227-0689-00L System Identification **21 November 2018**

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

We will not provide solutions for the MATLAB parts of the exercises. These will instead be discussed in the exercises sessions.

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?

Solution

Let

$$v(k) = \frac{1}{A(z)}e(k) \qquad x(k) = \frac{B(z)}{A(z)}u(k) \tag{7.8}$$

The following underlies the **assumption:** $E\{u(k)e(k-\tau)\} = 0$, i.e. the noise and the input are uncorrelated. Then one can write

$$
\begin{aligned}
\Sigma = E &\left\{ \begin{bmatrix} x(k-1) \\ \vdots \\ x(k-n) \\ u(k-1) \\ \vdots \\ u(k-m) \end{bmatrix} [x(k-1) \ \dots \ x(k-n) \quad u(k-1) \ \dots \ u(k-m)] \right\} \\
+ E &\left\{ \begin{bmatrix} v(k-1) \\ \vdots \\ v(k-n) \\ 0 \\ \vdots \\ 0 \end{bmatrix} [v(k-1) \ \dots \ v(k-n) \quad 0 \ \dots \ 0] \right\} \qquad (7.9) \\
= &\begin{bmatrix} A & B \\ B^T & C \end{bmatrix} + \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix} \qquad (7.10)
\end{aligned}
$$

where A, B, C are the $n \times n$, $n \times m$, $m \times m$ sub-matrices in the first summand of equation (7.9) and D the $n \times n$ sub-matrix in the second summand of equation (7.9). Clearly, the condition $C > 0$ is necessary for $\Sigma > 0$. Under the **assumption:** $D > 0$, i.e. the noise is persistently exciting, this condition is also sufficient, as from the Schur complement we then find

$$A - BC^{-1}B^T \geq 0 \qquad (7.11)$$

and

$$\text{rank } \Sigma = m + \text{ rank } (A + D - BC^{-1}B^T) \qquad (7.12)$$

Thus

$$\text{rank}\Sigma = n + m \qquad (7.13)$$

This shows that under the mild assumptions on the noise $e(k)$ $(D > 0)$, $\Sigma > 0$ if $u(k)$ is persistently exciting of order m.

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,

$$\hat{\theta} = Z^\top Y, \tag{7.14}$$

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 (7.14) to be unbiased we require that $Z^\top \Phi = I$.
2. The covariance matrix of any linear unbiased estimator of the form (7.14) 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. $\operatorname{cov}\left\{\hat{\theta}_Z\right\} \le \operatorname{cov}\left\{\hat{\theta}\right\}$.

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.

Solution

1. For a linear estimator of the form (7.14) to be unbiased we require that

   $$\theta = E\left\{\hat{\theta}\right\}.$$

   Hence, for zero-mean Gaussian noise, ϵ, and fixed Φ we get

   $$\theta = E\left\{Z^\top Y\right\} = E\left\{Z^\top\left(\Phi\theta + \epsilon\right)\right\} = Z^\top \Phi\theta,$$

   which requires that

   $$Z^\top \Phi = I.$$

2. The covariance matrix of any linear unbiased estimator of the form (7.14) is

$$
\begin{aligned}
\operatorname{cov}\left\{\hat{\theta}\right\} &= E\left\{\left(\hat{\theta}-\theta\right)\left(\hat{\theta}-\theta\right)^\top\right\} && \left[\hat{\theta} = Z^\top Y\right] \\
&= E\left\{\left(Z^\top Y-\theta\right)\left(Z^\top Y-\theta\right)^\top\right\} && \left[Y = (\Phi\theta+\epsilon)\right] \\
&= E\left\{\left(Z^\top(\Phi\theta+\epsilon)-\theta\right)\left(Z^\top(\Phi\theta+\epsilon)-\theta\right)^\top\right\} && \left[Z^\top\Phi = I\right] \\
&= Z^\top E\left\{\epsilon\epsilon^\top\right\} Z && \left[E\left\{\epsilon\epsilon^\top\right\} = R\right] \\
&= Z^\top R Z.
\end{aligned}
$$

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

$$
\begin{aligned}
\operatorname{cov}\left\{\hat{\theta}_Z\right\} &= Z^\top R Z && \left[Z^\top = \left(\Phi^\top R^{-1}\Phi\right)^{-1}\Phi^\top R^{-1}\right] \\
&= \left(\Phi^\top R^{-1}\Phi\right)^{-1}\Phi^\top R^{-1} R R^{-1}\Phi\left(\Phi^\top R^{-1}\Phi\right)^{-1} \\
&= \left(\Phi^\top R^{-1}\Phi\right)^{-1} && .
\end{aligned}
$$

4. For $\hat{\theta}_Z$ to be the BLUE in the class of all unbiased estimators, we require that $\operatorname{cov}\left\{\hat{\theta}\right\} - \operatorname{cov}\left\{\hat{\theta}_Z\right\} \geq 0$:

$$
\begin{aligned}
\operatorname{cov}\left\{\hat{\theta}\right\} - \operatorname{cov}\left\{\hat{\theta}_Z\right\} &= Z^\top R Z - \left(\Phi^\top R^{-1}\Phi\right)^{-1} && \left[Z^\top\Phi = I\right] \\
&= Z^\top R Z - Z^\top\Phi\left(\Phi^\top R^{-1}\Phi\right)^{-1}\Phi^\top Z \\
&= Z^\top\left[R - \Phi\left(\Phi^\top R^{-1}\Phi\right)^{-1}\Phi^\top\right] Z && \left[F = R - \Phi\left(\Phi^\top R^{-1}\Phi\right)^{-1}\Phi^\top\right] \\
&= Z^\top F Z.
\end{aligned}
$$

Hence, we need to show that F is positive semidefinite so that $Z^\top F Z$ is positive semidefinite as well. We can show that

$$
\begin{aligned}
F^\top R^{-1} F &= \left(R - \Phi\left(\Phi^\top R^{-1}\Phi\right)^{-1}\Phi^\top\right)^\top R^{-1}\left(R - \Phi\left(\Phi^\top R^{-1}\Phi\right)^{-1}\Phi^\top\right) \\
&= \left(R - \Phi\left(\Phi^\top R^{-1}\Phi\right)^{-1}\Phi^\top\right)\left(I - R^{-1}\Phi\left(\Phi^\top R^{-1}\Phi\right)^{-1}\Phi^\top\right) \\
&= R - \Phi\left(\Phi^\top R^{-1}\Phi\right)^{-1}\Phi^\top - \Phi\left(\Phi^\top R^{-1}\Phi\right)^{-1}\Phi^\top + \\
&\quad \Phi\left(\Phi^\top R^{-1}\Phi\right)^{-1}\Phi^\top R^{-1}\Phi\left(\Phi^\top R^{-1}\Phi\right)^{-1}\Phi^\top \\
&= R - \Phi\left(\Phi^\top R^{-1}\Phi\right)^{-1}\Phi^\top \\
&= F.
\end{aligned}
$$

Since R is positive definite, R^{-1} is positive definite, we can conclude that $F = F^\top R^{-1} F$ has to be positive semidefinite and therefore $\operatorname{cov}\left\{\hat{\theta}_Z\right\} \leq \operatorname{cov}\left\{\hat{\theta}\right\}$. Hence, $\hat{\theta}_Z$ is the BLUE with the smallest variance, but the assembly of Z requires knowledge of the error variance, R.

MATLAB **exercise:**

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.

Solution hints

1. To generate the random variable $w(t)$ for a realisation use *randn* and *rand* for the normal and uniform distribution respectively. The amplitude of the uniform distribution around zero mean needs to be computed for the variance of 0.2. Compute the dynamics of the disturbed model for each distribution. Assemble the resulting data set of inputs y and outputs u in least squares form, as illustrated in slides 9.25-26 of Lecture 9. Solving the least squares problem leads to estimates of a and b for each realisation. Plot the histogram of these estimates using MATLAB command *hist*. It should look similar (only similar because of different noise and input realisations) to Figure 7.1 which we generated for 1000 realisations of $w(t)$ of lengths $N = \{2^6, 2^8, 2^{10}, 2^{12}, 2^{14}\}$. For larger N the estimates for uniform distribution converge to the given parameters of a and b.

2. Compute the residual sum squared (RSS) error for each realisation and divide it by the variance 0.2. Compute the χ distribution using the MATLAB command *chi2rnd*, where we have 2 parameters for the present problem. Finally plot the histogram of the RSS error, the χ distribution and its approximation $\mathcal{N}(N - p, 2(N - p))$. Note the good approximation for larger N in Figure 7.2.

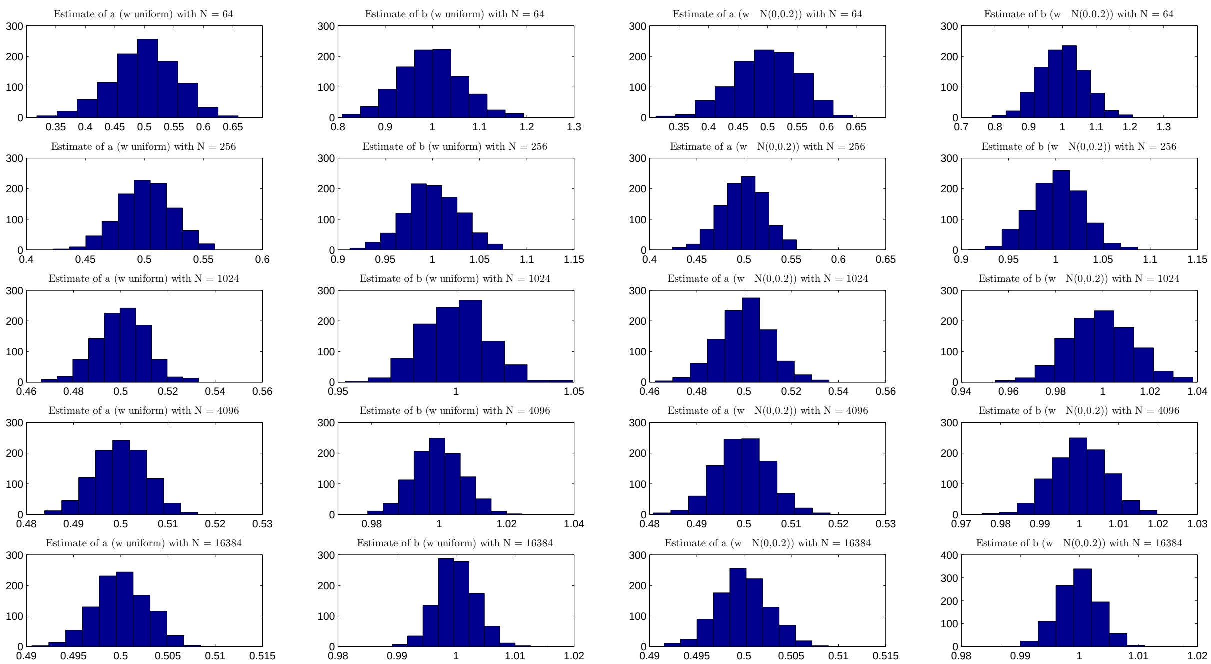

Figure 7.1: Histograms of the least squares estimates across different sample lengths N and for $\mathcal{N}(0, 0.2)$ and a uniform distribution with zero mean.

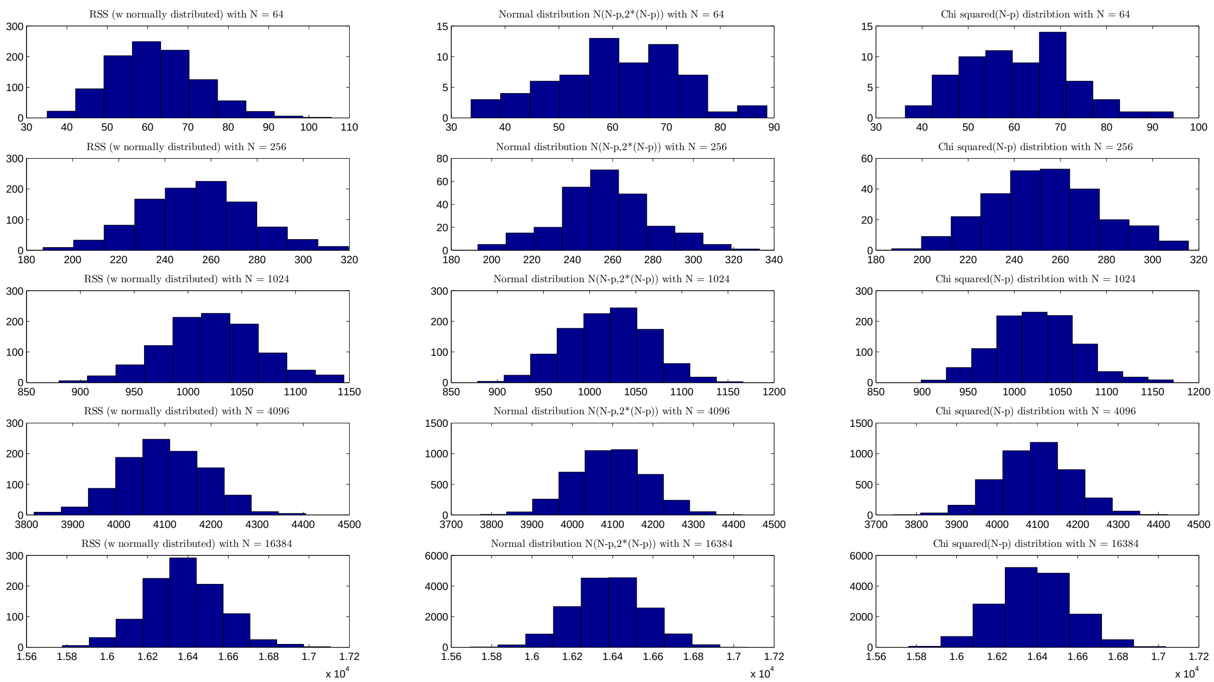

Figure 7.2: Histograms of (left) RSS error for normally distributed disturbance, (center) approximation $\mathcal{N}(N-p, 2(N-p))$ and (right) χ distribution for $p = 2$.