:

$$\hat{G}_N(e^{i\omega}) \to \frac{\phi_{yu}(\omega)}{\phi_{uu}(\omega)} \tag{1}$$

Hence, the we need to find the assymptotic estimates of $\phi_{yu}(\omega)$ and $\phi_{uu}(\omega)$ as functions of $\phi_{yy}(\omega)$ and $\phi_{vv}(\omega)$, the external signals. Now,

$$u(t) = -F(q)y(t) + r(t)\text{and } \ y(t) = G_0(q)u(t) + v(t) \tag{2}$$

which gives

$$u(t) = -F(G_0(q)u(t) + v(t)) + r(t) \ \ \Leftrightarrow \ \ (1 + FG_0)u(t) = -F(t)v(t) + r(t) \tag{3}$$

Correlating (3) with $u(t)$ and transforming gives

$$(1 + FG_0)\phi_{uu} = -F\phi_{vu} + \phi_{ru} \tag{4}$$

Correlating (3) with $v(t)$ and transforming gives

$$(1 + FG_0)\phi_{uv} = -F\phi_{vv} + \phi_{rv}$$

equivalent to

$$\phi_{uv} = \frac{-F(e^{i\omega})}{(1 + F(e^{i\omega})G_0(e^{i\omega}))}\phi_{vv}$$

Changing the order of the signals gives the complex conjugate

$$\phi_{vu} = \frac{-F(e^{-i\omega})}{(1 + F(e^{-i\omega})G_0(e^{-i\omega}))}\phi_{vv}$$

Now, correlate (3) with $r(t)$

$$(1 + FG_0)u(t)r(t) = -Fv(t)r(t) + r(t)r(t)$$

transform

$$(1 + FG_0)\phi_{ur} = -F\phi_{vr} + \phi_{rr} \Leftrightarrow \phi_{ur} = \frac{1}{(1 + F(e^{i\omega})G_0(e^{i\omega}))}\phi_{rr}$$

Inserting in the spectra into (4) gives

$$(1+F(e^{i\omega})G_0(e^{i\omega}))\phi_{uu} = -F(t)\frac{-F(e^{-i\omega})}{(1+F(e^{-i\omega})G_0(e^{-i\omega}))}\phi_{vv}+\frac{1}{(1+F(e^{-i\omega})G_0(e^{-i\omega}))}\phi_{rr}$$

from which we get

$$\phi_{uu} = \frac{|F(e^{i\omega})|^2\phi_{vv}+\phi_{rr}}{(1+F(e^{-i\omega})G_0(e^{-i\omega}))(1+F(e^{i\omega})G_0(e^{i\omega}))}$$

Correlating (2) with $u(t)$ and transforming

$$\phi_{yu} = G_0(e^{i\omega})\phi_{uu}+\phi_{vu}$$

We can now form the estimate (1)

$$\begin{aligned}
\hat{G}_N(e^{i\omega}) &= \frac{\phi_{yu}(\omega)}{\phi_{uu}(\omega)} = \\
&= \frac{G_0(e^{i\omega})\frac{|F(e^{i\omega})|^2\phi_{vv}+\phi_{rr}}{(1+F(e^{-i\omega})G_0(e^{-i\omega}))(1+F(e^{i\omega})G_0(e^{i\omega}))}+\frac{-F(e^{-i\omega})}{(1+F(e^{-i\omega})G_0(e^{-i\omega}))}\phi_{vv}}{\frac{|F(e^{i\omega})|^2\phi_{vv}+\phi_{rr}}{(1+F(e^{-i\omega})G_0(e^{-i\omega}))(1+F(e^{i\omega})G_0(e^{i\omega}))})} = \\
&= \frac{G_0(e^{i\omega})(|F(e^{i\omega})|^2\phi_{vv}+\phi_{rr})-F(e^{-i\omega})(1+F(e^{i\omega})G_0(e^{i\omega}))\phi_{vv}}{|F(e^{i\omega})|^2\phi_{vv}+\phi_{rr}} = \\
&= \frac{G_0(e^{i\omega})\phi_{rr}-F(e^{-i\omega})\phi_{vv}}{|F(e^{i\omega})|^2\phi_{vv}+\phi_{rr}}
\end{aligned}$$

In the case if $\phi_{rr}=0$, that is, there is no excitation of the system by the reference signal,

$$\hat{G}_N(e^{i\omega}) = \frac{-F(e^{-i\omega})\phi_{vv}}{|F(e^{i\omega})|^2\phi_{vv}} = \frac{-F(e^{-i\omega})}{F(e^{-i\omega})F(e^{i\omega})} = -\frac{1}{F(e^{i\omega})}$$

Instead of obtaining a model of the plant, we get the inverse of the controller. This is due to that the direction of time is lost when the correlation is estimated. Both the plant and the feedback are causal transfer functions. The causality is indicated by the arrows in the control loop. The inverse of the feeback is hence the dependence of the input on the output "backward" in time since it goes against the direction of dependens in the control loop.

Ff $F=0$, ie the system is open-loop, there is no feedaback,

$$\hat{G}_N(e^{i\omega}) = \frac{G_0(e^{i\omega})\phi_{rr}}{\phi_{rr}} = G_0(e^{i\omega})$$

So without feedback we get the correct estimate.

We conclude that if we have no excitation of the reference signal then we obtain no information about the plant in the spectral estimate. If the energy of the feedback is low compared to the energy in the reference signal, then the estimate can be useful, although there is feedback giving a bias in the estimate.

4 Back to problem

(a) We have

$$y(t) = A\sin(\omega t + \phi) + A_\nu \sin(\omega_\nu t)$$

How will the second term affect the estimate of the Bode plot?

For s_T

$$s_T = \int_0^{2\pi n/\omega} [A\sin(\omega t + \phi)\sin(\omega t) + A_\nu \sin(\omega_\nu t)\sin(\omega t)]dt$$

Second term

$$A_\nu \int_0^{2\pi n/\omega} \sin(\omega_\nu t)\sin(\omega t)dt$$

if $\omega_\nu \neq \omega$ becomes

$$= \frac{1}{2}A_\nu \int_0^{2\pi n/\omega} \cos((\omega_\nu - \omega)t) - \cos((\omega_\nu + \omega)t)dt =$$

$$= \frac{1}{2}A_\nu \left[\frac{\sin((\omega_\nu - \omega)t)}{(\omega_\nu - \omega)} - \frac{\sin((\omega_\nu + \omega)t)}{(\omega_\nu + \omega)}\right]_0^{2\pi n/\omega}$$

$$= \frac{1}{2}A_\nu (\frac{\sin(-2\pi n + \frac{\omega_\nu}{\omega}2\pi n)}{(\omega_\nu - \omega)} - \frac{\sin(2\pi n + \frac{\omega_\nu}{\omega}2\pi n)}{(\omega_\nu + \omega)})$$

$$= \frac{1}{2}A_\nu \sin(\frac{\omega_\nu}{\omega}2\pi n)(\frac{1}{(\omega_\nu - \omega)} - \frac{1}{(\omega_\nu + \omega)}) = \frac{A_\nu \omega \sin(\frac{\omega_\nu}{\omega}2\pi n)}{\omega_\nu^2 - \omega^2}$$

Now if $\omega_\nu = \omega$ we have

$$A_\nu \int_0^{2\pi n/\omega} \sin(\omega t)^2 dt = A_\nu \left[\frac{t}{2} - \frac{1}{4\omega}\sin(2\omega t)\right]_0^{2\pi n/\omega} = \frac{A_\nu T}{2}$$

Similarily for $c_T(\omega)$:

$$c_T = \int_0^{2\pi n/\omega} [A\sin(\omega t + \phi)\cos(\omega t) + A_\nu \sin(\omega_\nu t)\cos(\omega t)]dt$$

The second term becomes, assuming $\omega \neq \omega_\nu$

$$A_\nu \int_0^{2\pi n/\omega} \sin(\omega_\nu t)\cos(\omega t)dt = A_\nu \left[-\frac{\cos((\omega_\nu - \omega)t)}{2(\omega_\nu - \omega)} - \frac{\cos((\omega_\nu + \omega)t)}{2(\omega_\nu + \omega)} \right]_0^{2\pi n/\omega} =$$
$$= -\frac{1}{2}A_\nu \cos(\frac{\omega_\nu}{\omega}2\pi n)(\frac{1}{(\omega_\nu - \omega)} + \frac{1}{(\omega_\nu + \omega)}) = -\frac{A_\nu \omega_\nu \cos(\frac{\omega_\nu}{\omega}2\pi n)}{\omega_\nu^2 - \omega^2}$$

Finally if:

$$\omega_\nu = \omega$$

Then

$$A_\nu \int_0^{2\pi n/\omega} \sin(\omega t)\cos(\omega t) = \frac{A_\nu}{\omega}\sin^2(2\pi n) = 0$$

So from this follows, for $\omega_\nu \neq \omega$:

$$s_T(\omega) = \frac{1}{2}T|G(i\omega)|u_1 \cos(\phi(\omega)) + \frac{A_\nu \omega \sin(\frac{\omega_\nu}{\omega}2\pi n)}{\omega_\nu^2 - \omega^2}$$

$$c_T(\omega) = \frac{1}{2}T|G(i\omega)|u_1 \sin(\phi(\omega)) - \frac{A_\nu \omega_\nu \cos(\frac{\omega_\nu}{\omega}2\pi n)}{\omega_\nu^2 - \omega^2}$$

We see that the term from the noise does not depend on T so for sufficiently large T the noise term will become relatively insignificant. However if $\omega = \omega_\nu$ then

$$s_T(\omega) = \frac{1}{2}T|G(i\omega)|u_1 \cos(\phi(\omega)) + \frac{A_\nu T}{2}$$

So the noise term will also increase with T. This means that the influence of the disturbance will not be smaller when more measurments are collected.

(b) For periodic input signals, and if the noise is broadbanded, then the frequency responce metod gives unbiased estimates and the variance decrease as $1/N$, where N is the number of data. Hence, to half the variance, the observation length need to be doubled, ie, $2\,T$.

(c) For stability it is important to know the frequency function around the intended bandwidth of the controled system. Hence, given design goals on the bandwidth, use most of the time for experimenting near that frequency.

5 Back to problem The solution can be obtained in the following way: The input can be desribed as a sum of two steps, a positive at $t = 0$ and a negative at $t = T$. The output is a convolution of the unknown impulse response and the steps. Formulate the convolution for $t = 0, 1, \ldots$, each one giving one equation when set equal to observed value $y(t)$. The impulse responce is the solution to these equations.

6 Back to problem

(a) The second data set, from the second period, has correct initial conditions since it is the second period of a periodic function. The first data set has a transient due to that the inital condition is incorrect. In The plot we see one estimate, dashed - redish, being smooth, less disturbed, and the other one seems to be "noisy". Hence, it is the dashed estimate which comes from the second set.

(b) The first ETFE seems to have less variance as the number of data increase, which indicates non-periodic disturbance. This is clearer at high frequencies where the signal energy is lower. It isnot be absolutely sure that the second ETFE has a periodic disturbancy from looking at its ETFE. But by comparing with the first ETFE it is very likely that there is a sinosoidal disturbance approximately at frequency 0.3 rad.

7 Back to problem

(a) We have:

$$y_k + a\, y_{k-1} = b\, u_{k-1} + w_k + c\, w_{k-1}$$

which can re-writen as

$$y_k = -a\, y_{k-1} + b\, u_{k-1} + w_k + c\, w_{k-1}$$

and use the model

$$y_k + \hat{a}\, y_{k-1} = \hat{b}\, u_k$$

We want to find the $\hat{a}_k$ and $\hat{b}_k$ that minimize:

$$\begin{aligned} \mathrm{E}\,(y_k - \hat{y}_k)^2 &= \mathrm{E}\,(y_k + \hat{a}\, y_{k-1} - \hat{b}\, u_{k-1})^2 = \\ &= \mathrm{E}\,(y_k^2 + 2\hat{a}\, y_k y_{k-1} - 2\hat{b}\, y_k u_{k-1} + \hat{a}^2\, y_{k-1}^2 - 2\hat{a}\, \hat{b}\, y_{k-1} u_{k-1} + \hat{b}^2\, u_{k-1}^2) \end{aligned} \quad (1)$$

Introduce

$$\mathrm{E}\,(y_k^2) = r_0$$

We need to express (1) in terms of σ^2 and r_0. Start with

$$\begin{aligned}\mathrm{E}\,(y_k y_{k-1}) &= \mathrm{E}\,(-a\,y_{k-1}y_{k-1} + b\,u_k y_{k-1} + w_k y_{k-1} + c\,w_{k-1}y_{k-1}) = \\ &\mathrm{E}\,(-a\,y_{k-1}y_{k-1}) + \mathrm{E}\,(c\,w_{k-1}y_{k-1}) = -a\,r_0 + \mathrm{E}\,(c\,w_{k-1}y_{k-1}) \quad (2)\end{aligned}$$

and

$$\mathrm{E}\,(w_{k-1}y_{k-1}) = \mathrm{E}\,(w_{k-1}(-a\,y_{k-2} + b\,u_{k-2} + w_{k-1} + c\,w_{k-2})) = \sigma^2$$

Inserting in (2) gives

$$\mathrm{E}\,(y_k y_{k-1}) = -a\,r_0 + c\,\sigma^2$$

Next term

$$\mathrm{E}\,(y_k u_{k-1}) = \mathrm{E}\,(-a\,y_{k-1}u_{k-1} + b\,u_{k-1}u_{k-1} + w_k u_{k-1} + c\,w_{k-1}u_{k-1}) = b\sigma^2 \quad (3)$$

Next

$$\mathrm{E}\,(y_{k-1}u_{k-1}) = 0$$

Now, inserting in (1)

$$\mathrm{E}\,(y_k - \hat{y}_k)^2 = r_0(1 - 2\hat{a}\,a + \hat{a}^2) + 2\hat{a}\,c\,\sigma^2 - 2\hat{b}\,b\,\sigma^2 + \hat{b}^2\sigma^2 \quad (4)$$

Minimize for $\hat{b}$,

$$0 = -2b\,\sigma^2 + 2\hat{b}\,\sigma^2 \quad \Rightarrow \quad b = \hat{b}$$

and for $\hat{a}$,

$$\begin{aligned}&r_0(1 - 2\hat{a}\,a + \hat{a}^2) + 2\hat{a}\,c\,\sigma^2 - 2\hat{b}\,b\,\sigma^2 + \hat{b}^2\sigma^2 = \\ &\qquad - 2r_0 a + 2r_0\hat{a} + 2c\,\sigma^2 = 0 \quad \Rightarrow \quad \hat{a} = a - \frac{c\,\sigma^2}{r_0}\end{aligned}$$

(b) Take (4) for $\hat{a} = a$ and $\hat{b} = b$ minus (4) for the terms minimizing (4),

$$\begin{aligned}&r_0(1 - 2a\,a + a^2) + 2a\,c\,\sigma^2 - 2b\,b\,\sigma^2 + b^2\sigma^2 \\ &\quad - r_0(1 - 2(a - \frac{c\,\sigma^2}{r_0})\,a + (a - \frac{c\,\sigma^2}{r_0})^2) - 2(a - \frac{c\,\sigma^2}{r_0})\,c\,\sigma^2 + 2b\,b\,\sigma^2 - b^2\sigma^2 = \\ &\qquad - r_0(\frac{2a\,c\,\sigma^2}{r_0} + \frac{c^2\,\sigma^4}{r_0^2} - \frac{2a\,c\,\sigma^2}{r_0}) + \frac{2c^2\sigma^4}{r_0} = \frac{c^2\sigma^4}{r_0} > 0\end{aligned}$$

(c)

$$r_0 = \mathrm{E}\, y_k y_k = \mathrm{E}\,(-a\, y_{k-1} + b\, u_{k-1} + w_k + c\, w_{k-1}) =$$
$$a^2\, \mathrm{E}\, y_{k-1}^2 + b^2\, \mathrm{E}\, u_{k-1}^2 + \mathrm{E}\, w_k^2 + c^2\, \mathrm{E}\, w_{k-1}^2 - 2a\, c \mathrm{E}\, y_k w_{k-1} =$$
$$a^2\, r_0 + b^2\, \sigma^2 + \sigma^2 + c^2\, \sigma^2 - 2a\, c\, \sigma^2 \quad \Rightarrow$$

$$r_0 = \sigma^2 \frac{b^2 + c^2 + 1 - 2a\, c}{1 - a^2}$$

8 Back to problem

(a) The estimate of K and τ are the values minimizing the sum of squared errors:

$$\sum_{t=0}^{4} (h(0.2t) - K\, e^{-0.2t/\tau})^2 =$$
$$(3.4 - K)^2 + (2.3 - K\, e^{-0.2/\tau})^2 + (1.7 - K\, e^{-0.4/\tau})^2 + (1.2 - K\, e^{-0.6/\tau})^2 + (0.9 - K\, e^{-0.8/\tau})^2$$

Taking the derivative with respect of K and τ and solving for when they are zero gives the estimate. This then gives $K = 3.358$, $\tau] = 0.583$.

(b) We have from the problem

$$Y = X\theta + E$$

where $\theta = [\log K\ 1/\tau]^T$ and

$$X = \begin{bmatrix} 1 & 0 \\ 1 & -0.2 \\ 1 & -0.4 \\ 1 & -0.6 \\ 1 & -0.8 \end{bmatrix}$$

and

$$Y = \log(h) = \begin{bmatrix} 1.22 \\ 0.83 \\ 0.53 \\ 0.18 \\ -0.11 \end{bmatrix}$$

The least-squares estimate becomes

$$\hat{\theta} = (X^T X)^{-1} X^T Y = \begin{bmatrix} 1.19 \\ 1.65 \end{bmatrix}$$

$$\Rightarrow \quad K = e^{1.19} = 3.3, \quad \tau = 1/1.65 = 0.606$$

(c) Minimizing the squared error using the original parameterization, as in (a) gives a minimization problem that is not quadratic in the parameters and, hence, it is computational more complicated. Re-parameterization, as in (b), converts the model to becoming linear in the parameters and the criterion is then quadratic (linear regression) so it is computationally easier. These are the computational apects. The criterion which is minimized is also changed by the transform so the minimum, ie, the estimate is not the same. Fot the transformed problem, the data are assumed to be generatyed as

$$\log y(t) = \log K - t/\tau + e(t)$$

which can be re-written as

$$y(t) = K\, e^{-t/\tau} e^{e(t)}$$

Compared to the original problem where the assumption is

$$y(t) = K\, e^{-t/\tau} + e(t)$$

From this we see that there is no difference in the noise free case, $e(t) = 0$ but as soon as the model cannot fit the data exactly, ie, there is noise, then the estimate will be different. Also, the uncertainty of the parameter estimate will be in different in the two cases.

9 Back to problem

(a)

$$\hat{\theta}-\theta_0 = (\Phi_N^T \Phi_N)^{-1}\Phi_N^T Y_N - \theta_0 = (\Phi_N^T \Phi_N)^{-1}\Phi_N^T(\Phi_N\theta_0 + E_N) - \theta_0 = (\Phi_N^T \Phi_N)^{-1}\Phi_N^T E_N$$

(b) The bias error is the expectation (with respect to the noise, E_N) of the total error, ie,

$$\mathrm{E}[(\Phi_N^T \Phi_N)^{-1}\Phi_N^T E_N]$$

and the variance part is then the difference between the total error and the bias error,

$$(\Phi_N^T \Phi_N)^{-1}\Phi_N^T E_N - \mathrm{E}[(\Phi_N^T \Phi_N)^{-1}\Phi_N^T E_N]$$

Note that the variance error depends on the particular realization of the error and it will hence, be different each time the estimate is calculated on *new* data, meanwhile the bias error does not change.

10 Back to problem

The LS solution is

11 Back to problem

(a) We have the system

$$y(t) = b_1 u(t-1) + ... + b_m u(t-m) + v(t)$$

and the model

$$\hat{y}(t) = \hat{b}_1 u(t-1) + ... + \hat{b}_m u(t-m) = \theta^T \varphi(t)$$

where $\theta = [b_1, \dots, b_m]^T$ and $\varphi(t) = [u(t-1), \dots, u(t-m)]^T$.
The least squares estimate is

$$\hat{\theta} = [\frac{1}{N}\sum_{t=1}^{N}\varphi(t)\varphi^T(t)]^{-1}\frac{1}{N}\sum_{t=1}^{N}\varphi(t)y(t)$$

(b) The LS estimate in (a) can be expressed as

$$\begin{aligned}\hat{\theta} = [\frac{1}{N}\sum_{t=1}^{N}\varphi(t)\varphi^T(t)]^{-1}\frac{1}{N}\sum_{t=1}^{N}\varphi(t)y(t) &= [\frac{1}{N}\sum_{t=1}^{N}\varphi(t)\varphi^T(t)]^{-1}\frac{1}{N}\sum_{t=1}^{N}\varphi(t)(\varphi^T(t)\theta + v(t)) \\ &= \theta + [\frac{1}{N}\sum_{t=1}^{N}\varphi(t)\varphi^T(t)]^{-1}\frac{1}{N}\sum_{t=1}^{N}\varphi(t)v(t)\end{aligned}$$

and hence

$$\hat{\theta} - \theta = [\frac{1}{N}\sum_{t=1}^{N}\varphi(t)\varphi^T(t)]^{-1}\frac{1}{N}\sum_{t=1}^{N}\varphi(t)v(t)$$

This expression goes to zero (assuming $u(t)$ is exciting the system) if

$$\frac{1}{N}\sum_{t=1}^{N}\varphi(t)v(t) \to 0 \text{ when } N \to \infty$$

This means that $v(t)$ must be uncorrelated with the components of $\varphi(t)$ for the estimate to be consistent, and this is independent on if $v(t)$ is white or colored.

- To have a consistent estimate in the case $v(t)$ is white, $u(t)$ cannot depend on future $v(t)$ but it can be correlated with past $v(t)$. That is, feedback is possible.
- In the case $v(t)$ is colored, then it must be uncorrelated with $u(t)$ since a colored $v(t)$ cannot be correlated with *only* future $u(t)$. That is, it must be an open-loop experiment.

12 Back to problem

(a) For a general linear system

$$y(t) = G(q)u(t) + H(q)e(t)$$

we have the one step ahead predictor

$$\hat{y}(t|t-1) = H^{-1}(q)G(q)u(t) + (1 - H^{-1}(q))y(t)$$

In our case

$$G(q) = \frac{B(q)}{A(q)} = \frac{b_2 q^{-2}}{1 - a_1 q^{-1} - a_2 q^{-2}}, \quad H(q) = \frac{1}{A(q)} = \frac{1}{1 - a_1 q^{-1} - a_2 q^{-2}}$$

So the 1-step prediction as transfer function expression becomes

$$\hat{y}(t|t-1) = B(q)u(t) + (1 - A(q))y(t) = b_2 q^{-2} u(t) + (a_1 q^{-1} + a_2 q^{-2})y(t)$$

and in equations it becomes

$$\begin{aligned} \hat{y}(t|t-1) = b_2 q^{-2} u(t) + (1 - (1 - a_1 q^{-1} - a_2 q^{-2}))y(t) \\ = \hat{y}(t|t-1) = b_2 u(t-2) + a_1 y(t-1) + a_2 y(t-2) \end{aligned}$$

(b) We have generally for k-step predictor

$$\hat{y}(t|t-k) = \bar{H}_k(q)H(q)^{-1}G(q)u(t) + (1 - \bar{H}_k(q)H(q)^{-1})y(t)$$

where $\bar{H}_k(q)$ is the k first terms of the impulse response. So for a two step predictor

$$\hat{y}(t|t-2) = \bar{H}_2(q)H(q)^{-1}G(q)u(t) + (1 - \bar{H}_2(q)H(q)^{-1})y(t) \tag{1}$$

To obtain $\bar{H}_2(q)$ we expand $\frac{1}{A(q)}$ using the geometric sum

$$\frac{1}{1+x} = \sum_{i=0}^{\infty} (-x)^i$$

Hence, we obtain

$$\begin{aligned} H(q) = \frac{1}{1 - a_1 q^{-1} - a_2 q^{-2}} = \sum_{i=0}^{\infty} (a_1 q^{-1} + a_2 q^{-2})^i = \\ 1 + a_1 q^{-1} + a_2 q^{-2} + \sum_{i=2}^{\infty} (a_1 q^{-1} + a_2 q^{-2})^i \end{aligned}$$

and identify

$$\bar{H}_2 = 1 + a_1 q^{-1}$$

So the 2-step predictor as transfer function becomes, inserting in (1),

$$\hat{y}(t|t-2) = \bar{H}_2(q)H(q)^{-1}G(q)u(t) + (1 - \bar{H}_2(q)H(q)^{-1})y(t) =$$
$$(1 + a_1q^{-1})b_2q^{-2}u(t) + (1 - (1 + a_1q^{-1})(1 - a_1q^{-1} - a_2q^{-2}))y(t) =$$
$$(b_2q^{-2} + b_2a_1q^{-3})u(t) + ((a_2 + a_1^2)q^{-2} + a_1a_2q^{-3})y(t)$$

or written out as equations

$$\hat{y}(t|t-2) = b_2u(t-2) + b_2a_1u(t-3) + (a_2 + a_1^2)y(t-2) + a_1a_2y(t-3)$$

(c) (Reference, S& S p 232.) From the one-step predictor one have the We can define

$$\hat{y}(t|t+n) = y(t) \qquad (2))$$

for $n \leq 0$ since $y(t)$ is available at time t and also later times. Then, using the process definition,

$$y(t) = b_2u(t-2) + a_1y(t-1) + a_2y(t-2) + e(t)$$

and linearity of the expectation, the k-step prediction can be expressed as

$$\hat{y}(t|t-k) = \mathrm{E}_{Y_{t-k}}y(t) = b_2u(t-2) + a_1\mathrm{E}_{Y_{t-k}}y(t-1) + a_2\mathrm{E}_{Y_{t-k}}y(t-2) =$$
$$b_2u(t-2) + a_1\hat{y}(t-1|t-k) + a_2\hat{y}(t-2|t-k)$$

This can be interpreted as we on the right side use "best available estimate" as replacement for the true values $y(t-1)$ and $y(t-2)$, respectively. To obtain the m step predictor, the equation needs to be iterated for $k = 1,\ 2, \ldots,\ k$.

Let us "test" this predictor for some k and compare with the earlier result.

$k = 1$

$$\hat{y}(t|t-1) = b_2u(t-2) + a_1\hat{y}(t-1|t-1) + a_2\hat{y}(t-2|t-1) =$$
$$b_2u(t-2) + a_1y(t-1) + a_2y(t-2) \qquad (3)$$

where (2)) was used.

$k = 2$

$$\hat{y}(t|t-2) = b_2u(t-2) + a_1\hat{y}(t-1|t-2) + a_2\hat{y}(t-2|t-2) =$$
$$b_2u(t-2) + a_1(b_2u(t-3) + a_1y(t-2) + a_2y(t-3)) + a_2y(t-2) =$$
$$b_2u(t-2) + b_2a_1u(t-3) + (a_2 + a_1^2)y(t-2) + a_1a_2y(t-3)$$

where (2)) and (3) were used. Comparing these prediction with the expression in (a) and (b) concludes that the result is the same. We also see that outputs up to time $t-2$ but not later are used.

13 Back to problem

(a) We write the model as

$$\hat{y}(t) = \frac{\hat{B}(q)}{\hat{A}(q)}u(t) + \frac{1}{\hat{A}(q)}e(t)$$

Note that $\hat{y}(t)$ is not a prediction since $e(t)$ is used. The true system is:

$$y(t) = \frac{B(q)}{A(q)}u(t) + \frac{C(q)}{A(q)}e(t)$$

The error is the difference

$$\begin{aligned} ||y(t) - \hat{y}(t)|| = ||\frac{B(q)}{A(q)}u(t) + \frac{C(q)}{A(q)}e(t) - \frac{\hat{B}(q)}{\hat{A}(q)}u(t) - \frac{1}{\hat{A}(q)}e(t)|| = \\ ||(\frac{B(q)}{A(q)} - \frac{\hat{B}(q)}{\hat{A}(q)})u(t) + (\frac{C(q)}{A(q)} - \frac{1}{\hat{A}(q)})e(t)|| = \\ ||(\frac{B(q)}{A(q)} - \frac{\hat{B}(q)}{\hat{A}(q)})u(t)|| + ||(\frac{C(q)}{A(q)} - \frac{1}{\hat{A}(q)})e(t)|| \end{aligned} \tag{1}$$

where the last step follows from that $u(t)$ and $e(t)$ are uncorrelated (independent). Also, the norm is the expectation of the absolute value.

(b) The parameters of $\hat{A}(q)$ are present in both terms, but since it is of high order, it contains parameters so that both terms can be, almost, zero. To see that, we factorize $\hat{A}(q)$ so that the first term becomes zero, ie, write

$$\hat{B}(q) = B(q)\tilde{A}(q)$$

$$\hat{A}(q) = A(q)\tilde{A}(q)$$

Insert this in (1)

$$\begin{aligned} ||(\frac{B(q)}{A(q)} - \frac{B(q)\tilde{A}(q)}{A(q)\tilde{A}(q)})u(t)|| + ||(\frac{C(q)}{A(q)} - \frac{1}{A(q)\tilde{A}(q)})e(t)|| = \\ ||(\frac{C(q)\tilde{A}(q) - 1}{A(q)\tilde{A}(q)})e(t)|| \end{aligned} \tag{2}$$

From this we see that if $\tilde{A}(q)$ can be chosen equal to $C^{-1}(q)$ then also the second term becomes zero. This is not possible, but we can make it close to zero. Assume $\tilde{A}(q)$ has degree n and let it be the n first terms of the expansion of $C^{-1}(q)$, ie,

$\tilde{A}(q) = C^{-1}(q) - \sum_{i=n+1}^{\infty}(-c)^i q^{-i}$ then, using that $e(t)$ is white, the numerator of (2) becomes

$$= ||(C(q)(C^{-1}(q) - \sum_{i=n+1}^{\infty}(-c)^i q^{-i}) - 1)e(t)|| =$$

$$||C(q)\sum_{i=n+1}^{\infty}(-c)^i q^{-i}e(t)|| \leq ||C(q)|| \cdot \sum_{i=n+1}^{\infty}||c^i q^{-i}e(t)|| \leq (1+|c|)\frac{|c|^n}{1-|c|}||e(t)|| \tag{3}$$

This choice of $\tilde{A}(q)$ will not minimize the criterion, but we see from (3) that it gives a smaller value of the criterion for growing n. Hence, for large n this $\tilde{A}(q)$ will be close to the minimum.

(c) Many model based control design methods gives a controler of the same degree as the model. On the same time, if not necessary on wants to avoid high-order controlers. Hence, this speaks against the high order ARX model and a low-order OE model can be better. An approach could be to apply model-order reduction on the high order ARX model before the controller is designed.

14 Back to problem

We have:

$$y(t) = a_1 y(t-1) + e(t)$$

Using the standardform

$$y(t) = G(q)u(t) + H(q)e(t)$$

we identify

$$G(q) = 0$$

and

$$H(q) = \frac{1}{(1 - a_1 q^{-1})} = \sum_{i=0}^{\infty}(a_1 q^{-1})^i$$

where we also Taylor expanded $H(q)$. This expansion can be divided for a k step predictor

$$H(q) = \sum_{i=0}^{k-1}(a_1 q^{-1})^i + \sum_{i=k}^{\infty}(a_1 q^{-1})^i$$

where first part is defined with

$$\bar{H}_k(q) = \sum_{i=0}^{k-1} (a_1 q^{-1})^i$$

We have for a k step predictor:

$$\hat{y}(t|t-k) = \bar{H}_k(q)H(q)^{-1}G(q)u(t) + (1 - \bar{H}_k(q)H(q)^{-1})y(t) =$$

$$(1 - (\sum_{i=0}^{k-1}(a_1q^{-1})^i(1 - a_1q^{-1})))y(t) = (1 - \sum_{i=0}^{k-1}(a_1q^{-1})^i + a_1q^{-1}\sum_{i=0}^{k-1}(a_1q^{-1})^i)y(t) =$$

$$(\sum_{i=1}^{k-1}(a_1q^{-1})^i + \sum_{i=1}^{k}(a_1q^{-1})^i)y(t) = (a_1q^{-1})^k y(t) = a_1^k y(t-k)$$

15 Back to problem

We have:

$$y(t) = c_1 e(t-1) + e(t)$$

Using the standardform

$$y(t) = G(q)u(t) + H(q)e(t)$$

we identify

$$H = (q^{-1}c_1 + 1)$$

and

$$G(q) = 0$$

To form the predictor we identify the k first terms in the impulse response of $H(q)$

$$\bar{H}_1(q) = 1$$

and

$$\bar{H}_{k\geq 2}(q) = 1 + q^{-1}c_1$$

So for a k step predictor:

$$\hat{y}(t|t-k) = \bar{H}_k(q)H(q)^{-1}G(q)u(t) + (1 - \bar{H}_k(q)H(q)^{-1})y(t) =$$

$$(1 - \bar{H}_k(q)H(q)^{-1})y(t) = (1 - \bar{H}_k(q)\frac{1}{(q^{-1}c_1 + 1)})y(t) = (1 - \bar{H}_k(q)(\frac{1}{(q^{-1}c_1 + 1)})y(t)$$

For $k = 1$:

$$\hat{y}(t|t-1) = (1 - \bar{H}_1(q)(\frac{1}{(q^{-1}c_1 + 1)})y(t) =$$

$$(1 - \bar{H}_1(q)(\sum_{i=0}^{\infty}(-c_1q^{-1})^i)y(t) = (1 - \sum_{i=0}^{\infty}(-c_1q^{-1})^i)y(t) = -\sum_{i=1}^{\infty}(-c_1q^{-1})^i y(t)$$

For $k > 1$

$$\hat{y}(t|t-k) = (1 - \bar{H}_{k\geq 2}(q)\frac{1}{q^{-1}c_1 + 1})y(t) = (1 - \frac{q^{-1}c_1 + 1}{q^{-1}c_1 + 1})y(t) = 0$$

16 Back to problem

17 Back to problem

We have:

$$v(t) = y(t) + a\,y(t-1) - b\,u(t-1) = y(t) - \theta\varphi(t)$$

where $\varphi^T(t) = [-y(t-1)\ u(t-1)]$ and $\theta = [a\ b]$. We have the probability function:

$$f(x) = \mu\, e^{-\mu\, x} \quad x \geq 0$$

So using Maximum Likelihood estimation with the data points, we get the definition of the estimate as

$$\arg\max_{\theta} \prod_{t=2}^{N} \mu\, e^{-\mu(y(t)-\theta\varphi(t))} = \mu^N e^{-\mu \sum_{t=2}^{N}(y(t)-\theta\varphi(t))}$$

Take the logarithm of this gives

$$\arg\max_{\theta} N \log \mu - \mu \sum_{t=2}^{N}(y(t) - \theta\,\varphi(t))$$

If μ is known we then want to maximize

$$\max_{\theta} -\mu \sum_{t=2}^{N}(y(t) - \theta, \varphi(t))$$

Subject to:

$$y(t) - \theta\,\varphi(t) \geq 0, \quad t = 2, \ldots, N$$

This can be simplified and expressed on matrix form

$$\max_{\theta} \theta \sum_{t=2}^{N} \varphi(t)$$

Subject to

$$\Phi\,\theta^T \leq Y$$

with

$$Y = \begin{bmatrix} y(2) \\ \vdots \\ y(N) \end{bmatrix}, \quad X = \begin{bmatrix} -y(1) & u(1) \\ \vdots & \vdots \\ -y(N-1) & u(N-1) \end{bmatrix}$$

If μ also needs to be estimated, the optimization problem becomes

$$\max_{\theta,\mu}\; N \log \mu - \mu \sum_{t=2}^{N} y(t) + \mu\,\theta \sum_{t=2}^{N} \varphi(t)$$

Subject to

$$\Phi\,\theta^T \leq Y, \quad \mu > 0$$

which is a much nastier problem.

18 Back to problem

(a) We have

$$y(t) = a_0 + b(t)e(t)$$

and

$$e(t) = \frac{y(t) - a_0}{b(t)}$$

Now, $e(t)$ is normally distributed (Gaussian) so it follows the distribution,

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}}$$

Put in our values for e(t)

$$f(\frac{y(t)-a_0}{b(t)}) = \frac{1}{\sqrt{2\pi\sigma_e^2}} e^{-\frac{(\frac{y(t)-a_0}{b(t)})^2}{2\sigma_e^2}}$$

We want the maximum probability for all N measurements of $y(t)$:

$$\max_\theta p(y|\theta) = \max_{a_0,\sigma_e} \prod_{t=1}^{N} \frac{1}{\sqrt{2\pi\sigma_e^2}} e^{-\frac{(\frac{y(t)-a_0}{b(t)})^2}{2\sigma_e^2}} = \max_{a_0,\sigma_e} (\frac{1}{\sqrt{2\pi\sigma_e^2}})^N e^{\sum_{t=1}^{N} -\frac{(y(t)-a_o)^2}{2b^2(t)\sigma_e^2}}$$

We take the logarithm of this

$$\max_\theta \log p(y|\theta) = \max_{a_0,\sigma_e} \log(\frac{1}{\sqrt{2\pi\sigma_e^2}})^N + \sum_{t=1}^{N} -\frac{(y(t)-a_o)^2}{2b^2(t)\sigma_e^2} =$$

$$\max_{a_0,\sigma_e} N\log(\frac{1}{\sqrt{2\pi}}) - \frac{N}{2}\log(\sigma_e^2) - \sum_{i=1}^{N} \frac{(y(t)-a_o)^2}{2b^2(t)\sigma_e^2}$$

Take the derivative with respect to $a_{0,}\sigma_e^2$: For a_0:

$$\frac{d\ \log p(y|\theta)}{da_0} = 2\sum_{t=1}^{N} \frac{(y(t)-a_o)}{b^2(t)\sigma_e^2} = 0 \quad \Leftrightarrow$$

$$\sum_{t=1}^{N} \frac{y(t)}{b^2(t)} = \sum_{t=1}^{N} \frac{a_o}{b^2(t)} \quad \Leftrightarrow$$

$$a_0 = \frac{\sum_{t=1}^{N} \frac{y(t)}{b^2(t)}}{\sum_{t=1}^{N} \frac{1}{b^2(t)}}$$

For σ_e^2 (derivative with respect to σ_e^2)

$$\frac{d\ \log p(y|\theta)}{d\sigma_e^2} = -\frac{N}{2\sigma_e^2} + \sum_{t=1}^{N} \frac{(y(t)-a_o)^2}{2b^2(t)\sigma_e^4} = 0 \quad \Leftrightarrow$$

$$N\sigma_e^2 = \sum_{t=1}^{N} \frac{(y(t)-a_o)^2}{b^2(t)} \quad \Leftrightarrow$$

$$\sigma_e^2 = \frac{1}{N}\sum_{t=1}^{N} \frac{(y(t)-a_o)^2}{b^2(t)}$$

(b) The ML estimate is efficient (S& S p 210), so the asymptotic estimate of the variance converges to the Cramér-Rao lower bound. That is,

$$\operatorname{Cov}\hat{\theta} = M^{-1}$$

where M, is the fisher information matrix, so for our case:

$$\theta = \begin{bmatrix} a_0 \\ \sigma_e^2 \end{bmatrix}$$

and

$$M = -E \begin{bmatrix} \frac{d^2 \log p(y|\theta)}{da_0^2} & \frac{d^2 \log p(y|\theta)}{da_0 d\sigma_e^2} \\ \frac{d^2 \log p(y|\theta)}{d\sigma_e^2 da_0} & \frac{d^2 \log p(y|\theta)}{d(\sigma_e^2)^2} \end{bmatrix}$$

From before

$$\frac{d \log p(y|\theta)}{da_0} = 2\sum_{t=1}^{N} \frac{(y(t) - a_o)}{2b^2(t)\sigma_e^2}$$

Taking the derivative once more gives

$$\frac{d^2 \log p(y|\theta)}{da_0^2} = -2\sum_{t=1}^{N} \frac{1}{2b^2(t)\sigma_e^2} = -\sum_{t=1}^{N} \frac{1}{b^2(t)\sigma_e^2}$$

This is a deterministic expression, so taking the expectation does not change anything, ie,

$$E\left[\frac{d^2 \log p(y|\theta)}{da_0^2}\right] = -\sum_{t=1}^{N} \frac{1}{b^2(t)\sigma_e^2}$$

Now to obtain $\frac{d^2 \log p(y|\theta)}{da_0 d\sigma_e^2}$, we use $\frac{d \log p(y|\theta)}{da_0}$ and take the derivative with respect to σ_e^2

$$\frac{d^2 \log p(y|\theta)}{da_0 d\sigma_e^2} = -2\sum_{t=1}^{N} \frac{(y(t) - a_o)}{2b^2(t)\sigma_e^4}$$

Taking the expectation gives

$$E\left[\frac{d^2 \log p(y|\theta)}{d\sigma_e^2}\right] = -E\left[2\sum_{t=1}^{N} \frac{(y(t) - a_o)}{2b^2(t)\sigma_e^4}\right] - \frac{1}{\sigma_e^4}\sum_{t=1}^{N} \frac{E\ [e(t)]}{b(t)} = 0$$

And for $\frac{d^2 \log p(y|\theta)}{d(\sigma_e^2)^2}$, first derivative

$$\frac{d \log p(y|\theta)}{d\sigma_e^2} = -\frac{N}{2\sigma_e^2} + \sum_{t=1}^{N} \frac{(y(t) - a_o)^2}{2b^2(t)(\sigma_e^2)^2}$$

Taking the derivative once more gives

$$\frac{d^2 \ \log p(y|\theta)}{d(\sigma_e^2)^2} = \frac{N}{2\sigma_e^4} - 2\sum_{t=1}^{N} \frac{(y(t)-a_o)^2}{2b^2(t)\sigma_e^6}$$

And then apply the expectation to this

$$E\left[\frac{d^2 \ \log p(y|\theta)}{d(\sigma_e^2)^2}\right] = \frac{N}{2\sigma_e^4} - \frac{1}{\sigma_e^6}\sum_{t=1}^{N} E\left[\frac{(y(t)-a_o)^2}{b^2(t)}\right] =$$

$$\frac{N}{2\sigma_e^4} - \frac{1}{\sigma_e^6}\sum_{t=1}^{N} E\left[e(t)^2\right] = \frac{N}{2\sigma_e^4} - \frac{N\sigma_e^2}{\sigma_e^6} = -\frac{N}{2\sigma_e^4}$$

If we put this together we obtain

$$M = \begin{bmatrix} \sum_{t=1}^{N} \frac{1}{b^2(t)\sigma_e^2} & 0 \\ 0 & \frac{N}{2\sigma_e^4} \end{bmatrix}$$

So asymptotically:

$$\operatorname{Cov}\hat{\theta} = M^{-1} = \begin{bmatrix} \frac{1}{\sum_{t=1}^{N} \frac{1}{b^2(t)\sigma_e^2}} & 0 \\ 0 & \frac{2\sigma_e^4}{N} \end{bmatrix}$$

19 Back to problem

Given:

20 Back to problem

We have

$$\hat{r} = \frac{1}{\sqrt{N}}\sum_{t=1}^{N} e(t) \begin{bmatrix} u(t-s-1) \\ u(t-s-2) \\ u(t-s-3) \\ \vdots \\ u(t-s-M) \end{bmatrix}$$

Consider one term of the sum ($t=1$), scaled $\sqrt{N}$,

$$\hat{r}_1 = e(1) \begin{bmatrix} u(1-s-1) \\ u(1-s-2) \\ u(1-s-3) \\ \vdots \\ u(1-s-M) \end{bmatrix}$$

Since $e(t)$ is normally distributed and white, this is a normal distribution with variance

$$\mathrm{Var}(\hat{r}_1) =$$
$$\lambda \begin{bmatrix} u(1-s-1)^2 & u(1-s-1)u(1-s-2) & & & \\ u(1-s-1)u(1-s-2) & u(1-s-2)^2 & & & \\ & & u(1-s-3)^2 & & \\ & & & \ddots & \\ & & & & u(1-s-M) \end{bmatrix}$$

As example, the (1,2) term is obtained as,

$$\mathrm{Cov}(u(1-s-1)e(t), u(1-s-2)e(t))) =$$
$$u(1-s-1)u(1-s-2)\mathrm{Cov}(e(t), e(t)) = u(1-s-1)u(1-s-2)\lambda$$

Since $e(t)$ is white, its mean is zero and, hence, the expected value of the normal distribution is zero.

We observe that $\mathrm{Var}(\hat{r}_1)$ can be written as

$$\mathrm{Var}(\hat{r}_1) = \lambda\zeta(1)\zeta^T(1)$$

and then

$$\hat{r}_1 \in \mathcal{N}(0, \lambda\zeta(1)\zeta^T(1))$$

Generalizing for each term in the sum

$$\hat{r}_t \in \mathcal{N}(0, \frac{\lambda}{N}\zeta(t)\zeta^T(t))$$

A sum of independent Gaussian stochastic variables is also Gaussian where the variance is the sum of the variances of the individual variables. That is, in our case

$$\hat{r} \in \mathcal{N}(0, \lambda\frac{1}{N}\sum_{t=1}^{N}\zeta(t)\zeta^T(t)) = \hat{r} \in \mathcal{N}(0, \lambda\hat{R}_u)$$

We are almost done, it is the distribution of $\frac{1}{\lambda}\hat{r}^T\hat{R}_u^{-1}\hat{r}$ we need. Define the stochastic variable

$$\mu = \frac{1}{\sqrt{\lambda}}\hat{R}_u^{-0.5}\hat{r}$$

for which $\mu \in \mathcal{N}(0, I)$. Hence, μ is a M-dimensional vector of independent Gaussian components with unit-variance, and $\mu^T\mu$ is a sum of M squared Gaussian variables with unit variance. This means that

$$\mu^T\mu \in As\chi^2(M)$$

To think about: Will we have the same result without the assumption of normal distributed $e(t)$?

21 Back to problem

(a) Since the system is within the model class the estimates $\hat{a}$ and $\hat{c}$ tend assymtopically to their true values a and c, ie,

$$\mathrm{E}\,[\hat{a}-\hat{c}[= a-c$$

The variance become

$$\mathrm{Var}\,(\hat{a}-\hat{c}) = \mathrm{Var}\,\hat{a}+\mathrm{Var}\,\hat{b}-2\mathrm{Cov}(\hat{a},\hat{b}) = \mathrm{Cov}(\hat{\theta}_N)_{11}+\mathrm{Cov}(\hat{\theta}_N)_{22}-2\mathrm{Cov}(\hat{\theta}_N)_{12} = \frac{\sigma^2}{N(c-a)^2}((1-a^2)(1-a\,c)^2+(1-c^2)(1-a\,c)^2-2(1-c^2)(1-a^2)(1-a\,c)) = \frac{\sigma^2(1-ac)}{N(c-a)^2}((1-ac)(1-a^2+1-c^2)-2(1-c^2)(1-a^2)) = \frac{\sigma^2(1-a^2c^2)}{N}$$

(b) With variance from (a), and estimates of $\hat{a}$ and $\hat{c}$ a 95% confidence interval of $a-c$ is

$$(\hat{a}-\hat{c}) \pm 1.96 \times \sqrt{\frac{\sigma^2(1-\hat{a}^2\hat{c}^2)}{N}}$$

and if 0 is NOT within this interval, then the null hypothesis can be rejected with 95% certainty. Somewhat simpler this can be expressed as if $|\hat{a}-\hat{c}| > 1.96 \times \sqrt{\frac{\sigma^2(1-\hat{a}^2\hat{c}^2)}{N}}$ then the null hypothesis is rejected in favor of the larger model.

(c) F-test, see, eg, S&S 11.5, p440. We don't have a mean square error V_1 and V_2 based on data, but we can calculate the expected result of these. The test statistic is

$$\tau_f = \frac{V_N^1 - V_N^2}{V_N^2} \cdot \frac{N-p_2}{p_2-p_1}$$

Now

$$V_N^2 = \sigma_2^2 = \sigma^2$$

and we need to calculate σ_1^2. Model 1 has the predictor

$$\hat{y}(t) = 0$$

So the assymptotic estimate becomes

$$\hat{\sigma}_1^2 = \mathrm{E}\,(y(t)-\hat{y}^2(t)) = \mathrm{E}\,(y^2(t)) = \mathrm{E}\,[-a\,y(t)y(t-1)+y(t)e(t)+c\,y(t)e(t-1)] \tag{1}$$

Note that this holds also when $c=a$ is true.

So we look at the terms of (1) seperately,

$$\mathrm{E}\,[y(t)e(t)] = \mathrm{E}\,[-a\,y(t-1)e(t) + e^2(t) + c\,e(t-1)e(t)] = \mathrm{E}\,e^2(t) = \sigma^2$$

since $e(t)$ is white. Now looking at

$$\mathrm{E}\,[y(t)e(t-1)] = \mathrm{E}\,[-a\,y(t-1)e(t-1) + e(t)e(t-1) + c\,e^2(t-1)] = \sigma^2(c-a)$$

Finally for

$$\begin{aligned}\mathrm{E}\,[y(t)y(t-1)] = \mathrm{E}\,[-a\,y(t-1)y(t-1) + e(t)y(t-1) + c\,e(t-1)y(t-1)]\\ = \mathrm{E}\,[-a\,y(t-1)y(t-1)] + c\sigma^2\end{aligned}$$

Inserting in (1) now gives

$$\begin{aligned}\mathrm{E}\,y^2(t) = \mathrm{E}\,[-a\,y(t)y(t-1) + y(t)e(t) + c\,y(t)e(t-1)]\\ = -a(\mathrm{E}\,[-a\,y^2(t-1)] + c\,\sigma^2) + \sigma^2 + c\,\sigma^2(c-a)]\end{aligned}$$

Since $\mathrm{E}\,y^2(t) = \mathrm{E}\,y^2(t-1)$ we obtain

$$(1-a^2)\mathrm{E}\,y^2(t) = \sigma^2(1 + c^2 - 2a\,c)$$

and

$$\mathrm{E}\,y^2(t) = \frac{\sigma^2(1 + c^2 - 2a\,c)}{1-a^2}$$

which is the assymptotic estimate of σ_1^2 (and V_N^1).

This can now be used in an F test. We reject the null hypothesis if

$$\tau_f = \frac{V_N^1 - V_N^2}{V_N^2} \cdot \frac{N - p_2}{p_2 - p_1} \geq F_\alpha(p_2 - p_1, N - p_2)$$

where $p_2 = 2$ and $p_1 = 0$ since model 1 has no parameters and model 2 has 2.

With $N = 102$ measurements we obtain

$$F_{0.05}(2, 100) = 3.09$$

and the test statistic becomes

$$\tau_f \;=\; \frac{V_N^1 - V_N^2}{V_N^2} \cdot \frac{N - p_2}{p_2 - p_1} \;=\; \frac{\sigma_1^2 - \sigma_2^2}{\sigma_2^2} \cdot \frac{100}{2} \;=\; \frac{\frac{\sigma^2(1+c^2-2a\,c)}{1-a^2} - \sigma^2}{\sigma^2} \cdot 50 \;\geq\; 3.09$$

which is equivalent to

$$\frac{(1 + c^2 - 2a\,c)}{1-a^2} \geq 1.0618$$

In the problem two cases are given.

(1) With $a = 0.2$ and $c = 0.1$ we obtain

$$1.01 \geq 1.0618$$

which is not true and model 2 is rejected.

(2) If $a = 0.7$, $c = 0.2$ then the inequality becomes

$$1.49 \geq 1.0618$$

which is true. That is, model 2 is accepted.

Remark 1: normally, the hypothesis test is based on data on which the test statistic is evaluated. In this exercises we made use of the true system generating the data and we calculated the expectation of the hypothesis tests. That is, When the tests are performed on data, then the test statistics is a stochastic variable and there is a risk that an incorrect decision is taken.

Remark 2: often the χ^2 distribution is used instead of the F distribution. When $N \to \infty$ the F distribution converge towards a χ^2 distribution which is a bit easier to work with.

Remark 3: So what is the point with this exercise? Well, depending on the number of estimation data, the model selection method indicate models of different complexity. It is, again, the bias-variance trade-off. Here we had an example where the number of data were fixed and we considered instead the difference between the simple model and the large model. If there was small difference between them, then the simple model was prefered. If the difference was larger, then there was enough evidence in the data to indicate that the null hypothesis should be rejected in favor of the alternative larger model.

22 Back to problem

Let $p = 2\,n + 1$ denote the number of parameters in the models $\mathcal{M}$ and N the number of data points.

(a) The Akaike information criterion gives

$$AIC(p) = \log V(\theta_N) + \frac{2\,p}{N} = [0.500\ 0.05\ 0.0395\ 0.0387\ 0.0481]$$

with lowest value for $n = 4$, ie, $p = 9$.

(b) The Final Prediction Error becomes

$$FPE(p) = \frac{N+p}{N-p} V(\theta_N) = [1.63\ 1.0305\ 1.00959\ 0.9987\ 0.9981]$$

which has its lowest value for $n = 5$, ie, $p = 11$.

Hence, for this problem AIC and FPE suggest different model order. Note also, since FPE has its lowest value at the end of the interval of tested model orders, normally one would then test for larger orders until the criterion starts to increase.

23 Back to problem

(a) We have the sequences and wish to calculate its covariance function:

$$\mu_u = \lim_{N\to\infty} \frac{1}{N} \sum_{t=1}^{N} u(t)$$

$$C_{uu}(\tau) = \lim_{N\to\infty} \frac{1}{N} \sum_{t=1}^{N} (u(t+\tau) - \mu_u)(u(t) - \mu_u)^T$$

For C_2 the sequence repeats after 3 elements, this means that $u(t)$ repeats after 3 elements, so the sum

$$\mu_u = \lim_{N\to\infty} \frac{1}{N} \sum_{t=1}^{N} u(t) = \lim_{N\to\infty} \frac{1}{N} \frac{N}{3} \sum_{t=1}^{3} u(t) = \frac{1}{3}$$

The same can be done for the covariance

$$C_2(\tau) = \lim_{N\to\infty} \frac{1}{N} \sum_{t=1}^{N} (u(t+\tau) - \mu_u)(u(t) - \mu_u)^T =$$

$$\lim_{N\to\infty} \frac{1}{N} \frac{N}{3} \sum_{t=1}^{3} (u(t+\tau) - \mu_u)(u(t) - \mu_u)^T = \frac{1}{3} \sum_{t=1}^{3} (u(t+\tau) - \mu_u)(u(t) - \mu_u)^T$$

With the same logic for C_3 which gives

$$\mu_u = \frac{1}{7} \sum_{t=1}^{7} u(t) = \frac{1}{7}$$

and

$$C_3(\tau) = \frac{1}{7}\sum_{t=1}^{7}(u(t+\tau)-\mu_u)(u(t)-\mu_u)^T$$

Generalizing gives for C_N

$$\mu_u = \frac{1}{2^N-1}\sum_{t=1}^{2^N-1} u(t)$$

and

$$C_N(\tau) = \frac{1}{2^N-1}\sum_{t=1}^{2^N-1}(u(t+\tau)-\mu_u)(u(t)-\mu_u)^T$$

(b) The power spectra is defined as

$$S_{uu} = \sum_{\tau=-\infty}^{\infty} C(\tau)e^{-i\,\tau\,\omega}$$

For C_2 it becomes

$$S_{uu}^2 = \sum_{\tau=-\infty}^{\infty} C_2(\tau)e^{-i/,\tau\,\omega} = \sum_{l=-\infty}^{\infty}\sum_{\tau=0}^{2} C_2(\tau+3l)e^{-i\,\tau\,\omega}e^{-i3l\,\omega}$$

We use the fact that C_2 is periodic with period 3, ie $C_2(\tau+3l) = C_2(\tau)\,\forall\, l$ so

$$S_{uu}^2 = \sum_{l=-\infty}^{\infty}\sum_{\tau=0}^{2} C_2(\tau)e^{-i\tau\omega}e^{-i3l\,\omega} = \sum_{\tau=0}^{2} C_2(\tau)e^{-i\,\tau\,\omega}\sum_{l=-\infty}^{\infty} e^{-i3l\,\omega}$$

Now the function

$$\sum_{l=-\infty}^{\infty} e^{-i3l\,\omega}$$

is zero if we do not take steps of $3\omega = 2\pi$, in which case it becomes infinite. It can be expressed with the Dirac delta function

$$\sum_{l=-\infty}^{\infty} e^{-i2l\,\omega} = \frac{2\pi}{3}\sum_{k=0}^{2}\delta(\omega - 2\pi k/3)$$

Using this we obtain

$$S_{uu}^2 = \frac{2\pi}{3}\sum_{k=0}^{2}\sum_{\tau=0}^{2} C_2(\tau)e^{-i\,\tau\,\omega}\delta(\omega - 2\pi k/3)$$

For C_N the spectra becomes

$$S_{uu}^N = \sum_{\tau=-\infty}^{\infty} C_N(\tau) e^{-i\tau\, omega} = \sum_{l=-\infty}^{\infty} \sum_{\tau=0}^{2^N-2} C_N(\tau + (2^N-1)l) e^{-i\tau\,\omega} e^{-i\,l(2^N-1)\omega}$$

With equivalent mathematics as before this can be expressed as

$$S_{uu} = \frac{2\pi}{2^N-1} \sum_{k=0}^{2^N-2} \sum_{\tau=0}^{2^N-2} C_N(\tau) e^{-i\tau\,\omega} \delta(\omega - \frac{2\pi k}{2^N-1})$$

24 Back to problem

We have

$$\gamma_{yu}(\omega) = \frac{|S_{yu}(i\omega)|}{\sqrt{S_{uu}(i\omega) S_{yy}(i\omega)}}$$

Now assuming the noise is zero mean, from the system definition one obtains

$$S_{yu}(i\omega) = G(i\omega) S_{uu}(i\omega)$$

and

$$S_{yy}(i\omega) = |G(i\omega)|^2 S_{uu}(i\omega) + S_{vv}(i\omega)$$

Now:

$$\gamma_{yu}^2(\omega) = \frac{|S_{yu}(i\omega)|^2}{S_{uu}(i\omega) S_{yy}(i\omega)} = \frac{|G(i\omega)|^2 S_{uu}(i\omega)^2}{S_{uu}(i\omega)(|G(i\omega)|^2 S_{uu}(i\omega) + S_{vv}(i\omega))} = \frac{1}{1 + \frac{S_{vv}(i\omega)}{S_{uu}(i\omega)|G(i\omega)|^2}}$$

On this form the dependence on the signal to noise ratio ($\frac{S_{vv}(i\omega)}{S_{uu}(i\omega)}$) is obvious. Good signal to noise ration gives $\gamma_{yu}^2(\omega)$ close to 1 and a bad ratio gives a value close to 0.

The expression also shows that if $G(i\omega)$ is small, then it will be hard to identify.

25 Back to problem

(a) We have the system

$$y(t) + a\,y(t-1) = b\,u(t-1) + w(t) + c\,w(t-1) \quad \Rightarrow$$

$$y(t) = -a\,y(t-1) + b\,u(t-1) + w(t) + c\,w(t-1)$$

and the controler

$$u(t) = -K\,y(t)$$

Inserting the controler in the system equation we obtain

$$\begin{aligned} y(t) &= -a\,y(t-1) - b\,K\,y(t-1) + w(t) + c\,w(t-1) \\ &= -(a + b\,K)y(t-1) + w(t) + c\,w(t-1) = \alpha\,y(t-1) + w(t) + c\,w(t-1) \end{aligned}$$

which describes the output in the extrernal, exciting signal, w_k. The variable $\alpha = -(a + b\,K)$ can be identified, but we cannot decide a and b from that equation.

(b) We now use a model

$$\hat{y}(t) = \hat{\alpha}\,y(t-1)$$

to identify this process.

This is formulated as finding the $\hat{\alpha}$ which minimizes

$$\mathrm{E}\,(y(t) - \hat{y}(t))^2 = \mathrm{E}\,(y(t) - \hat{\alpha}y(t-1))^2 = \mathrm{E}\,y^2(t) + \hat{\alpha}^2 y^2(t-1) - 2\hat{\alpha}y(t)y(t-1)$$

where we need to find the three correlations expressed in σ^2. First we notice that, since we have stationary signals, $\mathrm{E}\,y^2(t-1) = \mathrm{E}\,y^2(t)$, so there are in fact two correlations to be found.

Then, noticing $\mathrm{E}\,w(t)y(t-1) = 0$ and $\mathrm{E}\,w(t-1)y(t-1) = \sigma^2$, due to whitness of w_k. Further,

$$\mathrm{E}\,y(t)w(t-1) = \alpha \mathrm{E}\,y(t-1)w(t-1) + w(t)w(t-1) + c\,w(t-1)w(t-1) = \alpha\,\sigma^2 + c\,\sigma^2$$

using the system equation. Now we obtain for $\mathrm{E}\,y(t)y(t-1)$,

$$\mathrm{E}\,y(t)y(t-1) = \alpha \mathrm{E}\,y(t-1)y(t-1) + w(t)y(t-1) + c\,w(t-1)y(t-1) = \alpha\,\mathrm{E}\,y^2(t) + c\,\sigma^2$$

Then for $\mathrm{E}\,y(t)y(t)$,

$$\begin{aligned} \mathrm{E}\,y(t)y(t) &= \alpha\,y(t)y(t-1) + y(t)w(t) + c\,y(t)w(t-1) \\ &= \alpha(\alpha\,\mathrm{E}\,y(t)y(t) + c\,\sigma^2) + \sigma^2 + c\,y(t)w(t-1) = \\ &\qquad \alpha(\alpha\,\mathrm{E}\,y(t)y(t) + c\,\sigma^2) + \sigma^2 + c(\alpha\,\sigma^2 + c\,\sigma^2) \end{aligned}$$

Solving with respect to $\mathrm{E}\,y(t)y(t)$ then gives

$$\mathrm{E}\,y(t)y(t) = \frac{2\alpha c\sigma^2 + \sigma^2 + c^2\sigma^2}{1 - \alpha^2}$$

Inserting all this in the criterion

$$\mathrm{E}\, y^2(t) + \hat{\alpha}^2 y^2(t-1) - 2\hat{\alpha}\, y(t)y(t-1) = \mathrm{E}\, y(t)y(t)(1+\hat{\alpha}^2) - 2\hat{\alpha}(\alpha r_0 + c^2\,\sigma^2)$$

where we kept $\mathrm{E}\ y(t)y(t)$ for readability, and since it does not depend on $\hat{\alpha}$. Minimize with respect to $\hat{\alpha}$ gives

$$2\mathrm{E}\, y(t)y(t)\,\hat{\alpha} - 2(\alpha\,\mathrm{E}\, y(t)y(t) + c^2\,\sigma^2) = 0 \quad \Rightarrow$$

$$\hat{\alpha} = \alpha + \frac{c\,\sigma^2}{\mathrm{E}\, y(t)y(t)} = -(a + b\,K) + \frac{c^2\,\sigma^2}{\mathrm{E}\, y(t)y(t)} \qquad (1))$$

where $\mathrm{E}\ y(t)y(t)$ was given a few lineas above. The estimate is a combination the three system parameters a, b and c and how much it "gets" from each of them depends on the size of the feedback K. The influence from c is due to that the model does not have the correct disturbance model.

(c) By repeating the identification for values of K we obtain several estimates like (1). To be able to solve for a and b we need four equations since there are four unknowns (a, b, c and σ^2). Hence, four experiements with K_1, K_2, K_3 and K_4 need to be performed.

26 Back to problem

We have persistantly exiting of order n if the following limits exist,

$$\mu_u = \lim_{N\to\infty} \frac{1}{N} \sum_{t=1}^{N} u(t)$$

$$\hat{C}_{uu}(\tau) = \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} u(t)u_{k-\tau}^T$$

and if the correlation matrix is positive definite

$$R_{uu}(n) = \begin{pmatrix} \hat{C}_{uu}(0) & \hat{C}_{uu}(1) & \hat{C}_{uu}(n-1) \\ \hat{C}_{uu}(-1) & \hat{C}_{uu}(0) & \hat{C}_{uu}(n-2) \\ & & \\ \hat{C}_{uu}(1-n) & & \hat{C}_{uu}(0) \end{pmatrix}$$

In our case,

$$\hat{C}_{uu}(\tau) = 1\ \forall\ \tau$$

From which follows that

$$R_{uu}(1) = 1$$

is positive (definite), but

$$R_{uu}(2) = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$$

is not positive definite, as it does not have full rank. Hence, the input is persistantly exiting of order 1 but not greater.

For consistency estimate of the parameters, we need to investigate conditions for $a_N \rightarrow a$ and $b_N \rightarrow b$ when $N \rightarrow \infty$. Hence, we form the PE criterion (quadratic norm for simplicity, it does not matter which one we choose) and minimize with respect to the parameters and investigate the limit. Introduce the symbol

$$\overline{\sum} = \frac{1}{N}\sum_{t=1}^{N}$$

The control signal $u(t) = 1$ inserted in the system equation gives

$$y(t) = -a\,y(t-1) + b + w(t)$$

The model is

$$\hat{y}(t) = -\hat{a}\,y(t-1) + \hat{b}$$

The estimate, $\hat{a}_N$ and $\hat{b}_N$ that minimize (index N surpressed)

$$\begin{gathered} \overline{\sum}(y(t) - \hat{y}(t))^2 = \overline{\sum}(y(t) + \hat{a}\,y(t-1) - \hat{b})^2 = \\ \overline{\sum}\,y^2(t) + \hat{a}^2\overline{\sum}\,y^2(t-1) + 2\hat{a}\,\overline{\sum}\,y(t)y(t-1) + \hat{b}^2 - 2\,\hat{b}\,\overline{\sum}\,y(t) - 2\,\hat{b}\,\hat{a}\,\overline{\sum}\,y(t-1) \end{gathered} \tag{1}$$

From whiteness of w_k, $\overline{\sum} w(t)y(t-1) = 0$ follows. The correlations need to be expressed in σ^2 and other system properties, typically the parameters a and b.

There are three correlations (well one of them is a mean-value), introduce the follwing symbols

$$\overline{\sum}y^2(t) = r_0, \quad \overline{\sum}y(t)y(t-1) = r_1, \quad \overline{\sum}y(t) = \bar{y} \tag{2}$$

Since signals are stationary $\overline{\sum}y^2(t-1) = r_0$. The criterion to be minimize, (1) then becomes

$$r_0 + \hat{a}^2 r_0 + 2\hat{a}\,r_1 + \hat{b}^2 - 2\,\hat{b}\,\bar{y} - 2\,\hat{b}\,\hat{a}\,\bar{y} \tag{3}$$

Take the derivative with respect to $\hat{a}$ gives

$$2\hat{a}\, r_0 + 2\, r_1 - 2\,\hat{b}\,\bar{y} = 0$$

and with respect to $\hat{b}$,

$$2\,\hat{b} - 2\,\bar{y} - 2\,\hat{a}\,\bar{y} = 0$$

This gives us the following equations for $\hat{a}$ and $\hat{b}$

$$\begin{aligned} (r_0 - \bar{y}^2)\hat{a} &= \bar{y}^2 - r_1 \\ \hat{b} &= \bar{y}(1 + \hat{a}) \end{aligned} \tag{4}$$

The solution depends on the correlation, and we see that $(r_0 - \bar{y}^2) \neq 0$ is a condition for a solution.

Let us now calculat the correlations. From the system equation we obtain the follwing expresions

$$r_1 = \overline{\sum} y(t)y(t-1) = -a\,\overline{\sum} y(t-1)y(t-1) + \overline{\sum}\, w(t)y(t-1) + b\,\overline{\sum}\, y(t-1) = -a\, r_0 + b\,\bar{y}$$

$$r_0 = -a\, r_1 + b\,\bar{y} + \sigma^2$$

$$\bar{y} = -a\,\bar{y} + b$$

With the follwing solutions

$$r_1 = \frac{b^2}{(1+a)^2} - \frac{a\,\sigma^2}{1-a^2}$$

$$r_0 = \frac{b^2}{(1+a)^2} + \frac{\sigma^2}{1-a^2}$$

$$\bar{y} = \frac{b}{1+a}$$

From this we see that the condition for a solution for the estimates becomes

$$r_0 - \bar{y}^2 = \frac{b^2}{(1+a)^2} + \frac{\sigma^2}{1-a^2} - \frac{b^2}{(1+a)^2} = \frac{\sigma^2}{1-a^2} \neq 0$$

That is $\sigma^2 \neq 0$, which means that the noise cannot be zero.

It remains to show that the solution of the estimate is the true parameter values. By inserting the values for the correlations in (4) you obtain

$$\hat{a} = a$$

$$\hat{b} = b$$

That is, the assymptotic estimate is consistent if the noise is non-zero.

This is an example where identifiability is obtained by excitation by the noise. The plant and the disturbance model share the parameter a, and it is identifiable thanks to the noise, giving us the first equation in (4). The input signal is persistent exciting to order 1 and gives us the second equation in (4). It contains a linear combination of the two parameters.

27 Back to problem

Given:

28 Back to problem

Sinusoids are persistently exiting of order 2 and we have three unknowns, so we do not expect it to be possible to identify all three parameters. This can also easily be seen with simple trigonometry:

$$y(t) = b_1 \sin(t-1) + b_2 \sin(t-2) + b_3 \sin(t-3)$$

We now use the fact that:

$$\sin(a-b) = \sin(a)\cos(b) - \sin(b)\cos(a)$$

We use that so that we have only time dependet terms $\sin(t-1)$ in the system equation. This gives,

$$y(t) = b_1 \sin(t-1) + b_2(\sin(t-1)\cos(1) - \sin(1)\cos(t-1)) + b_3(\sin(t-1)\cos(2) - \sin(2)\cos(t-1))$$

Rewriting this we get:

$$y(t) = \sin(t-1)(b_1 + b_2\cos(1) + b_3\cos(2)) - \cos(t-1)(b_2\sin(1) + b_3\sin(2))$$

So while we can identify the coefficients of the $\sin(t-1)$ and $cos(t-1)$ term, this only gives us two equations. We have three unknowns and it is not possible to uniquely identify all the parameters as there will be multiple parameter combinations that give the same result.

29 Back to problem

(a) The closed-loop system is defined by

$$G_{cl}(s) = \frac{G_P(s)}{1 + G_C(s)G_P(s)}$$

gives

$$G_P(s) = \frac{G_{cl}(s)}{1 - G_{cl}(s)G_C(s)}$$

Now, use this equation and use the estimated closed-loop system to form an estimate of the plant

$$\hat{G}_P(s) = \frac{(s+1)(s+3)}{(s+2)(s+3)(s+4) + s + 1 - (s+1)} = \frac{s+1}{(s+2)(s+4)}$$

(b) The main disadvantage with indirect identification is that any error in $G_C(s)$ (including deviation from a linear regulator, due to input saturations or anti-windup measurements) will be incorporated in the estimate of $G_P(s)$.

30 Back to problem

(a) We look at the system

$$\begin{aligned} A &= \begin{bmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{bmatrix} \\ B &= \begin{bmatrix} b \\ 0 \end{bmatrix} \\ C &= \begin{bmatrix} 1 & 0 \end{bmatrix} \end{aligned}$$

and calculate the transfer function:

$$G = C(qI - A)^{-1}B = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} q - a_{11} & -a_{12} \\ -a_{12} & q - a_{22} \end{bmatrix}^{-1} \begin{bmatrix} b \\ 0 \end{bmatrix} =$$
$$\frac{1}{(q - a_{11})(q - a_{22}) + a_{12}^2} \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} q - a_{22} & a_{12} \\ a_{12} & q - a_{11} \end{bmatrix} \begin{bmatrix} b \\ 0 \end{bmatrix} =$$
$$\frac{(q - a_{22})b}{(q - a_{11})(q - a_{22}) + a_{12}^2} = \frac{b\,q - a_{22}b}{q^2 - (a_{11} + a_{22})q + a_{11}a_{22} + a_{12}^2}$$

With an equation for the transfer function as:

$$\frac{k_1 + k_2 q}{k_3 + k_4 q + q^2}$$

With:

$$\begin{aligned}
k_1 &= -a_{22} b \\
k_2 &= b \\
k_3 &= a_{12}^2 + a_{11} a_{22} \\
k_4 &= -a_{11} - a_{22}
\end{aligned}$$

We can uniquely determine everything but the sign of a_{12}.

(b) Now:

$$C = \begin{bmatrix} 0 & 1 \end{bmatrix}$$

The transfer function becomes:

$$\frac{1}{(q - a_{11})(q - a_{22}) + a_{12}^2} \begin{bmatrix} 0 & 1 \end{bmatrix} \begin{bmatrix} q - a_{22} & a_{12} \\ a_{12} & q - a_{11} \end{bmatrix} \begin{bmatrix} b \\ 0 \end{bmatrix} =$$
$$\frac{a_{12} b}{(q - a_{11})(q - a_{22}) + a_{12}^2} = \frac{a_{12} b}{q^2 - (a_{11} + a_{22}) q + a_{11} a_{22} + a_{12}^2}$$

4 unknowns and 3 equations so we cannot uniquely determine the parameters.

(c) We now have:

$$\frac{(q - a_{22}) b}{q^2 - (a_{11} + a_{22}) q + a_{11} a_{22} + a_{12}^2}$$

and

$$\frac{a_{12} b}{q^2 - (a_{11} + a_{22}) q + a_{11} a_{22} + a_{12}^2}$$

First transfer function can be used to determine everything except sign on a_{12}, which can then easily be obtained from the second transfer function. Hence, the model is identifiable.

31 Back to problem

The system is described by

$$\begin{aligned} x(t+1) &= a\,x(t) + b\,u(t) + \xi(t) \\ y(t) &= x(t) + e(t) \end{aligned} \tag{1}$$

and we want to describe it on the form

$$y(t) = G(q,\theta)u_2(t) + H(q,\theta)e_2(t)$$

where $\theta = [a\ b]$. We need to find $G(q,\theta)$ and $H(q,\theta)$ and we might need to re-define input and noise, that is why we introduced new notation, $u_2(t)$ and $e_2(t)$.

Since $y(t) = x(t) + e(t)$ we start by re-writing $x(t)$,

$$x(t) = \frac{b\,q^{-1}}{1 - a\,q^{-1}}u(t) + \frac{q^{-1}}{1 - a\,q^{-1}}\xi(t)$$

Insert in (1) we obtain

$$y(t) = \frac{b\,q^{-1}}{1 - a\,q^{-1}}u(t) + e(t) + \frac{q^{-1}}{1 - a\,q^{-1}}\xi(t) \tag{2}$$

from were we can identify

$$G(q,\theta) = \frac{b\,q^{-1}}{1 - a\,q^{-1}}$$

and it is concluded that the input does not need to be re-defined. It remains to describe the disturbance. It should be driven by a white signal $e_2(t)$. Hence, we investigate the correlation of the disturbance since $H(q,\theta)$ should describe the color of it. Define

$$w(t) = e(t) + \frac{q^{-1}}{1 - a\,q^{-1}}\xi(t)$$

or as equation

$$w(t) = a\,w(t-1) + e(t) + \xi(t-1) \tag{3}$$

Multiply (3) with $w(t)$ and $w(t-1)$, respectively and take the expectation. Then we obtain

$$\begin{aligned} r_0 &= E\,w^2(t) = a\,r_1 - a\,E\,w(t)e(t-1) + E\,w(t)\xi(t-1) + E\,w(t)e(t) \\ r_1 &= E\,w(t)w(t-1) \\ &= a\,r_0 - a\,E\,w(t-1)e(t-1) + E\,w(t-1)\xi(t-1) + E\,w(t-1)e(t) \end{aligned}$$

By multiplying (3) with $e(t+1)$, $e(t)$, $e(t-1)$, $\xi(t)$ and $\xi(t-1)$, and taking the expectation one obtains $E\,w(t)e(t) = \sigma^2$, $E\,w(t)e(t-1) = 0$, $E\,w(t)e(t+1) = E\,w(t-1)e(t) = 0$, $E\,w(t)e\xi(t) = 0$, and $E\,w(t)\xi(t-1) = \lambda$. Inserting this gives

$$
\begin{aligned}
r_0 &= a\,r_1 + \lambda + \sigma^2 \\
r_1 &= a\,r_0 - a\,\sigma^2
\end{aligned}
$$

with solution

$$
\begin{aligned}
r_0 &= \sigma^2 + \frac{\lambda}{1-a^2} \\
r_1 &= a\frac{\lambda}{1-a^2}
\end{aligned} \tag{4}
$$

It is fairly easy to veryfy that $r_k = E\,w(t)w(t+k) = a\,r_{k-1}$ for $k \leq 2$.

Now we propose a $H(q,\theta)$ and a $e_2(t)$ so that $w_2(t) = H(q,\theta)e_2(t)$ and the correlations above remains the same. From r_k, $k \leq 2$ it follows that a pole in a is needed, and from $k \leq 2$ it follows that one zero is needed to fit the first two corellations. Hence, we propose

$$
w_2(t) = H(q,\theta)e_2(t) = \frac{1+\beta\,q^{-1}}{1-\alpha\,q^{-1}}e_2(t) \tag{5}
$$

and $w_2(t)$ should have the same correlations as $w(t)$. That is, we need to determine α, β, and $\sigma_2^2 = E\,e_2^2(t)$ to fullfil this. We do this by computing the correlations for $w_2(t)$ in the same way as above for $w_(t)$. Re-write (5) to equation form

$$
w_2(t) = \alpha\,w_2(t-1) + e_2(t) + \beta\,e_2(t-1) \tag{6}
$$

By multiplying (6) with $e_2(t)$, $e_2(t-1)$, and taking the expectation one obtains $E\,w_2(t)e_2(t) = E\,e_2^2(t) = \sigma_2^2$ and $E\,w_2(t)e_2(t-1) = \alpha\,E\,w_2(t-1)e_2(t-1) + E\,e_2(t)e_2(t-1) + \beta\,E\,e_2^2(t-1) = (\alpha+\beta)\sigma_2^2$. By multiplying (6) with $w_2(t)$, $w_2(t-1)$, and taking the expectation one obtains

$$
\begin{aligned}
E\,w_2(t)w_2(t) &= \alpha\,E\,w_2(t)w_2(t-1) + E\,w_2(t)e_2(t) + \beta\,E\,w_2(t)e_2(t-1) = \\
&\qquad \alpha\,E\,w_2(t)w_2(t-1) + \sigma_2^2 + \beta(\beta+\alpha)\sigma_2^2 \\
E\,w_2(t)w_2(t-1) &= \alpha\,E\,w_2(t-1)w_2(t-1) + E\,w_2(t-1)e_2(t) + \\
&\qquad \beta\,E\,w_2(t-1)e_2(t-1) = \alpha\,E\,w_2(t)w_2(t) + \beta\,\sigma_2^2
\end{aligned} \tag{7}
$$

Solving with respect to $E\,w_2(t)w_2(t)$ and $E\,w_2(t)w_2(t-1)$ gives

$$
\begin{aligned}
E\,w_2(t)w_2(t) &= \frac{1+\beta^2+2\,\beta\,\alpha}{1-\alpha^2}\sigma_2^2 \\
E\,w_2(t)w_2(t-1) &= \frac{\alpha+\alpha\,\beta^2+\beta\,\alpha^2+\beta}{1-\alpha^2}\sigma_2^2
\end{aligned} \tag{8}
$$

and by multiplying (6) with $w_2(t-k)$ one obtain $E\, w_2(t)w_2(t+k) = \alpha\, E\, w_2(t)w_2(t+k-1)$ for $k \leq 2$. We now fit α, β and σ_2^2 so that these expression becomes equal to r_0, r_1, and r_k for $k \leq 2$, (4). This gives

$$\alpha = a$$
$$\sigma_2^2 = \frac{\lambda + (1-a^2)\sigma^2}{1+\beta^2+2\,\beta\, a}$$

and the following, somewhat messy, equation for β

$$\beta^2 + (a + \frac{1}{a} + \frac{\lambda}{a\,\sigma^2})\beta + 1 = 0 \tag{9}$$

Since this is a quadratic equation, it has two solutions. For the problem to make sense, $|a| < 1$, $0 < \lambda < 1$ and $\sigma^2 > 0$, and if we belive the theory of the course, (9) should have at least one of its solution inside the unit circle. Infact, you can easily show that it is exactly one inside and the other is outside.

To have a numeric example, assume $a = \lambda = 0.5$ and $\sigma^2 = 1$. Then (9) becomes

$$\beta^2 + 3.5\,\beta + 1 = 0$$

with solutions $\beta = -0.31$ and $\beta = -3.19$, and the zero inside the unit circle gives $\sigma_2^2 = 1.59$.

We have managed to write the system on the standard form with a plant model and a disturbance model. However, we have only specified second order properties of the noise sequence $e_2(t)$, no distribution. We cannot. This means that if we identify the system with a PEM, it will not be eqivalent with Maximum likelihood (ML) since we do not have any "true" distribution of $e_2(t)$. The theory tells us that if we estimate with PEM, the estimates of G and H will be concistent, ie, converge, with probability one, towards the true transfere functions when the numher of data goes to infinity, these G and H corrspond to the state space model given in the problem. The uncertainty of the estimates will not be as small as possible, since PEM is not ML. It is possible to formulate the ML estimator of the system, but it is complicated. It is often like that, ML estimate can be formulated, but not computed and research efforts are spent on finding good approxiumate solutions.

32 Back to problem

The transfer function becomes:

$$C(sI-A)^{-1}B = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} s-a_{11} & -a_{12} & 0 \\ -a_{21} & s-a_{22} & 0 \\ 0 & 1 & s \end{bmatrix}^{-1} \begin{bmatrix} b_{11} \\ b_{21} \\ 0 \end{bmatrix}$$

gives

$$C(sI - A)^{-1}B = -\frac{a_{21}b_{11} - a_{11}b_{21} + b_{21}s}{s(-a_{12}a_{21} + a_{11}a_{22} - (a_{11} + a_{22})s + s^2)}$$

Hence, all models in the model set have the form

$$\frac{k_1 + k_2 s}{s(k_3 + k_4 s + s^2)}$$

That is, can be described by 4 parameters and by comparing, it is clear that there are several combinations of the 6 unknowns give the same transfere function. Therefore the model structure is not identifiable.

33 Back to problem

We have:

$$\begin{aligned} \dot{x} &= A\,x + B_1 u + B_2 v \\ y &= C\,x \end{aligned}$$

$$A = \begin{bmatrix} 0 & K \\ 0 & 0 \end{bmatrix}, \quad B_1 = \begin{bmatrix} K \\ 0 \end{bmatrix}, \quad B_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 0 \end{bmatrix}$$

We wish to discretise this system (S & S (6.40)), and we can use zero order hold:

We will write the discrete system in the form of:

$$x(t+h) = \phi\,x(t) + \Gamma_0 u(t) + v(t)$$

Where h is the sampling time. Now:

$$\phi = e^{A\,h} = I + A\,h = \begin{bmatrix} 1 & K\,h \\ 0 & 1 \end{bmatrix}$$

As $A^2 = 0$. Moreover

$$\Gamma_0 = \int_0^h e^{A\,s} ds\, B_1 = \int_0^h \begin{bmatrix} 1 & K\,s \\ 0 & 1 \end{bmatrix} ds\, B_1 = \begin{bmatrix} h & K\frac{h^2}{2} \\ 0 & h \end{bmatrix} B_1 = \begin{bmatrix} h\,K \\ 0 \end{bmatrix}$$

The covariance of $v(t)$ becomes

$$R_1 = \int_0^h e^{A\,s} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} e^{A^T s} ds = ...$$

So we have the system:

$$x(t+h) = \begin{bmatrix} 1 & K\,h \\ 0 & 1 \end{bmatrix} x(t) + \begin{bmatrix} h\,K \\ 0 \end{bmatrix} u(t) + v(t)$$

$$y = \begin{bmatrix} 1 & 0 \end{bmatrix} x$$

34 Back to problem

(a) Since everything except the white noise is available at $t-1$, the one step ahead predictor is given by

$$\hat{y}_W(t|t-1) = \frac{b_1 u(t-1) + b_2 u(t-2)}{1 + a_1 y^2(t-1)}$$

where we added a subscript W to indicate that it is the predictor for the case of white disturbance. We will use it further down.

(b) Compared to (a), the one-step ahead-prediction now needs to be adjusted for the colored noise. Hence, we need to include $\hat{v}(t|t-1)$ to the answer in (a). Actually, $\hat{v}(t|t-1)$ was included in (a), but it was zero and we made no remark about this.

If we have a noise term (ARMA, 1st order)

$$v(t) = -d_1 q^{-1} v(t) + (1 + c_1 q^{-1}) e(t)$$

with $e(t)$ being white noise. This can be re-written as

$$v(t) = \frac{1 + c_1 q^{-1}}{1 + d_1 q^{-1}} e(t) = H(q^{-1}, d_1, c_1) e(t)$$

Then, the prediction becomes

$$\hat{v}(t|t-1) = (1 - H(q^{-1}, d_1, c_1)^{-1}) v(t) = \frac{(c_1 - d_1) q^{-1}}{1 + c_1 q^{-1}} v(t)$$

and, from the system definition we have

$$v(t) = y(t) - \frac{b_1 u(t-1) + b_2 u(t-2)}{1 + a_1 y^2(t-1)} \tag{1}$$

Note that only information up to $t-1$ is needed. The predictor now becomes

$$\hat{y}(t|t-1) = \hat{y}_W(t|t-1) + \hat{v}(t|t-1) = \frac{b_1 u(t-1) + b_2 u(t-2)}{1 + a_1 y^2(t-1)} + \hat{v}(t|t-1)$$

(c) So for a) we have:

$$\hat{y}_W(t) = \frac{b_1u(t-1) + b_2u(t-2)}{1 + a_1y^2(t-1)} \tag{2}$$

To reduce the amount of writing, by noticing that b_1 and b_2 enters in similar way, only $u(t-1)$ need to be replaced by $u(t-2)$, we skip calculating $\frac{d\hat{y}(t)}{db_2}$. Straight forward derivation of (2) gives

$$\frac{d\hat{y}_W(t)}{d\theta} = \begin{bmatrix} \frac{d\hat{y}(t)}{da_1} \\ \frac{d\hat{y}_W(t)}{db_1} \\ \frac{d\hat{y}_W(t)}{db_2} \end{bmatrix} = \begin{bmatrix} -\frac{b_1u(t-1)+b_2u(t-2)}{(1+a_1y^2(t-1))^2}y^2(t-1) \\ \frac{u(t-1)}{1+a_1y^2(t-1)} \\ \dots \end{bmatrix} \tag{3}$$

for b) we have:

$$\hat{v}(t|t-1) = (1 - H(q^{-1}, d_1, c_1)^{-1})e(t) = \frac{(c_1 - d_1)q^{-1}}{1 + c_1q^{-1}}v(t)$$

and the predictor

$$\hat{y}(t) = \frac{b_1u(t-1) + b_2u(t-2)}{1 + a_1y^2(t-1)} + \frac{c_1 - d_1}{(1 + c_1q^{-1})}(y(t-1) - \frac{b_1u(t-2) + b_2u(t-3)}{1 + a_1y^2(t-2)}) =$$
$$\hat{y}_W(t) + \frac{c_1 - d_1}{(1 + c_1q^{-1})}(y(t-1) - \hat{y}_W(t-1))$$

By comparing with the case of white disturbance, we can express the derivative with respect to a_1, as

$$\frac{d\hat{y}(t)}{da_1} = \frac{d\hat{y}_W(t)}{da_1} - \frac{(c_1 - d_1)q^{-1}}{1 + c_1q^{-1}}\frac{d\hat{y}_W(t)}{da_1} \tag{4}$$

and corresponding for b_1 (and b_2). For the two parameters in the disturbance model, it is enough to filter $v(t)$, (1) through the derivative of the filter, ie,

$$\frac{d}{dc_1}\frac{(c_1 - d_1)q^{-1}}{1 + c_1q^{-1}} = \frac{1}{(1 + c_1q^{-1})} - \frac{(c_1 - d_1)q^{-1}}{(1 + c_1q^{-1})^2} = \frac{(1 + d_1)q^{-1}}{(1 + c_1q^{-1})^2}$$
$$\frac{d}{dd_1}\frac{(c_1 - d_1)q^{-1}}{1 + c_1q^{-1}} = -\frac{1}{(1 + c_1q^{-1})}$$

Putting this together we obtain

$$\frac{d\hat{y}(t)}{d\theta} = \begin{bmatrix} \frac{d\hat{y}(t)}{da_1} \\ \frac{d\hat{y}(t)}{db_1} \\ \frac{d\hat{y}(t)}{db_2} \\ \frac{d\hat{y}(t)}{dc_1} \\ \frac{d\hat{y}(t)}{dd_1} \end{bmatrix} = \begin{bmatrix} \frac{d\hat{y}_W(t)}{da_1} - \frac{(c_1-d_1)q^{-1}}{1+c_1q^{-1}}\frac{d\hat{y}_W(t)}{da_1} \\ \frac{d\hat{y}_W(t)}{db_1} - \frac{(c_1-d_1)q^{-1}}{1+c_1q^{-1}}\frac{d\hat{y}_W(t)}{db_1} \\ \dots \\ \frac{(1+d_1)q^{-1}}{(1+c_1q^{-1})^2}(y(t-1) - \hat{y}_W(t-1)) \\ \frac{1}{(1+c_1q^{-1})}(y(t-1) - \hat{y}_W(t-1)) \end{bmatrix}$$

where $\hat{y}_W(t)$ and its derivatives are given further up.

Inserting all subexpressions gives

$$\frac{d\hat{y}(t)}{d\theta} = \begin{bmatrix} \frac{d\hat{y}(t)}{da_1} \\ \frac{d\hat{y}(t)}{db_1} \\ \frac{d\hat{y}(t)}{db_2} \\ \frac{d\hat{y}(t)}{dc_1} \\ \frac{d\hat{y}(t)}{dd_1} \end{bmatrix} =$$

$$\begin{bmatrix} -\frac{b_1u(t-1)+b_2u(t-2)}{(1+a_1y^2(t-1))^2}y^2(t-1) + \frac{c_1-d_1}{(1+c_1q^{-1})}(\frac{b_1u(t-2)+b_2u(t-3)}{(1+a_1y^2(t-2))^2})y^2(t-2) \\ \frac{u(t-1)}{1+a_1y^2(t-1)} - \frac{c_1-d_1}{(1+c_1q^{-1})}\frac{u(t-2)}{1+a_1y^2(t-2)} \\ \frac{u(t-2)}{1+a_1y^2(t-1)} - \frac{c_1-d_1}{(1+c_1q^{-1})}\frac{u(t-3)}{1+a_1y^2(t-2)} \\ \frac{1}{(1+c_1q^{-1})}(y(t-1) - \frac{b_1u(t-2)+b_2u(t-3)}{1+a_1y^2(t-2)}) - \frac{(c_1-d_1)q^{-1}}{(1+c_1q^{-1})^2}(y(t-1) - \frac{b_1u(t-2)+b_2u(t-3)}{1+a_1y^2(t-2)}) \\ -\frac{1}{(1+c_1q^{-1})}(y(t-1) - \frac{b_1u(t-2)+b_2u(t-3)}{1+a_1y^2(t-2)}) \end{bmatrix}$$

The expressions become quite messy. It is nice to have computer software calculating the expressions and applying them to the data in a tool for system identification.

35 Back to problem According to the problem description the system is described as

$$y(t) = -a_1y(t-1) - a_2y(t-2) + b_1f(u(t-1)) + b_2f(u(t-2)) + e(t) \qquad (1)$$

and the predictor can be described as

$$\hat{y}(t|t-1) = B(q)f(u(t)) + [1 - A(q)]y(t)$$

with a natural definition of $B(q)$ and $A(q)$. The function $f(\cdot)$ is given in the problem. With

$$B(q) = b_1q^{-1} + b_2q^{-2}$$

we have nine different cases, for instance:

$$\begin{cases} \hat{y}(t|t-1) = \theta_1\theta_2(b_1q^{-1} + b_2q^{-2}) + [1 - A(q)]y(t) & u(t-1), u(t-2) > \theta_2 \\ \hat{y}(t|t-1) = \theta_1 b_1 u(t-1) + b_2\theta_1\theta_2 + [1 - A(q)]y(t) & |u(t-1)| < \theta_2, u(t-2) > \theta_2 \\ \vdots \end{cases}$$

etc, for all combinations of $u(t-1)$, $u(t-2)$.

36 Back to problem

(a) Given the system description and that $\{v(t)\}$ is white, the predictor becomes $\hat{y}(t,\theta) = x(t)$. It remains writing $x(t)$ as an inner product between $\theta = [a_1,\ a_2,\ b_1,\ b_2,\ c_1]^T$ and a pseudo-regressor. Looking at the expresion for $x(t)$ gives

$$\varphi^T(t,\theta) = [-x(t-1)\ -x(t-2)\ u(t-1)\ u(t-2)\ x(t-1)u(t-1)]$$

where $x(t-1)$ and $x(t-2)$ are generated with the system equation.

(b) In a), the best prediction of $\{v(t)\}$ was 0. Now a better prediction can be obtained using the knowlwedge that it is a first order ARMA process. That is,

$$v(t) = -\alpha\, v(t-1) + \beta\, e(t-1) + e(t)$$

This gives

$$v(t) = \frac{1+\beta\, q^{-1}}{1+\alpha\, q^{-1}} e(t) = H(q^{-1})e(t)$$

and

$$\hat{v}(t|t-1) = (1 - H^{-1}(q^{-1}))v(t) = \frac{(\alpha-\beta)q^{-1}}{1+\alpha\, q^{-1}} v(t)$$

Hence,

$$\hat{v}(t|t-1) = -\alpha\, \hat{v}(t-1|t-2) + (\alpha-\beta)v(t-1)$$

where $v(t-1) = y(t-1) - x(t)$ and the one step predictor can be put together as

$$\hat{y}(t,\theta) = x(t) + \hat{v}(t|t-1)$$

37 Back to problem

(a) We have the system

$$\ddot{y} + a\, y(t)\dot{y}(t) + b\, y(t) = u(t)$$

which we re-write as

$$\ddot{y} = -a\, y(t)\dot{y}(t) - b\, y(t) + u(t)$$

and then integrate

$$\int_{t_0}^{t_1} \ddot{y}(t)dt = -a\int_{t_0}^{t_1} y(t)\dot{y}(t)dt - b\int_{t_0}^{t_1} y(t)dt + \int_{t_0}^{t_1} u(t)dt \Leftrightarrow$$

$$\dot{y}(t_1) - \dot{y}(t_0) = -a\int_{t_0}^{t_1} y(t)\dot{y}(t)dt - b\int_{t_0}^{t_1} y(t)dt + \int_{t_0}^{t_1} u(t)dt$$

By dividing the data range into intarvals where we apply this integral we obtain a set of equations

$$\begin{bmatrix} \dot{y}(t_1) - \dot{y}(t_0) \\ \dot{y}(t_2) - \dot{y}(t_1) \\ \vdots \\ \dot{y}(t_N) - \dot{y}(t_{N-1}) \end{bmatrix} = \begin{bmatrix} \int_{t_0}^{t_1} y(t)\dot{y}(t)dt & \int_{t_0}^{t_1} y(t)dt & \int_{t_0}^{t_1} u(t)dt \\ \int_{t_1}^{t_2} y(t)\dot{y}(t)dt & \int_{t_1}^{t_2} y(t)dt & \int_{t_1}^{t_2} u(t)dt \\ \vdots & \vdots & \vdots \\ \int_{t_{N-1}}^{t_N} y(t)\dot{y}(t)dt & \int_{t_{N-1}}^{t_N} y(t)dt & \int_{t_{N-1}}^{t_N} u(t)dt \end{bmatrix} \begin{bmatrix} -a \\ -b \\ 1 \end{bmatrix} \Leftrightarrow$$

$$\begin{bmatrix} \dot{y}(t_1) - \dot{y}(t_0) - \int_{t_0}^{t_1} u(t)dt \\ \dot{y}(t_2) - \dot{y}(t_1) - \int_{t_1}^{t_2} u(t)dt \\ \vdots \\ \dot{y}(t_N) - \dot{y}(t_{N-1}) - \int_{t_{N-1}}^{t_N} u(t)dt \end{bmatrix} = \begin{bmatrix} \int_{t_0}^{t_1} y(t)\dot{y}(t)dt & \int_{t_0}^{t_1} y(t)dt \\ \int_{t_1}^{t_2} y(t)\dot{y}(t)dt & \int_{t_1}^{t_2} y(t)dt \\ \vdots & \vdots \\ \int_{t_{N-1}}^{t_N} y(t)\dot{y}(t)dt & \int_{t_{N-1}}^{t_N} y(t)dt \end{bmatrix} \begin{bmatrix} -a \\ -b \end{bmatrix}$$

Note that all integrals can be calculated given the values of $\dot{y}$, y and u.

This is easily solvable using least squares,

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

With:

$$\hat{\theta} = \begin{bmatrix} -\hat{a} \\ -\hat{b} \end{bmatrix} \quad Y = \begin{bmatrix} \dot{y}(t_1) - \dot{y}(t_0) - \int_{t_0}^{t_1} u(t)dt \\ \dot{y}(t_2) - \dot{y}(t_1) - \int_{t_1}^{t_2} u(t)dt \\ \vdots \\ \dot{y}(t_N) - \dot{y}(t_{N-1}) - \int_{t_{N-1}}^{t_N} u(t)dt \end{bmatrix}$$

and

$$\phi = \begin{bmatrix} \int_{t_0}^{t_1} y(t)\dot{y}(t)dt & \int_{t_0}^{t_1} y(t)dt \\ \int_{t_1}^{t_2} y(t)\dot{y}(t)dt & \int_{t_1}^{t_2} y(t)dt \\ \vdots & \vdots \\ \int_{t_{N-1}}^{t_N} y(t)\dot{y}(t)dt & \int_{t_{N-1}}^{t_N} y(t)dt \end{bmatrix}$$

Remark: Given an observation interval, into how many sub-intervals should one divide the integration? A shorter integration interval gives more "data", equations in the LS estimation. More data is generally good, but a shorter integral interval decrease the attenuation of the noise in the data, hence noisier data. It is not obvious how one should balance these two effects to optimize the accuracy of the estimate.

(b) Uncertainty in the initial condition could mean that $\dot{y}(t_0)$ is unknown. A possible modification that can be done is to use modulating functions changing the system to

$$\ddot{y}(t)p_k(t) = -ay(t)\dot{y}(t)p_k(t) - by(t)p_k(t) + u(t)p_k(t)$$

Proceeding as in (a) we get

$$\begin{bmatrix} \int_{t_0}^{t_1} \ddot{y}(t)p_1(t)dt \\ \int_{t_1}^{t_2} \ddot{y}(t)p_2(t)dt \\ \vdots \\ \int_{t_{N-1}}^{t_N} \ddot{y}(t)p_N(t)dt \end{bmatrix} = \begin{bmatrix} \int_{t_0}^{t_1} y(t)\dot{y}(t)p_1(t)dt & \int_{t_0}^{t_1} y(t)p_1(t)dt & \int_{t_0}^{t_1} u(t)p_1(t)dt \\ \int_{t_1}^{t_2} y(t)\dot{y}(t)p_2(t)dt & \int_{t_1}^{t_2} y(t)p_2(t)dt & \int_{t_1}^{t_2} u(t)p_2(t)dt \\ \vdots & \vdots & \vdots \\ \int_{t_{N-1}}^{t_N} y(t)\dot{y}(t)p_N(t)dt & \int_{t_{N-1}}^{t_N} y(t)p_N(t)dt & \int_{t_{N-1}}^{t_N} u(t)p_N(t)dt \end{bmatrix} \begin{bmatrix} -a \\ -b \\ 1 \end{bmatrix}$$

This gives us freedom do modify the properties of the solution and sensitivity to initial conditions. However, it is not straight-forward to how to choose $p_k(\cdot)$.

To remove the dependence on, a possibly wrong value, of $\dot{y}(t_0)$ the integral of the first row can be re-written

$$\int_{t_0}^{t_1} \ddot{y}(t)p_1(t)dt = [\dot{y}(t)p_1(t)]_{t_0}^{t_1} - \int_{t_0}^{t_1} \dot{y}(t)\dot{p}_1(t)dt =$$

$$[\dot{y}(t)p_1(t)]_{t_0}^{t_1} - [y(t)\dot{p}_1(t)]_{t_0}^{t_1} + \int_{t_0}^{t_1} y(t)\ddot{p}_1(t)dt$$

So if $p_1(\cdot)$ chosen so that $p_1(t_0) = 0$ and $\dot{p}_1(t_0) = 0$ we can minimize the impact of the initial conditions. However, it would also be possible to exclude the first equation to obtain this.

38 Back to problem

(a) The weightning of data decay as λ^τ where τ is the age of the data measured in time steps. Hence, the weight after 100 steps should be less then 0.15, ie,

$$\lambda^{100} < 0.15$$

which gives

$$100\ln(\lambda) < \ln(0.15) \Rightarrow \lambda < e^{\frac{\ln(0.15)}{100}} = 0.9812$$

(b) Typically, the initial conditions will not be correct and this gives a bias in the estimate. Due to the forgetting-factor, its influence decay as λ^t. That is, it dissappear assymptotically.

39 Back to problem

(a) The general least square solution based on data up to t is

$$\hat{\theta}(t) = (\sum_{i=1}^{t} \varphi(i)\varphi(i)^T)^{-1} \sum_{j=1}^{t} \varphi(j)y(j)$$

Since we have a rectangular window of length m, only the m last data should be included hence our estimate is

$$\hat{\theta}(t) = (\sum_{i=t-m+1}^{t} \varphi(i)\varphi(i)^T)^{-1} \sum_{j=t-m+1}^{t} \varphi(j)y(j) \tag{1}$$

What remains is to re-write this estimate to the form given in the problem. Let's start with the matrix and call it $P(t)$, ie,

$$P(t) = (\sum_{i=t-m+1}^{t} \varphi(i)\varphi(i)^T)^{-1}$$

Note, we cannot be sure that this is the $P(t)$ given in the problem, so we make a guess at this point. Further down we will see if we are right. If not, notation should be changed.

To write $P(t)$ on a recursive form, it should be expressed as a function of $P(t-1)$ and the update. We will also have use of the following matrix relation, called the *Matrix Inversion Lemma*

$$(A + B\,C\,D)^{-1} = A^{-1} - A^{-1}B(C^{-1} + D\,A^{-1}B)^{-1}D\,A^{-1}$$

It looks like magic. Verify that it is true!

Hence $P(t)$ becomes

$$\begin{aligned} P(t) = \left(\sum_{i=t-m}^{t-1} \varphi(i)\varphi(i)^T + \varphi(t)\varphi^T(t) - \varphi(t-m)\varphi^T(t-m) \right)^{-1} &= \\ \left(P^{-1}(t-1) + \varphi(t)\varphi^T(t) - \varphi(t-m)\varphi^T(t-m)\right)^{-1} &= \\ P(t-1) - P(t-1)\begin{bmatrix}\varphi(t) & \varphi(t-m)\end{bmatrix} & \\ \left(I + \begin{bmatrix}\varphi^T(t) \\ -\varphi^T(t-m)\end{bmatrix} P(t-1) \begin{bmatrix}\varphi(t) & \varphi(t-m)\end{bmatrix}\right)^{-1} \begin{bmatrix}\varphi^T(t) \\ -\varphi^T(t-m)\end{bmatrix} P(t-1) &= \\ P(t-1) - \begin{bmatrix}K_1(t) & K_2(t)\end{bmatrix} \begin{bmatrix}\varphi^T(t) \\ -\varphi^T(t-m)\end{bmatrix} P(t-1) & \end{aligned} \tag{2}$$

where we used the Matrix Inversion Lemma in the second last step with

$$A = P^{-1}(t-1), \qquad B = \begin{bmatrix}\varphi(t) & \varphi(t-m)\end{bmatrix},$$
$$C = 1, \qquad D = \begin{bmatrix}\varphi^T(t) \\ -\varphi^T(t-m)\end{bmatrix}$$

and in the last step we identified

$$\begin{bmatrix}K_1(t) & K_2(t)\end{bmatrix} = P(t-1)\begin{bmatrix}\varphi(t) & \varphi(t-m)\end{bmatrix}\left(I + \begin{bmatrix}\varphi^T(t) \\ -\varphi^T(t-m)\end{bmatrix} P(t-1)\begin{bmatrix}\varphi(t) & \varphi(t-m)\end{bmatrix}\right)^{-1} \tag{3}$$

according to the definition in the problem.

Before moving on, we can obtain the following relation between $P(t)$ and $[K_1(t)\ K_2(t)]$ which is useful later on. We do this by multiplying (2) with $\begin{bmatrix}\varphi(t) & \varphi(t-m)\end{bmatrix}$ from the right side, add and subtract $[K_1(t)\ K_2(t)]$, and using (3) in the last step,

$$\begin{aligned}
&P(t)\begin{bmatrix}\varphi(t) & \varphi(t-m)\end{bmatrix} = \\
&P(t-1)\begin{bmatrix}\varphi(t) & \varphi(t-m)\end{bmatrix} - \begin{bmatrix}K_1(t) & K_2(t)\end{bmatrix}\begin{bmatrix}\varphi^T(t) \\ -\varphi^T(t-m)\end{bmatrix} P(t-1)\begin{bmatrix}\varphi(t) & \varphi(t-m)\end{bmatrix} = \\
&P(t-1)\begin{bmatrix}\varphi(t) & \varphi(t-m)\end{bmatrix} + \begin{bmatrix}K_1(t) & K_2(t)\end{bmatrix} \\
&- \begin{bmatrix}K_1(t) & K_2(t)\end{bmatrix}(I + \begin{bmatrix}\varphi^T(t) \\ -\varphi^T(t-m)\end{bmatrix} P(t-1)\begin{bmatrix}\varphi(t) & \varphi(t-m)\end{bmatrix}) = \\
&P(t-1)\begin{bmatrix}\varphi(t) & \varphi(t-m)\end{bmatrix} + \begin{bmatrix}K_1(t) & K_2(t)\end{bmatrix} - P(t-1)\begin{bmatrix}\varphi(t) & \varphi(t-m)\end{bmatrix} = \\
&\begin{bmatrix}K_1(t) & K_2(t)\end{bmatrix}
\end{aligned} \tag{4}$$

Next, let us formulate the parameter update in a recursive way. In the problem the form of the update is given, so we kno in which way we are moving. Starting from

(1) we obtain,

$$
\begin{aligned}
\hat{\theta}(t) = P(t) \sum_{j=t-m+1}^{t} \varphi(j)y(j) = \\
P(t)(\sum_{j=t-m}^{t-1} \varphi(j)y(j) + \varphi(t)y(t) - \varphi(t-m)y(t-m)) = \\
P(t)(P^{-1}(t-1)\hat{\theta}(t-1) + \varphi(t)y(t) - \varphi(t-m)y(t-m)) = \\
P(t)((P^{-1}(t)-\varphi(t)\varphi^T(t)+\varphi(t-m)\varphi^T(t-m))\hat{\theta}(t-1)+\varphi(t)y(t)-\varphi(t-m)y(t-m)) = \\
\hat{\theta}(t-1)+P(t)\varphi(t)(y(t)-\varphi^T(t)\hat{\theta}(t-1))-P(t)\varphi(t-m)(y(t-m)-\varphi^T(t-m)\hat{\theta}(t-1)) = \\
\hat{\theta}(t-1) + P(t)\varphi(t)\varepsilon(t,\hat{\theta}(t-1)) - P(t)\varphi(t-m)\varepsilon(t-m,\hat{\theta}(t-1)) = \\
\hat{\theta}(t-1) + \begin{bmatrix} K_1(t) & K_2(t) \end{bmatrix} \begin{bmatrix} \varepsilon(t,\hat{\theta}(t-1)) \\ -\varepsilon(t-m,\hat{\theta}(t-1)) \end{bmatrix} \quad (5)
\end{aligned}
$$

To conclude (2), (3), and (5) gives the recursive algorithm,

$$
\begin{aligned}
\hat{\theta}(t) &= \hat{\theta}(t-1) + K_1(t)\varepsilon(t,\hat{\theta}(t-1)) - K_2(t)\varepsilon(t-m,\hat{\theta}(t-1)) \\
(K_1(t)\ \ K_2(t)) &= P(t-1)(\varphi(t)\ \ \varphi(t-m)) \\
&\quad \cdot \left\{ I + \begin{bmatrix} \varphi^T(t) \\ -\varphi^T(t-m) \end{bmatrix} P(t-1)(\varphi(t)\ \ \varphi(t-m)) \right\}^{-1} \\
P(t) &= P(t-1) - (K_1(t)\ \ K_2(t)) \begin{bmatrix} \varphi^T(t) \\ -\varphi^T(t-m) \end{bmatrix} P(t-1)
\end{aligned}
$$

(b) Assume $\dim\varphi(t) = d$. $P(t)$ requires $d(d+1)/2$ parameters (symmetric), $\hat{\theta}$ requires d, ε requires m, φ requires $d \cdot m$, K requires $2 \cdot d$ which makes

$$d^2/2 + 5\,d/2 + (d+1)m$$

RLS, on the other hand $P(t)$ requires $d(d+1)/2$ parameters (symmetric), $\hat{\theta}$ requires d, ε requires 1, φ requires d, K requires d which makes

$$d^2/2 + 7\,d/2 + 1$$

The rectangular window requires memory increasing linearly with m. The memory need of RLS is independent on the value of the forgetting factor λ.

(c) A rectangular window gives an algorithms that totally forgotten values outside the window. After a step change of the parameters, an unbiased estimate is obtained after m samples. The RLS, however, never forget totally. On the other hand, the

RLS makes a quicker response to s change since the recent data have larger influence. Depending on the problem, one of the algorithms will be best. If the parameters are changing all the time and in small steps, RLS probably make sense since also old data are relevant. On the other hand, a situation where the parameters change only at some occations, and if the new values are independent of the previous ones, then the feature that old data are totally forgotten, can be a good option.

Remark: there are more alternatives. For the situation where there are sudden parameter changes, a detection algorithm which re-initialize the algorithm and totally forgets old information can be a good idea.

40 Back to problem

The plant is given by

$$G(s) = \frac{b}{s(1+s)} \tag{1}$$

and the controller by

$$u(t) = -k\,y(t) + k\,r(t) \tag{2}$$

The reference model for the closed-loop system is

$$G_c(s) = \frac{1}{s^2+s+1} \tag{3}$$

We start by forming the closed-loop system

$$Y(s) = G(s)U(s) = \frac{b(-k\,Y(s) + k\,R(s))}{s(1+s)} \Leftrightarrow$$

$$Y(s) = \frac{b\,k}{s^2+s+b\,k}R(s)$$

So this gives us our model if $k = \frac{1}{b}$ but b is unknown. k should be adapted using the MIT rule,

$$\frac{dk}{dt} = -\gamma e \frac{de}{dk} \tag{4}$$

With y_m indicating the output from the reference model e becomes

$$\begin{aligned} E(s) &= Y - Y_m \\ &= \left(\frac{bk}{s^2+s+bk} - \frac{1}{s^2+s+1}\right)R(s) \end{aligned} \tag{5}$$

and its derivative

$$\frac{dE}{dk} = \left(\frac{b}{s^2+s+b\,k} - \frac{b^2k}{(s^2+s+b\,k)^2}\right)R(s) = b\frac{1}{s^2+s+b\,k}(R(s) - Y(s))$$

where we identify $Y(s)$, and in that way we can use the measured plant output. However, we cannot realise this as we do not know b. But if we assume that we are close to the correct estimate and aproximate $b\,k = 1$, we obtain

$$\frac{de}{dk} = b\frac{1}{s^2+s+1}(R(s) - Y(s)) \tag{6}$$

This means that the adaption rule will not be strictly the MIT rule when $k \neq 1/b$. There is still a b in the numerator of (6). We will combine it with the adaption parameter γ which we can set to any value. That is,we define $\gamma\, b = \gamma_b$.

To build a simulation model we need to define an input signal $r(t)$ and a time-varying plant (1) depending on parameter $b(t)$. Then we need to model (2-6), note that they do not depend on the unknown parameter b. Properly connected according to the notations, we have a simulation model of the adaptive controlled system. Signals to display to illustrate the result are $k(t)$ and $1/b(t)$ together, $e(t)$, and $y(t)$.

With some abuse of notation, mixing time-domain and Laplace domain, we can insert the expressions in (6) to obtain the differential equation for k,

$$\frac{dk}{dt} = -\gamma_b\left[\frac{1}{s^2+s+1}(R(s) - Y(s))\right]e(t)$$

This is a nonlinear differential equation where $(R(s) - Y(s))$ enters through a linear filter, indicated with the Laplace part. The result of that part is then multiplied by $e(t)$. k dependency is hidden in $Y(s)$ and in $e(t)$. Properties of the equation depend on γ_b which need to be chosen so the adaption is stable... Try different values in simulation. What is a good value? Can it be too high or too low? What happens in these cases? Use the Simulink model `adaptivety2referencefollowing_AdaptiveControl1.slx`, which can be found under “ExerciseSolutions” on the course web, to experiment with γ_b.

41 Back to problem

We have adaption law,

$$\theta(t) = -\gamma_1 r(t)e(t) - \gamma_2\int^t r(\tau)e(\tau)d\tau \tag{1}$$

plant

$$Y(s) = \frac{\theta}{s}R(s)$$

where θ is the unknown, possibly varying, parameter, and the reference model

$$Y_m(s) = \frac{\theta_0}{s}R(s)$$

The error is defined as the difference between the plant output and the output of the reference model

$$e = y - y_m \tag{2}$$

Since we are looking for the differencial equation of $e(t)$, we take the derivative of (2)

$$\frac{de}{dt} = (\theta(t) - \theta_0)r(t)$$

once more gives

$$\frac{d^2e}{dt^2} = \dot{\theta}(t)r(t) + (\theta(t) - \theta_0)\dot{r}(t) \tag{3}$$

Derivative of (1),

$$\dot{\theta}(t) = -\gamma_1\dot{r}(t)e(t) - \gamma_1 r(t)\dot{e}(t) - \gamma_2 r(t)e(t)$$

Insert this in (3)

$$\frac{d^2e}{dt^2} = (-\gamma_1\dot{r}(t)e(t) - \gamma_1 r(t)\dot{e}(t) - \gamma_2 r(t)e(t))r(t) + (\theta(t) - \theta_0)\dot{r}(t)$$

Re arrange this

$$\frac{d^2e}{dt^2} + \gamma_1 r^2(t)\dot{e}(t) + (\gamma_1\dot{r}(t)r(t) + \gamma_2 r^2(t))e(t) = (\theta(t) - \theta_0)\dot{r}(t)$$

With a constant reference this becomes

$$\frac{d^2e}{dt^2} + \gamma_1 r^2(t)\dot{e}(t) + \gamma_2 r^2(t)e(t) = 0$$

So the error has second order dynamics whose properties are defined by γ_1, γ_2 and the constant value of $r(t)$.

42 Back to problem We have the plant

$$G(s) = \frac{b}{(s+a)(s+p)} \tag{1}$$

The general contral law is given by

$$J\,U(s) = -K\,Y(s) + T\,R(s) \tag{2}$$

Inserting the controller

$$Y(s) = \frac{b}{(s+a)(s+p)J}(-K\,Y(s) + T\,R(s))$$

and then reformulate, gives closed loop system

$$Y(s) = \frac{b\,T}{(s+a)(s+p)J + K\,b}R(s) \tag{3}$$

The decired closed-loop system is given by

$$Y_m(s) = \frac{\omega^2}{s^2 + 2\zeta\,\omega\,s + \omega^2}R(s) \tag{4}$$

Make (3) and (4) equal, ie,

$$\frac{b\,T}{(s+a)(s+p)J + K\,b} = \frac{\omega^2}{s^2 + 2\zeta\,\omega\,s + \omega^2} \Longleftrightarrow$$

$$(s+a)(s+p)J + K\,b\,\omega^2 = b\,T(s^2 + 2\zeta\,\omega\,s + \omega^2)$$

For the controler to be proper the degree of J must be at least the same as K and R. Let us assume first order polynomials for all three of them, and evaluate if that is sufficient. In T we also introduce a parameter so it can ensure the correct gain. Hence, introduce,

$$J = (j_0 + s) \qquad K = k_0 s + k_1 \qquad T = t_0(s + a_0)$$

We can conclude correct t_0 will be

$$t_0 \approx \frac{\omega^2}{b}$$

Inserting that we obtain

$$(s+a)(s+p)(s+j_0) + (k_0 s + k_1)b = (s + a_0)(s^2 + 2\zeta\,\omega\,s + \omega^2)$$

The polynomial T should be the observer dynamics. You can see that by the fact that $R(s)$ is multiplied by T but in the decired closed loop T is not present. Since p is known we can use this to do a pole cancellation and decide the values for a_0 and k_1,

$$a_0 = p \qquad\qquad\qquad k_1 = k_0 p$$

Inserting this gives

$$(s+a)(s+p)(s+j_0) + (s+p)k_0 b = (s+p)(s^2 + 2\zeta\,\omega\,s + \omega^2) \Longleftrightarrow$$

$$(s+a)(s+j_0) + k_0 b = (s^2 + 2\zeta\omega s + \omega^2)$$

This equation is always solvable in the 2 parameters, k_0 and j_0. This means that the suggested control structure will give the decired closed loop system if the parameters are tuned to the correct values. It remains to find the tuning law for t_0, k_0, and j_0.

Inserting the controller polynomials in the closed loop system (3) gives

$$Y(s) = \frac{b\,t_0(s+p)}{(s+a)(s+p)(s+j_0)+(s+p)k_0 b}R(s) = \frac{b\,t_0}{(s+a)(s+j_0)+k_0 b}R(s)$$

The control error is

$$e = y - y_m = (\frac{b\,t_0}{(s+a)(s+j_0)+k_0 b} - \frac{\omega^2}{s^2+2\zeta\,\omega\,s+\omega^2})$$

Taking the derivative with respect to the three parameters gives

$$\frac{de}{dt_0} = \frac{b}{(s+a)(s+j_0)+k_0 b}R(s) \approx \frac{b}{s^2+2\zeta\,\omega\,s+\omega^2}R(s)$$
$$\frac{\partial e}{\partial k_0} = -\frac{b^2 t_0}{((s+a)(s+j_0)+k_0 b)^2}R(s) = -\frac{b}{(s+a)(s+j_0)+k_0 b}Y(s) \approx -\frac{b}{s^2+2\zeta\,\omega\,s+\omega^2}Y(s)$$
$$\frac{\partial e}{\partial j_0} = -\frac{(s+a)b\,t_0}{((s+a)(s+j_0)+k_0 b)^2}R(s) = -\frac{(s+a)}{(s+a)(s+j_0)+k_0 b}Y(s) \approx -\frac{(s+a)}{s^2+2\zeta\,\omega\,s+\omega^2}Y(s)$$

For the two first parameters, the unknown factor b in the numerator can be inegrated into the adaption parameter γ. For the last ona, however, we have a zero at the unknown value a. This can be handled by using the plant relation (1) which has a pole in a,

$$Y(s) = \frac{b}{(s+a)(s+p)}U(s)$$

Inserting this gives us

$$\frac{\partial e}{\partial j_0} \approx \frac{b}{(s^2+2\zeta\omega s+\omega^2)(s+p)}U(s)$$

which works fine since p is known. Putting this together, the MIT-rule gives us the following adaption

$$\frac{dt_0}{dt} = -e(t)\frac{\gamma}{s^2+2\zeta\omega s+\omega^2}R(s)$$
$$\frac{dk_0}{dt} = e(t)\frac{\gamma}{s^2+2\zeta\omega s+\omega^2}Y(s)$$
$$\frac{dj_0}{dt} = e(t)\frac{\gamma}{(s^2+2\zeta\omega s+\omega^2)(s+p)}U(s)$$

This adaption law can be implemented using the signals $r(t)$, $u(t)$, and $y(t)$. Note that the answer is not given on a strict correct way since we mix Laplace notation and time notation for non-linear differential equations.

43 Back to problem

Given: