INT. J. CONTROL, 1984, VOL. 40, NO. 5, 971–987

P/9

Structural identification of linear multivariable systems using overlapping forms: a new parametrization

MICHEL GEVERS† and AH CHUNG TSOI‡

The choice of a parametrization for the representation of a linear multivariable control system amounts to the selection of a basis of the rows of the Hankel matrix of Markov elements. The so-called 'overlapping' or 'pseudo-canonical' forms are traditionally obtained by imposing two selection rules: a block selection rule and a chain selection rule. In this paper, these constraints are relaxed to requiring only a chain selection rule. This allows for more flexibility in selecting numerically well-conditioned parametrizations.

1. Introduction

A problem which has been the subject of many studies in linear multivariable systems theory is the determination of the structure of a state space or an autoregressive moving average (ARMA) model for which the parameters of the model are uniquely identifiable. The first approach is to use canonical forms (state space or ARMA) (Luenberger 1967, Dehnam 1974, Rissanen 1974, Guidorzi 1981). However, it was shown by Van Overbeek and Ljung (1982) that this parametrization may lead to an ill-posed computational problem. An alternative approach, namely the overlapping parametrization (also called pseudo-canonical form) has been proposed recently. This concept was first suggested by Glover and Willems (1974), and studied by Ljung and Rissanen (1976), Van Overbeek and Ljung (1982), Picci (1980), Rissanen (1981), Deistler and Hannan (1981), Wertz *et al.* (1982), Gevers and Wertz (1982), Guidorzi and Beghelli (1982), and Corrêa and Glover (1982). It has been shown that the set of all finite-dimensional linear systems can be represented by a finite number of uniquely identifiable parametrizations. Each parametrization can be represented by a set of integers known as structure indices. Each system may be represented by more than one such parametrization, and any two parametrizations describing the same system are related.

The determination of a possible set of structure indices is obtained by selecting, in a suitable way, a basis for the rows of the Hankel matrix of impulse responses (or Markov parameters). In order to obtain a representation that contains a small number of uniquely identifiable parameters, certain rules must be imposed for the selection of this basis. All overlapping forms described so far have involved the following two selection rules (Rissanen and Ljung 1975, Guidorzi and Beghelli 1982, Corrêa and Glover 1982, Wertz, Gevers and Hannan 1982, Van Overbeek and Ljung 1982, Gevers and Wertz 1982):

(i) a block selection rule, i.e. if the dimension of the output vector is p, then an entire block of p rows is chosen;

Received 13 March 1984.

† Department of Systems Engineering, Research School of Physical Sciences, Australian National University, Canberra, A.C.T. 2600, Australia.

‡ Department of Electrical and Electronic Engineering, University of Auckland, Auckland, New Zealand.

(ii) a chain selection rule, i.e. if a vector is in the basis, then its corresponding predecessor vector is also in the basis (the predecessor of a vector in the Hankel matrix is the one that is located p rows above that vector).

For canonical (as opposed to overlapping) forms, a third selection rule was added that made the selected basis unique (Denham 1974, Rissanen 1974, Guidorzi 1981). Both Van Overbeek and Ljung (1982) and Wertz *et al.* (1982) proposed that the first p rows (namely corresponding to the first block) be chosen. Unfortunately, it was shown by Gevers and Wertz (1982, 1984) that this selection rule leads to a corresponding matrix fraction description (MFD) form which may have a singular leading coefficient matrix. Instead, they proposed that the qth block row be chosen, where $q = \max(n_i) - p + 1$ and n_i, $i = 1, \ldots, p$ are the proposed structure indices. They showed that, with this selection rule, the corresponding MFD form has an identity leading coefficient matrix.

It is logical to ask the question : Is it possible to relax the selection rule that an entire block row needs to be chosen ? If so, what are the properties of the corresponding parametrization ?

In this paper, we will show that it is possible to relax selection rule (i) above. The resulting parametrization gives rise to both a minimal state-space representation and a non-minimal state-space representation. The relationship between these two representations will be obtained. It will be shown that the non-minimal representation is completely observable, but may not be completely controllable. In addition, it will be shown how the corresponding MFD form can be determined.

2. Multivariable system parametrizations

We consider a p-dimensional stationary full-rank zero-mean stochastic process $\{y_t\}$ with rational spectrum. The linear least-squares predictor of $\{y_t\}$, given the past history of the process, is of full rank. Then, $\{y_t\}$ may be described up to second-order statistics by a state-space representation

$$\left.\begin{aligned} x_{t+1} &= Fx_t + Ke_t \\ y_t &= Hx_t + e_t \end{aligned}\right\} \tag{2.1}$$

where x_t is an n-dimensional state vector, y_t is a p-dimensional observation vector, e_t is a p-dimensional white noise sequence with covariance matrix Q and F, K, H are constant matrices of appropriate dimensions. F is assumed to be stable. Alternatively, the process $\{y_t\}$ may be represented by an input–output representation

$$A(z)y_t = B(z)e_t \tag{2.2}$$

where $zy_t \triangleq y_{t+1}$, $A(z)$ and $B(z)$ are $p \times p$ matrix polynomials in z

$$A(z) = A_0 + A_1 z^{-1} + \ldots + A_{\bar{n}} z^{-\bar{n}}$$

$$B(z) = B_0 + B_1 z^{-1} + \ldots + B_{\bar{n}} z^{-\bar{n}}$$

where $\bar{n}$ is the maximum lag in the process. It is assumed that

$$\det A(z) \neq 0 \quad \text{for } |z| \geqslant 1$$

and

$$\lim_{z\to\infty} A^{-1}(z)B(z) = I$$

Remark 2.1

In this paper, in contradistinction to Gevers and Wertz (1984), we do not assume n in (2.1) to be minimum ; equivalently, we do not assume $A(z)$ and $B(z)$ to be left coprime.

From (2.1) it is simple to show that

$$y_t = \sum_{i=0}^{\infty} R_i e_{t-i} \tag{2.3}$$

where the R_i are $p \times p$ matrices, known either as the Markov parameters, or the impulse response matrices. Furthermore

$$R_i = HF^{i-1}K, \quad i = 1, 2, \ldots, R_0 = I$$

Note that the Markov parameters are determined by

$$R_j = E[y_t y_{t-j}^{\mathrm{T}}] \tag{2.4}$$

By demanding that the causal inverse of y_t exists, i.e.

$$e_t = \sum_{i=0}^{\infty} N_i y_{t-i}$$

where $N_0 = I_p$, and

$$N(z) = \sum_{i=0}^{\infty} N_i z^{-i}$$

has no poles outside the unit circle, it is possible to find that

$$e_t \triangleq y_t - \hat{y}_{t/t-1}$$

where $\hat{y}_{t/t-k}$ is the linear least squares k-step ahead predictor of y_t. Note that

$$\hat{y}_{t+j/t-1} = \sum_{i=j+1}^{\infty} R_i e_{t+j-i}, \qquad j = 0, 1, 2, \ldots$$

Similarly

$$\hat{y}_{t+j+1/t} = \sum_{i=j+1}^{\infty} R_i e_{t+j+1-i}, \qquad j = 0, 1, 2, \ldots$$

Let

$$\hat{Y}_N(t) \triangleq \begin{bmatrix} \hat{y}_{t/t-1} \\ \hat{y}_{t+1/t-1} \\ \vdots \\ \hat{y}_{t+N-1/t-1} \end{bmatrix} = \begin{bmatrix} R_1 & R_2 & R_3 & \cdots \\ R_2 & R_3 & R_4 & \cdots \\ \vdots & \vdots & \vdots & \\ R_N & R_{N+1} & \vdots & \cdots \end{bmatrix} \begin{bmatrix} e_{t-1} \\ e_{t-2} \\ \vdots \\ \vdots \end{bmatrix} = H_{N,\infty} \begin{bmatrix} e_{t-1} \\ e_{t-2} \\ \vdots \\ \vdots \end{bmatrix} \tag{2.5}$$

Hence

$$\hat{Y}_N(t+1) = \begin{bmatrix} \hat{y}_{t+1/t} \\ \hat{y}_{t+2/t} \\ \vdots \\ \vdots \\ y_{t+N/t} \end{bmatrix}$$

$$= \begin{bmatrix} R_1 & R_2 & R_3 & \cdots \\ R_2 & R_3 & R_4 & \cdots \\ R_3 & R_4 & R_5 & \cdots \\ \vdots & \vdots & \vdots & \\ R_N & R_{N+1} & \vdots & \cdots \end{bmatrix} \begin{bmatrix} e_t \\ e_{t-1} \\ e_{t-2} \\ \vdots \\ \vdots \end{bmatrix}$$

$$= \begin{bmatrix} R_2 & R_3 & R_4 & \cdots \\ R_3 & R_4 & R_5 & \cdots \\ R_4 & R_5 & R_6 & \cdots \\ \vdots & \vdots & \vdots & \\ R_{N+1} & R_{N+2} & \vdots & \cdots \end{bmatrix} \begin{bmatrix} e_{t-1} \\ e_{t-2} \\ e_{t-3} \\ \vdots \\ \vdots \end{bmatrix} + \begin{bmatrix} R_1 \\ R_2 \\ R_3 \\ \vdots \\ R_N \end{bmatrix} e_t \qquad (2.6)$$

Comparing (2.6) with (2.1) and (2.5), it is obvious that the state vector x_t forms a basis of $\hat{Y}_N(t)$.

Let the set of selected basis components be denoted by $I = \{i_1, i_2, \ldots, i_n\}$. Also, let the $[j+(k-1)p]$th component of $\hat{Y}_N(t)$ be denoted by $\hat{y}_{jk}(t-1)$. Let R^j denote the jth block row of $H_{N,\infty}$, and r_{ij} the ith row of R^j, i.e.

$$R^j = [R_j \;\; R_{j+1} \ldots] = \begin{bmatrix} r_{1j} \\ r_{2j} \\ \vdots \\ r_{pj} \end{bmatrix}$$

Then, $r_{ij} = [i+(j-1)p]$th row of $H_{N,\infty}$.

Let $r_{ij}(k)$ = row p-vector made up of the kth set of p elements of row r_{ij}. Thus

$$H_{N,\infty} = \begin{bmatrix} r_{11} \\ r_{21} \\ \vdots \\ r_{p1} \\ \cdots\cdots \\ r_{12} \\ \vdots \\ r_{p2} \\ \cdots\cdots \\ \vdots \end{bmatrix}$$

Let x_t be the basis.

Definition 2.1

For given i, $i = 1, 2, \ldots, p$, denote

(*a*) $$S_i \triangleq \{\hat{y}_{ik}(t-1) \mid \hat{y}_{ik}(t-1) \in x_t \quad \text{for some } k \geqslant 1\}$$

(*b*) Let $n_i \triangleq$ number of elements in S_i. Then $n_1, n_2, \ldots, n_p$ will be called the structure indices. In addition,

$$n = \sum_{i=1}^{p} n_i$$

is the dimension of x_t.

(*c*) $$m_i \triangleq \min_{m=1,2,\ldots} \{m \mid \hat{y}_{im}(t-1) \in x_t\}, \quad i = 1, 2, \ldots, p$$

(*d*) $$s \triangleq \max_{i=1,2,\ldots,p} \{m_i + n_i - 1\}$$

Example

Suppose $p = 2$, $n = 5$, $I = \{1, 4, 5, 6, 8\}$, then

$$S_1 = \{\hat{y}_{11}(t-1), \hat{y}_{13}(t-1)\}$$

$$S_2 = \{\hat{y}_{22}(t-1), \hat{y}_{23}(t-1), \hat{y}_{24}(t-1)\}$$

and

$$n_1 = 2, \quad n_2 = 3, \quad m_1 = 1, \quad m_2 = 2, \quad s = 4$$

We wish to impose the following selection rules :

Rule 1

$S_i \neq \varnothing$, or equivalently, $n_i \geqslant 1$, $i = 1, 2, \ldots, p$, i.e. every one of the p components of $y(t)$ appears at least once.

Rule 2

For $i = 1, 2, \ldots, p$

$$\{\hat{y}_{i,m_i}(t-1), \hat{y}_{i,m_i+1}(t-1), \ldots, \hat{y}_{i,m_i+n_i-1}(t-1)\} \in x_t$$

i.e. the n_i components of x_t whose first index is i appear in n_i successive blocks. This is the chain rule.

Rule 3

$m_i = 1$, for at least one i in $\{1, 2, \ldots, p\}$.

Remark 2.2

This is a relaxation of the selection rules imposed by Van Overbeek and Ljung (1982), Wertz *et al.* (1982) and Gevers and Wertz (1984) in that we do not require one full block $\hat{y}_{jk}(t-1)$, $j = 1, 2, \ldots, p$ in the basis x_t. In Van Overbeek and Ljung (1982) and Wertz *et al.* (1982) $k = 1$, and in Gevers and Wertz (1984) $k = \max_i \{n_i\}$.

Remark 2.3

Rules 1–3 cannot be relaxed further. If they were not imposed, then the resultant parametrization would be over-parametrized. Equivalently, the matrix F might become full.

Example

With $p=2$, $n=5$, the following are a few candidate selections $\{1, 2, 3, 4, 5\}$, $\{1, 3, 4, 6, 8\}$, $\{1, 3, 6, 8, 10\}$, $\{1, 4, 6, 8, 10\}$, $\{1, 6, 8, 10, 12\}$, $\{11, 13, 2, 4, 6\}$.

We will now define two state vectors which will lead to two different state-space representations, the first one minimal, the second one not.

3. State-space representations

Representation 1

$$x_t = [\hat{y}_{1,m_1}(t-1), \hat{y}_{1,m_1+1}(t-1), \ldots, \hat{y}_{1,m_1+n_1-1}(t-1) \vdots \ldots \vdots \hat{y}_{p,m_p}(t-1), \ldots, \hat{y}_{p,m_p+n_p-1}(t-1)]^{\mathrm{T}}$$

The dimension of x_t is

$$n = \sum_{i=1}^{p} n_i$$

The state-space representation is characterized by the matrices H, F, K as shown in (2.1). They are obtained by expressing

$$\hat{y}_{1,m_1+n_1}(t-1), \hat{y}_{2,m_2+n_2}(t-1), \ldots, \hat{y}_{p,m_p+n_p}(t-1)$$

as a function of x_t, or equivalently, by expressing

$$r_{1,m_1+n_1}, r_{2,m_2+n_2}, \ldots, r_{p,m_p+n_p}$$

as a function of the corresponding basis rows of $H_{N,\infty}$. The corresponding rows of $H_{N,\infty}$ may be represented as

$$R = \begin{bmatrix} r_{1,m_1} \\ \vdots \\ r_{1,m_1+n_1-1} \\ \vdots \\ r_{p,m_p} \\ \vdots \\ r_{p,m_p+n_p-1} \end{bmatrix}$$

Thus, for $i=1, 2, \ldots, p$, there exists unique γ_{ikl} such that

$$r_{i,m_i+n_i} = \sum_{k=1}^{p} \sum_{l=m_k}^{m_k+n_k-1} \gamma_{ikl} r_{kl} \tag{3.1}$$

By the Hankel structure, it follows that

$$r_{i,m_i+n_i+q} = \sum_{k=1}^{p} \sum_{l=m_k}^{m_k+n_k-1} \gamma_{ikl} r_{k,l+q}, \quad q \geq 0 \tag{3.2}$$

In addition

$$\hat{y}_{i,m_i+n_i}(t-1) = \sum_{k=1}^{p} \sum_{l=m_k}^{m_k+n_k-1} \gamma_{ikl} \hat{y}_{kl}(t-1) \tag{3.3}$$

Hence, we have

$$\left.\begin{aligned} x_{t+1} &= Fx_t + Ke_t \\ y_t &= Hx_t + e_t \end{aligned}\right\} \tag{3.4}$$

where

$$F = \left[\begin{array}{c|c|c|c}
\begin{matrix}0\\ \vdots \\ 0\end{matrix} \quad I_{n_1-1} & \bigcirc & \cdots & \bigcirc \\
\gamma_{11m_1} \cdots \gamma_{11m_1+n_1-1} & \gamma_{12m_2} \cdots \gamma_{12m_2+n_2-1} & & \gamma_{1pm_p} \cdots \gamma_{1pm_p+n_p-1} \\ \hline
\bigcirc & \begin{matrix}0\\ \vdots \\ 0\end{matrix} \quad I_{n_2-1} & \cdots & \bigcirc \\
\gamma_{21m_1} \cdots \gamma_{21m_1+n_1-1} & \gamma_{22m_2} \cdots \gamma_{22m_2+n_2-1} & & \gamma_{2pm_p} \cdots \gamma_{2pm_p+n_p-1} \\ \hline
\vdots & & & \\ \hline
\bigcirc & \bigcirc & \cdots & \begin{matrix}0\\ \vdots \\ 0\end{matrix} \quad I_{n_p-1} \\
\gamma_{p1m_1} \cdots \gamma_{p1m_1+n_1-1} & \gamma_{p2m_2} \cdots \gamma_{p2m_2+n_2-1} & & \gamma_{ppm_p} \cdots \gamma_{ppm_p+n_p-1}
\end{array}\right]$$

is an $n \times n$ matrix, and

$$H = \begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_p \end{bmatrix}$$

is a $p \times n$ matrix, and where

$$h_i = [\underbrace{0 \ldots 0}_{n_1} \vdots \ldots \vdots \underbrace{1\ 0 \ldots 0}_{n_i} \vdots \ldots \vdots \underbrace{0 \ldots 0}_{n_p}] \quad \text{if } m_i = 1$$

and

$$h_i = [\beta_{i11m_1} \cdots \beta_{i11m_1+n_1-1} \vdots \beta_{i12m_2} \cdots \beta_{i12m_2+n_2-1} \vdots \cdots \vdots \beta_{i1pm_p} \cdots \beta_{i1pm+n_p-1}] \quad \text{if } m_i > 1$$

The coefficients β_{ijkl} are obtained from the following $ps-n$ relationships for $i=1, 2, \ldots, p$, $j=1, \ldots, m_i-1$, and $j=m_i+n_i, \ldots, s$

$$r_{ij} = \sum_{k=1}^{p} \sum_{=m_k}^{m_k+n_k-1} \beta_{ijkl} r_{kl} \tag{3.5}$$

$$K = \begin{bmatrix} r_{1m_1}(1) \\ \vdots \\ r_{1m_1+n_1-1}(1) \\ \vdots \\ r_{pm_p}(1) \\ \vdots \\ r_{pm_p+n_p-1}(1) \end{bmatrix}$$

is an $n \times p$ matrix.

Remark 3.1

The total number of ' free ' parameters for this representation in F is pn, in K is pn, and in H is vn, where v = number of m_i that are >1 $(0 \leqslant v \leqslant p-1)$. Hence, the number of parameters involved lies between $2np$ and $3np-n$.

Remark 3.2

The possibility of a non-canonical H makes the conversion of state-space representation to the input–output representation rather difficult.

Representation 2

The state vector is defined as

$$\bar{x}_t = [\hat{y}_{11}(t-1), \hat{y}_{12}(t-1), \ldots, \hat{y}_{1s}(t-1) \vdots \ldots \vdots \hat{y}_{p1}(t-1), \hat{y}_{p2}(t-1), \ldots, \hat{y}_{ps}(t-1)]^{\mathrm{T}}$$

The dimension of $\bar{x}_t$ is ps. The corresponding state-space representation is characterized by the matrices $\bar{H}$, $\bar{F}$, $\bar{K}$. These are obtained by expressing $\hat{y}_{1,s+1}(t-1), \hat{y}_{2,s+1}(t-1), \ldots, \hat{y}_{p,s+1}(t-1)$ as a function of x_t, or equivalently, by expressing $r_{1,s+1}, \ldots, r_{p,s+1}$ as a function of the corresponding rows of $H_{N,\infty}$, i.e. for $i=1, \ldots, p$, there exists unique α_{ikl} such that

$$r_{i,s+1} = \sum_{k=1}^{p} \sum_{l=m_k}^{m_k+n_k-1} \alpha_{ikl} r_{kl} \tag{3.6}$$

Again by the Hankel structure, this implies

$$r_{i,s+q} = \sum_{k=1}^{p} \sum_{l=m_k}^{m_k+n_k-1} \alpha_{ikl} r_{k,l+q-1}, \quad q \geqslant 1 \tag{3.7}$$

Also, we have

$$\hat{y}_{i,s+1}(t-1) = \sum_{k=1}^{p} \sum_{l=m_k}^{m_k+n_k-1} \alpha_{ikl} \hat{y}_{k,l}(t-1) \tag{3.8}$$

The rows of $H_{N,\infty}$ corresponding to $\bar{x}_t$ are given by

$$\bar{R} = \begin{bmatrix} r_{11} \\ \vdots \\ r_{1s} \\ \vdots \\ r_{p1} \\ \vdots \\ r_{ps} \end{bmatrix}$$

From these, we have

$$\left.\begin{aligned} \bar{x}_{t+1} &= \bar{F}\bar{x}_t + \bar{K}e_t \\ y_t &= \bar{H}\bar{x}_t + e_t \end{aligned}\right\} \tag{3.9}$$

where

$$\bar{F} = \left[\begin{array}{c|c|c|c}
\begin{matrix} \begin{matrix}0\\ \vdots \\ 0\end{matrix} & I_{s-1} \\ \multicolumn{2}{c}{\underbrace{0 \ldots 0}_{m_1-1}\ \alpha_{11m_1} \cdots \alpha_{11m_1+n_1-1}\ \underbrace{0 \ldots 0}_{s-(m_1+n_1-1)}} \end{matrix} & \begin{matrix} O \\ 0 \ldots 0\ \alpha_{12m_2} \cdots \alpha_{12m_2+n_2-1}\ 0 \ldots 0 \end{matrix} & \cdots & \begin{matrix} O \\ 0 \ldots 0\ \alpha_{1pm_p} \cdots \alpha_{1pm_p+n_p-1}\ 0 \ldots 0 \end{matrix} \\ \hline
\begin{matrix} O \\ 0 \ldots 0\ \alpha_{21m_1} \cdots \alpha_{21m_1+n_1-1}\ 0 \ldots 0 \end{matrix} & \begin{matrix} \begin{matrix}0\\ \vdots \\ 0\end{matrix} & I_{s-1} \\ \multicolumn{2}{c}{0 \ldots 0\ \alpha_{22m_2} \cdots \alpha_{22m_2+n_2-1}\ 0 \ldots 0} \end{matrix} & \cdots & \begin{matrix} O \\ 0 \ldots 0\ \alpha_{2pm_p} \cdots \alpha_{2pm_p+n_p-1}\ 0 \ldots 0 \end{matrix} \\ \hline
\vdots & \vdots & & \vdots \\ \hline
\begin{matrix} O \\ 0 \ldots 0\ \alpha_{p1m_1} \cdots \alpha_{p1m_1+n_1-1}\ 0 \ldots 0 \end{matrix} & \begin{matrix} O \\ 0 \ldots 0\ \alpha_{p2m_2} \cdots \alpha_{p2m_2+n_2-1}\ 0 \ldots 0 \end{matrix} & & \begin{matrix} \begin{matrix}0\\ \vdots \\ 0\end{matrix} & I_{s-1} \\ \multicolumn{2}{c}{0 \ldots 0\ \alpha_{ppm_p} \cdots \alpha_{ppm_p+n_p-1}\ 0 \ldots 0} \end{matrix}
\end{array}\right]$$

is a $ps \times ps$ matrix

$$\bar{H} = \begin{bmatrix} 1\ 0 \dots 0 & & & \\ & 1\ 0 \dots 0 & & \\ & & \dots & \\ & & & 1\ 0 \dots 0 \end{bmatrix}$$

is a $p \times ps$ matrix

$$\bar{K} = \begin{bmatrix} r_{11}(1) \\ \vdots \\ r_{1s}(1) \\ \vdots \\ r_{p1}(1) \\ \vdots \\ r_{ps}(1) \end{bmatrix}$$

is a $ps \times p$ matrix.

Remark 3.3

The total number of 'free' parameters in this representation in $\bar{F}$ is np, in $\bar{H}$ is 0, and in $\bar{K}$ is $p^2 s$. Hence, the total number of parameters is $np + p^2 s$. It is rather difficult to compare the total number of parameters involved in Representations 1 and 2.

Representation 1 : $2np \leqslant N_1 \leqslant 3np - n$.

Representation 2 : $N_2 = np + p^2 \left[\max_i \, (m_i + n_i - 1) \right]$.

It is interesting to consider how Representation 1 is related to Representation 2. To this end, for $i = 1, \dots, p$, denote

$$\beta_{ij} = [\beta_{ij1m_1} \cdots \beta_{ij1m_1+n_1-1} \vdots \cdots \vdots \beta_{ijpm_p} \cdots \beta_{ijpm_p+n_p-1}]$$
$$j = 1, \dots, m_i - 1, \quad j = m_i + n_i, \dots, s \qquad (3.10)$$

$$\alpha_i = [\alpha_{i1m_1} \cdots \alpha_{i1m_1+n_1-1} \vdots \cdots \vdots \alpha_{ipm_p} \cdots \alpha_{ipm_p+n_p-1}] \qquad (3.11)$$

$$\gamma_i = [\gamma_{i1m_1} \cdots \gamma_{i1m_1+n_1-1} \vdots \cdots \vdots \gamma_{ipm_p} \cdots \gamma_{ipm_p+n_p-1}] \qquad (3.12)$$

The following results are trivial.

Lemma 3.1

γ_i coincides with either α_i or β_{i,m_i+n_i}, i.e.

$$\gamma_i = \alpha_i \qquad \text{if } m_i + n_i - 1 = s \qquad (3.13\,a)$$

$$\gamma_i = \beta_{i,m_i+n_i} \qquad \text{if } m_i + n_i - 1 < s \qquad (3.13\,b)$$

Proof

This follows trivially from the definitions of α_i, β_{ij} and γ_i, see (3.6), (3.5) and (3.1). □

The parameters of $\bar{H}$, $\bar{F}$, $\bar{K}$ can be computed from the parameters of H, F, K and some of the β_{ijkl} as follows.

Lemma 3.2

(*a*)
$$\alpha_i = \gamma_i F^{s-(m_i+n_i-1)}$$

(*b*) n of the ps rows of $\bar{K}$ appear directly in K. The other rows of $\bar{K}$ are related to those of K via the β_{ijkl}

$$r_{ij}(1) = \sum_{k=1}^{p} \sum_{l=m_k}^{m_k+n_k-1} \beta_{ijkl} r_{kl}(1)$$

$i = 1, \dots, p$, $j = 1, \dots, m_i - 1$, and $j = m_i + n_i, \dots, s$.

Proof

(*a*) The result follows from the definition of γ (3.1)–(3.2) and the Hankel structures of F and $\bar{F}$. Note in particular that, if $m_i + n_i - 1 = s$, then (3.13 *a*) follows.

(*b*) This follows trivially by comparing the structures of K and $\bar{K}$, and by noting the definition of β_{ijkl} (3.5).

Remark 3.4

To construct $\bar{K}$, we will usually need some β_{ijkl} which do not appear in F or H, i.e. they are neither γ_i nor α_i. In fact, if $s > m_i + n_i$, we need β_{ijkl}, for $j = m_i + n_i + 1, \dots, s$, and if $m_i > 2$, we need β_{ijkl} for $j = 2, \dots, m_i - 1$.

Example

Let $p = 2$, $n = 4$ and $I = \{1, 3, 6, 8\}$. Thus $n_1 = 2$, $m_1 = 1$, $n_2 = 2$, $m_2 = 3$ and $s = 4$.

Representation 1

$$F = \begin{bmatrix} 0 & 1 & 0 & 0 \\ \gamma_{111} & \gamma_{112} & \gamma_{123} & \gamma_{124} \\ 0 & 0 & 0 & 1 \\ \gamma_{211} & \gamma_{212} & \gamma_{223} & \gamma_{224} \end{bmatrix}$$

$$K = \begin{bmatrix} r_{11}(1) \\ r_{12}(1) \\ r_{23}(1) \\ r_{24}(1) \end{bmatrix}$$

$$H = \begin{bmatrix} 1 & 0 & 0 & 0 \\ \beta_{2111} & \beta_{2112} & \beta_{2123} & \beta_{2124} \end{bmatrix}$$

and x_t contains the components indexed (1, 3, 6, 8) of $\hat{Y}_N(t)$.

Representation 2

$$
\bar{F} = \left[\begin{array}{cccc:cccc}
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
\alpha_{111} & \alpha_{112} & 0 & 0 & 0 & 0 & \alpha_{123} & \alpha_{124} \\
\hdashline
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
\alpha_{211} & \alpha_{212} & 0 & 0 & 0 & 0 & \alpha_{223} & \alpha_{224}
\end{array}\right] \qquad
\bar{K} = \left[\begin{array}{c}
r_{11}(1) \\ r_{12}(1) \\ r_{13}(1) \\ r_{14}(1) \\ \hdashline r_{21}(1) \\ r_{22}(1) \\ r_{23}(1) \\ r_{24}(1)
\end{array}\right]
$$

and

$$
\bar{H} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}
$$

and $\bar{x}_t$ contains the components indexed (1, 3, 5, 7, 2, 4, 6, 8) of $\hat{Y}_N(t)$.

In this case

$$\gamma_1 = \beta_{13}, \quad \gamma_2 = \alpha_2, \quad \alpha_1 = \gamma_1 F^2$$

Finally, we have the following result concerning Representation 2.

Lemma 3.3

Consider the Representation 2 defined as before, with dim $\bar{x}_t = ps$. Then

(i) $$\text{rank}\,[\bar{K}, \bar{F}\bar{K}, \bar{F}^2\bar{K}, \ldots] = n$$

where n is the order of the system.

(ii) $$\text{rank}\,[\bar{H}^T, \bar{F}^T\bar{H}^T, (\bar{F}^T)^2\bar{H}^T, \ldots] = ps$$

Proof

(i) From the form of $\bar{F}$ and $\bar{K}$ in (3.9), we have

$$
\bar{F}\bar{K} = \left[\begin{array}{c}
r_{12}(1) \\ r_{13}(1) \\ \vdots \\ r_{1s}(1) \\ r_{1,s+1}(1) \leftarrow \\ \hdashline r_{22}(1) \\ \vdots \\ r_{2,s+1}(1) \leftarrow \\ \hdashline \vdots \\ \hdashline r_{p2}(1) \\ \vdots \\ r_{p,s+1}(1) \leftarrow
\end{array}\right], \qquad
\bar{F}^2\bar{K} = \left[\begin{array}{c}
r_{13}(1) \\ \vdots \\ r_{1,s+1}(1) \\ r_{1,s+2}(1) \leftarrow \\ \hdashline r_{23}(1) \\ \vdots \\ r_{2,s+2}(1) \leftarrow \\ \hdashline \vdots \\ \hdashline r_{p3}(1) \\ \vdots \\ r_{p,s+2}(1) \leftarrow
\end{array}\right]
$$

where the rows indicated by the arrows are obtained by applying (3.7) with $q = 1$ and 2, respectively. Hence $[\bar{K}, \bar{F}\bar{K}, \bar{F}^2\bar{K}, \ldots] = R$. Thus,

$$\text{rank}\,[\bar{K}, \bar{F}\bar{K}, \bar{F}^2\bar{K}, \ldots] = n$$

The proof of (ii) is trivial as the observability matrix is a matrix obtained by permuting the rows of the identity matrix.

Remark 3.5

This theorem tells us that $\bar{F}$, $\bar{K}$, $\bar{H}$ is completely observable. However, it will only be completely controllable if $ps = n$; otherwise it is not completely controllable.

4. Input–output representations

As remarked earlier, it is rather difficult to obtain an input–output representation from Representation 1. This is mainly because of the non-canonical form of H. However, it is relatively easy to obtain an input–output representation from Representation 2.

Consider the input–output model

$$A(z)y_t = B(z)e_t \tag{4.1}$$

where $zy_t \triangleq y_{t+1}$, and $A(z)$, $B(z)$ are $p \times p$ matrix polynomials.

The construction of $A(z)$ and $B(z)$ from Representation 2 is a trivial modification of the results given in § 4 of Gevers and Wertz (1984). Hence, we will omit the derivations. The results are summarized in the following proposition.

Proposition 4.1

The input–output representation

$$A(z)y_t = B(z)e_t$$

may be obtained from Representation 2

$$\bar{x}_{t+1} = \bar{F}\bar{x}_t + \bar{K}e_t$$

$$y_t = \bar{H}\bar{x}_t + e_t$$

where $\bar{F}$, $\bar{K}$, $\bar{H}$ are as defined before. Then

$$a_{ii}(z) = z^s - \alpha_{iim_i+n_i-1}z^{m_i+n_i-2} - \alpha_{iim_i+n_i-2}z^{m_i+n_i-3} - \ldots - \alpha_{iim_i}z^{m_i-1}$$

$$a_{ij}(z) = -\alpha_{ijm_j+n_j-1}z^{m_j+n_j-2} - \alpha_{ijm_j+n_j-2}z^{m_j+n_j-3} - \ldots - \alpha_{ijm_j}z^{m_j-1} \quad \text{for } i \neq j$$

$$b_{ii}(z) = z^s + \beta_{iis}z^{s-1} + \ldots + \beta_{ii1}$$

$$b_{ij}(z) = \beta_{ijs}z^{s-1} + \ldots + \beta_{ij1}, \quad i \neq j$$

Remark 4.1

$A_0 = B_0 = I_p$, hence $A(z)$ and $B(z)$ are both row and column proper. In addition, we have the following proposition for computing $B(z)$ from $A(z)$ and $\bar{K}$.

Proposition 4.2

Let

$$\bar{B}=\begin{bmatrix} B^1 \\ B^2 \\ \vdots \\ B^p \end{bmatrix}$$

be a $p(s+1)\times p$ matrix where

$$B^i=\begin{bmatrix} \beta_{i11} & \cdots\cdots & \beta_{ip1} \\ \vdots & & \\ \beta_{i1s} & \cdots\cdots & \beta_{ips} \\ 0 \ldots 0 & 1 & 0 \ldots 0 \end{bmatrix}$$

$\uparrow$ ith position

is an $(s+1)\times p$ matrix

$$\tilde{K}=\begin{bmatrix} K^1 \\ K^2 \\ \vdots \\ K^p \end{bmatrix}$$

is a $p(s+1)\times p$ matrix where

$$K^i=\begin{bmatrix} u_i \\ k_{i1} \\ \vdots \\ k_{is} \end{bmatrix}$$

is an $(s+1)\times p$ matrix

$$u_i=[0 \ldots 0 \;\; 1 \;\; 0 \ldots 0]$$

$\uparrow$ ith position

Then $\bar{B}=M\tilde{K}$, where M is a $p(s+1)\times p(s+1)$ matrix and where

$$M_{ii}=\begin{bmatrix} 0 \ldots 0 & -\alpha_{iim_i} \cdots -\alpha_{iim_i+n_i-1} & 0 \ldots 0 & 1 \\ \vdots & \ddots & \ddots & \\ 0 & & & \\ -\alpha_{iim_i} & \ddots & & \\ \vdots & & \ddots & \\ -\alpha_{iim_i+n_i-1} & & & \\ 0 & & & \bigcirc \\ \vdots & & & \\ 0 \;\; 1 & & & \\ 1 & & & \end{bmatrix}$$

is an $(s+1)\times(s+1)$ matrix and

$$M_{ij}=\begin{bmatrix} 0 \ldots 0 & -\alpha_{ijm_j} \cdots -\alpha_{ijm_j+n_j-1} & 0 \ldots 0 \\ \vdots & \ddots & \ddots \\ 0 & & \\ -\alpha_{ijm_j} & \ddots & \\ \vdots & & \\ -\alpha_{ijm_j+n_j-1} & & \bigcirc \\ 0 & & \\ \vdots & \ddots & \\ 0 & & \end{bmatrix}$$

is an $(s+1)\times(s+1)$ matrix.

Proof

The proof follows by the same derivations as those of § 4 of Gevers and Wertz (1984).

Example

We continue the earlier example. It is possible to express the Representation 2 in input–output form as

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} y_{t+4}+\begin{bmatrix} 0 & -\alpha_{124} \\ 0 & -\alpha_{224} \end{bmatrix} y_{t+3}+\begin{bmatrix} 0 & -\alpha_{123} \\ 0 & -\alpha_{223} \end{bmatrix} y_{t+2} +\begin{bmatrix} -\alpha_{112} & 0 \\ -\alpha_{212} & 0 \end{bmatrix} y_{t+1}+\begin{bmatrix} -\alpha_{111} & 0 \\ -\alpha_{211} & 0 \end{bmatrix} y_t$$

$$=\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \epsilon_{t+4}+\begin{bmatrix} \beta_{114} & \beta_{124} \\ \beta_{214} & \beta_{224} \end{bmatrix} \epsilon_{t+3}+\begin{bmatrix} \beta_{113} & \beta_{123} \\ \beta_{213} & \beta_{223} \end{bmatrix} \epsilon_{t+2} +\begin{bmatrix} \beta_{112} & \beta_{122} \\ \beta_{212} & \beta_{222} \end{bmatrix} \epsilon_{t+1}+\begin{bmatrix} \beta_{111} & \beta_{121} \\ \beta_{112} & \beta_{122} \end{bmatrix} \epsilon_t$$

where β_{ijk} may be obtained from α_i, and $\bar{K}$.

Remark 4.2

From the special structure of $A(z)$, it is possible to write $A(z)$ as

$$A(z)=\begin{bmatrix} z^{s-m_1+1}-\ldots-\alpha_{11m_1} & -(\alpha_{12m_2+n_2-1}z^{n_2-1}+\ldots+\alpha_{12m_2}) & \cdots & -(\alpha_{1pm_p+n_p-1}z^{n_p-1}+\ldots+\alpha_{1pm_p}) \\ -(\alpha_{21m_1+n_1-1}z^{n_1-1}+\ldots+\alpha_{21m_1}) & z^{s-m_2+1}-\ldots-\alpha_{22m_2} & \cdots & -(\alpha_{2pm_p+n_p-1}z^{n_p-1}+\ldots+\alpha_{2pm_p}) \\ \vdots & \vdots & & \vdots \\ -(\alpha_{p1m_1+n_1-1}z^{n_1-1}+\ldots+\alpha_{p1m_1}) & -(\alpha_{p2m_2+n_2-1}z^{n_2-1}+\ldots+\alpha_{p2m_2}) & \cdots & z^{s-m_p+1}-\ldots-\alpha_{ppm_p} \end{bmatrix}$$

$$\times \begin{bmatrix} z^{m_1-1} & & & \\ & z^{m_2-1} & & \\ & & \ddots & \\ & & & z^{m_p-1} \end{bmatrix} \triangleq A_1(z) \begin{bmatrix} z^{m_1-1} & & & \\ & z^{m_2-1} & & \\ & & \ddots & \\ & & & z^{m_p-1} \end{bmatrix}$$

Thus, the characteristic equation of $A(z)$ is obtained as

$$\det A(z) = \det A_1(z) \left(z^{\sum_{i=1}^{p} m_i - p} \right)$$

Note that from the definition of s, the matrix polynomial $A_1(z)$ is both row and column proper. Note also that the diagonal elements of $A_1(z)$ are of degree $(s-m_i+1)$, and the (i, j)th off-diagonal elements are of order (n_j-1).

Remark 4.3

As is proved in Lemma 3.3, the Representation 2 may not be controllable. Equivalently, this implies that the matrix polynomials $A(z)$ and $B(z)$ may not be right coprime.

5. Conclusions

In this paper, we have relaxed the selection rules of the basis of a state-space representation of a multivariable control system. In Van Overbeek and Ljung (1982), Wertz *et al.* (1982) and Gevers and Wertz (1984), a full block row had to be selected and the chain rule had to be observed. However, here, we have examined the possibility of relaxing the rules by deleting the selection of one full block row. We have examined the representations arising from these selection rules, one minimal and another non-minimal. We have shown the relationships between the minimal representation and the non-minimal representation. Furthermore, from the non-minimal representation, we can obtain the corresponding input–output representation. The non-minimal representation is found to be completely observable, but may not be completely controllable. Equivalently, the input–output representation may not be right coprime.

We may ask how useful this relaxation of selection rules is in practice? The answer to this may lie in the desire to obtain an orthogonal (or as orthogonal as possible) basis for the state-space representation. In Ljung and Rissanen (1976), Van Overbeek and Ljung (1982) and Wertz *et al.* (1982) the choice of the basis for the state-space is based on either complexity issues or on orthogonality. The imposition of the selection of a block row seems to be rather arbitrary. By relaxing this constraint, we allow the basis to be selected among a wider set of candidate components, thereby increasing the chances of obtaining a well-conditioned basis. Effectively, we allow the data to inform us of what the most orthogonal basis should be, subject only to the restriction of the chain rule.

In practice, it may happen that imposing both the chain rule and the block selection rule would result in an ARMA model of relatively short AR part. However, relaxing the block selection rule may result in an ARMA model which has a long AR part (the lag is $s = \max_i (m_i + n_i - 1)$ which may have

generically many zero elements. One way to interpret this may be that the underlying time series consists of a number of dynamical processes of different time scales. These and other possible interpretations of this new parametrization are currently under investigation.

References

Corrêa, G. O., and Glover, K., 1982, *Proc. 6th I.F.A.C. Symp. on Identification and System Parameter Estimation*, Washington, pp. 1110–1115.

Deistler, M., and Hannan, E. J., 1981, *J. mult. Analysis*, **11,** 474.

Denham, M. J., 1974, *I.E.E.E. Trans. autom. Control*, **19,** 646.

Gevers, M., and Wertz, V., 1982, *Proc. 6th I.F.A.C. Symp. on Identification and System Parameter Estimation*, Washington, pp. 287–292.

Gevers, M., and Wertz, V., 1984, *Automatica*, **20.3.**

Glover, K., and Willems, J. C., 1974, *I.E.E.E. Trans. autom. Control*, **19,** 640.

Guidorzi, R., 1981, *Automatica*, **17,** 117.

Guidorzi, R., and Beghelli, S., 1982, *Proc. 6th I.F.A.C. Symp. on Identification and System Parameter Estimation*, Washington, pp. 461–466.

Ljung, L., and Rissanen, J., 1976, *Proc. 4th I.F.A.C. Symp. on Identification and System Parameter Estimation*, Tbilisi.

Luenberger, D. G., 1967, *I.E.E.E. Trans. autom. Control*, **12,** 290.

Picci, G., 1980, *Proc. Workshop on Numerical Methods for System Engineering Problems*, Lexington.

Rissanen, J., 1974, *Automatica*, **10,** 175.

Rissanen, J., and Ljung, L., 1975, *Proc. CNR CISM Symp. on Algebraic Systems Theory*, Udine.

Rissanen, J., 1981, *Proc. Int. Workshop on Rational Approximations for Systems*, Leuven, Belgium.

Van Overbeek, A. J. M., and Ljung, L., 1982, *Automatica*, **18,** 529.

Wertz, V., Gevers, M., and Hannan, E. J., 1982, *I.E.E.E. Trans. autom. Control*, **27,** 1200.