System Identification
Exercise session 3

Sholeh Yasini

April 25, 2016
Least squares (LS) method

Linear regression for dynamical models

\[ A(q^{-1})y_t = B(q^{-1})u_t + e_t \]

where

\[ A(q^{-1}) = 1 + a_1 q^{-1} + ... + a_n a q^{-n}a \]
\[ B(q^{-1}) = b_1 q^{-1} + ... + b_n b q^{-n}b \]

it can equivalently be written as

\[ y_t = \phi_t^\top \theta + e_t \]

where

\[ \phi_t^\top = (-y_{t-1} - y_{t-2} \ldots y_{t-n_a} u_{t-1} u_{t-2} \ldots u_{t-n_b}) \]
\[ \theta = (a_1 \ldots a_{n_a} b_1 \ldots b_{n_b})^\top \]

minimizing sum of squared errors

\[ V(\theta) = \sum_{t=1}^{n} e_t^2 \]

gives

\[ \hat{\theta} = \left( \frac{1}{n} \sum_{t=1}^{n} \phi_t \phi_t^\top \right)^{-1} \left( \frac{1}{n} \sum_{t=1}^{n} \phi_t y_t \right) \]

Exercise: derive the normal equation.
LS method (linear regression)

- **Assumption 1** assume that data are generated by ("the true system:")

\[ y_t = \phi^\top \theta_0 + \epsilon_t \]

Thus \( \theta_0 \) is the true parameter.

- **Assumption 2** assume that \( \epsilon_t \) is white noise with zero mean and variance \( \lambda \)

If the estimate \( \hat{\theta} \) is 'good', it should be closed to the true parameter \( \theta_0 \). Let's derive the estimation error

\[
\hat{\theta} - \theta_0 = \left( \frac{1}{n} \sum_{t=1}^{n} \phi_t \phi_t^\top \right)^{-1} \left( \left( \frac{1}{n} \sum_{t=1}^{n} \phi_t y_t \right) - \left( \frac{1}{n} \sum_{t=1}^{n} \phi_t \phi_t^\top \right) \theta_0 \right) \\
= \left( \frac{1}{n} \sum_{t=1}^{n} \phi_t \phi_t^\top \right)^{-1} \left( \frac{1}{n} \sum_{t=1}^{n} \phi_t \epsilon_t \right)
\]

Thus \( \hat{\theta} \) tends to \( \theta_0 \) (asymptotically) if

- **Assumption 3** \( \mathbb{E} \phi_t \phi_t^\top \) is nonsingular
- **Assumption 4** \( \mathbb{E} \phi_t \epsilon_t = 0 \), i.e., the regressor vector is not influenced by noise
LS (linear regression)

if all assumptions 1-4 are satisfied

- $\hat{\theta}$ is an unbiased estimate of $\theta_0$, $E\hat{\theta} = \theta_0$
- the uncertainty of the least squares estimate is given by

$$P = cov\hat{\theta} = E(\hat{\theta} - \theta_0)(\hat{\theta} - \theta_0)^\top = \lambda \left(\sum_{t=1}^{n} \phi_t \phi_t^\top\right)^{-1}$$

Note that for the Assumption 3 to be satisfied the input must be PE
Prediction error method

- A model obtained by identification can be used in many ways, depend of the purpose of modeling. In many applications the model is used for prediction.
- PEM The goal is to minimize the prediction error

\[ \epsilon_t = y_t - \hat{y}_{t|t-1} \]

where \( \hat{y}_{t|t-1} \) is the best prediction of \( y_t \) based on past observations.
Optimal predictor

- **AR model** what is the best prediction of $y_t$ given past observations $\{y_1, \ldots, y_{t-1}\}$

$$A(q^{-1})y_t = e_t$$

$$\hat{y}_{t|t-1} = (1 - A(q^{-1})) y_t$$

- **MA model** what is the optimal predictor of MA model given past observations $\{y_1, \ldots, y_{t-1}\}$

$$y_t = C(q^{-1})e_t$$

$$\hat{y}_{t|t-1} = \left(\frac{C(q^{-1}) - 1}{C(q^{-1})}\right) y_t$$

- **ARMA model** what is the best optimal predictor for ARMA model given past data $\{y_1, \ldots, y_{t-1}\}$

$$A(q^{-1})y_t = C(q^{-1})e_t$$

$$\hat{y}_{t|t-1} = \left(\frac{C(q^{-1}) - A(q^{-1})}{C(q^{-1})}\right) y_t$$
Exercise 3.1

consider the predictor

\[ \hat{y}_t = \frac{1 - \alpha}{1 - \alpha q^{-1}} y_{t-1} \]

where \(|\alpha| < 1\) is a constant.

▶ for which ARMA model is the predictor optimal?
▶ show that if \(y_t = m\) for all \(t\), the predictor will in steady-state be equal to \(m\).
Exercise 3.2

cross correlation test for the LS method
consider the standard LS estimate of the parameters in an ARX model

\[ A(q^{-1})y_t = B(q^{-1})u_t + e_t \]

the estimate of cross correlation between residuals and input is given by

\[ \hat{R}_{\epsilon u}(\tau) = \frac{1}{n} \sum_{t=1}^{n} \epsilon_t u_{t-\tau} \]

where \( \epsilon_t = y_t - \hat{y}_t = y_t - \phi_t^\top \hat{\theta} \) is the prediction error. Show that LS estimate gives

\[ \hat{R}_{\epsilon u}(\tau) = 0, \quad \tau = 1, 2, \ldots, n_b \]
The variance increases if more parameters than needed are estimated.

Assume that data from an AR(1) process is collected:

\[ y_t + a_0 y_{t-1} = e_t \]

where \( e_t \) is white noise with variance \( \lambda \) and where \( |a_0| < 1 \). Assume that we fit this system the following two candidate models:

\[ M_1 : y_t + ay_{t-1} = \epsilon_t \]
\[ M_2 : y_t + a_1 y_{t-1} + a_2 y_{t-2} = \epsilon_t' \]

Let \( \hat{a} \) denote the LS estimate of \( a \) in \( M_1 \), and let \( \hat{a}_1 \) and \( \hat{a}_2 \) be the LS estimate in \( M_2 \). What are the asymptotic variance of these estimators?

**Hint:**

\[ R_y(1) = -a R_y(0), \quad R_y(0) = \frac{\lambda}{1 - a^2} \]
Exercise 3.4

variance of estimator in an estimated ARX
Assume that data from the following ARX process is collected:

\[ y_t + a_0 y_{t-1} = b_0 u_{t-1} + e_t \]

where \( e_t \) is white noise with variance \( \lambda \) and where \( |a_0| < 1 \).

calculate the asymptotic variance of the parameter estimates.

Hint: \( R_y(0) = \mathbb{E} y_t^2 = \frac{b_0^2 \sigma + \lambda}{1 - a_0^2} \)
The end