Final Exam in System Identification for F and STS

Answers and Brief Solutions

Date: March 16, 2007
Examiner: Magnus Mossberg

1. a) \[ \lim_{N \to \infty} \hat{\theta}_N - \theta_0 = \left[ \begin{array}{c} E\{y^2(t)\} \\ -E\{y(t)u(t)\} \\ E\{u^2(t)\} \end{array} \right] \left[ \begin{array}{c} -E\{y(t-1)w(t)\} \\ E\{u(t-1)w(t)\} \end{array} \right]^{-1} \neq 0 \]

b) For example,
\[ \hat{\theta}_N = \left( \frac{1}{N} \sum_{t=1}^{N} z(t) \phi^T(t) \right)^{-1} \left( \frac{1}{N} \sum_{t=1}^{N} z(t)y(t) \right) \]

where
\[ z(t) = [-y(t-2) \ u(t-1)]^T \]

2. a) \[ \text{mse}(\hat{d}) = \text{var}(\hat{d}) + (\text{bias}(\hat{d}))^2 = \text{var}(\hat{d}) = \frac{\lambda^2}{N} \]

b) \[ \text{mse}(\hat{d}) = \text{var}(\bar{d}) + (\text{bias}(\bar{d}))^2 = \frac{a^2 \lambda^2}{N} + d^2(a-1)^2 \]

where
\[ a_{\text{opt}} = \frac{d^2}{d^2 + \lambda^2/N} \]

minimizes \( \text{mse}(\bar{d}) \). The estimator \( \bar{d} \) is not realizable with this choice of \( a \), since \( a_{\text{opt}} \) depends on the unknown parameter \( d \).

c) \[ \hat{d}(t) = \left( \sum_{k=1}^{t} \lambda^{t-k} \right)^{-1} \sum_{k=1}^{t} \lambda^{t-k} x(k) \]
\[ \hat{d}(t) = \hat{d}(t-1) + \frac{1 - \lambda}{1 - \lambda^t} (x(t) - \hat{d}(t-1)) \]
3. a) Not identifiable; predictor ˆ\(y(t|b_1, b_2) = (b_1 + b_2)u_0\).
   
b) Not identifiable; predictor ˆ\(y(t|a, b) = (-bc - a)y(t - 1)\).
   
   Use, for example, \(u(t) = -cy(t - 1)\) or use two different values of \(c\).

4. a) One resonance peak is given by a complex conjugated pair of poles. Start with model order four.
   
b) ARX; linear regression.
   
c) Overfit.
   
d) The result is often given as a table or as a curve.
   
e) The \(A\)-polynomial of the ARX-model must describe the disturbance dynamic through \(1/A\). This can result in a slightly erroneous description of the system dynamic. It is easier for the ARMAX-model to describe both the system- and the disturbance dynamic due to the \(C\)-polynomial.
   
f) A useful and good model can describe new data with high enough accuracy. Such a test can also reduce the risk for overfit.

5. Predictor
   
   \[\hat{y}(t|\theta) = \begin{bmatrix} -y(t - 1) & -y(t - 2) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \varphi^T(t) \theta\]
   
a) 
   \[E((\hat{\theta}_N - \theta_0)(\hat{\theta}_N - \theta_0)^T) \approx \frac{\lambda^2}{N} (E\{\varphi(t)\varphi^T(t)\})^{-1} = \frac{\lambda^2}{N} \begin{bmatrix} r_y(0) & r_y(1) \\ r_y(1) & r_y(0) \end{bmatrix}^{-1} = \frac{1}{N} \begin{bmatrix} 1 & a_0 \\ a_0 & 1 \end{bmatrix}\]
   
   so \(\text{var} (\hat{a}_1) = \text{var} (\hat{a}_2) = 1/N\).
   
b) 
   \[\lim_{N \to \infty} \hat{\theta}_N = (E\{\varphi(t)\varphi^T(t)\})^{-1} E\{\varphi(t)y(t)\}\]
   
   \[= \begin{bmatrix} r_y(0) & r_y(1) \\ r_y(1) & r_y(0) \end{bmatrix}^{-1} \begin{bmatrix} -r_y(1) \\ -r_y(2) \end{bmatrix} = \begin{bmatrix} a_0 \\ 0 \end{bmatrix}\]
   
   \[\lim_{N \to \infty} \hat{a}_1 = a_0\]. The estimate is correct.
   
   \[\lim_{N \to \infty} \hat{a}_2 = 0\]. This makes sense when comparing the model with the system.

6. a) 
   \[\hat{G}(e^{i\omega}) = \sum_{k=1}^{\infty} \hat{g}(k)e^{-i\omega k} = \frac{1}{\alpha} \sum_{k=1}^{\infty} y(k)e^{-i\omega k} = \frac{1}{\alpha} \sqrt{N} Y_N(\omega) = \frac{Y_N(\omega)}{U_N(\omega)} = \hat{G}_N(e^{i\omega})\]
where it is used that

\[ U_N(\omega) = \frac{1}{\sqrt{N}} \alpha \]

in the second last equality.

b)

\[
\tilde{g}(t) = \hat{g}(t) - g_0(t) = \frac{v(t) - v(t-1)}{\beta}
\]

\[
\text{var}(\tilde{g}(t)) = \frac{2\lambda^2}{\beta^2}
\]

7. Linear regression

\[
-\hat{r}(\tau) = \begin{bmatrix} \hat{r}(\tau-1) & \cdots & \hat{r}(\tau-na) \end{bmatrix} \begin{bmatrix} a_1 \\ \vdots \\ a_{na} \end{bmatrix}
\]

where

\[
\hat{r}(\tau) = \frac{1}{N} \sum_{t=\tau}^{N} y(t)y(t-\tau)
\]

\[ \tau = nc + 1, nc + 2, \ldots, nc + na \] give

\[
\begin{bmatrix}
\hat{r}(nc) & \cdots & \hat{r}(nc+1-na) \\
\vdots & \ddots & \vdots \\
\hat{r}(nc+na-1) & \cdots & \hat{r}(nc)
\end{bmatrix}
\begin{bmatrix}
\hat{a}_1 \\
\vdots \\
\hat{a}_{na}
\end{bmatrix}
= 
\begin{bmatrix}
-\hat{r}(nc+1) \\
\vdots \\
-\hat{r}(nc+na)
\end{bmatrix}
\]

This can (essentially, apart from different start indexes in sums) be written as

\[
\left( \frac{1}{N} \sum_{t=1}^{N} z(t)\varphi^T(t) \right) \tilde{\theta}_N^\text{IV} = \frac{1}{N} \sum_{t=1}^{N} z(t)y(t)
\]

where

\[ \varphi(t) = [-y(t-1) \cdots -y(t-na)]^T \]

(the AR-part is estimated) and

\[ z(t) = [-y(t-(nc+1)) \cdots -y(t-(nc+na))]^T \]
8. a) Let

$$\hat{\lambda}_N = \frac{1}{N} \sum_{t=1}^{N} \varepsilon^2 (t, \hat{\theta}^{(p)}_N)$$

The $C_p$-criterion can then be written as

(1) $$C_p = \frac{N\hat{\lambda}_N}{\hat{s}_N^2} - (N - 2p)$$

When selecting $p$, the quantities $N$ and

$$\hat{s}_N^2 = \frac{1}{N} \sum_{t=1}^{N} \varepsilon^2 (t, \hat{\theta}^{(p_{\text{max}})}_N)$$

are seen as constants. Minimizing (1) with respect to $p$ is therefore equivalent to minimizing

(2) $$N\hat{\lambda}_N + 2p\hat{s}_N^2$$

with respect to $p$. Compare with AIC, which minimizes

(3) $$N\hat{\lambda}_N \left(1 + \frac{2p}{N}\right) = N\hat{\lambda}_N + 2p\hat{\lambda}_N$$

The only difference between (2) and (3) is that $\hat{\lambda}_N$ in (3) is replaced by $\hat{s}_N^2$. From an operational point of view, this difference is minor.

b) Assume that the denominator is

$$C(z) = (z - p_1)(z - p_2)$$

so

$$C(e^{i\omega}) = (e^{i\omega} - p_1)(e^{i\omega} - p_2)$$

Resonance peaks at $\omega = \pm 1$ means that $C(e^{i\omega})$ has minimum values for $\omega = \pm 1$:

$$e^i - p_1 = 0$$
$$e^{-i} - p_2 = 0$$

so $p_1 = e^i$ and $p_2 = e^{-i}$.

c) Minimum point of $Q(x)$? Solve $Q'(x) = 0$ to get

$$x = x_k - (J''(x_k))^{-1} J'(x_k)$$

Take $x_{k+1}$ as the minimum point:

$$x_{k+1} = x_k - (J''(x_k))^{-1} J'(x_k)$$