Identification of Parsimonious Model Structure

ZHANG Zhiyong

(Institute of Automation, Fuzhou University·Fuzhou, 350002, PRC)

Abstract: The projection technique is used to identify a parsimonious model structure of canonical vector difference equations for an actual system. The structure consists of observability indices, autoregression orders and a parsimonious structure of parameters to be estimated. A new procedure, with a whitening projection operator, is proposed to estimate observability indices as well as autoregression orders in the case of coloured equation errors, which can be modelled with AR. Moreover, an algorithm is presented to determine the parsimonious parameter structure, each element of which has a significant contribution to improve the quality of the model.

Key words: identification; model structure; parameter estimation

1 Introduction

Among the various existing methods of structural identification, the procedures based on the range error test or the residual error test are still widely used since they are simple, direct and easily understood. But a limitation of these methods is that they give satisfactory results only if the equation error is a white-noise process or if it has known statistics. Moreover, though several approaches to determine a model set with the smallest possible number of parameters to be estimated can be found in the literature, they can be used only after the computation of parameter values.

In section 3, a new procedure will be proposed to determine the structure indices of ARARX model sets, directly based on the input and output data. In section 4, the determination of so-called parsimonious parameter structure will be delt with. The proposed approach can be used prior to parameter estimation.

2 Formulation of the Problem

Consider the following multivariable discrete-time model described by the canonical vector difference equations (VDE)

$$y(t) = A(q^{-1})y(t) + B(q^{-1})u(t) = \sum_{k=1}^{\nu_m} A^{(k)}y(t-k) + \sum_{k=1}^{\nu_m} B^{(k)}u(t-k), \tag{1}$$

where $y(\cdot)\in R^m, u(\cdot)\in R^r$, q^{-1} is the backward shift operator, i. e. $q^{-1}y(t)=y(t-1)$. If the observability indices are arranged in increasing order, $\nu_1\leqslant\nu_2\leqslant\cdots\leqslant\nu_m$, without any loss of generality, the coefficient matrices $A^{(k)}(k=1, \cdots, m)$ are all lower triangular matrices, and the number of non-zero i-th rows in $A^{(k)}(k=1, \cdots, m)$ is precisely equal to the observability index

Manuscript received Nov. 8, 1990, revised Jun. 24, 1991.

V_i [1].

In the noisy case, VDE (1) can be rewritten in the form

$$y_i(t) = A_i(q^{-1})y(t) + B_i(q^{-1})u(t) + \omega_i(t), \quad i = 1, \cdots, m, \tag{2}$$

in which

$$A_i(q^{-1}) = \Big[\Big(\sum_{k=\nu_i-1}^{0} a_{i1k}q^{k-\nu_i}\Big), \cdots, \Big(\sum_{k=\nu_i-1}^{0} a_{i(i-1)k}q^{k-\nu_i}\Big), \Big(\sum_{k=\nu_i-1}^{0} a_{iik}q^{k-\nu_i}\Big), \cdots, \Big(\sum_{k=\nu_i-1}^{0} a_{imk}q^{k-\nu_i}\Big)\Big], \tag{3}$$

$$B_i(q^{-1}) = \Big[\Big(\sum_{k=\nu_i-1}^{0} b_{i1k}q^{k-\nu_i}\Big), \cdots, \Big(\sum_{k=\nu_i-1}^{0} b_{irk}q^{k-\nu_i}\Big)\Big], \tag{4}$$

$$\omega_i(t) = [1 - D_i(q^{-1})]\omega_i(t) + e_i(t), \quad D_i(q^{-1}) = 1 - d_{i1}q^{-1} - \cdots - D_{in_{di}}q^{-n_{di}}. \tag{5}$$

In (5), $e_i(t)$ is a white noise. From (2) and (5), a prediction model set can be constructed in the folloing form

$$\hat{y}_i(t) = \hat{A}_i(q^{-1})y(t) + \hat{B}_i(q^{-1})u(t) + [1 - \hat{D}_i(q^{-1})]\hat{\omega}_i(t), \tag{6}$$

$$\hat{\omega}_i(t) = y_i(t) - [\hat{A}_i(q^{-1})y(t) + \hat{B}_i(q^{-1})u(t)]. \tag{7}$$

The topic of this paper is to choose a feasible model structure ($\hat{\nu}_i$, $\hat{n}_{di}$, $\hat{\Theta}_{1i}$, $i=1, \cdots, m$) of model sets (6) and (7) for an actual system by using input and output data without the explicit estimation of parameters. Here $\hat{\nu}_i$ and $\hat{n}_{di}$ are, respectively, the observability index and autoregression order of the i-th submodel. $\hat{\Theta}_{1i}$ is a set which consists of those nonzero parameters in $\hat{A}_i(\cdot)$ and $\hat{B}_i(\cdot)$ that have to be estimated. The triple ($\hat{\nu}_i$, $\hat{n}_{di}$, $\hat{\Theta}_{1i}$) is the parsimonious structure[2] of i-th submodel sets (6) and (7) for an actual system in the least square sense.

3 Determination of the Structure Indices

Using output and input data, the following equations can be written

$$\mathbf{y}_i(t) = \sum_{j=1}^{m}\sum_{k=\nu_i-1}^{0} a_{ijk}\mathbf{y}_j(t+k-\nu_i) + \sum_{j=1}^{r}\sum_{k=\nu_i-1}^{0} b_{ijk}\mathbf{u}_j(t+k-\nu_i) + \mathbf{w}_i(t), \tag{8}$$

$$\mathbf{w}_i(t) = \sum_{s=1}^{n_{di}} d_{is}\mathbf{w}_i(t-s) + \mathbf{e}_i(t), \tag{9}$$

where

$$\mathbf{y}_j(t+k) = [y_j(t+k), \cdots, y_j(t+k+N-1)]^{\mathrm{T}},$$
$$\mathbf{u}_j(t+k) = [u_j(t+k), \cdots, u_j(t+k+N-1)]^{\mathrm{T}},$$
$$\mathbf{w}_i(t) = [\omega_i(t), \cdots, \omega_i(t+N-1)]^{\mathrm{T}}.$$

Let

$$\begin{aligned} H_n = [&\mathbf{y}_1(t-1), \cdots, \mathbf{y}_m(t-1), \mathbf{u}_1(t-1), \cdots, \mathbf{u}_r(t-1),\\ &\mathbf{y}_1(t-2), \cdots, \mathbf{y}_m(t-2), \mathbf{u}_1(t-2), \cdots, \mathbf{u}_r(t-2),\\ &\vdots\\ &\mathbf{y}_1(t-n_i), \cdots, \mathbf{y}_m(t-n_i), \mathbf{u}_1(t-n_i), \cdots, \mathbf{u}_r(t-n_i)], \end{aligned}$$

in which n_i is the test observability index.

Projecting $y_i(t)$ onto the null space of the space $S_n = \text{span}_{col}(H_n)$, it can be shown that there is a residual vector r_n such that $r_n = M_n y_i(t)$ where $M_n = (I - H_n H_n^+)$ and H_n^+ is a pseudoniverse of H_n. M_n is the complement projection operator of S_n.

If the test observability index n_i is not larger than the real one ν_i, i. e. $n_i \leqslant \nu_i$, the space $S_\nu = \text{span}_{col}\{H_\nu\}$ associated with ν_i can be decomposed as a direct sum of S_n and its complement $\overline{S}_n$ in S_ν. That is $S_\nu = S_n \oplus \overline{S}_n$. And relation $M_n = M_\nu + P_{\nu-n}$ holds, in which $P_{\nu-n}$ is a projection operator of subspace $\overline{S}_n$. Using this relation and the orthogonality of projection operators, the following can be shown

$$\|r_n\|^2 = \|r_\nu\|^2 + \|P_{\nu-n} y_i(t)\|^2, \quad (n_i \leqslant \nu_i).$$

The second term on the righthand side will be decreasing to zero with $(\nu_i - n_i)$ decreasing to zero[3,4].

If $n_i \geqslant \nu_i$, it follows that $M_n y_i(t) = M_n w_i(t)$ since

$$\text{span}_{col}\{H_\nu\} \subseteq \text{span}_{col}\{H_n\}.$$

Using equations (9) and (8), the following can be obtained

$$r_n = \Delta(n_i) + e_i(t), \quad (n_i \geqslant \nu_i) \tag{10}$$

and then

$$\|r_n\|^2 = \|\Delta(n_i)\|^2 + \|e_i(t)\|^2, \quad (n_i \geqslant \nu_i),$$

where

$$\Delta(n_i) = \sum_{s=1}^{n_{di}} d_{is} M_n \Big\{ y_i(t-s) - \Big[\sum_{j=1}^{m} \sum_{k=-\nu_j-1}^{0} a_{ijk} y_j(t-s+k-\nu_i)$$

$$+ \sum_{j=1}^{r} \sum_{k=-\nu_j-1}^{0} b_{ijk} u_j(t-s+k-\nu_i) \Big] \Big\} + e_i(t), \quad (n_i \geqslant \nu_i). \tag{11}$$

In the special case of $d_{is} = 0 (\forall \ s)$, the algorithms proposed in [3, 4] are efficient in determining the observability indices since the argument $\|\Delta(n_i)\|^2$ equals zero. In general cases, however, $d_{is} \neq 0$, and $\|\Delta(n_i)\|^2$ will be a function of the test observability index n_i. From equation (11), it can be seen clearly that the projection operator M_n will nullify more and more vectors of $y_j(\cdot)$ and $u_j(\cdot)$ in (11) with the increase of n_i because the $\text{span}_{col}(H_n)$ becomes larger and larger. When n_i grows to be equal to $(\nu_i + n_{di})$ or more, the $\text{span}_{col}\{H_n\}$ will incorporate all of the vectors $y_j(\cdot)$ and $u_j(\cdot)$, and then M_n will annihilate all of them, and $\|\Delta(n_i)\|^2$ will become zero.

From the above analysis, a conclusion can be drawn as follows

$$V(n_i) = \begin{cases} \|P_{\nu-n} y_i(t)\|^2 + \|\Delta(\nu_i)\|^2 + \|e_i(t)\|^2, & n_i < \nu_i, \\ \|\Delta(\nu_i)\|^2 + \|e_i(t)\|^2, & n_i = \nu_i, \\ \|\Delta(\nu_i)\|^2 - \|P_{\nu-n} w_i(t)\|^2 + \|e_i(t)\|^2, & \nu_i < n_i < \nu_i + n_{di}, \\ \|e_i(t)\|^2 \sim \sigma^2 \chi^2(N - L_0), & n_i \geqslant \nu_i + n_{di}. \end{cases} \tag{12}$$

Remarks

- The argument $V(n_i)$ is constructed in the form $V(n_i) = \|r_n\|^2$.
- In equation (12), relation $\|r_\nu\|^2 = \|\Delta(\nu_i)\|^2 + \|e_i(t)\|^2$ is used.

- $P_{n-\nu}$ is a projection operator of $\overline{S}_\nu$, which is subject to the relation of the decomposition

$$S_n = \text{span}_{\cos}\{H_n\},$$

$$S_n = S_\nu \oplus \overline{S}_\nu, \quad \nu_i < n_i < \nu_i + n_{di}.$$

- In equation (12), N is the dimension of data vectors, and L_0 is the number of vectors, $y_j(\cdot)$ and $u_j(\cdot)$, in equation (8) with the test observability index n_i.
- If the argument $V(n_j)$ is plotted versus n_i, its typical shape is similar to the dashed lines in Fig. 1. From this plot, (ν_i+n_{di}) is obtained as the smallest integer n_i for which the part of the residual plot for $n_i \geqslant (\nu_i + n_{di})$ is almost flat.

In order to determine the observability index ν_i, an additional whitening projection operator will be adapted[5,6]. Assume that the residual error r_n in the case of $n_i \geqslant \nu_i$ can be modelled with AR, i. e.

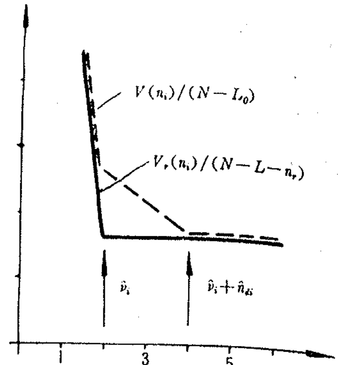

Fig. 1 Determination of ν_i and n_{di}

$$r_n = \sum_{s=1}^{n_{r0}} c_s r_n(t-s) + e_i(t), \quad n_i \geqslant \nu_i, \tag{13}$$

where

$$r_n(t-s) = [r_n(t-s), r_n(t-s+1), \cdots, r_n(t-s+N-1)]^T.$$

Construct a matrix M_r, a complement projection operator of the null space of

$$\text{span}\{r_n(t-s), s=1,2,\cdots,n_r \text{ and } n_r \geqslant n_{r0}\}.$$

Multiplying equation (13) by the operator M_r from the left, it follows that

$$M_r r_n = e_i(t), \quad n_i \geqslant \nu_i \text{ and } n_r \geqslant n_{r0},$$

and

$$\|M_r r_n\|^2 = \|e_i(t)\|^2, \quad n_i \geqslant \nu_i \text{ and } n_r \geqslant n_{r0}.$$

This means that the residual error r_n is whitened by the operator M_r in the case of $n_i \geqslant \nu_i$. But for $n_i < \nu_i$, the other relations can be derived from equations (11) and (12)

$$\|M_r r_n\|^2 = \|M_r[P_{\nu-n} y_i(t) + \sum_{s=1}^{n_{r0}} c_s r_\nu(t-s)]\|^2 + \|e_i(t)\|^2.$$

If $V_r(n_i)$ denotes the squared length of the so-called whitening residual error $(M_r\ r_n)$, i. e. $V_r(n_i) = \|M_r\ r_n\|^2 = \|M_r\ M_n\ y_i(t)\|^2$, an expression can be summed up as follows

$$V_r(n_i) = \begin{cases} \|M_r[P_{\nu-n} y_i(t) + \sum_{s=1}^{n_{r0}} c_s r_\nu(t-s)]\|^2 + \|e_i(t)\|^2, & n_i < \nu_i, \\ \|e_i(t)\|^2 \sim \sigma^2 \chi^2(N-L-n_r), & n_i \geqslant \nu_i. \end{cases} \tag{14}$$

Remarks:

- An example of the plot $V_r(n_i)$ is shown with the solid lines in Fig. 1. From these plots, it can be seen that the argument $V_r(n_i)$ reduces rapidly with the increase of n_i when $n_i < \nu_i$ and it almost keeps the same value when $n_i \geqslant \nu_i$. Thus the smallest value of n_i in the almost flat part of the solid line plot can be taken as the estimate of the the i-th observability index. Consequently,

$\hat{\nu}_i$ and $\hat{n}_{di}$ can be determined from the plots $V(n_i)$ and $V_r(n_i)$.

In expression (14), L is the number of vectors $y_j(\cdot)$ and $u_j(\cdot)$ in equation (8) with the test observability index n_i, and n_r is the dimension of span$\{r_*(t-s)$, $s=1$, $\cdots$, n_r and $n_r \geqslant n_{r0}\}$.

4 Determination of a Parsimonious Parameter Structure

In some cases a parameter struture $\hat{\theta}_L$ in model (6), associated with the estimated index $\hat{\nu}_i$, may be extravagant since the values of some elements of $\hat{\theta}_L$, i. e. some coefficients in $\hat{A}_i(\cdot)$ and $\hat{B}_i(\cdot)$ are near or even equal to zero. This means that some column vectors in the data matrix H_ν with the index $\hat{\nu}_i$, in equation (8) have small or even no contribution to the reduction of the squared length of the residual error vector.

Denote with $\{h_1, \cdots, h_L\}$ all column vectors of H_ν and with $\{h_{s_1}, \cdots, h_{s_l}\}$ all vectors which have a significant contribution to the reduction of the squared sum of residual errors. A parameter structure, corresponding to $\{h_{s_1}, \cdots, h_{s_l}\}$, is called as a parsimonious parameter structure of model (6), and it will be denoted with $\hat{\theta}_{li}$.

A procedure to select $\{h_{s_1}, \cdots, h_{s_l}\}$, from H_ν is proposed as follows.

1) Let $k=1$, $M^{(0)}=I$, $y_r^{(0)}=y_i(t)$, $h_i^{(0)}=h_j$, $(j=1, \cdots, L)$.

Set an integer set $S=\{s_j=0 \mid j=1, \cdots, L\}$. Assign n_r to be a large enough value.

2) Compute

$$P_j^{(k)} = \frac{[h_j^{(k-1)}][h_j^{(k-1)}]^{\mathrm{T}}}{[h_j^{(k-1)}]^{\mathrm{T}}[h_j^{(k-1)}]}, \quad j = 1, \cdots, L \text{ and } j \neq s_1, \cdots, s_{k-j},$$

$$y_j^{(k)} = p_j^{(k)} y_j^{(k-1)},$$

$$c_j^{(k)} = [y_j^{(k)}]^{\mathrm{T}}[y_j^{(k)}],$$

where

$P_j^{(k)}$ is a projection operator associated with $h_j^{(k-1)}$;

$y_j^{(k)}$ is a projection of $y_j^{(k-1)}$ onto $h_j^{(k-1)}$;

$c_j^{(k)}$ is the potential contribution of $h_j^{(k-1)}$ to the reduction of the squared length of y_j^{k-1};

3) Select the direction asssociated with the maximum contribution as the optimum projection direction. Define

$$c_{s_k}^{(k)} = \max\{cx_j^{(k)}, j = 1, \cdots, L;\ j \neq s_1, \cdots, s_{k-1}\},$$

with setting its corresponding j into the unit s_k of the integer set S. Denote with $h_{s_k}^{(k)}$ and $P_{s_k}^{(k)}$ the vector and its projection operator associated with $c_{s_k}^{(k)}$.

4) Compute

$$M^{(k)} = M^{(k-1)} - P_{s_k}^{(k)},$$

$$y_r^{(k)} = M^{(k)} y_r^{(k-1)},$$

where

$M^{(k)}$ is a complement projection operator of span $\{h_{s_1}(1), \cdots, h_{s_k}(k)\}$; $y^{(k)}$ is a residual vector of $y_i(t)$.

5) Compute the whitening residual error. Construct

$$r(-k_r) = [0, 0, \cdots, 0, y_r^{(k)}(1), \cdots, y_r^{(k)}(N-k_r)]^T$$

nd

$$\Psi = [r(-1), r(-2), \cdots, r(-n_r)].$$

Compute

$$M_r = I - \Psi(\Psi^T\Psi)^{-1}\Psi^T,$$
$$e^{(k)} = M_r y_r^{(k)},$$
$$V^{(k)} = [e^{(k)}]^T[e^{(k)}],$$

where

M_r is a whitening projection operator with the order n_r; $e^{(k)}$ is a whitening residual error vector.

6) Compute

$$F(k) = \frac{V^{(k)} - V^{(L)}}{V^{(L)}} \cdot \frac{N-L-n_r}{L-k}.$$

If $F(k) < F_a(L-k, N-L-n_r)$, then $l=k$. The vector set $\{h_{s_1}, \cdots, h_{s_l}\}$ and its corresponding parsimonious parameter structure $\hat{\Theta}_{1i}$ can be determined. Otherwise, proceed.

7) Set

$$h_j^{(k)} = M^{(k)} h_j^{(k-1)}, \quad \text{for } j = 1, \cdots, L;\ j \neq s_1, \cdots, s_k.$$

Let $k=k+1$ and go to step 2.

About the above algorithm, the following aspects are noteworthy.

- After l stages of the above, the procedure $h_{s_1}^{(1)}, \cdots, h_{s_l}^{(l)}$ are orthogonal derections.

- Step 5 does the residual error whitening. It is the key to make the follwing relation hold

$$F(1) = \frac{V^{(l)} - V^{(L)}}{V^{(L)}} \cdot \frac{N-L-n_r}{L-1} \sim F_a(L-l, N-L-n_r)$$

under the hypotheses: $\overline{\Theta}_{1i}$ is empty. $\overline{\Theta}_{1i}$ is the complement set of $\overline{\Theta}_{1i}$ in $\overline{\Theta}_L$.

- In step 5, the order of a whitening filter is assigned a priorily. But, in fact, an alternative can be easily made up as follows.

Set a subloop;

Project $r(0)$ onto $r(-1)$, $r(-2)$, $\cdots$ one by one until the $v^{(k)}$ has no significant change.

5 Conclusion

A parsimonious structure identification scheme has been presented. Although this scheme is applicable for the ARARX, it can have its merits in the case in which output data are corrupted. Moreover, the proposed algorithm in section 4 seems suitable to be used independently to identify the parsimonious structure of a model set.

Refrences

[1] Bokor, J. and Keviczky, L.. Structure Properties and Structure Estimation of Vector Difference Equations. Int. J.

Control, 1982, 36:461—475

[2] Ljung, L. and Soderström, T.. Theory and Practice of Recursive Identification. MIT Press, 1983

[3] Guidorzi, R. P., Losito, M. P. and Muratori, T.. The Range Error Test in the Structural Identification of Linear Multivariable Systems. IEEE Trans. Automat. Contr., 1982, AC-27:1044—1053

[4] Sinha, N. K. and Kuszta, B.. Modeling and Identification of Dynamic Systems. Van Nostrand Reinhold Company, 1983

[5] Zhang Zhiyong. Whitening Residual Error Technique to Determine the Structure of VDE. Control Theory and Applications, 1988. 5(3):13—19 (in Chinese)

[6] Zhang Zhiyong. The Structure Identification of Multivariable Systems with Time-Lags. Acta Automatica Sinica, 1989, 15 (4):352—356 (in Chinese)

简练模型结构的辨识

张志涌

(福州大学自动化所,350002)

摘要: 本文采用投影技术辨识系统的标准向量差分方程的简练结构——可观性指数,噪声自回归阶数以及需估的最小模型参数位置.本文提出的方法可在有色噪声下同时把可观性指数与AR模型噪声的阶估计出来.此外,算法还可确定那些对改善模型质量贡献最大的参数的位置.

关键词: 辨识;模型结构;参数估计

本文作者简介

张志涌 1944年生.1967年毕业于清华大学自动控制系.1982年获上海交通大学电工计算机系现代控制专业硕士;1989至1990年进修于荷兰 Eindhoven 大学 Eykhoff 教授研究室.现在福州大学自动化所从事科研工作.主要研究兴趣为系统辨识和电子线路故障诊断的理论及应用.