Räkneövningar Empirisk modellering

Bengt Carlsson
Systems and Control
Dept of Information Technology, Uppsala University

25th February 2009

Abstract

Räkneuppgifter samt lite kompletterande teori.

Contents

1 L1-Linear regression

1.1 Problems

Problem 1. A linear trend model
a) Consider a linear regression model

$$\hat{y}(t) = a + bt$$

Calculate the least squares estimate for the following two cases

1. The data are $y(1), y(2), \ ... \ y(N)$. For this case, use $S_o = \sum_{t=1}^{N} y(t)$, $S_1 = \sum_{t=1}^{N} ty(t)$
2. The data are $y(-N), y(-N+1), \ ... \ y(N)$. For this case, use $S_o = \sum_{t=-N}^{N} y(t)$, $S_1 = \sum_{t=-N}^{N} ty(t)$

Hints:

$$\begin{aligned} \sum_{t=1}^{N} t &= \frac{N(N+1)}{2} \\ \sum_{t=1}^{N} t^2 &= \frac{N(N+1)(2N+1)}{6} \end{aligned}$$

b) Suppose that the parameter a is estimated with

$$\begin{aligned} \hat{a} &= \frac{S_o}{N}, \quad \text{case 1 above} \\ \hat{a} &= \frac{S_o}{2N+1}, \quad \text{case 2 above} \end{aligned}$$

The parameter b is estimated by the least squares method using the model structure:

$$\hat{y}(t) - \hat{a} = bt$$

Calculate $\hat{b}$ for the two cases and compare with the estimate obtained in (a).

c) Assume that the data $y(1), y(2), \ ... \ y(N)$ is generated by

$$y(t) = a_o + b_o t + e(t)$$

where $e(t)$ is white noise with variance λ. Calculate the variance of the quantity $s(t) = \hat{a} + \hat{b}t$. What is the variance for $t = 1$ and $t = N$? For which t is the variance minimal?
Hint:
Let $\theta = (a \ \ b)^T$ and $\varphi(t) = (1 \ \ t)^T$. Then

$$\text{var } s(t) = \varphi(t)^T P \varphi(t)$$

where $P = \text{var } \hat{\theta}$.

2) Some accuracy results for linear trend models

a) Assume that the data $y(1), y(2),$ ... $y(N)$ is generated by

$$y(t) = a_o + b_o t + e(t)$$

where $e(t)$ is white noise with variance λ. The parameters in a linear trend model

$$\hat{y}(t) = a + bt$$

are estimated with the least squares method. Calculate the variance of $\hat{b}$.
b) Assume that we differenced the data and introduce a new signal

$$z(t) = y(t) - y(t-1) \quad t = 2, 3, ...N$$

We then have that the data $z(t)$ obeys

$$z(t) = b_o + w(t) \tag{1}$$

where the new noise source $w(t) = e(t) - e(t-1)$. The parameter b_o may then be estimated from the following model

$$\hat{z}(t) = b$$

Calculate the variance of $\hat{b}$ and compare with the accuracy obtained in (2a).
Hint:
Note that the noise $w(t)$ in (1) is not white. Hence the expression (**??**) need to be used when calculating the variance.

3. The problem of collinearity.
Consider the following model

$$\hat{y}(t) = au_1(t) + bu_2(t)$$

Here $u_1(t)$ and $u_2(t)$ are two measured input signals. Suppose that the data is generated by

$$y(t) = a_o u_1 + bu_2$$

where $u_1 = K$ and $u_2 = L$, that is two constant signals with amplitude K and L are used as input signals. Show that $\det \Phi^T\Phi = 0$, and hence the least squares method can not be used.
Remark:
The columns in Φ must be linearly independent for $(\Phi^T\Phi)^{-1}$ to exist.

1.2 Solutions

1) In general we have

$$\hat{\theta} = [\sum_{t=1}^{N} \varphi(t)\varphi^T(t)]^{-1} \sum_{t=1}^{N} \varphi(t)y(t)) = [\Phi^T\Phi]^{-1}\Phi^T Y \tag{2}$$

For the considered model $\theta = (a \;\; b)^T$, and $\varphi = (1 \;\; t)^T$.

Case (i). Data $y(1), \ldots, y(N)$.

$$\begin{aligned}\hat{\theta} &= \begin{bmatrix} \sum_{t=1}^N 1 & \sum_{t=1}^N t \\ \sum_{t=1}^N t & \sum_{t=1}^N t^2 \end{bmatrix}^{-1} \begin{bmatrix} S_o \\ S_1 \end{bmatrix} = \begin{bmatrix} N & \frac{N(N+1)}{2} \\ \frac{N(N+1)}{2} & \frac{N(N+1)(2N+1)}{6} \end{bmatrix}^{-1} \begin{bmatrix} S_o \\ S_1 \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{N(N-1)}[2(2N+1)S_o - 6S_1] \\ \frac{6}{N(N-1)(N+1)}[2S_1 - (N+1)S_o] \end{bmatrix}\end{aligned}$$

Case (ii). Data $-y(N), \ldots, y(N)$. This gives $2N+1$ data points. All sums will have the form $\sum_{t=-N}^{N}$.

$$\hat{\theta} = \begin{bmatrix} 2N+1 & 0 \\ 0 & \frac{N(N+1)(2N+1)}{3} \end{bmatrix}^{-1} \begin{bmatrix} S_o \\ S_1 \end{bmatrix} = \begin{bmatrix} \frac{1}{2N+1} S_o \\ \frac{3}{N(N+1)(2N+1)} S_1 \end{bmatrix} \tag{3}$$

b)
The model $y(t) - \hat{a} = bt$ gives the least squares estimate

$$\hat{b} = \frac{\sum t(y(t) - \hat{a})}{\sum t^2} = \frac{S_1 - \hat{a} \sum t}{\sum t^2}$$

case (i):

$$\hat{b} = \frac{3}{N(N+1)(2N+1)}[2S_1 - (N+1)S_o]$$

which is not equal to the solution in a) and is therefore wrong.

case (ii):

$$\hat{b} = \frac{S_1 - 0}{\sum t^2} = \frac{3}{N(N+1)(2N+1)} S_1$$

which is equal to the solution in a) and is therefore correct.

c)

$$P = \lambda[\sum_{t=1}^{N} \varphi(t)\varphi^T(t)]^{-1} = \lambda \begin{bmatrix} \sum_{t=1}^N 1 & \sum_{t=1}^N t \\ \sum_{t=1}^N t & \sum_{t=1}^N t^2 \end{bmatrix}^{-1} = \lambda \begin{bmatrix} N & \frac{N(N+1)}{2} \\ \frac{N(N+1)}{2} & \frac{N(N+1)(2N+1)}{6} \end{bmatrix}^{-1} \tag{4}$$

Straightforward calculations using the P above and $\varphi = (1 \;\; t)^T$ give

$$\text{var } s(t) = \varphi(t)^T P \varphi(t) = \frac{12\lambda}{N(N+1)(N-1)}[t^2 - t(N+1) + \frac{(N+1)(2N+1)}{6}]$$

and

$$\text{var } s(1) = \frac{12\lambda}{N(N+1)(N-1)}[1^2 - 1(N+1) + \frac{(N+1)(2N+1)}{6}] \approx \frac{4\lambda}{N} \text{ when N is large}$$

$$\text{var } s(N) = \frac{12\lambda}{N(N+1)(N-1)}[N^2 - N(N+1) + \frac{(N+1)(2N+1)}{6}] = \text{var } s(1) \approx \frac{4\lambda}{N} \text{ when N is large}$$

The t giving minimal variance is obtained by setting

$$\frac{d}{dt}\text{var } s(t) = \frac{12\lambda}{N(N+1)(N-1)}[2t - (N+1)] = 0$$

which gives $t = \frac{N+1}{2}$. Hence, the minimal variance is in the middle of the observation interval. Furthermore

$$\text{var } s(\frac{N+1}{2}) = \frac{12\lambda}{N(N+1)(N-1)}[\frac{(N+1)^2}{4} - \frac{(N+1)^2}{2} + \frac{(N+1)(2N+1)}{6}] = \frac{\lambda}{N}$$

For large data sets, the standard deviation decreases from $2\sqrt{(\lambda/N)}$ at $s(1)$ and $s(N)$ to $\sqrt{(\lambda/N)}$ in the middle of the interval.

2) For a linear regression we have that

$$\text{cov}\hat{\theta} = \lambda[\Phi^T\Phi]^{-1} = \lambda[\sum_{t=1}^{N} \varphi(t)\varphi^T(t)]^{-1}$$

For a linear trend $\varphi = (1 \;\; t)^T$. This gives

$$\text{cov}\hat{\theta} \;\; = \;\; \lambda \begin{bmatrix} \sum_{t=1}^N 1 & \sum_{t=1}^N t \\ \sum_{t=1}^N t & \sum_{t=1}^N t^2 \end{bmatrix}^{-1} = \lambda \begin{bmatrix} N & \frac{N(N+1)}{2} \\ \frac{N(N+1)}{2} & \frac{N(N+1)(2N+1)}{6} \end{bmatrix}^{-1}$$

The variance of $\hat{b}$ is then found (the matrix above need to be inverted) to be $\lambda\frac{12}{N(N^2-1)}$.

b) When the data is differentiated we have that the system is

$$z(t) = b_o + w(t)$$

where $w(t) = e(t) - e(t-1)$. We have

$$\begin{aligned} R_w(0) &= E\{w^2(t)\} = E\{[e(t) - e(t-1)]^2\} = 2\lambda \\ R_w(1) &= E\{w(t+1)w(t)\} = E\{(e(t+1) - e(t))(e(t) - e(t-1))\} = -\lambda \\ R_w(k) &= 0 \;\; k > 1 \end{aligned}$$

The noise is not white and in order to calculate the variance we need to calculate $R = E\{ww^T\}$ where $w = (w(1), w(2), \ldots, w(N-1))^T$. See Section 4.3 in "Linear regression". Note that when the data is differentiated one data point is lost. We have

$$R = \lambda \begin{bmatrix} 2 & -1 & \ldots & 0 \\ -1 & 2 & \ldots & 0 \\ \vdots & & \ddots & -1 \\ 0 & \ldots & -1 & 2 \end{bmatrix}$$

For the model $\hat{z(t)} = b$ we have $\varphi(t) = 1$ and $\Phi = (1, \ldots, 1)^T$ (with $N-1$ rows).

We can now calculate the variance of $\hat{b}$ (the least squares estimate) from the covariance matrix (which becomes the variance since $\hat{\theta}$ is a scalar)

$$\text{cov}\hat{\theta} = \text{var}\hat{b} = (\Phi^T\Phi)^{-1}\Phi^T R\Phi(\Phi^T\Phi)^{-1} = \lambda\frac{1}{N-1}2\frac{1}{N-1} = \lambda\frac{2}{(N-1)^2} \qquad (5)$$

It is easily seen that this expression is larger than the one in a).

3) We have $\varphi(t) = (u_1(t) \;\; u_2(t))^T$, with $u_1(t) = K$ and $u_2(t) = L$. We assume the number of data to be N. This gives the $(N|2)$ matrix:

$$\Phi = \begin{bmatrix} K & L \\ \vdots & \vdots \\ K & L \end{bmatrix}$$

and

$$\Phi^T \Phi = \begin{bmatrix} NK^2 & NKL \\ NKL & NL^2 \end{bmatrix}$$

We then see that $\det(\Phi^T \Phi) = 0$ and hence the LS method can not be used. It is not possible to determine the parameters uniquely from this data set (which also should be intiutively clear).

2 L2-Stochastic processes and discrete time system

2.1 Problems

1. Static gain and stability.
Consider the discrete time system

$$y(t) = H(q)u(t), \quad \text{where } H(q) = \frac{0.1 + 1q^{-1}}{1 - 0.2q^{-1}}$$

Calculate, poles, zeros and static gain of the system. Is the system stable?

2. Spectrum.
The following stochastic process is given:

$$y(t) - 0.2y(t-1) = e(t) - 0.1e(t-1)$$

where $e(t)$ is white noise with zero mean and variance λ.

a) Determine the spectrum of y; $\Phi_y(\omega)$.

b) Determine the cross spectrum $\Phi_{ye}(\omega)$.

3. Covariance function and spectrum.
Consider the following system

$$y(t) = H(q)u(t), \quad \text{where } H(q) = \frac{bq^{-1}}{1 + aq^{-1}}$$

The input signal has the following covariance function $R_u(0) = 1$, $R_u(1) = R_u(-1) = 0.5$, $R_u(\tau) = 0$ for $|\tau| > 1$. Calculate the spectrum of the output signal.

4. Covariance functions
Calculate the covariance function for the following stochastic processes where $e(t)$ is white noise with variance λ.

a)

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

b)

$$y(t) + ay(t-1) = e(t) \quad |a| < 1$$

2.2 Solutions

1)

$$H(q) = \frac{0.1 + 1q^{-1}}{1 - 0.2q^{-1}} = \frac{0.1q + 1}{q - 0.2}$$
$$H(z) = \frac{0.1z + 1}{z - 0.2}$$

We immediately see that the system has one zero in $z = -10$ and one pole in $z = 0.2$. Informally we could as well solve the roots with the q-operator (but not with q^{-1}-operator).
The static gain is obtained by setting $q = 1$ (or $z = 1$). We have $H(1) = \frac{0.1+1}{1-0.2} = 1.1/0.8$
The system is stable since all poles are inside the unit circle.

2) We can write the system as

$$H(q) = \frac{1 - 0.1q^{-1}}{1 - 0.2q^{-1}} = \frac{q - 0.1}{q - 0.2}$$

a) First note that since $e(t)$ is white noise whith variance λ $\Phi_e(\omega) = \lambda$. By using

$$\Phi_y(\omega) = |H(e^{iw})|^2 \Phi_e(\omega)$$

we get

$$\Phi_y(\omega) = |\frac{e^{i\omega} - 0.1}{e^{i\omega} - 0.2}|^2 \lambda = \frac{(e^{i\omega} - 0.1)(e^{-i\omega} - 0.1)}{(e^{i\omega} - 0.2)(e^{-i\omega} - 0.2)} \lambda = \frac{1.01 - 0.2\cos(\omega)}{1.04 - 0.4\cos(\omega)} \lambda$$

b) The cross spectra for a discrete time system $H(q)$ is given by

$$\Phi_{ye}(\omega) = H(e^{iw}) \Phi_e(\omega)$$

which directly gives

$$\Phi_{ye}(\omega) = \frac{e^{i\omega} - 0.1}{e^{i\omega} - 0.2} \lambda$$

3) The definition of spectral density (spectrum) is

$$\Phi_u(\omega) = \sum_{k=-\infty}^{k=\infty} R_u(k) e^{-i\omega k}$$

The given covariance function gives

$$\Phi_u(\omega) = 0.5e^{-i\omega} + 1 + 0.5e^{i\omega} = 1 + \cos(\omega)$$

and the output spectral density is

$$\Phi_y(\omega) = |H(e^{iw}|^2\Phi_u(\omega) = H(e^{iw})H(e^{-iw})\Phi_u(\omega) = \frac{b}{e^{iw}+a}\frac{b}{e^{-iw}+a}(1+\cos(\omega)) = \frac{b^2(1+\cos(\omega))}{1+a^2+2a\cos(\omega)}$$

4 a) The MA(1) process is

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

and we directly get

$$\begin{aligned}
R_y(0) &= Ey^2(t) = E(e(t)+ce(t-1))(e(t)+ce(t-1)) = (1+c^2)\lambda \\
R_y(1) &= Ey(t+1)y(t) = E(e(t+1)+ce(t))(e(t)+ce(t-1)) = c\lambda \\
R_y(k) &= Ey(t+k)y(t) = E(e(t+k)+ce(t+k-1))(e(t)+ce(t-1)) = 0, \quad \text{for } k > 1
\end{aligned}$$

b) The covariance function for the AR(1) process

$$y(t) - ay(t-1) = e(t) \quad |a| < 1$$

can be solved (this also holds for a general AR(n) process) by the so called Yule Walker equations (not a course requirement). Basically the idea is to multiply the AR process with $y(t-k)$ and take expectations. We have to distinguish between $k = 0$, and $k > 0$. For $k = 0$ we get

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

Taking expectations give

$$R_y(0) + aR_y(1) = \lambda$$

For $k > 0$ we get

$$y(t-k)(y(t) + ay(t-1)) = e(t)y(t-k)$$

Taking exectations give

$$R_y(k) + aR_y(k-1) = 0$$

We hence end up with a lset of linear equations. For $k = 0,\ 1$ we get

$$\begin{pmatrix} 1 & a \\ a & 1 \end{pmatrix}\begin{pmatrix} R_y(0) \\ R_y(1) \end{pmatrix} = \begin{pmatrix} \lambda \\ 0 \end{pmatrix} \tag{6}$$

with solution

$$\begin{aligned}
R_y(0) &= \frac{\lambda}{1-a^2} \\
R_y(1) &= (-a)\frac{\lambda}{1-a^2}
\end{aligned}$$

It is easy to see that

$$R_y(k) = (-a)^{|k|}\frac{\lambda}{1-a^2}$$

2.3 Egenskaper kovariansfunktioner

Låt $x(t)$ vara en stokastisk stationär[1] process med medelvärde $\mathrm{E}x(t) = 0$. Processens kovariansfunktion definieras av

$$R(\tau) = E[x(t+\tau)x(t)]$$

Följande relationer gäller

- $R(0) \pm R(\tau) \geq 0$.

  Bevis: $E[x(t+\tau) \pm x(t)]^2 = E[x(t+\tau)]^2 + 2E[x(t)]^2 \pm 2E[x(t+\tau)x(t)] = R(0) + R(0) \pm 2R(\tau) = 2(R(0) \pm R(\tau))$. Uttrycket $E[x(t+\tau) \pm x(t)]^2$ är alltid positivt alltså är även $R(0) \pm R(\tau)$ positivt varav påståendet följer.

- $R(0) \geq |R(\tau)|$. Följer direkt av ovanstående bevis.
- $R(-\tau) = R(\tau)$, dvs kovariansfunktionen är symmetrisk kring origo.

  Bevis $R(-\tau) = E[x(t-\tau)x(t)] = E[x(t)x(t-\tau)] =$ sätt $t = s + \tau = E[x(s+\tau)x(s)] = R(\tau)$.

- Om $|R(\tau)| = R(0)$ för något $\tau \neq 0$ så är x(t) periodisk.Bevis: utlämmnas.

Korskovariansfunktionen beskriver samvariatonen mellan två stokastiska processer $x(t)$ och $y(t)$ och definieras

$$R_{xy}(\tau) = E[x(t+\tau)y(t)]$$

Vi ger följande relationer utan bevis:

- $R_{xy}(\tau) = R_{yx}(-\tau)$
- $R_{xy}(\tau) \neq R_{xy}(-\tau)$ i allmänhet.
- $R_{xy}(\tau) = 0$ för alla $\tau \Rightarrow x$ och y är okorrelerade.

[1]En process är stationär om dess egenskaper (fördelningar) ej beror av absolut tid.

2.4 Exempel kovariansfunktioner

Det är inget tentakrav att kunna räkna ut komplicerade kovariansfunktioner. Till tentan ska man kunna räkna ut kovariansfunktionen för en MA process (av godtycklig ordning). Mera komplicerade kovariansuttryck ges som ledning.

Nedan ges ett par exempel på några mera komplicerade kaovariansuutryck. Som vanligt betecknar $e(t)$ vitt brus med medelvärde 0 och varians λ. Vi använder också $R_y(k) = Ey(t+k)y(t)$.

ARMA(1,1) processen

$$y(t) + ay(t-1) = e(t) + ce(t-1)$$

$$\begin{aligned}
R_y(0) &= \lambda \frac{1+c^2-2ac}{1-a^2} \\
R_y(1) &= \lambda \frac{(c-a)(1-ac)}{1-a^2} \\
R_y(k) &= \lambda (-a)^{k-1} \frac{(c-a)(1-ac)}{1-a^2}, \quad k > 1
\end{aligned}$$

ARMAX(1,1,1) processen

$$y(t) + ay(t-1) = bu(t-1) + e(t) + ce(t-1)$$

Insignalen är vit, med medelvärde 0 och varians $Eu^2 = \sigma$

$$\begin{aligned}
R_y(0) &= \frac{b^2\sigma + (1+c^2-2ac)\lambda}{1-a^2} \\
R_y(1) &= \frac{-ab^2\sigma + (c-a)(1-ac)\lambda}{1-a^2} \\
Ey(t)u(t) &= 0 \\
Ey(t)u(t-1) &= b\sigma
\end{aligned}$$

Notera att $b = 0$ ger ARMA(1,1) processen och $c = 0$ ger en ARX(1,1) process.

3 L3 and L4-Parameter estimations

3.1 Problems

1) Criteria with constraints
Consider the following scalar non-linear minimization problem

$$\min_{\theta} V_N(\theta)$$

where

$$V_N(\theta) = \frac{1}{N}\sum_{j=1}^{N}(y(t) - \hat{y}(t,\theta))^2$$

The following constraint is also given:

$$0 \leq \theta \leq 1$$

Assume that solutions to $\frac{dV_N(\theta)}{d\theta} = 0$ have been found. Describe how the minimization problem should be solved in principle.

2) Calculating the least squares estimate for ARX models.
Consider the following ARX model:

$$y(t) + ay(t-1) = bu(t-1) + e(t) \tag{7}$$

($e(t)$ is white noise with zero mean)

Assume that available data are : $y(1),\ u(1),\ y(2), u(2),\ \ldots\ , y(102),\ u(102)$ and the following sums have been calculated:

$$\sum_{t=2}^{102} y^2(t-1) = 5.0,\ \ \sum_{t=2}^{102} y(t-1)u(t-1) = 1.0,\ \ \sum_{t=2}^{102} u^2(t-1) = 1.0,$$

$$\sum_{t=2}^{102} y(t)y(t-1) = 4.5,\ \ \sum_{t=2}^{102} y(t)u(t-1) = 1.0,$$

Which value of $\theta = (a\ \ b)^T$ minimizes the quadratic criteria

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

where $\hat{y}(t,\theta)$ is the predictor obtained from the ARX model (7)?

3) Data with non zero mean.
Assume that the data from a system (normally the data is also noise corrupted but that is not considered in this example) is given by

$$A(q)y(t) = B(q)u(t) + K \tag{8}$$

where $A(q) = 1 + a_1q^{-1}+,\ \ldots\ , a_nq^{-n},\ ,\ B(q) = b_1q^{-1}+,\ \ldots\ , b_nq^{-n}$, and K is an unknown constant.

a) Show that by using the following transformation of the data

$$\begin{aligned}\Delta(q)u(t) &= (1-q^{-1})u(t) = u(t) - u(t-1)\\ \Delta(q)y(t) &= (1-q^{-1})y(t) = y(t) - y(t-1)\end{aligned}$$

as new input and output signals, the standard LS (least squares) method can be used to find the parameters in $A(q)$ and $B(q)$.

b) Show that the constant K easily can be included in the LS estimate for the model (8).

c) What is the standard procedure to deal with data with non zero mean?

4) The problem with feedback.
Consider the following system:

$$y(t) + ay(t-1) = bu(t-1) + e(t) \tag{9}$$

($e(t)$ is white noise with zero mean)

Assume that the system is controlled with a proportional controller

$$u(t) = -Ky(t)$$

Show that $P = \sum_{t=1}^{N} \varphi(t)\varphi^T(t)$ becomes singular!

5) Optimal input signal.
The following system is given:

$$y(t) = b_o u(t) + b_1 u(t-1) + e(t)$$

$e(t)$ is white noise with zero mean and variance λ.

The parameters are estimated with the least squares method. Consider the case when the number of data points N goes to infinity (in practice, this means that we have many data points available)

a) Show that var($\hat{b}_o$) and var($\hat{b}_1$) only depends on the following values of the covariance function:

$$\begin{aligned}R_u(0) &= Eu^2(t) = \lim_{N\to\infty}\frac{1}{N}\sum_{k=1}^{N} u^2(k)\\ R_u(1) &= Eu(t)u(t-1) = \lim_{N\to\infty}\frac{1}{N}\sum_{k=1}^{N} u(k)u(k-1)\end{aligned}$$

b) Assume that the energy of the input signal is constrained to

$$R_u(0) = Eu^2(t) = \lim_{N\to\infty}\frac{1}{N}\sum_{k=1}^{N} u^2(k) \le 1$$

Determine $R_u(0)$ and $R_u(1)$ so that the variance of the parameter estimates are minimized.

6) The following system is given:

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

$e(t)$ is white noise with zero mean and variance λ.
Assume that the number of data points goes to infinity.

a) Assume that $u(t)$ is white noise[2] with variance σ and zero mean. Show that the least squares estimate converge to the true system parameters.

b) Assume that $u(t)$ is a unit step: $u(t) = 0$, $t \leq 0$, and $u(t) = 1$, $t \geq 1$, show that the matrix $\bar{R} = \lim_{N\to\infty} \frac{1}{N} \sum_{t=1}^{N} \varphi(t)\varphi^T(t)$ becomes singular.

7) The following system is given:

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

$e(t)$ is white noise with zero mean and variance λ.

The following predictor

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

is used to estimate the parameter b with the least squares (LS) method.

Calculate the LS estimate of b (expressed in b_1 and b_2) as the number of data points goes to infinity for the cases:

a) The input signal $u(t)$ is white noise.

b) The input signal is a sinusoidal $u(t) = A\sin(\omega_1 t)$ wich has the covariance function $R_u(\tau) = \frac{1}{2}A^2\cos(\omega_1\tau)$

[2] In general, we will also assume that $e(t)$ and $u(t)$ are uncorrelated if not explicitely stated otherwise.

3.2 Solutions

1) Let θ_i be the parameters which gives $\frac{dV_N(\theta)}{d\theta} = 0$. Check the values of $V_N(\theta_i)$, $V_N(0)$ and $V_N(1)$ and select the values of θ which minimizes V.

2) The predictor for the ARX model is $\hat{y}(t) = \varphi^T(t)\theta$ where $\varphi(t) = (-y(t-1) \;\; u(t-1))^T$ and $\theta = (a \;\; b)^T$. The least squares estimate is

$$\begin{aligned}
\hat{\theta} &= [\sum_{t=1}^{N} \varphi(t)\varphi^T(t)]^{-1} \sum_{t=1}^{N} \varphi(t)y(t)) \\
&= \begin{bmatrix} \sum_{t=2}^{102} y^2(t-1) & -\sum_{t=2}^{102} y(t-1)u(t-1) \\ -\sum_{t=2}^{102} y(t-1)u(t-1) & \sum_{t=2}^{102} u^2(t-1) \end{bmatrix}^{-1} \begin{bmatrix} -\sum_{t=2}^{102} y(t-1)y(t) \\ \sum_{t=2}^{102} u(t-1)y(t) \end{bmatrix} \\
&= \begin{bmatrix} 5 & -1 \\ -1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} -4.5 \\ 1 \end{bmatrix} \\
&= \begin{bmatrix} -0.875 \\ 0.125 \end{bmatrix}
\end{aligned}$$

3a) Multiplying the left and right hand side of the system

$$A(q)y(t) = B(q)u(t) + K$$

with $\Delta(q)$ gives

$$\Delta(q)A(q)y(t) = \Delta(q)B(q)u(t) + \Delta(q)K$$

but since K is a constant, $\Delta(q)K(1-q^{-1})K == K - K = 0$ and hence

$$A(q)(\Delta(q)y(t)) = B(q)(\Delta(q)u(t))$$

Thus if we use the differentiated input and output signals we can use the standard LS estimate to estimate A and B.

Remark: Any filter $L(q)$ with $L(1) = 0$ would remove the constant K. Note that $L(1) = 0$ means zero steady state gain.

b) Use the regression vector

$$\varphi(t) = (-y(t-1) \;\; -y(t-2) \ldots -y(t-n) \;\; u(t-1) \;\; u(t-2) \ldots u(t-n) \;\; 1)^T$$

and

$$\theta = (a_1 \;\; a_2 \ldots a_n \;\; b_1 \;\; b_2 \ldots b_n \;\; K)^T$$

Note that we can view the system consisting of two input signals; $u(t)$ and 1.

c) Remove the mean from the data, that is use the new signals:

$$\begin{aligned}
\bar{y}(t) &= y(t) - \frac{1}{N}\sum_{k=1}^{N} y(k) \\
\bar{u}(t) &= u(t) - \frac{1}{N}\sum_{k=1}^{N} u(k)
\end{aligned}$$

4a) We directly get

$$\begin{aligned}
P &= [\sum_{t=1}^{N} \varphi(t)\varphi^T(t)] \\
&= \begin{bmatrix} \sum_{t=1}^{N} y^2(t-1) & -\sum_{t=1}^{N} y(t-1)u(t-1) \\ -\sum_{t=1}^{N} y(t-1)u(t-1) & \sum_{t=1}^{N} u^2(t-1) \end{bmatrix} \\
&= \begin{bmatrix} \sum_{t=1}^{N} y^2(t-1) & K\sum_{t=1}^{N} y^2(t-1) \\ K\sum_{t=1}^{N} y^2(t-1) & K^2\sum_{t=1}^{N} y^2(t-1) \end{bmatrix}
\end{aligned}$$

In the third equality we have used $u(t) = -Ky(t)$.

It is directly seen that the matrix is singular ($\det P = 0$) hence this experimental condition can not be used to estimate the parameters uniquely. This can also be seen if we look at the predictor $\hat{y}(t) = -ay(t-1) + bu(t-1)$. With the controller $u(t) = -Ky(t)$ the predictor becomes $\hat{y}(t) = -ay(t-1) - bKy(t-1) = -(bK+a)y(t-1)$ and we see that the predictor does not uniqely depends on a and b.

5) We have that (see Linear Regression)

$$\begin{aligned}
\text{cov}(\hat{\theta}) &= \lambda[\sum_{t=1}^{N} \varphi(t)\varphi^T(t)]^{-1} = \frac{\lambda}{N}[\frac{1}{N}\sum_{t=1}^{N} \varphi(t)\varphi^T(t)]^{-1} \rightarrow_{N\rightarrow\infty} \frac{\lambda}{N}[[E\{\varphi(t)\varphi^T(t)\}]^{-1} \\
&= \frac{\lambda}{N}(\bar{R})^{-1}
\end{aligned}$$

With $\varphi(t) = (u(t) \;\; u(t-1))^T$ we get

$$\bar{R} = \begin{bmatrix} R_u(0) & R_u(1) \\ R_u(1) & R_u(0) \end{bmatrix}$$

and as $N \rightarrow \infty$

$$\text{cov}(\hat{\theta}) = \frac{\lambda}{N}\begin{bmatrix} R_u(0) & R_u(1) \\ R_u(1) & R_u(0) \end{bmatrix}^{-1}$$

Hence,

$$\text{var}(\hat{b}_o) = \text{var}(\hat{b}_1) = \frac{\lambda}{N}\;\frac{R_u(0)}{R_u^2(0) - R_u^2(1)}$$

b) It is seen directly (note that $R_u(0) \geq |R_u(\tau)|$) that the variances are minimized for $R_u(0) = 1$ and $R_u(1) = 0$. One example of a signal that fulfills this condition is white noise with unit variance.

6) The predictor is given by $\hat{y}(t) = \varphi^T(t)\theta$ where
$\varphi(t) = (u(t-1) \;\; u(t-2))^T$ and $\theta = (b_1 \;\; b_2)^T$.
The least squares estimate is (we normalize with $\frac{1}{N}$ since we then get a feasible expression as $N \rightarrow \infty$)

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

As $N \rightarrow \infty$

$$\hat{\theta}_\infty = (\bar{R})^{-1}E\{\varphi(t)y(t)\}$$

where $\bar{R} = E\{\varphi(t)\varphi^T(t)\}$ cf the previous problem.

$$\hat{\theta} = \begin{bmatrix} R_u(0) & R_u(1) \\ R_u(1) & R_u(0) \end{bmatrix}^{-1} \begin{bmatrix} R_{yu}(1) \\ R_{yu}(2) \end{bmatrix}$$

Since $u(t$ is white noise $R_u(1) = 0$ and

$$\begin{aligned} R_{yu}(1) &= Ey(t)u(t-1) = E\{[b_1u(t-1+b_2u(t-1)+e(t)]u(t-1)\} = b_1R_u(0) \\ R_{yu}(2) &= Ey(t)u(t-1) = E\{[b_1u(t-1+b_2u(t-1)+e(t)]u(t-2)\} = b_2R_u(0) \end{aligned}$$

the estimate converges to

$$\hat{\theta}_\infty = \begin{bmatrix} 1/R_u(0) & 0 \\ 0 & 1/R_u(0) \end{bmatrix} \begin{bmatrix} R_u(0)b_1 \\ Ru(0)b_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

which is expected since we have a FIR model (wich could be interpreted as a linear regression model) and model structure is correct.

b) We first calculate

$$\begin{aligned} \frac{1}{N}\sum_{t=1}^{N}\varphi(t)\varphi^T(t) &= \frac{1}{N}\begin{bmatrix} \sum_{t=1}^N u^2(t-1) & \sum_{t=1}^N u(t-1)u(t-2) \\ \sum_{t=1}^N u(t-1)u(t-2) & \sum_{t=1}^N u^2(t-1) \end{bmatrix} \\ &= \frac{1}{N}\begin{bmatrix} N-1 & N-2 \\ N-2 & N-2 \end{bmatrix} = \begin{bmatrix} \frac{N-1}{N} & \frac{N-2}{N} \\ \frac{N-2}{N} & \frac{N-2}{N} \end{bmatrix} \end{aligned}$$

and we see that

$$\bar{R} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

which is singular. This means that a step gives too poor excitation of the system, asymptotically the matrix (which should be inverted in the least squares method) becomes singular.

7) In this case $\varphi(t) = (u(t-1))$ and $\theta = b$. Asymptotically in N we have

$$\hat{\theta}_\infty = [E\{\varphi(t)\varphi^T(t)\}]^{-1}E\{\varphi(t)y(t)\}$$

For this simple predictor we have:

$$\begin{aligned} E\{\varphi(t)\varphi^T(t)\} &= Eu^2(t-1) = R_u(0) \\ E\{\varphi(t)y(t)\} &= Ey(t)u(t-1) = R_{yu}(1) \end{aligned}$$

and by using the system generating the data

$$R_{yu}(1) = Ey(t)u(t-1) = E\{[b_1u(t-1+b_2u(t-1)+e(t)]u(t-1)\} = b_1R_u(0)+b_2R_u(1)$$

which gives

$$\hat{\theta}_\infty = \hat{b}_\infty = b_1 + b_2\frac{R_u(1)}{R_u(0)}$$

a) If $u(t)$ is white noise $R_u(1) = 0$ and we get $\hat{b}_\infty = b_1$.

b) For $u(t)$ beeing a sinusoid, $\hat{b}_\infty = b_1 + b_2 \cos w_1$

4 L5- Some additional problems

4.1 Complementary theory - Analysis of the least squares estimate

4.1.1 Results from ”Linear Regression”

The accuracy result is based on the following assumptions:

Assumption A1.
Assume that the data are generated by (”the true system”):

$$y(t) = \varphi^T(t)\theta_o + e(t) \quad t = 1, \ldots, N \tag{10}$$

where $e(t)$ is a nonmeasurable disturbances term to be specified below. In matrix form, (10) reads

$$Y = \Phi\theta_o + \mathbf{e} \tag{11}$$

where $\mathbf{e} = [e(1) \ \ldots e(N)]^T$.

Assumption A2.
It is assumed that $e(t)$ is a white noise process[3] with variance λ.

Assumption A3.
It is finally assumed that $E\{\varphi(t)e(s)\} = 0$ for all t and s. This means that the regression vector is not influenced (directly or indirectly) by the noise source $e(t)$

In the material ”Linear Regression” it was shown that if Assumptions A1-A3 hold then

1. The least squares estimate $\hat{\theta}$ is an unbiased estimate of θ_o, that is $E\{\hat{\theta}\} = \theta_o$.

2. The uncertainty of the least squares estimate as expressed by the covariance matrix P is given by

$$\begin{aligned} P &= \operatorname{cov} \hat{\theta} = E\{(\hat{\theta} - E\hat{\theta})(\hat{\theta} - E\hat{\theta})^T\} = E\{(\hat{\theta} - \theta_o)(\hat{\theta} - \theta_o)^T\} = \lambda(\Phi^T\Phi)^{-1} \\ &= \lambda[\sum_{t=1}^{N} \varphi(t)\varphi^T(t)]^{-1} \end{aligned}$$

4.1.2 Results for the case when A3 does not hold

For the case when A3 does *not* hold we have that

1. The least squares estimate $\hat{\theta}$ is consistent:

$$\hat{\theta} \to \theta_o \text{ as } N \to \infty \tag{12}$$

   where N is the number of data points.

[3]A white noise process $e(t)$ is a sequence of random variables that are *uncorrelated*, have mean zero, and a constant finite variance. Hence, $e(t)$ is a white noise process if $E\{e(t)\} = 0$, $E\{e^2(t)\} = \lambda$, and $E\{e(t)e(j)\} = 0$ for t not equal to j.

2. The covariance matrix P is given by

$$P = \text{cov}\,\hat{\theta} \rightarrow \frac{\lambda}{N}[E\{\varphi(t)\varphi^T(t)\}]^{-1} \text{ as } N \rightarrow \infty \tag{13}$$

Remarks:

- Two typical examples when A3 does not hold are when the system is an AR-process or an ARX-process (we then have values of the output in the regressor vector.
- The results only holds asymptotically in N. In practice this means that we need to have many data (some hundreds are typically enough) points for the estimate to be reliable (and also to get reliable estimate of the covariance matrix).
- If the noise is not white the estimate will in general not be consistent (in contrast to when A3 holds- see "Linear Regression").
- Results for general linear models are presented on page 297-299 in the text book.

4.1.3 Bias, variance and mean squared error

Let $\hat{\theta}$ be a scalar estimate of the true parameter θ_o. The bias is defined as

$$\text{bias}(\hat{\theta}) = E\{\hat{\theta}\} - \theta_o \tag{14}$$

The variance is given by

$$\text{var}(\hat{\theta}) = E\{(\hat{\theta} - E\{\hat{\theta}\})^2\} \tag{15}$$

The Mean Squared Error (MSE) is given by

$$\text{MSE}(\hat{\theta}) = E\{(\hat{\theta} - \theta_o)^2\} = \text{var}(\hat{\theta}) + [\text{bias}(\hat{\theta})]^2 \tag{16}$$

In practice we want to have an estimate with as small MSE as possible. In some cases this may mean that we accept a small bias if the variance of the estimate can be reduced.

4.2 Problems

1) Predictor using exponential smoothing

A simple predictor for a signal $y(t)$ is the so-called exponential smoothing which is given by

$$\hat{y}(t) = \frac{1-\alpha}{1-\alpha q^{-1}} y(t-1)$$

a) Show that if $y(t) = m$ for all t, the predictor will in steady state (stationarity) be equal to m.

b) For which ARMA model is the predictor optimal?
Hint: Rewrite the predictor in the form $\hat{y}(t) = L(q)y(t)$ and compare with the predictor for an ARMA model $A(q)y(t) = C(q)y(t)$.

2) Cross correlation for LS estimate of ARX parameters.
Consider the standard least squares estimate of the parameters in an ARX model:

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

where $A(q) = 1 + a_1 q^{-1} + ... + a_{na} q^{-na}$ och $B(q) = b_1 q^{-1} + ... + b_{nb} q^{-nb}$. The estimate of the cross correlation between residuals and inputs is given by

$$\hat{R}_{\epsilon u}(\tau) = \frac{1}{N} \sum_{t=1}^{N} \epsilon(t) u(t-\tau))$$

where $\epsilon(t) = y(t) - \hat{y}(t) = y(t) - \varphi^T(t)\hat{\theta}$ is the prediction error. Show that the least squares estimate gives

$$\hat{R}_{\epsilon u}(\tau) = 0 \qquad \tau = 1, 2...nb$$

Hint: Show that the least squares estimate gives $\sum_{t=1}^{N} \varphi(t)\epsilon(t) = 0$.

3) The variance increases if more parameters than needed are estimated!
Assume that data from an AR(1) process (”The system”) is collected:

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

where $e(t)$ is white noise with zero mean and variance λ. The system is stable wich means that $|a_o| < 1$

Consider the following two predictors

$$\text{M1} \quad \hat{y}(t) = -ay(t-1)$$

$$\text{M2} \quad \hat{y}(t) = -a_1 y(t-1) - a_2 y(t-2)$$

where the parameters for each predictor is estimated with the least squares method.

It can easily be shown that for M1: $\hat{a} \to a_o$ and for M2: $\hat{a}_1 \to a_o$ and $\hat{a}_2 \to 0$, as the number of data points $N \to \infty$.

Hence both estimate gives consistent estimate. Show that the ”price to pay” for estimating too many parameters is that $\text{var}(\hat{a}_1) > \text{var}(\hat{a})$ as $N \to \infty$.
Hint: For the AR(1) process we have that $R_y(k) = Ey(t+k)y(t) = (-a_o)^k R_y(0)$, $k = 1, 2..$ where $R_y(0) = \frac{\lambda}{1-a_o^2}$

4) Variance of parameters in an estimated ARX model.
Assume that data was collected from the following ARX process (”The system”)

$$y(t) + a_o y(t-1) = b_o u(t-1) e(t)$$

where $e(t)$ is white noise with zero mean and variance λ. The system is stable wich means that $|a_o| < 1$. The input signal is uncorrelated with $e(t)$, and is white noise with zero mean and variance σ.

The parameters in the following predictor

$$\hat{y}(t|\theta) = -ay(t-1) + bu(t-1)$$

are estimated with the least squares method.
Calculate the asymptotic (in number of data points N) variance of the parameter estimates.
Hint:

$$R_y(0) = Ey^2(t) = \frac{b_o^2\sigma + \lambda}{1 - a_o^2}$$

4.3 Solutions

1a) We rewrite the predictor in standard form

$$\hat{y}(t) = \frac{q^{-1}(1-\alpha)}{1-\alpha q^{-1}} y(t) = L(q)y(t)$$

The static gain is given by $L(1)$ and since we have $L(1) = 1$ the steady state value of the predictor will be $\hat{y}(t) = m$.

b) For an ARMA process

$$A(q)y(t) = C(q)e(t)$$

the optimal predictor is

$$\hat{y}(t) = (1 - \frac{A(q)}{C(q)})y(t) = (\frac{C(q) - A(q)}{C(q)})y(t)$$

Hence, we need to find $A(q)$ and $C(q)$ so that

$$\frac{C(q) - A(q)}{C(q)} = \frac{q^{-1}(1-\alpha)}{1-\alpha q^{-1}}$$

which gives $C(q) = 1 - \alpha q^{-1}$ and $A(q) = 1 - q^{-1}$.

2) The least squares estimate is given by

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

which can be written as

$$[\sum_{t=1}^{N} \varphi(t)\varphi^T(t)]\hat{\theta} = \sum_{t=1}^{N} \varphi(t)y(t)) \tag{17}$$

With $\epsilon(t) = y(t) - \varphi^T(t)\hat{\theta}$ we get

$$\begin{aligned}
\sum_{t=1}^{N} \varphi(t)\epsilon(t) &= \sum_{t=1}^{N} \varphi(t)[y(t) - \varphi^T(t)\hat{\theta}] = \sum_{t=1}^{N} \varphi(t)y(t) - [\sum_{t=1}^{N} \varphi(t)\varphi^T(t)]\hat{\theta} \\
&= \sum_{t=1}^{N} \varphi(t)y(t) - \sum_{t=1}^{N} \varphi(t)y(t) = 0
\end{aligned}$$

In the last equality (17) was used. We thus have

$$0 = \sum_{t=1}^{N} \varphi(t)\epsilon(t) = \sum_{t=1}^{N} \begin{bmatrix} -y(t-1) \\ -y(t-2) \\ \vdots \\ -y(t-na) \\ u(t-1) \\ u(t-2) \\ \vdots \\ u(t-nb) \end{bmatrix} \epsilon(t)$$

This means that all estimates

$$\hat{R}_{\epsilon u}(\tau) = \frac{1}{N}\sum_{t=1}^{N} \epsilon(t)u(t-\tau)) = 0 \quad \text{for } \tau = 1, 2...nb$$

Therefore, the values of the estimated cross correlation function $\hat{R}_{\epsilon u}(\tau)$ for $\tau = 1, 2...nb$ can not be used to determine if an estimated ARX model is good or bad. They will always be zero. See also the text book on page 368!

3) In general we have for estimated AR-parameters that

$$P = \text{cov}\ \hat{\theta} \to \frac{\lambda}{N}[E\{\varphi(t)\varphi^T(t)\}]^{-1} \text{ as } N \to \infty$$

For M1 we have $\varphi(t) = -y(t-1)$ and $E\{\varphi(t)\varphi^T(t)\} = Ey^2(t-1) = R_y(0) = \frac{\lambda}{1-a_o^2}$ (see the Hint). Thus

$$\text{var}(\hat{a}) \to \frac{\lambda}{N}\frac{1-a_o^2}{\lambda} = \frac{1}{N}(1-a_o^2) \text{ as } N \to \infty$$

For M2 we have $\varphi(t) = [-y(t-1) \quad -y(t-2)]^T$ and therefore

$$E\{\varphi(t)\varphi^T(t)\} = \begin{pmatrix} R_y(0) & R_y(1) \\ R_y(1) & R_y(0) \end{pmatrix}$$

This gives

$$[E\{\varphi(t)\varphi^T(t)\}]^{-1} = \frac{1}{R_y(0)^2 - R_y(1)^2}\begin{pmatrix} R_y(0) & -R_y(1) \\ -R_y(1) & R_y(0) \end{pmatrix}$$

and

$$\begin{aligned} \text{var}(\hat{a}_1) &\to \frac{\lambda}{N}\frac{R_y(0)}{R_y(0)^2 - R_y(1)^2} = \frac{\lambda}{N}\frac{R_y(0)}{R_y(0)^2 - a_o^2R_y(0)^2} = \frac{\lambda}{N}\frac{1}{R_y(0)(1-a_o^2)} \\ &= \frac{1}{N} \text{ as } N \to \infty \end{aligned}$$

We have thus shown that $\text{var}(\hat{a}_1) > \text{var}(\hat{a})$ as $N \to \infty$.

4) The predictor gives $\varphi(t) = [-y(t-1) \ \ u(t-1)]^T$ and $\theta = [a \ \ b]^T$. This gives

$$E\{\varphi(t)\varphi^T(t)\} = \begin{pmatrix} R_y(0) & -R_{yu}(0) \\ -R_{yu}(1) & R_u(0) \end{pmatrix}$$

The cross covariance between y and u is

$$R_{yu}(0) = E\{(-a_oy(t-1) + b_ou(t-1) + e(t))(u(t))\} = 0$$

since $u(t)$ is white noise (uncorrelated with $e(t)$). The (asymptotic) covariance matrix is

$$P = \text{cov}\ \hat{\theta} \to \frac{\lambda}{N}\begin{pmatrix} \frac{b_o^2\sigma+\lambda}{1-a_o^2} & 0 \\ 0 & \sigma \end{pmatrix}^{-1} = \frac{\lambda}{N}\begin{pmatrix} \frac{1-a_o^2}{b_o^2\sigma+\lambda} & 0 \\ 0 & \frac{1}{\sigma} \end{pmatrix} \text{ as } N \to \infty$$

From the diagonal elements we get the variances: $\text{var}(\hat{a}) = \frac{\lambda}{N}\frac{1-a_o^2}{b_o^2\sigma+\lambda}$ and $\text{var}(\hat{b}) = \frac{\lambda}{N\sigma}$ (as $N \to \infty$).