KYBERNETIKA — VOLUME *19* (1983), NUMBER 2

ALGORITHMS FOR DETERMINING THE MODEL STRUCTURE OF A CONTROLLED SYSTEM

MIROSLAV KÁRNÝ

Determination of the model structure of a controlled system is the necessary first step for the design of majority of sophisticated controllers including self-tuners. The Bayesian theory of system classification proved to be suitable for this task. However, the set of possible hypothesis given by orders, delays, numbers of inputs, additional measured variables including external ones etc. is too large to be fully compared. To obtain a practical tool some procedure selecting the tested hypotheses is needed. Such a procedure as well as all substantial supporting algorithms for computer-aided structure determination of a multivariate regression model is described in the present paper.

1. INTRODUCTION

The structure as well as the parameter values of an optimal controller are fully determined by the performance criterion and the model of the controlled system. Self-tuning controllers, determining the parameter values on-line on the basis of the measured data, decrease the demands for a prior knowledge of the model substantially, Real-time parameter estimation moreover admits to track slow changes of a system, increasing in this way the flexibility of the model. However, the need for the prior knowledge of the model structure remains. This implies the necessity of a preliminary step of the model structure determination before an implementation of self-tuners.

The theory of the Bayesian system classification elaborated in [6], [7] has appeared to be a suitable tool to solve *this* problem (cf. [2]). The effectivity of this approach has been confirmed by many experiments with simulated as well as real data.

The algorithms described in the paper try to extend this approach in the two directions:

- to restrict the need for a detailed understanding of the used theory by non-specialized users
- to generate automatically new hypotheses which are worth of being tested.

The first point lies just in the line of the whole development of self-tuners, to minimize the number of tuned factors which are not directly connected with the observable closed loop behaviour. The second one tries, in this special case, to overcome the computational burden resulting from the dimensionality of the problem. The proposed algorithm concatenates the Bayesian classification theory presented in [6], [7] with the LDL^T-decomposition technique of [1], [5] and with a rather simple hill-climbing technique in order to generate different sets of hypotheses.

Special attention is paid to the numerical behaviour of algorithms (the problem is as a rule ill-conditioned) as well as to the computation speed.

2. PRELIMINARIES

Some results, which try to keep the paper self-contained, are given in this section. The notations are also introduced in this way.

2.1. Elimination of the unknown parameters in the fundamental model

Let the controlled system be described by a multivariate normal linear-in-parameters regression model, called fundamental model for brevity,

$$p(y_{(t)} \mid d^{(t-1)}, u_{(t)}, \Theta) = N_{y_{(t)}}(\hat{y}_{(t)}, \Omega^{-1}), \quad t \geqq t_0 \geqq 1 \tag{2.1}$$

where

$p(\alpha \mid \beta)$ – probability density function of α conditioned on β

$N_y(\hat{y}, \Omega^{-1}) = \left|\frac{\Omega}{2\pi}\right|^{1/2} \exp\left[-\frac{1}{2}(y - \hat{y})^T \Omega (y - \hat{y})\right]$

t – discrete time

$y_{(t)}$ – ν-vector of system outputs

$u_{(t)}$ – μ-vector of system inputs

$d_{(t)} = \begin{bmatrix} y_{(t)} \\ u_{(t)} \end{bmatrix}$ – λ-vector of measured data, $\lambda = \nu + \mu$

$d^{(t)} = (d_{(1)}, d_{(2)}, \ldots, d_{(t)})$.

The linearity-in-parameters means that

$$\hat{y}_{(t)} = P^T z_{(t)}$$

where ϱ-dimensional vector statistic $z(\cdot)$ maps

$$z(\cdot) : d^{(t-1)}, u_{(t)} \to z_{(t)}, \quad t \geqq t_0 \geqq 1. \tag{2.2}$$

The unknown parameters Θ of the model are P and Ω where Ω is the inversion of the covariance matrix in the normal model (2.1).

The structure of the model (2.1), fixing linearity and normality, is determined by the structure of the statistic $z(\cdot)$ in (2.2). The structure is assumed to be fixed in the present section. This fact will be denoted by conditioning on the hypothesis H_z.

The class of input generators satisfying the natural conditions of control (see [7])

$$p(u_{(t)} \mid d^{(t-1)}, \Theta, H_z) = p(u_{(t)} \mid d^{(t-1)}) \tag{2.3}$$

will be assumed throughout the paper. The probability density function of the observed data is then given by

$$p(d^{(t)} \mid H_z) = \prod_{\tau=t_0}^{t} [p(u_{(\tau)} \mid d^{(\tau-1)})] \cdot p(d^{(t_0-1)} \mid H_z) \cdot \int \prod_{\tau=\tau_0}^{t} [p(y_{(\tau)} \mid d^{(\tau-1)}, u_{(\tau)}, \Theta, H_z)]\, p(\Theta \mid d^{(t_0-1)}, H_z)\, \mathrm{d}\Theta \tag{2.4}$$

where the prior probability density of the unknown parameters $p(\Theta \mid d^{(t_0-1)}, H_z)$ has to be specified. Denoting

$$\mathscr{L}_{(t)} = \prod_{\tau=t_0}^{t} [p(u_{(\tau)} \mid d^{(\tau-1)})]\, p(d^{(t_0-1)} \mid H_z)$$

$$\mathscr{L}_{(t)}(\Theta) = \prod_{\tau=t_0}^{t} [p(y_{(\tau)} \mid d^{(\tau-1)}, u_{(\tau)}, \Theta, H_z)]\, p(\Theta \mid d^{(t_0-1)}, H_z) \tag{2.5}$$

then Eq. (2.4), for the model (2.1) and for the assumed conjugate probability density function $p(\Theta \mid d^{(t_0-1)}, H_z)$ [3], i.e. having the functional form of the product $\prod_{\tau=t_0}^{t} [\cdot]$ in $\mathscr{L}$ cf. Eq. (2.5), can be rewritten

$$p(d^{(t)} \mid H_z) = \mathscr{L}_{(t)} \frac{\iint \left|\frac{\Omega}{2\pi}\right|^{\vartheta_{(t)}/2} \exp\left\{-\tfrac{1}{2} \operatorname{tr}\left[\Omega \begin{bmatrix} P \\ -I \end{bmatrix}^{\mathrm{T}} V_{(t)} \begin{bmatrix} P \\ -I \end{bmatrix}\right]\right\} \mathrm{d}P\, \mathrm{d}\Omega}{\iint \left|\frac{\Omega}{2\pi}\right|^{\vartheta_{(t_0-1)}/2} \exp\left\{-\tfrac{1}{2} \operatorname{tr}\left[\Omega \begin{bmatrix} P \\ -I \end{bmatrix}^{\mathrm{T}} V_{(t_0-1)} \begin{bmatrix} P \\ -I \end{bmatrix}\right]\right\} \mathrm{d}P\, \mathrm{d}\Omega} \tag{2.6}$$

where

$$\vartheta_{(t)} = \vartheta_{(t-1)} + 1, \quad t \geqq t_0 \tag{2.7}$$

$$V_{(t)} = V_{(t-1)} + \begin{bmatrix} z_{(t)} \\ y_{(t)} \end{bmatrix} \begin{bmatrix} z_{(t)} \\ y_{(t)} \end{bmatrix}^{\mathrm{T}} = \begin{bmatrix} V_{z(t)} & V_{zy(t)} \\ V_{zy(t)}^{\mathrm{T}} & V_{y(t)} \end{bmatrix}, \quad t \geqq t_0 \tag{2.8}$$

The initial conditions of the recursions of the sufficient statistic (2.7), (2.8) are the parameters of $p(\Theta \mid d^{(t_0-1)}, H_z)$. The poor prior information is assumed which

can be modelled by

$$\vartheta_{(t_0-1)} = \varrho + v + 1 \tag{2.9}$$

$$V_{(t_0-1)} = \varepsilon I\,, \quad \varepsilon > 0\,, \quad \varepsilon \text{ small} \tag{2.10}$$

The integrals of the same type in the numerator and the denominator of Eq. (2.6) take the form [7]

$$\mathcal{I}_{(k)} = \left|V_{z(k)}\right|^{-v/2} \left|\Lambda_{(k)}\right|^{-(\vartheta_{(k)}-\varrho-v-1)/2} \mathcal{H}(v, \vartheta_{(k)} - \varrho)\,, \quad k = t_0 - 1\,, \quad t \tag{2.11}$$

where (compare (2.8))

$$\Lambda = V_y - V_{zy}^{\mathrm{T}} V_z^{-1} V_{zy} \tag{2.12}$$

and $\mathcal{H}$ is some function dependent only on t and v for the choice (2.9).
Using (2.9), (2.10), (2.11) and denoting

$$\varkappa_{(t)} = \vartheta_{(t)} - \varrho - v - 1 = \varkappa_{(t-1)} + 1\,, \quad \varkappa_{(t_0-1)} = 0 \tag{2.13}$$

and

$$\bar{\mathcal{I}}_{(t)} = \tilde{\mathcal{I}}_{(t)} \mathcal{H}(v, \varkappa_{(t)} + v + 1) / \mathcal{H}(v, v + 1) \tag{2.14}$$

the final form of Eq. (2.4) reads

$$p(d^{(t)} \mid H_z) = \left|\frac{V_{z(t)}}{\varepsilon}\right|^{-v/2} \left|\Lambda_{(t)}\right|^{-\varkappa_{(t)}/2} \bar{\mathcal{I}}_{(t)}\,. \tag{2.15}$$

2.2. Fast accumulation of the sufficient statistic

The solved problem requires to assume the statistic (2.2) of a high dimensionality. The computational burden connected with the recursion (2.8) even for a medium-amount data is then considerable.

The substantial reduction may be achieved when the statistics z in (2.2) has the "shifting" structure as it is typical of dynamical systems assumed. The algorithm given below is presented just for its ability to decrease the computations required. To simplify notations $t_0 = 1$ is set without loss of generality. Let

$$V_{(t)} = V_{(t-1)} + \bar{f}_{(t)} \bar{f}_{(t)}^{\mathrm{T}} = \begin{bmatrix} V_f & V_{f1} \\ V_{f1}^{\mathrm{T}} & \underbrace{V_1}_{1} \end{bmatrix} \begin{matrix} \} \lambda \\ \} 1 \end{matrix}\,, \quad V_{(0)} = \varepsilon I \tag{2.16}$$

where the vector $\bar{f}$ has the structure

$$\bar{f}_{(t)} = \begin{bmatrix} f_{(t)} \\ f_{(t-1)} \\ \vdots \\ f_{(t-n)} \\ 1 \end{bmatrix} \begin{matrix} \\ \\ \\ \\ \} 1 \end{matrix} \lambda \tag{2.17}$$

for some $n > 0$. The unit is the only non-shifted part of $\bar{f}_{(t)}$. Taking $(r, s) = (i\lambda + k, j\lambda + l)$ entry of $V_{(t)}$, with $0 \leqq j \leqq i \leqq n$, $1 \leqq k, l \leqq \lambda$, the following identities are easily derived

$$V_{(t)r,s} = V_{(0)r,s} + \sum_{\tau=1}^{t} \bar{f}_{(\tau)r}\bar{f}_{(\tau)s} = V_{(0)r,s} +$$

$$(\sum_{\tau=j+1}^{t+j} + \sum_{\tau=1}^{j} - \sum_{\tau=t+1}^{t+j}) f_{(\tau-i)k} f_{(\tau-j)l} = V_{(0)r,s} + \tag{2.18}$$

$$\sum_{\tau=1}^{t} f_{(\tau-(i-j))k} f_{(\tau)l} + \sum_{\tau=1}^{j} [f_{(0-(i-\tau))k} f_{(0-(j-\tau))l} - f_{(t-(i-\tau))k} f_{(t-(j-\tau))l}] =$$

$$V_{(0)r,s} + F_{(t)(i-j)\lambda+k,l} + \sum_{\tau=1}^{j} [\bar{f}_{(0)(i-\tau)\lambda+k} \bar{f}_{(0)(j-\tau)\lambda+l} -$$

$$- \bar{f}_{(t)(i-\tau)\lambda+k} \bar{f}_{(t)(j-\tau)\lambda+l}]$$

where

$$F_{(t)} = F_{(t-1)} + \begin{bmatrix} f_{(t)} \\ f_{(t-1)} \\ \vdots \\ f_{(t-n)} \end{bmatrix} f_{(t)}^{\mathrm{T}}, \quad F_{(0)} = 0. \tag{2.19}$$

The accumulation of F (2.19) instead of V_f requires (approximately) at any time t only $n\lambda^2$ multiplications instead of $\frac{1}{2}(n+1)^2\lambda^2$. The similar algebra shows that for $s < (n+1)\lambda + 1$

$$V_{(t)s,(n+1)\lambda+1} = \bar{f}_{(t)l} + \sum_{\tau=1}^{j} [\bar{f}_{(0)(j-\tau)\lambda+l} - \bar{f}_{(t)(j-\tau)\lambda+l}] \tag{2.20}$$

where

$$\bar{f}_{(t)} = \bar{f}_{(t-1)} + f_{(t)}, \quad \bar{f}_{(0)} = 0. \tag{2.21}$$

It remains to notice that the recursions of $\vartheta_{(t)}$ and $V_{1(t)}$ (2.16) coincide.

2.3. LDL^{T}-decomposition as a tool for $p(d^{(t)} \mid H_z)$ computation

The positive definite matrix (2.8) can be decomposed (cf. [1]) in

$$V = LDL^{\mathrm{T}} \tag{2.22}$$

where L is the unique lower triangular matrix with units on the diagonal and D is the positive diagonal matrix. The definition (2.22) directly implies that

$$|V_z| = \prod_{i=1}^{\varrho} D_i \tag{2.23}$$

and it can be shown that

$$|A| = \prod_{i=\varrho+1}^{\varrho+\nu} D_i \tag{2.24}$$

i.e. having the decomposition (2.22) the value of the probability density function $p(d^{(t)} \mid H_z)$ (2.15) can be easily computed. The algorithms which make possible to determine (2.22) for different models form the core of this paper. But they need to have (2.22) for some initial model. The required algorithm resulting directly from definition (2.22) has become the standard part of software packages but the following modification is needed for our purposes. Let us assume that the matrix V is stored in the strict upper triangular matrix, say V, and the diagonal part in the vector, say VG. The fixed structure of the vector statistic $\vec{f}_{(t)}$ (2.17) is assumed and (2.22) is determined for the permuted vector

$$\vec{f}'_i = \vec{f}_{q_i} \tag{2.25}$$

where the permutation vector q permutes the numbers $\{1, 2. \ldots, \lambda(n+1)+1\}$. The resulting decomposition of a submatrix of the type (ω, ω), storing L into the strict lower triangular part of V and D on the diagonal part of V, is computed as follows

$$\begin{aligned}
&i = 1, 2, \ldots, \omega \\
&\quad i' = q_i \\
&\quad A = VG_{i'} \\
&\quad k = 1, 2, \ldots, i-1 \\
&\qquad R_k = V_{ik} * V_{kk} \\
&\qquad A = A - V_{ik} * R_k \\
&\quad \textbf{end of } k \\
&\quad V_{ii} = A \\
&\quad j = i+1,\ i+2, \ldots, \omega \\
&\qquad j' = q_j \\
&\qquad \textbf{if } i' > j' \textbf{ then } B = V_{j'i'} \\
&\qquad\qquad \textbf{else } B = V_{i'j'} \\
&\qquad k = 1, 2, \ldots, i-1 \\
&\qquad\quad B = B - R_k * V_{jk} \\
&\qquad \textbf{end of } k \\
&\qquad\quad V_{ji} = B/A \\
&\quad \textbf{end of } j \\
&\textbf{end of } i
\end{aligned} \tag{2.26}$$

Notice that the ω-vector R is introduced to save multiplications. The do-loops, in the pseudo-programming language used, are assumed to be empty when the lower bound is greater than the upper one.

3. COMPARISON OF STRUCTURES OF THE MODELS

The finite amount of possible structures of models, given by z^i, $i = 1, 2, \ldots, N$, is assumed in this section. The prior probabilities of the hypotheses $H_i = H_{z^i}$, $i = 1, 2, \ldots, N$ are corrected by the observed data into the aposterior ones according to the Bayes rule

$$p(H_i \mid d^{(t)}) = \frac{p(d^{(t)} \mid H_i)\, p(H_i)}{\sum_{k=1}^{N} p(d^{(t)} \mid H_k)\, p(H_k)} \tag{3.1}$$

The following simplifying but reasonable assumptions are accepted:

(i) The initial time t_0 is the same for all hypotheses and the initial data $d^{(t_0-1)}$ are not used to correct the probabilities of hypotheses, i.e.

$$p(H_i \mid d^{(t_0-1)}) = p(H_i) \tag{3.2}$$

or equivalently

$$p(d^{(t_0-1)} \mid H_i) = p(d^{(t_0-1)}) \tag{3.3}$$

The relation (3.3) implies the functional independency of $\tilde{\mathscr{D}}_{(t)}$ on H_i in (2.5).

(ii) The initial uncertainty of unknown parameters within any structure is modelled by the conjugate prior distribution given by the parameters in (2.9), (2.10) having the same ε for all H's.

(iii) No reason exists to expect one hypothesis to be more probable than the others, i.e.

$$p(H_i) = \frac{1}{N}\,. \tag{3.4}$$

The assumptions (i), (ii), (iii) and the natural conditions of control (2.3) imply that Eq. (3.1) takes the form

$$p(H_i \mid d^{(t)}) = \frac{\gamma_{i(t)}}{\sum_{k=1}^{N} \gamma_{k(t)}} \tag{3.5}$$

where

$$\gamma_{i(t)} = \left|V_{z^i(t)}\right|^{-\nu/2} \left|\Lambda_{i(t)}\right|^{-\varkappa(t)/2} \varepsilon^{e\nu/2} \tag{3.6}$$

$$\Lambda_{i(t)} = V_y - V_{z^i y}^{\mathrm{T}} V_{z^i}^{-1} V_{z^i y} \tag{3.7}$$

4. SELECTION OF CLASSES OF THE COMPARED MODELS

Let us take a collection of all scalar functions of the past process history which are the candidates to be incorporated into the regressor (2.2). If the number of such

functions is $\bar{\varrho}$ then the number of possible hypotheses about the model structure is $2^{\bar{\varrho}}$, being too large to be taken as N in Section 3, i.e. too large to be completely compared.

The following strategy to produce automatically "reasonable" and sufficiently small sets of hypotheses will be adapted here instead. Let the "maximal" regressor $\bar{z}(\cdot)$ mentioned above be divided into two parts

$$z = \begin{bmatrix} z^a \\ z^r \end{bmatrix} \begin{matrix} \} \varrho \\ \} \bar{\varrho} \end{matrix} \tag{4.1}$$

where z^a (a stands for accepted) is a reasonable guess of the regressor structure and z^r (r-reject) is the complement of z^a to $\bar{z}$. The two sets of the tested hypotheses are formed:

Extension – the best system description lies in the set formed by z^a concatenated with just one entry from z^r.

Restriction – the best system description lies in the set formed by z^a with just one entry removed.

The extension and restriction, concatenated with the hypothesis that z^a appears to be best regressor, contain $\bar{\varrho} + 1$ hypotheses. This small number admits to chain the pairs of extensions and restrictions in such a way that the most probable partition (4.1) from the last step forms the initial partition for the new one. This process is stopped when either z^a does not change after the pair of extension and restriction or some number of steps is exhausted.

The above procedure can be viewed as some sort of a hill-climbing when the maximum likelihood hypothesis is searched. The full search is performed in the neighbourhood of the current guess z^a, the neighbourhood being defined through the restriction and extension.

It is known, and it has been confirmed experimentally, that at least for small and medium amount of data the likelihood function is a multimodal one. It follows that the search has to be repeated for different initial guesses z^a, e.g. $\varrho = 0$ and $\varrho = \bar{\varrho}$ are usually the reasonable starting points when the prior information is poor. However, a guess of an expert has to be preferred when it is at disposal.

Notice that the above way of forming of new hypotheses does not cover directly the problem of the selection between two disjoint groups, having more than one member, of the regressor entries. This problem occurs, e.g., when for a dynamic system one of two inputs has to be selected. However, two separate runs selecting the best guess in each group can be compared when comparing the final values of the likelihood functions.

5. SYSTEM DESCRIPTION FOR NESTED STRUCTURES

Let us assume two hypotheses about the model structure $H_{z^i} = H_i$, $i = 1, 2$ where

$$z^2 = \begin{bmatrix} z^1 \\ \cdot \end{bmatrix} \begin{matrix} \} \varrho \\ \} 1 \end{matrix} \tag{5.1}$$

then the submatrices of $V(2.8)$ fulfill

$$V_{z^2} = \underbrace{\begin{bmatrix} V_{z^1} & \vdots \\ \cdots & \cdot \end{bmatrix}}_{\varrho \quad 1} \begin{matrix} \} \varrho \\ \} 1 \end{matrix} \tag{5.2}$$

$$V_{z^2 y} = \underbrace{\begin{bmatrix} V_{z^1 y} \\ \cdots \end{bmatrix}}_{v} \begin{matrix} \} \varrho \\ \} 1 \end{matrix} . \tag{5.3}$$

Let the LDL^{T} decomposition of the matrix V corresponding to z^2 is at disposal, i.e.

$$V = \begin{bmatrix} V_z & V_{zy} \\ V_{zy}^{\mathrm{T}} & V_y \end{bmatrix} = \begin{bmatrix} L_z & 0 \\ L_{zy} & L_y \end{bmatrix} \begin{bmatrix} D_z & 0 \\ 0 & D_y \end{bmatrix} \underbrace{\begin{bmatrix} L_z^{\mathrm{T}} & L_{zy}^{\mathrm{T}} \\ 0 & L_y^{\mathrm{T}} \end{bmatrix}}_{\varrho + 1 \quad v} \begin{matrix} \} \varrho + 1 \\ \} v \end{matrix} \tag{5.4}$$

The identities (5.2), (5.3) imply that

$$L_{z^2} = \underbrace{\begin{bmatrix} L_{z^1} & 0 \\ \cdots & \cdot \end{bmatrix}}_{\varrho \quad 1} \begin{matrix} \} \varrho \\ \} 1 \end{matrix}, \quad D_{z^2} = \underbrace{\begin{bmatrix} D_{z^1} & 0 \\ 0 & \beta \end{bmatrix}}_{\varrho \quad 1} \begin{matrix} \} \varrho \\ \} 1 \end{matrix}, \quad L_{z^2 y} = \underbrace{\left[\begin{array}{c|c} L_{z^1 y} & \delta \end{array}\right]}_{\varrho \quad 1} \} v \tag{5.5}$$

and some algebra confirms that

$$\Lambda_2 = L_y D_y L_y^{\mathrm{T}} . \tag{5.6}$$

Moreover, see (5.5),

$$\Lambda_2 = \Lambda_1 - \beta \delta \delta^{\mathrm{T}} . \tag{5.7}$$

It means that to arrive at LDL^{T} decomposition of Λ_1 matrix, only a one-rank modification of Λ_2 decomposition is needed. The definition of the LDL^{T} decomposition can be used to derive the algorithm performing this task for Λ_i, $i = 1, 2$ in dependence on the sign of $\alpha = \pm\beta$. The modified decomposition is assumed to be stored in the lower triangular part of some matrix, say G, D_y is stored in its diagonal part. Notice that the algorithm below destroys the content of the vector δ.

(5.8)

$S = 1$
$j = 1, 2, \ldots, v$
$H = \delta_j$
$A = \alpha * H * S$
$B = G_{jj}$
$C = B + H * A$
$G_{jj} = C$
$S = S * B/C$
$A = A/C$
$i = j + 1, j + 2, \ldots, v$
$B = G_{ij}$
$\delta_i = \delta_i - B * H$
$G_{ij} = B + A * \delta_i$
end of i
end of j

The above algorithm and some details can be found elsewhere (see [1]).

To proceed further, let us assume the case when the LDT^{T} decomposition of the statistic V belonging to $\bar{z} = \begin{bmatrix} z^{\mathrm{a}} \\ z^{\mathrm{r}} \end{bmatrix}$ is available. The matrices L, D are divided accordingly

(5.9)
$$D = \begin{bmatrix} D_{z^{\mathrm{a}}} & 0 \\ & \tilde{D}_{z^{\mathrm{r}}} \\ 0 & & D_y \end{bmatrix} \begin{matrix} \} \varrho \\ \\ \} v \end{matrix} \Bigg\} \bar{\varrho}$$
$$\underbrace{\qquad\qquad}_{\bar{\varrho}} \quad \underbrace{\quad}_{v}$$

(5.10)
$$L = \begin{bmatrix} L_{z^{\mathrm{a}}} & 0 \\ L_{z^{\mathrm{a}}z^{\mathrm{r}}} & \tilde{L}_{z^{\mathrm{r}}} \\ L_{z^{\mathrm{a}}y} & \tilde{L}_{z^{\mathrm{r}}y} & L_y \end{bmatrix} \begin{matrix} \} \varrho \\ \\ \} v \end{matrix} \Bigg\} \bar{\varrho}$$
$$\underbrace{\qquad\qquad}_{\bar{\varrho}} \quad \underbrace{\quad}_{v}$$

The foregoing discussion confirmes that L_z and D_z really lie in the denoted subfields. Moreover, the LDT^{T} factorization of A is easily obtained by the successive application of the algorithm (5.8) to the $L_y D_y L_y^{\mathrm{T}}$ with columns of $\tilde{L}_{z^{\mathrm{r}}y}$ in the role of δ's and with diagonal entries of $\tilde{D}_{z^{\mathrm{r}}}$ in the role of α's (with plus sign).

The accepted strategy of searching for the best structure, described in Section 4,

requires to move some entry from z^a to z^r or in the opposite direction. The algorithm proposed for such a task in [5] will be briefly reviewed now.

Let the ith entry, $i \leqq \varrho$, of ϱ-vector z^a has to be moved to z^r. It can be shown that it is reasonable firstly permute $z_i \leftrightarrow z_{i+1}$ then $z_i \leftrightarrow z_{i+2} \ldots z_i \leftrightarrow z_\varrho$ and to decrease ϱ by unity. Each of these permutations corresponds to some permutation in V and destroys the LDT^T factorization. The reason to perform the permutation $z_i \leftrightarrow z_\varrho$ in the above-mentioned way lies in the fact that the algorithm, say $\mathscr{R}$, which reconstructs LDT^T decomposition is rather simple. The solution of this task, which produces description of one member of the restriction set, can be formalized

$$z_i \leftrightarrow z_{i+1}\mathscr{R} z_i \leftrightarrow z_{i+2}\mathscr{R} \ldots z_i \leftrightarrow z_\varrho \mathscr{R}, \text{ set } \varrho = \varrho - 1 \tag{5.11}$$

Similar is the case of forming of the extension set, $i > \varrho$,

$$z_i \leftrightarrow z_{i-1}\mathscr{R} z_i \leftrightarrow z_{i-2}\mathscr{R} \ldots z_i \leftrightarrow z_{\varrho+1}\mathscr{R}, \quad \text{set} \quad \varrho = \varrho + 1 \tag{5.12}$$

The algorithm $z_i \leftrightarrow z_{i+1}\mathscr{R}$, resulting directly from the definition of LDT^T factorization and from the form of the matrices L, D after permutation $z_i \leftrightarrow z_{i+1}$, is

(5.13)

```
k = 1, 2, ..., i - 1
    H = V_ik
    V_ik = V_i+1k
    V_i+1k = H
end of k
H = V_ii
S = V_i+1i+1
S1 = V_i+1i
H1 = S + H * S1 ** 2
C = S1 * H/H1
V_i+1i = C
V_ii = H1
H1 = S/H1
V_i+1i+1 = H1 * H
k = i + 2, i + 3, ..., ϱ̄ + ν
    H = V_ki
    S = V_ki+1
    V_ki = C * H + S * H1
    V_ki+1 = H - S * S1
end of k
```

This algorithm saves the substantial amount of computation time decreasing the number of operations from the order $\bar{\varrho}^3$ for the plain repetition of the LDT^T factorization to the order $\bar{\varrho}^2$.

6. SUMMARY OF THE ALGORITHMS AND USER-ORIENTED COMMUNICATION

The purpose of this section is to give overall view on the computer program and its use.

At the very beginning the user has to select the system output as well as the entries of the "maximal" regressor $\bar{z}$ which contains the union of all $\bar{z}$'s and to generate it in the shifting form (2.17) from the recorded data.

This part of data handling suits for subtraction of mean values from measured data. It restricts the possibility of numerical troubles by decreasing the numerical values of the data used.

Then the statistic V in (2.16) is formed (for full data record length available) with the help of the statistic (2.19), preferably in the separate task because different searches for the same V are usual. To construct V only ε (2.16) has to be specified.

A set of possible alternatives $\bar{z}$ and an initial guess z^a has to be selected at the next step. The following style of user-oriented communication seems to us to be satisfactory. The character string

$$(6.1)\quad [+\,+\,f_{1(t)} + +\,f_{2(t)} \cdots + +\,f_{v(t)}\,\&\,\&\,f_{v+1(t)} \cdots \&\,\&\,f_{\lambda(t)},\,\&\,\&\,f_{1(t-1)} \cdot \ldots \&\,\&\,f_{\lambda(t-1)},\,\ldots\,\&\,\&\,f_{1(t-\bar{n})} \cdots \&\,\&\,f_{\lambda(t-\bar{n})},\,\&\,\&\,\text{const.}]$$

is formed. The starting v-entries correspond to the system output and the last one to the constant term. The prefixes & & have to be filled by symbols $+\,+$, $+\,-$, $-\,+$, $-\,-$, having the meaning:

$+\,+$: this entry has to be used in the accepted model,
$+\,-$: this entry can be used in the accepted model and it is an entry of the initial guess,
$-\,+$: this entry can be used in the accepted model but it is not an entry of the initial guess,
$-\,-$: this entry must not be used in the accepted model.

These symbols are translated in the initial value of the permutation vector q (2.25) and in the dimensions ϱ of z^a and $\bar{\varrho}$ or $\bar{z}$. It has to produce the following re-ordering

$$(6.2)\qquad (\bar{z}, y) = \left\{ \bar{z} = \left\{ \begin{array}{|c|l|} \hline z^a \ldots & +\,+\ \text{part} \\ \cline{2-2} & +\,-\ \text{part} \\ \hline z^r & -\,+\ \text{part} \\ \hline y & +\,+\ \text{part} \\ \hline . & -\,-\ \text{part} \\ \hline \end{array} \right. \right.$$

The next program steps are the computation of the initial permuted LDT^{T} decomposition (2.26) with $\omega = \bar{\varrho} + \nu$ and the determination of LDL^{T} factorisation of Λ corresponding to z^{a}, i.e. successive application of the algorithm (5.8) for all columns of $\tilde{L}_{z^{r}y}$ (5.10). Then the principal part of the likelihood function γ_i (3.5) can be determined from the relation (2.23) and its analogy for the matrix Λ. The log-form of γ is more reasonable because of its exponential growth with length of data record:

$$\tilde{\gamma}^{a} = 2 \ln (\gamma^{a}_{(t)}) = -\nu \ln |V_{z^{a}(t)}| - \varkappa_{(t)} \ln |\Lambda_{a(t)}| + \varrho\nu \ln (\varepsilon) = \tag{6.3}$$

$$= -\sum_{i=1}^{e} \ln (V_{ii}) - \varkappa_{(t)} \sum_{i=1}^{\nu} \ln (G_{ii}) + \varrho\nu \ln (\varepsilon)$$

Having $\tilde{\gamma}^{a}$ the extension, see Section 4, is tried for all entries of z^{r}. The permutation vector q is of course permuted accordingly. If it is found some entry causing (maximal) growth the extension is taken as new z^{a}. Then the restriction is handled in the same way.

The computations are stopped if either no growth of $\tilde{\gamma}^{a}$ is accounted through the trial extension and restriction or a given number of steps is exhausted.

Partial as well as final results are displayed similarly as the input information. The character string of the type (6.1) corresponding to the current division of $\bar{z}$ in (6.2) is used. The entries of z^{a} are denoted by $+ +$ and of z^{r} by $- -$. The value of $\tilde{\gamma}^{a}$ is of interest when two runs differing in initial guess are compared.

7. ILLUSTRATIVE EXAMPLE

The purpose of this section is to demonstrate the behaviour of the program. The selected example illustrates that seemingly simple problems of structure determination fall in the class having too much members to be fully compared.

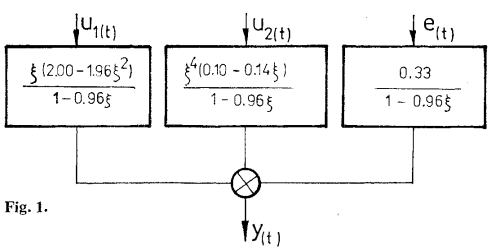

Fig. 1.

Let us assume single-input ($u_{2(t)}$) single-output ($y_{(t)}$) system with one external measurable variable ($u_{1(t)}$). The scheme of it is given in Fig. 1 (ξ – delay operator, $\xi(\cdot)_{(t)} = (\cdot)_{(t-1)}$) where $u_{i(t)}$, $e_{(t)}$ are mutually independent white Gaussian noises $N(0, 1)$. Notice that the common denominator and the properties of $e_{(t)}$ imply that the system in Figure 1 is the fundamental one.

The unknown structure parameters and their bounds are:

μ – number of inputs	$0 \leqq \mu \leqq 2$
δ_i, $i = 1, 2$ – delays	$0 \leqq \delta_i$
n_i, $i = 1, 2$ – orders of denominators	$0 \leqq \delta_i + n_i \leqq 7$
n – order of numerators	$0 \leqq n \leqq 7$

The maximal regressor concatenated with output and with the unit corresponding to the constant term contains subvectors

$$f_{(t-i)} = \begin{bmatrix} y_{(t-i)} \\ u_{1(t-i)} \\ u_{2(t-i)} \end{bmatrix}, \quad i = 0, 1, \ldots, 7, \quad \text{i.e.} \quad \lambda = 3, \quad \bar{\varrho} = 25.$$

It implies $2^{25} \approx 3 \cdot 10^7$ possible hypotheses.

Two runs for initial $\varrho = 0$ and $\varrho = \bar{\varrho}$, with 500 items of $f_{(t)}$ were performed. The structure was tested after each 125 steps starting with V (see equation (2.10)) given by $\varepsilon = 0.01$. The resulting partition of $\bar{z}$ formed the initial condition for continuation. The CPU time (IBM 370/135, PL-1 language, interactive mode), including two complete simulations and accumulations of the statistic V, was 4 min. 32 seconds.

The right structure was found at time 125 for the initial guess given by $\varrho = \bar{\varrho}$ and at time 375 for the initial $\varrho = 0$. The first case needed maximally 20 pairs of exsions and restrictions, the second one only six.

The different results at times 125 and 250 respectively were caused by multimodal character of likelihood function $\tilde{\gamma}$. The wrong guess implies values -18.03 and 17.90 in comparison with 177.68 and 415.41 for the right one. These values are, of course, sufficient for the selection of the better structure.

8. CONCLUSIONS

The proposed computer program for the Bayesian testing of the regressor structure in multivariate normal linear-in-parameters dynamic regression model has, according to our opinion, a rather wide range of applications in practice. It admits to determine not only delays, orders, significant inputs and/or external variables, but it might be used to confirm or reject whether some nonlinear function of the observed data is to be used in the regressor. The possibility to use the Bayesian theory to the given problem in practically met dimensions has been achieved by the algorithm generating chain of small sets of sufficiently probable hypotheses. It results in the several-minute runs in IBM 370/135 CPU time when the tested regressor has several tens of entries. The effectiveness of the way used was confirmed in simulated as well as real-data test cases.

Let us remark one important still explicitly unmentioned fact: the separation between the structure determination and the final control problem was enforced,

the aposteriori probabilities of hypotheses are not weighted by control-aimed loss function. This fact has to be taken into consideration when competitive hypotheses are of almost the same high probability.

The generally known observations about the behaviour of aposteriori probabilities are applicable to the solved problem in order to obtain some asymptotic results:

- if the right model structure lies in the class of the tested hypotheses, and
- if the observed data are sufficiently rich (in the terms of equivalency of some σ-algebras) or equivalently inputs and external variables are sufficiently exciting

then the aposteriori probabilities asymptotically converge to 1 for the right structure.

These results stress two fundamental facts:

- modeling, it means the selection of $\bar{z}(\cdot)$, is of the first class importance
- sufficiently "informative" data have to be collected.

At the end the two possible extensions are mentioned:

- overall algorithm can be done more sophisticated by exploiting the nesting property in the way similar to [4]
- the same type of algorithm can be used to determine sampling periods for simple models as an integer fraction of control period. It can be shown that for restricted model order the testing of models of the type

$$H_1 : y_{(t)} = \alpha_1^1 y_{(t-1)} + \alpha_2^1 y_{(t-2)} + \ldots$$

$$H_2 : y_{(t)} = \alpha_1^2 y_{(t-2)} + \alpha_2^2 y_{(t-4)} + \ldots$$

has a nontrivial solution.

(Received May 26, 1982.)

REFERENCES

[1] G. J. Bierman: Factorization Methods for Discrete Sequential Estimation. Academic Press, New York 1977.

[2] J. Fessl and M. Kárný: Choice of models for self-tuning regulators. Preprints 5th IFAC Symposium on Identification and Parameter Estimation, Darmstadt 1979, 1289—1186.

[3] M. H. DeGroot: Optimal Statistical Decisions. McGraw-Hill, New York 1970.

[4] M. Kárný: Bayesian estimation of model order. Problems of Control and Information Theory *9* (1980), 1, 33—46.

[5] M. Kárný: Recursive parameter estimation of regression model when the interval of possible values is given. Kybernetika *18* (1982), 1, 37—49.

[6] V. Peterka and M. Kárný: Bayesian system classification. Preprints 5th IFAC Symposium on Identification and Parameter Estimation, Darmstadt 1979, 349—356.

[7] V. Peterka: Bayesian approach to system identification. In: Trends and Progress in System Identification (P. Eykhoff, ed.), IFAC Series for Graduates, Research Workers and Practicing Engineers *1* (1981).

Ing. Miroslav Kárný, CSc., Ústav teorie informace a automatizace ČSAV (Institute of Information Theory and Automation — Czechoslovak Academy of Sciences), Pod vodárenskou věží 4, 182 08 Praha 8. Czechoslovakia.