# System Identification, Exercises 2

Kristiaan Pelckmans (IT/UU, 2338)

Course code: 1RT875, Report code: 61806,
F, FRI Uppsala University, Information Technology

January - March 2010

# Exercise 4.1

Consider the linear regression model

$$y(t) = a + bt + e(t)$$

Find the LS estimate of $a$ and $b$ in the following cases:

1. (a) The data are $y(1), y(2), \ldots, y(N)$. Set

$$S_0 = \sum_{t=1}^{N} y(t), \quad S_1 = \sum_{t=1}^{N} t y(t)$$

   (b) The data are $y(-N), \ldots, y(0), \ldots, y(N)$. Set

$$S_0' = \sum_{t=-N}^{N} y(t), \quad S_1' = \sum_{t=-N}^{N} t y(t)$$

Hint: $\sum_{t=1}^{N} t = \frac{1}{2}N(N-1)$ and $\sum_{t=1}^{N} t^2 = \frac{1}{6}N(N+1)(2N+1)$.

2. Next, suppose the parameter $a$ is first estimated as

$$\hat{a} = \frac{1}{N}S_0, \quad \hat{a}' = \frac{1}{2N+1}S_0'$$

then estimate $b$ using LS, in the model

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

What will $\hat{b}_{LS}$ become? Compare to (a).

Now assume $e(t)$ is white noise with variance $\lambda^2$.

1. Find the variance of $s(t) = \hat{a} - \hat{b}t$. What is this quantity for $t = 1$ and $t = N$? Where does it find its minimum?

2. Write the covariance matrix of $\theta = (\hat{a}\hat{b})^T$ in the form

$$P = \mathrm{cov}(\theta) = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$$

Find the asymptotic value of $\rho$.

3. Introduce the concentration ellipsoid

$$Q_\xi = \left\{\theta \;\middle|\; (\theta - \hat{\theta})^T P^{-1} (\theta - \hat{\theta}) \leq \xi \right\}$$

Roughly, vectors $\theta$ outside the ellipsoid $Q_\xi$ are unlikely to 'generate' the estimate $\hat{\theta}$ if $\xi \sim n$. In fact, one has for any $\theta \in \mathbb{R}^n$ that

$$\begin{aligned} & E[(\hat{\theta} - \theta)^T P^{-1} (\hat{\theta} - \theta)] \\ & \qquad = \operatorname{tr}\left(P^{-1} E\left[(\hat{\theta} - \theta_0)(\hat{\theta} - \theta_0)^T\right]\right) \\ & \qquad = \operatorname{tr}\left(P^{-1}\left((\theta - \theta_0)(\theta - \theta_0)^T + P\right)\right) \\ & \qquad = n + (\theta - \theta_0)^T (\theta - \theta_0) \approx n \end{aligned}$$

If $\hat{\theta}$ AN, then $(\hat{\theta} - \theta_0)^T P^{-1} (\hat{\theta} - \theta_0) \sim \chi^2$. Plot a the concentration ellipsoid when $\lambda^2 = 0.1$, $\xi = 2$ and (i) $N = 3$, (ii) $N = 8$.

# Exercise 4.5

Ill-conditioning of the normal equations in case of a polynomial trend model.

Given model

$$y(t) = a_0 + a_1 t + \cdots + a_r t^r + e(t)$$

Show that the condition number of the associated matrix $\Phi^T \Phi$ is ill-conditioned:

$$\mathrm{cond}(\Phi^T \Phi) \geq O(N^{2r}/(2r+1))$$

for large $N$, and where $r > 1$ is the polynomial order. Hint. Use the relations for a symmetric matrix $A$:

- $\lambda_{\max}(A) \geq \max_i A_{ii}$
- $\lambda_{\min}(A) \leq \min_i A_{ii}$

# Exercise 4.6

Fourier spectral analysis as a special case of regression analysis.

Assume a system

$$y(t) = \sum_{i=1}^{n} a_i \cos(\omega_k t) + \sum_{i=1}^{n} b_i \sin(\omega_k t)$$

with $\omega_i = \frac{2\pi i}{N}$ and $n \leq \lfloor \frac{N}{2} \rfloor$. Show that the LS estimates of $\theta = (a_1, \ldots, a_n, b_1, \ldots, b_n)^T \in \mathbb{R}^{2N}$ are equal to the Fourier coefficients for $k = 1, \ldots, n$ defined as

$$\begin{cases} a_k = \frac{2}{N} \sum_{t=1}^{N} y(t) \cos(\omega_k t) \\ b_k = \frac{2}{N} \sum_{t=1}^{N} y(t) \sin(\omega_k t) \end{cases}$$

Hint. Show first that the following inequalities hold:

$$\begin{cases} \sum_{t=1}^{N} \cos(\omega_k t)\cos(\omega_p t) = \frac{2}{N}\delta_{k,p} \\ \sum_{t=1}^{N} \sin(\omega_k t)\sin(\omega_p t) = \frac{2}{N}\delta_{k,p} \\ \sum_{t=1}^{N} \cos(\omega_k t)\sin(\omega_p t) = 0. \end{cases}$$

# Exercise 4.10

Conditions for the LS estimate to be BLUE.

For the (full) linear regression model

$$\mathbf{Y} = \Phi\theta + \mathbf{e},\ E[\mathbf{e}] = 0_N, E[\mathbf{e}\mathbf{e}^T] = \mathbf{R} > 0, (\Phi^T\Phi) > 0$$

show that conditions (i) and (ii) below are necessary and sufficient for LS to be BLUE.

1.
$$\Phi^T \mathbf{R}^{-1} \left( I_d - \Phi(\Phi^T\Phi)^{-1}\Phi^T \right) = 0$$

2.
$$\mathbf{R}\Phi = \Phi F,\ F \in \mathbb{R}^{n\times n}\ (\exists F^{-1})$$

(A) Give an example of $\mathbf{R}$ satisfying (1), (B) consider $\mathbf{R} = I_N + \alpha\Phi_k\Phi_k$ where $\alpha > 0$ is such that $\mathbf{R} > 0$

# Exercise 5.6

Spectral properties of a random wave.

Let $u_\alpha(t)$ be generated as $u_\alpha(1) = \pm a$, and

$$u_\alpha(t) = \begin{cases} u(t-1) & \text{With Probability } 1-\alpha \\ -u(t-1) & \text{With Probability } \alpha \end{cases}$$

where $0 < \alpha < 1$. The stochastic events are independent to past.

1. Derive the covariance function.

2. Derive the spectral densities. Show that the signal has mainly low-frequency character iff $\alpha \leq 0.5$.

# Exercise 5.9

Estimating impulse response with a PRBS.

Derive the solution of correlation analysis up to order $M-1$ when $u(t)$ is a PRBS with levels $\pm 1$ and period $N \geq M$, or

$$\hat{h}(k) = \frac{N}{(N+1)(N-M+1)} \left( \sum_{i=0, i\neq k}^{M-1} \hat{r}_{yu}(i) + (N-M+2)\hat{r}_{uy}(k) \right)$$

If $N \gg M$ show that this can be simplified to $\hat{h}(k) \approx \hat{r}_{uy}(k)$. If $M = N$, then $\hat{h}(k) \approx \hat{r}_{uy}(k) + \sum_{i=0}^{N-1} \hat{r}_{uy}(i)$. This might appear to be a contradiction to the fact that for large $m$ the covariance matrix of a PRBS with with unit variance converges to the identity matrix. Why is it not?

# Exercise 6.1

Stability boundary for a second-order system.

Consider the second-order AR model

$$y(t) + a_1 y(t-1) + a_2 y(t-2) = e(t)$$

Derive and plot the area in the $(a_1, a_2)$-plane for which the model is asymptotically stable.

# Exercise 7.4

Gradient calculation.

Consider the model structure:

$$y(t) = \frac{B(q^{-1})}{F(q^{-1})}u(t) + \frac{C(q^{-1})}{D(q^{-1})}e(t)$$

with the parameter vector

$$\theta = \Big(b_1, \ldots, b_{n_b}, \ldots, c_{n_c}, \ldots, d_{n_d}, \ldots f_{n_f}\Big)^T$$

what is the gradient $\dfrac{\partial \epsilon(t,\theta)}{\partial \theta}$ to be used in PEM?

# Exercise 7.5

Newton-Raphson minimization procedure.

Let $V(\theta)$ be an analytical cost function in terms of $\theta$. An iterative approach to finding the minimum to $V(\theta)$ can be found as follows. Let $\theta^{(k)}$ be the estimate at iteration $k$, then take $\theta^{(k+1)}$ as the minimum of the quadratic approximation to $V$ around $\theta^{(k)}$. Show that this principle leads to the Newton-Raphson procedure with $\alpha_k = 1$.

# Exercise 7.6

Gauss-Newton optimization procedure.

The Gauss-newton procedure for minimizing

$$V_N(\theta) = \sum_{t=1}^{N} \epsilon^2(t, \theta)$$

where $\epsilon(t, \theta)$ is differentiable wrt $\theta$. The optimum can be obtained from the Newton-Approximation procedure by making an approximation of the Hessian. It can also be obtained by 'quasilinearization', and in fact is sometimes referred to as the quasilinearization minimization method. To be more precise, consider the following linearization of $\epsilon$ around the current estimate $\theta^{(k)}$:

$$\tilde{\epsilon}(t, \theta) = \epsilon(t, \theta^{(k)}) - \psi^T(t, \theta^{(k)})(\theta - \theta^{(k)})$$

with $\psi(t,\theta) = -\frac{\partial^T \epsilon(t,\theta)}{\partial \theta}$. Then

$$\theta^{(k+1)} = \underset{\theta}{\operatorname{argmin}} \sum_{t=1}^{N} \tilde{\epsilon}^2(t,\theta)$$

Show that the obtained recursion is precisely the Gauss-Newton procedure.

# Exercise 7.7

Convergence-rate for the Newton-Raphson and Gauss-Newton procedures.

Consider the algorithms:

$$\begin{cases} A_1: & x^{(k+1)} = x^{(k)} - V''(x^{(k)})^{-1}V'(x^{(k)})^T \\ A_2: & x^{(k+1)} = x^{(k)} - SV'(x^{(k)})^T \end{cases}$$

for minimization of $V(x)$, the matrix $S$ is positive definite.

(a) Introduce a positive constant $\alpha > 0$ in $A_2$ for controlling the step length:

$$A_2': \quad x^{(k+1)} = x^{(k)} - \alpha SV'(x^{(k)})^T$$

Show that this algorithm has a decreasing sequence of of function values $V(x^{(k+1)}) \leq V(x^{(k)})$ if $\alpha$ is sufficiently small.

(b) Apply the algorithms to the function

$$V(x) = \frac{1}{2}x^T Ax + b^T x - c$$

where $A$ is (strictly) positive definite. The minimum point satisfies $Ax_* = -b$, for $A_1$, one has $x^{(1)} = x_*$. For $A_2$ one has

$$(x^{(k+1)} - x_*) = (I - SA)(x^{(k)} - x_*)$$

(Assuming that $(I - SA)$ has all eigenvalues inside the unit circle, $A_2$ will converge with a linear rate. In particular when $S = A^{-1} + Q$ and $Q$ is small, then convergence will be superlinear.)

# Exercise 7.22

The Steiglitz-McBride method.

Consider the output-error model

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

where both $A$ and $B$ are of degree $n$. Consider the following iterative scheme

$$(A^{(k+1)}, B^{(k+1)}) = \operatorname*{argmin}_{A,B} \sum_{t=1}^{N} \left( A(q^{-1}) \left( \frac{1}{A^{(k)}} y(t) \right) - B(q^{-1}) \left( \frac{1}{A^{(k)}} u(t) \right) \right)^2$$

Assume the system

$$A_0(q^{-1})y(t) = B_0(q^{-1})u(t) + v(t)$$

where $A_0, B_0$ are coprime and of order $n$, $u(t)$ is PE of order $2n$, and $v(t)$ is a stationary disturbance independent of the input. Consider the asymptotic case.

(1) Assume that $v(t)$ is white noise. Show that the only stationary solution to the algorithm is $A = A_0$ and $B = B_0$.

(2) Assume that $v(t)$ is colored noise, show that $A = A_0$ and $B = B_0$ is in general not a stationary solution to the algorithm.