ID3

ARX Identification

3.3 LEAST SQUARES ESTIMATE OF ARX MODELS

ARX models are mainly used for prediction and control so that the cost function introduced for their determination is the sum of the squares of prediction errors on the set of available data. Consider an assumed order, n, for an ARX model and the one–step–ahead predictor (3.2.1). If the input–output sequences $\{u(\cdot),\ y(\cdot)\}$ have been observed in the time interval $[1, L]$, predictor (3.2.1) can be used to compute predicted output values in the $N = L - n$ times $n+1, \dots, L$. The mean square prediction error on the available data is thus given, for a given set of parameters

$$\theta = [\alpha_1 \dots \alpha_n \ \beta_1 \dots \beta_n]^T, \tag{3.3.1}$$

by

$$J(\theta) = \frac{1}{N} \sum_{t=n+1}^{L} \big(y(t) - y(t|t-1) \big)^2 = \frac{\| y^\circ - y^p(\theta) \|_2^2}{N} = \frac{\varepsilon^T(\theta)\, \varepsilon(\theta)}{N} \tag{3.3.2}$$

where

$$y^p(\theta) = \big[\, y(n+1|n) \dots y(L|L-1) \,\big]^T \tag{3.3.3}$$

$$y^\circ = \big[\, y(n+1) \dots y(L) \,\big]^T \tag{3.3.4}$$

$$\varepsilon(\theta) = y^\circ - y^p(\theta)\,. \tag{3.3.5}$$

By introducing now the following Hankel matrices $H_n(y)$ and $H_n(u)$

$$H_n(y) = \begin{bmatrix} y(1) & \dots & y(n) \\ \vdots & & \vdots \\ y(L-n) & \dots & y(L-1) \end{bmatrix} \tag{3.3.6}$$

$$H_n(u) = \begin{bmatrix} u(1) & \dots & u(n) \\ \vdots & & \vdots \\ u(L-n) & \dots & u(L-1) \end{bmatrix} \tag{3.3.7}$$

it follows that

$$y^p(\theta) = \left[\, H_n(y)\; H_n(u) \,\right] \theta = H\theta\,. \tag{3.3.8}$$

The parameter vector θ minimizing cost function (3.3.2) is given by

$$\theta^\circ = H^+\, y^\circ \tag{3.3.9}$$

where H^+ denotes the pseudoinverse (or generalized inverse) of H. In fact, assuming $\theta = \theta^\circ$ we get

$$y^p(\theta^\circ) = H H^+\, y^\circ; \tag{3.3.10}$$

HH^+ performs the orthogonal projection of y° on im H; $y^p(\theta^\circ)$ is the vector of im H which minimizes the Euclidean norm of the error $\|y^\circ - y^p(\theta)\|_2$, and thus even cost function (3.3.2).

3.3.1 Algorithmic aspects

The pseudoinverse, H^+, of H can be computed by means of relation

$$H^+ = \left(H^T H\right)^{-1} H^T = S^{-1} H^T \tag{3.3.11}$$

when H has maximal rank, that is in our case, when its columns are linearly independent; in this case the solution minimizing cost function $J(\theta)$ (3.3.2) is unique and is given by the well–known least squares formula

$$\theta^\circ = S^{-1} H^T\, y^\circ\,. \tag{3.3.12}$$

Expression (3.3.12) can be directly deduced from (3.3.2) zeroing the gradient of $J(\theta)$, $\partial J(\theta)/\partial\theta$. We obtain, remembering that $\partial(A^T\theta)/\partial\theta = A$, $\partial(\theta^T A\theta)/\partial\theta = (A+A^T)\theta$, $\partial(\theta^T A)/\partial\theta = A$,

$$\begin{aligned} \frac{\partial J(\theta)}{\partial\theta} &= \frac{\partial}{\partial\theta}\left[\frac{\varepsilon^T(\theta)\,\varepsilon(\theta)}{N}\right] = \frac{1}{N}\,\frac{\partial}{\partial\theta}\left[\,(y^\circ - H\theta)^T (y^\circ - H\theta)\,\right] \\ &= \frac{1}{N}\,\frac{\partial}{\partial\theta}\left[\,\theta^T H^T H\theta - \theta^T H^T y^\circ - y^{\circ T} H\theta + y^{\circ T} y^\circ\,\right] \\ &= \frac{2}{N}\left(H^T H\theta - H^T y^\circ\right) \end{aligned} \tag{3.3.13}$$

which, under the assumption of invertibility for $H^T H$, leads to (3.3.12). In the practical implementation of the least squares algorithm it is not recommended to compute the inverse of $S = H^T H$ by means of standard algorithms because of its possible ill–conditioning. Since S is symmetrical and positive definite it is possible to use numerical approaches that take advantage of these properties to compute $\theta°$. Among suggested methods it is possible to mention Choleski factorization of S to compute S^{-1}, singular value decomposition (SVD) of H to solve (3.3.9) or Golub–Householder method.

SECTIONS MODULES QUESTIONS HOME PAGE
PREV. MODULE FAQ TUTOR NEXT MODULE