227-0689-00L System Identification **12 November 2018**

Exercise 7: Persistency of excitation, ARX models and least-squares

Background reading

The background material for this exercise is Sections 1.3, 4.2, 10.1, 13.2 and Appendix II of Ljung (*System Identification; Theory for the User*, 2nd Ed., Prentice-Hall, 1999).

Problem 1:

The output $y(k)$ of a linear, asymptotically stable, rational filter $\frac{B(z)}{A(z)}$ with input $u(k)$ can be written as

$$A(z)y(k) = B(z)u(k) + e(k), \tag{7.1}$$

where

$$A(z) = 1 + a_1 z^{-1} + ... + a_n z^{-n}, \tag{7.2}$$

$$B(z) = 1 + b_1 z^{-1} + ... + b_m z^{-m}, \tag{7.3}$$

and $e(k)$ denotes some random disturbance. With

$$\varphi(k) \equiv [-y(k-1) \ \ ... \ \ -y(k-n) \ \ u(k-1) \ \ ... \ \ u(k-m)]^T, \tag{7.4}$$

$$\theta \equiv [a_1 \ \ ... \ \ a_n \ \ b_1 \ \ ... \ \ b_m]^T, \tag{7.5}$$

equation (7.1) can be written as the linear regression

$$y(k) = \varphi^T(k)\theta + e(k). \tag{7.6}$$

The existence of a least squares estimate is equivalent to the nonsingularity (positive definiteness) of the covariance matrix

$$\Sigma = E\{\varphi(k)\varphi(k)^T\}. \tag{7.7}$$

Show that $\Sigma > 0$ if $u(k)$ is persistently exciting of order m. Which assumptions in relation to the disturbance must additionally hold?

Problem 2:

Consider the least-squares (LS) estimation problem:

$$Y = \Phi\theta + \epsilon,$$

where Φ is the regressor matrix and θ is the parameter vector to be estimated

$$Y := \begin{bmatrix} y(0) \\ \vdots \\ y(N-1)) \end{bmatrix}, \qquad \Phi := \begin{bmatrix} \varphi^\top(0) \\ \vdots \\ \varphi^\top(N-1) \end{bmatrix}, \qquad \theta := \begin{bmatrix} a_1 \\ \vdots \\ a_n \\ b_1 \\ \vdots \\ b_m \end{bmatrix}.$$

Assume that the noise, ϵ, is zero-mean Gaussian and correlated with $E\left\{\epsilon\epsilon^\top\right\} = R$. In this exercise we look for a linear estimator $\hat{\theta}$ of the form,

$$\hat{\theta} = Z^\top Y, \tag{7.8}$$

which is unbiased and minimizes its variance (cmp. lecture slide 9.25). For a given Φ show the following:

1. For a linear estimator of the form (7.8) to be unbiased we require that $Z^\top \Phi = I$.
2. The covariance matrix of any linear unbiased estimator of the form (7.8) is $\operatorname{cov}\left\{\hat{\theta}\right\} = Z^\top R Z$.
3. The covariance matrix of the best linear unbiased estimator (BLUE) $\hat{\theta}_Z$ with $\hat{\theta}_Z = \left(\Phi^\top R^{-1} \Phi\right)^{-1} \Phi^\top R^{-1} Y$ is $\operatorname{cov}\left\{\hat{\theta}_Z\right\} = \left(\Phi^\top R^{-1} \Phi\right)^{-1}$.
4. The best linear unbiased estimator $\hat{\theta}_Z$ exhibits the smallest variance in the class of all unbiased estimators, i.e. $\operatorname{cov}\left\{\hat{\theta}_Z\right\} \leq \operatorname{cov}\left\{\hat{\theta}\right\}$.

Hint: All covariance matrices are positive semi-definite and in our case we can assume that R is positive definite. The inverse of a positive definite matrix is also positive definite.

Matlab exercises:

Consider the ARX model

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

with $a = 1/2, b = 1$ and $w(t)$ is a sequence of independent and identically distributed (i.i.d.) random variables. Fix an input sequence $u(t)$ once and for all (e.g. i.i.d. random variables drawn from $\mathcal{N}(0, 1)$) and estimate a and b by least squares based on N observations. Repeat the experiment for different realisations of $w(t)$.

1. Plot the histograms of the least squares estimates across different sample lengths N and assume that $w(t)$ is drawn from (a) $\mathcal{N}(0, 0.2)$ and (b) a uniform distribution with zero mean and same variance. What can be observed for large N in the latter case?

2. Plot the histograms of the sum of squared residuals normalized by the noise variance for case (a). Note that this quantity is distributed according to $\chi^2(N-p)$, where p is the number of parameters and the distribution can well be approximated by $\mathcal{N}(N-p, 2(N-p))$, for large values of $N-p$. Verify this fact by plotting the probability density of χ^2 and normal distributions with appropriate parameters along with the corresponding histograms.