(71)

6th IFAC Symp. on Identification & Syst. Parameter Estimation, vol 2, pp 1403-1408

C18

ON THE PROBLEM OF STRUCTURE SELECTION FOR THE IDENTIFICATION OF STATIONARY STOCHASTIC PROCESSES.

M. GEVERS and V. WERTZ

Laboratoire d'Automatique et d'Analyse des Systèmes
Louvain University, Bâtiment Maxwell
B-1348 LOUVAIN-LA-NEUVE (BELGIUM).

ABSTRACT. Several "overlapping" but uniquely identifiable parametrizations can usually be fitted to the same multivariate stationary stochastic process. We show that these parametrizations are defined by a finite set of intrinsic invariants, and that they all give the same value to the determinant of the Fisher information matrix.

INTRODUCTION.

An important problem is that of determining the structure of the state-space or ARMA model for a multivariate stationary finite-dimensional stochastic process such that the model parameters become uniquely identifiable. Two different lines of thought have been followed for this problem. The first idea is to use canonical (state-space or ARMA) forms [1] - [2]. To any finite dimensional process one can associate in a unique way a set of "structural invariants" (e.g. the Kronecker invariants) which in turn uniquely determines a canonical form. The disadvantage with using canonical forms is that if those structural invariants are wrongly estimated, then the parameter estimation problem becomes ill-conditioned.

In recent years an alternative approach has been proposed, namely that of using "overlapping parametrizations" [3] - [8]. It has been recognized that the set of all finite dimensional systems can be represented by a finite number of parametrizations, each parametrization being uniquely identifiable. To each parametrization there corresponds a set of integers called "structure indices". Each of these parametrizations is able to represent almost all finite dimensional systems; each system can normally be represented by more than one such parametrization, and any two parametrizations for a given process are related by a linear transformation which corresponds to a coordinate transformation in Euclidean space; hence the word "overlapping" parametrizations. Now because a process can be represented in more than one overlapping form, the question naturally arises as to whether, for a given data set, any such form is better than the others in a statistical or numerical sense. So far no definite answer is available to this question. Different procedures have been proposed that select one out of several candidate parametrizations which is considered best in some ad hoc sence [4] - [7].

In this paper we first show that, for a n-th order process with a p-dimensional observation vector y_t, the parameters of any overlapping parametrization (either in state-space or in ARMA form) are obtained from a set of 2np "intrinsic invariants" which are determined from the Hankel matrix of impulse responses. To any choice of p structure indices, there corresponds, for a given system, a set of 2np parameters which completely specify this system. The choice of the p structure indices determines in which particular local coordinate space the system is described. From these 2np "intrinsic invariants", a unique state-space or ARMA parametrization can then be derived; these will belong to the set of overlapping parametrizations.

Next we compare different overlapping parametrizations in terms of asymptotic accuracy of the parameter estimates. We show that, if the determinant of the Fisher information matrix is used as a measure of asymptotic efficiency, then all overlapping parametrizations describing the same process are equivalent, in the sense that they will give the same value to this criterion.

Our result implies that if a process is modelled in state-space or ARMA form using a prediction error method, then the determinant of the covariance matrix of the parameter estimates will asymptotically be the same, whichever overlapping parametrization is used.

PARAMETRIZATION OF MULTIVARIATE SYSTEMS.

We consider throughout this paper a p-dimensional stationary full rank zero-mean stochastic process $\{y_t\}$ with rational spectrum. Then it is well known that $\{y_t\}$ can be described, up to second order statistics, by the following finite-dimensional representations :

State-space representation

$$x_{t+1} = Fx_t + Ke_t$$
$$y_t = Hx_t + e_t \qquad (2.1)$$

where the state x_t is an n-dimensional vector ; F, K and H are matrices of dimensions nxn, nxp, and pxn, F has all its eigenvalues strictly inside the unite circle, and $\{e_t\}$ is a p-dimensional white noise sequence with covariance matrix Q.

Input-output representation (ARMA model)

$$y_t+A_1y_{t-1}+\dots+A_ry_{t-r}=e_t+B_1e_{t-1}+\dots B_se_{t-s} \qquad (2.2a)$$

where $A_1,\dots A_r,B_1,\dots,B_s$ are pxp matrices and e_t is as before. This representation is equivalent with the following :

$$A(z)y_t = B(z)e_t \qquad (2.2b)$$

where A(z) and B(z) are square polynomial matrices in the variable z (z is the advance operator : $zy_t=y_{t+1}$), with det $A(z) \neq 0$ for $|z|\geq 1$, and $\lim_{z\to\infty} A^{-1}(z)\ B(z) = I$.

Without any loss of generality, we can make the following assumptions regarding these two representations.

Assumption 1a : The matrix triple (H,F,K) is of minimal order n, where n is the dimension of the state vector x_t, i.e. the pair (H,F) is observable and the pair (F,K) is controllable. n is then called the order of the process.

Assumption 1b : The polynomial matrices A(z) and B(z) are left coprime. It can then be shown that deg det A(z) = n, the order of the process.

Definition 1a : The set of all minimal triples (H,F,K) of order n will be denoted by S_n.

Definition 1b : The set of all left coprime polynomial pairs (A(z),B(z)) with deg det A(z)=n will be denoted by S_n^*.

Eliminating x_t in (2.1) or premultiplying (2.2) by $A^{-1}(z)$ leads to a third representation for the process $\{y_t\}$:

$$y_t = \sum_{i=0}^{\infty} H_i\, e_{t-i} = \underline{H}\ E^t \qquad (2.3)$$

where the pxp matrices H_i are called impulse response matrices (or Markov parameters). The infinite matrix $\underline{H}$ is defined as $\underline{H} = [H_0H_1H_2\dots]$ with $H_0=I_p$, $H(z)= \sum_{i=0}^{\infty} H_iz^i$ is analytic in $|z|\leq 1$. The infinite column vector E^t is defined as $E^t = [e_t^T, e_{t-1}^T, e_{t-2}^T, \dots]^T$. In addition we shall assume that the inverse $G(z) = H^{-1}(z) = \sum_{i=0}^{\infty} G_iz^i$ exists and is analytic in $|z|< 1$. Then $\{e_t\}$ is called the innovation process of $\{y_t\}$ and $e_t \triangleq y_t-\hat{y}_{t/t-1}$ where $\hat{y}_{t/t-1}$ is the linear least squares predictor of y_t given the past history Y^{t-1} of $\{y_t\}$.

The impulse response matrices are related to the representations (2.1) and (2.2) as follows:

$$H_0 = I,\quad H_i = H\ F^{i-1}\ K,\ i = 1,\ 2,\ \dots (2.4a)$$

$$\sum_{i=0}^{\infty} H_i\ z^{-i} = A^{-1}(z)\ B(z) \qquad (2.4b)$$

The impulse response representation (2.3) completely specifies the second-order statistics of the process $\{y_t\}$, namely the covariance function $R_y(k) = E\ \{y_ty_{t-k}^T\}$, $k = 0,\ 1,\ \dots$

We can now define identifiability up to second order statistics. Let θ be the vector of parameters in either the triple (H, F, K) or the pair (A(z), B(z)) and let Q be the covariance matrix of the white noise sequence $\{e_t\}$ in either (2.1) or (2.2).

Definition 2 : 2 parameter pairs (θ_1,Q_1) and (θ_2,Q_2) are undistinguishable if and only if

$$R_y(k\ ;\ \theta_1,Q_1) = R_y(k\ ;\ \theta_2,Q_2) \qquad k \geq 0 \qquad (2.5)$$

where $R_y(k\ ;\ \theta_i,\ Q_i)$ is the covariance function of the process $\{y_t\}$ generated by model i.

Note that if $\{y_t\}$ is Gaussian (or if a second order identification method is used such as a prediction error method), then the probability law (or the loss function) is completely determined by the second order moments, and Definition 2 can be replaced by the following

Definition 2' : For a Gaussian process 2 parameter pairs (θ_1,Q_1) and (θ_2,Q_2) are undistinguishable iff

$$p(Y_0^T\ ;\ \theta_1,Q_1) = p(Y_0^T\ ;\ \theta_2,Q_2) \qquad (2.6)$$
$$\forall\ Y_0^T \text{ and } \forall\ T > 0$$

Now it is easy to show that (θ_1,Q_1) and (θ_2,Q_2) are undistinguishable iff

$$Q_1 = Q_2 \text{ and } H_i(\theta_1)= H_i(\theta_2),\quad i=0,1,2\dots \qquad (2.7)$$

Because $Q_1 = Q_2$, we shall in the sequel drop the explicit dependence of $R_y(k)$ or $p(Y_0^T)$ on Q. The undistinguishability concept induces an equivalence relation on the sets S_n and S_n^*, which we shall denote by the symbol ∿. It follows from (2.4) and (2.7) that

$$\theta_1 \sim \theta_2 \iff H_i(\theta_1) = H_i(\theta_2)\ \forall\ i \qquad (2.8)$$

$$\iff H_2 = H_1T,\ F_2 = T^{-1}F_1T,\ K_2 = T^{-1}K_1 \text{ for some nonsingular matrix T.} \qquad (2.9)$$

$$\iff A_2(z) = M(z)A_1(z),\ B_2(z) = M(z)B_1(z) \text{ for some unimodu-}$$

lar matrix M(z). (2.10)

Two matrix triples (H_1, F_1, K_1) and (H_2, F_2, K_2) (resp. two polynomial matrix pairs $(A_1(z), B_1(z))$ and $(A_2(z), B_2(z))$ are called *equivalent* if the relations (2.9) (resp. (2.10)) hold.

The covariance function (or the probability law in the Gaussian case) of the process $\{y_t\}$ is completely determined by specifing(H, F, K, Q) or (A(z), B(z), Q). But because of the non-uniqueness induced by (2.8)-(2.9), in order to achieve identifiability, we have to find a *reparametrization* of the family $R_y(k ; \theta)$ or $p(Y_0^T ; \theta)$ in such a way that two different sets of parameters (in the reparametrized set) correspond to two different sequences of Markov parameters.

From now on we shall, for simplicity, asume that the process $\{y_t\}$ is Gaussian; identifiability is then defined by Definition 2'.
All statements hold, up to second order statistics, for non-Gaussian processes if $p(Y_0^T;\theta)$ is replaced by $R_y(k;\theta)$.

What is needed to achieve identifiability is a factorization of the map $p : \theta \to p(. ; \theta)$ in the following way :

$$\begin{tikzcd}
\mathcal{S}_n \arrow[rr, "p"] \arrow[dr, "f"'] & & P \\
& X_n \arrow[ur, "\hat{p}"'] &
\end{tikzcd} \qquad (2.11)$$

Here $\mathcal{S}_n$ is either S_n or S_n^* (see Definitions 1 above) ; p is the map defined by the probability law ; P is the image of p. The set X_n and the functions $f : \mathcal{S}_n \to X_n$ and $\hat{p} : X_n \to P^n$ must satisfy the following conditions :

a) for each $\theta \in \mathcal{S}_n$, $\xi = f(\theta)$ is finite-dimensional (2.12a)

b) $p(. ; \theta) = \hat{p}(. ; f(\theta))$ for all $\theta \in \mathcal{S}_n$ (2.12b)

c) $\hat{p}(. ; \xi_1) = \hat{p}(. ; \xi_2) \Rightarrow \xi_1 = \xi_2$ (2.12c)

The function f consists of a finite number of scalar components, say $f_1, \ldots, f_k$, which form a complete system of invariants (see [2]) for the equivalence relation (2.8), since by b) and c) :

$$\theta_1 \sim \theta_2 \Leftrightarrow \hat{p}(. ; f(\theta_1)) = \hat{p}(. ; f(\theta_2))$$
$$\Leftrightarrow f(\theta_1) = f(\theta_2) \qquad (2.13)$$

The set X_n can be identified with the quotient sets $S_n/\sim$ or $S_n^*/\sim$, or, equivalently, with the class of all impulse response sequences $\underline{H}$ admitting a minimal realization of order n.
Now it can be shown that, when p > 1, no single parametrization is able to describe all n-th order systems. Rather X_n can be described by a family of $\binom{n-1}{p-1}$ overlapping parametrizations of dimension 2np. Each set of 2np invariants (i.e. each local parametrization) is defined by specifying p integer valued numbers $n_1, \ldots, n_p$ called "structure indices" :

$$f_{n_1, \ldots, n_p} : \mathcal{S}_n(n_1, \ldots, n_p) \to X_n,$$
$$\text{with } X_n \in R^{2np} \qquad (2.14)$$

where $\mathcal{S}_n(n_1, \ldots, n_p)$ is a subset of $\mathcal{S}_n$.

Each map (2.14) is locally a complete system of surjective invariants of dimension 2np ; the subsets $\mathcal{S}_n(n_1, \ldots, n_p)$ for all possible choices of $n_1, \ldots, n_p$ overlap, and cover $\mathcal{S}_n$.

In the next section we shall define the structure indices and show how a choice of structure indices $n_1, \ldots, n_p$ defines a set of 2np invariants $f_{n_1, \ldots, n_p}$. These 2np invariants are computable functions of the impulse response matrices $H_0, H_1, H_2, \ldots$ and will be called intrinsic invariants. We shall then show how to construct overlapping state-space or ARMA parametrizations as a function of the 2np intrinsic invariants.

CONSTRUCTION OF A COMPLETE SYSTEM OF INVARIANTS.

From (2.3) we can write the linear least squares k-step ahead predictor $\hat{y}_{t+k|t}$ as follows :

$$\hat{y}_{t+k|t} = \sum_{i=k}^{\infty} H_i e_{t-i} \qquad k = 0,1,2,.. (3.1)$$

Therefore,

$$\hat{Y}_t \triangleq \begin{pmatrix} \hat{y}_{t+1|t} \\ \hat{y}_{t+2|t} \\ \cdot \\ \cdot \\ \cdot \end{pmatrix} = \begin{pmatrix} H_1 & H_2 & \cdots \\ H_2 & H_3 & \cdots \\ \cdot & & \\ \cdot & & \\ \cdot & & \end{pmatrix} \begin{pmatrix} e_t \\ e_{t-1} \\ \cdot \\ \cdot \end{pmatrix} = \mathcal{H} E^t \qquad (3.2)$$

Since the process is of order n, the rank of the Hankel matrix $\mathcal{H}$ is n. Therefore we can choose a set of linearly independent rows of $\mathcal{H}$, which will form a basis for the whole row space of $\mathcal{H}$. Now the structure indices will define which rows of $\mathcal{H}$ will form the basis, and we shall show how to construct a set of 2np invariants from $\mathcal{H}$ for a given choice of structure indices.

To any choice of n linearly independent rows of $\mathcal{H}$ we shall associate a multiindex $\underline{i} = (i_1, \ldots, i_n)$ where the numbers $i_1, \ldots, i_n$, arranged in increasing order, are the indices of the rows of $\mathcal{H}$ that form the basis. Two restrictive conditions will be imposed on the selection of the basis rows :

Condition 1 : if $j \in \underline{i}$, then $j-p \in \underline{i}$

Condition 2 : $1, 2, \ldots, p \in \underline{i}$

Condition 1 follows from the structure of the Hankel matrix : if the (j-p)-th row of

$\mathcal{H}$ is in the span of the preceding rows, so is the j-th row. Condition 2 results from the full rank assumption on $\{y_t\}$: it follows that the p components of $\hat{y}_{t+1/t}$ are linearly independent.

Definition 3 : If the selection of the basis vectors obeys conditions 1 and 2, then the corresponding multiindex is called "nice".

All nice multiindices correspond to a choice of the basis inside the first n-p+1 row blocks of $\mathcal{H}$. For given n and p, there are $\binom{n-1}{p-1}$ possible nice multiindices (Example : for a 2-dimensional vector process (p=2) of order 3 (n=3), there are only 2 choices : $\underline{i}_1 = (1, 2, 3)$ and $\underline{i}_2 = (1, 2, 4)$). There are however subsets of $\mathfrak{I}_n$ for which only one basis exists.

Definition of the structure indices

Let $\underline{i} = (i_1, \ldots, i_n)$ be a nice multiindex defining a basis for the rows of $\mathcal{H}$. For k = 1, ..., p, let n_k be the least natural number such that $(k + n_k p) \not\in \underline{i}$. Then $n_1, \ldots, n_p$ are called the "structure indices" corresponding to that basis; they specify which rows of $\mathcal{H}$ are taken in the basis. Note that $\sum_{i=1}^{p} n_i = n$. We can now define $\mathfrak{I}_n(n_1, \ldots, n_p)$ as the set of all n-th order systems for which the n rows specified by $n_1, \ldots, n_p$ in the Hankel matrix are linearly independent. $\mathfrak{I}_n(n_1,\ldots,n_p)$ is a proper subset of $\mathfrak{I}_n$. Consider now an element of $\mathfrak{I}_n(n_1, \ldots, n_p)$ specified by its Hankel matrix $\mathcal{H}$. We shall construct a complete system of 2np surjective invariants for this system, i.e. a reparametrization of this system using 2np parameters.

Let H^i be the i-th block of p rows of the infinite Hankel matrix $\mathcal{H}$ (e.g. $H^2 = [H_2 H_3 H_4 \ldots]$) and let

$$H^i = [h_{1i} \ldots h_{pi}]^T \qquad (3.3)$$

where h_{ki} are rows of infinite length. Since $\underline{H}$ is an element of $\mathfrak{I}_n(n_1, \ldots, n_p)$, the rows $h_{1(n_1+1)}, \ldots, h_{p(n_p+1)}$ can be expressed as :

$$h_{i(n_i+1)} = \sum_{j=1}^{p} \sum_{k=1}^{n_j} \alpha_{ijk} h_{jk} \quad i=1,\ldots,p \quad (3.4)$$

These relations define np scalar numbers α_{ijk}.

Now denote by $h_{ij}(k)$ the element in row i, column j of H_k. Then the 2np numbers $\{\alpha_{ijk}, k=1,\ldots,n_j; h_{ij}(k), k=1,\ldots,n_i; i,j=1,\ldots p\}$ (3.5) completely coordinatize $\mathfrak{I}_n(n_1,\ldots,n_p)$, i.e. they map that set in a one to one manner, on Euclidean space of dimension 2np. The impulse response sequence $H_1, H_2, H_3, \ldots$ is completely specified by the p structure indices and these 2np numbers. These 2np numbers constitute a complete system of surjective invariants, which will be called "intrinsic invariants" of the process.

In the notation of Section II :

$$f_{n_1,\ldots,n_p} : \mathfrak{I}_n(n_1,\ldots n_p) \longrightarrow \mathbb{R}^{2np}$$

$$: \{H_1, H_2, \ldots\} \to f_{n_1,\ldots,n_p} = \{\alpha_{ijk}, h_{ij}(k)\} \qquad (3.6)$$

We have thus constructed a family of functions $f_{n_1,\ldots,n_p}$, each of which is a complete system of surjective invariants mapping onto $\mathbb{R}^{2np}$. These functions are defined on the overlapping subsets $\mathfrak{I}_n(n_1 \ldots, n_p)$ which cover $\mathfrak{I}_n$. Next we show that one can construct corresponding overlapping (state-space or ARMA) parametrizations whose parameters are functions of the 2np intrinsic invariants just defined.

State-space parametrization

Consider an element of $\mathfrak{I}_n(n_1,\ldots,n_p)$ for a given set of structure indices, and let $\{\alpha_{ijk}, h_{ij}(k)\}$ be the intrinsic invariants of that element. Then the following is a state-space representation of that element :

$$H = \left.\left[\begin{array}{ccccc:c:ccccc} 1 & . & . & . & 0 & & 0 & . & . & . & 0 \\ 0 & . & . & . & . & & . & . & . & . & . \\ . & . & . & . & . & & . & . & . & . & . \\ . & . & . & . & . & \cdots & . & . & . & . & . \\ . & . & . & . & . & & . & . & . & . & . \\ 0 & . & . & . & 0 & & 1 & . & . & . & 0 \end{array}\right]\right\} p \qquad (3.7a)$$

$$\underbrace{\qquad\qquad}_{n_1} \qquad\qquad \underbrace{\qquad\qquad}_{n_p}$$

$$F = [F_{ij}] \quad i, j = 1,\ldots,p \qquad (3.7b)$$

$$F_{ii} = \underbrace{\begin{bmatrix} 0 & & \\ . & & \\ . & I_{n_i-1} & \\ . & & \\ 0 & & \\ \alpha_{ii1} & \cdots & \alpha_{iin_i} \end{bmatrix}}_{n_i} n_i, \quad F_{ij} = \underbrace{\begin{bmatrix} 0 & . & . & . & 0 \\ . & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ 0 & . & . & . & 0 \\ \alpha_{ij1} & & \cdots & & \alpha_{ijn_j} \end{bmatrix}}_{n_j} n_i \quad j \neq i$$

$$K = \begin{bmatrix} K_1 \\ . \\ . \\ . \\ K_p \end{bmatrix}, \text{ with } K_i = \begin{bmatrix} k_{i1} \\ . \\ . \\ . \\ k_{in_i} \end{bmatrix}, \text{ and } k_{ij} = [h_{i1}(j), \ldots, h_{ip}(j)], \text{ a row p-vector} \qquad (3.7c)$$

The proof is a straightforward but tedious verification that the relations (2.4a) hold with the α_{ijk} and $h_{ij}(k)$ defined from the H_i as in (3.4)-(3.5).

ARMA Parametrization

By an argument similar to that developed in [2] we obtain the following equations for the entries A(z) and B(z) of (2.2b) :

$$a_{ii}(z) = z^{n_i} - \alpha_{iin_i} z^{n_i-1} - \ldots - \alpha_{ii1} \qquad (3.12a)$$

$$a_{ij}(z) = -\alpha_{ijn_j} z^{n_j-1} - \ldots - \alpha_{ij1} \qquad (3.12b)$$

and

$$b_{ij}(z) = b_{ij\bar{n}+1}\, z^{\bar{n}} + b_{ij\bar{n}}\, z^{\bar{n}-1} + \ldots + b_{ij1} \qquad (3.12c)$$

with $\bar{n} \doteq \max\limits_{1 \leqslant i \leqslant p} n_i$. The α_{ijk} are defined by (3.4), while the b_{ijk} are defined as follows

$$\bar{B} = M\,\bar{K} \quad \text{where} \qquad (3.13)$$

$$\bar{B} = \begin{bmatrix} B^1 \\ \vdots \\ B^p \end{bmatrix}, \quad B^i = \begin{bmatrix} b_{i11} & & b_{ip1} \\ \vdots & & \\ b_{i1(\bar{n}+1)} & \cdots & b_{ip(\bar{n}+1)} \end{bmatrix} \qquad (3.14)$$

$$\underset{(n+p)\,\mathrm{x}p}{\bar{K}} = \begin{bmatrix} (1\ 0\ \ldots\ 0) \\ k_{11} \\ \vdots \\ k_{1n_1} \\ (0\ 1\ 0\ \ldots 0) \\ k_{21} \\ \vdots \\ k_{pn_p} \end{bmatrix} \qquad (3.15)$$

with k_{ij} defined by (3.7c)

$$M = [M_{ij}] \qquad (i,j = 1 \ldots p)$$

$$M_{ii} = \begin{array}{r|lcr|} & -\alpha_{ii1}\cdots-\alpha_{iin_i} & & 1 \\ & \vdots & & \\ & -\alpha_{iin_i} & & \\ & 1 & & \\ \bar{n}-n_i\Big\{ & 0 & \cdots\cdots & 0 \\ & 0 & \cdots\cdots & 0 \end{array}, \quad M_{ij} = \begin{array}{r|lcr|} & -\alpha_{ij1}\cdots-\alpha_{ijn_j} & & 0 \\ & \vdots & & \\ & -\alpha_{ijn_j} & & \\ & 0 & & \\ \bar{n}-n_j\Big\{ & 0 & \cdots\cdots & 0 \\ & 0 & \cdots\cdots & 0 \end{array} \qquad (3.16)$$

We can now summarize the main points of this section by extending the scheme (2.11). For every element of $\mathfrak{I}_n(n_1,\ldots, n_p)$, we have first defined the intrinsic invariants $f_{n_1,\ldots,n_p}$.

Next, by (3.7), (3.12) and (3.13), we have established two bijections $g_1(n_1,\ldots, n_p)$ and $g_2(n_1,\ldots, n_p)$ between the intrinsic invariants (defined on $\underline{H}$) and the quotient spaces $S_n/\sim$ and $S_n^*/\sim$

$$\begin{tikzcd}
\mathfrak{I}_n \arrow[rrrr, "p"] \arrow[drr, "f"'] \arrow[ddr, bend right, "h_1"'] \arrow[ddrrr, bend right=70, "h_2"'] & & & & P \\
& & X_n \arrow[urr, "\hat{p}"'] \arrow[dl, shift right, "g_1"'] \arrow[dr, shift left, "g_2"] & & \\
& S_n/\sim \arrow[ur, shift right, "g_1-1"'] & & S_n^*/\sim \arrow[ul, shift left, "g_2-1"] &
\end{tikzcd} \qquad (3.17)$$

Hence, by a property of complete sets of surjective invariants (see, e.g. [2]), $h_i(n_1, \ldots, n_p) = g_i(n_1, \ldots, n_p) \circ f_{n_1 \ldots n_p}$ are also locally complete sets of surjective invariants. This leads to the following factorizations :

$$\begin{tikzcd}
S_n \arrow[rr, "p"] \arrow[dr, "h_1"'] & & P & S_n^* \arrow[rr, "p"] \arrow[dr, "h_2"'] & & P \\
& S_n/\sim \arrow[ur, "\hat{p}\circ g_1^{-1}"'] & & & S_n^*/\sim \arrow[ur, "\hat{p}\circ g_2^{-1}"'] &
\end{tikzcd} \qquad (3.18)$$

In figs. (3.17)-(3.18) all quantities are to be indexed by $(n_1,\ldots, n_p)$. These factorizations of the probability map make the multivariable models identifiable. Given an arbitrary element (H, F, K) of $S_n(n_1,\ldots,n_p)$, which is parametrized by n^2+2np parameters that are not identifiable, we replace this element by an equivalent element of the quotient space $S_n/\sim$ via the map $h_1(n_1,\ldots,n_p)$ (see Fig. 3.18). This last element is parametrized by the system of 2np invariants defined by h_1 and appearing in the form (3.7). These 2np invariants are uniquely identifiable. The same can be said if an ARMA form is used.

ASYMPTOTIC EQUIVALENCE OF ALL OVERLAPPING FORMS.

In most cases an n-th order system can be represented in more than one of the overlapping forms, because different choices of nice multiindices can be made, which will define different sets of linearly independent rows of the Hankel matrix.

Example : Consider a bivariate process (p = 2) of order 3 (n = 3). Two nice multiindices exist, with the corresponding sets of structure indices : $n_1=2$, $n_2=1$ or $n_1=1$, $n_2=2$.

Now the state x_t of the state-space realization (2.1), with H, F, K defined by (3.7), is made up of the components of $\hat{Y}_{t-1}$ indexed by the element of the selected nice multiindex (see e.g. [6]). In our example;

- for $n_1=2$, $n_2=1$ – for $n_1=1$, $n_2=2$

$$\underline{i}_1=(1,2,3) \qquad\qquad \underline{i}_2=(1,2,4)$$

$$x_t \triangleq \begin{bmatrix} \hat{y}^1_{t/t-1} \\ \hat{y}^1_{t+1/t-1} \\ \hat{y}^2_{t/t-1} \end{bmatrix} \qquad\qquad x_t \triangleq \begin{bmatrix} \hat{y}^1_{t/t-1} \\ \hat{y}^2_{t/t-1} \\ \hat{y}^2_{t+1/t-1} \end{bmatrix}$$

In most cases, both choices are possible, but one might think that if $\hat{y}^1_{t+1/t-1}$ is close to the linear span of $\hat{y}^1_{t/t-1}$ and $\hat{y}^2_{t/t-1}$, then the choice of $\underline{i}_2$ would be preferable because the components of the state would be more orthogonal to one another, thereby making the ensuing parameter estimation problem numerically better behaved. More generally the question is whether any one of the overlapping parametrizations is optimal in some sense. We present a partial answer to this question.

<u>Theorem</u> : Given the intrinsic invariants

$\{\alpha_{ijk}, h_{ij}(k)\}$ and $\{\alpha^*_{ijk}, h^*_{ij}(k)\}$ corresponding to two different sets of structure indices $n_1, \ldots, n_p$ and $n^*_1, \ldots, n^*_p$, then the determinants of the information matrices corresponding to these two parametrizations are identical.

The proof of the theorem follows from the following lemma.

Lemma : Let $\{\alpha_{ijk}, h_{ij}(k)\}$ and $\{\alpha^*_{ijk}, h^*_{ij}(k)\}$ be the intrinsic invariants of a given n-th order system in $\mathcal{S}_n$ for two different choices of the structure indices. Then the Jacobian of the transformation between these two parameter vectors is unity.

The proof of the lemma can be found in [6]. The theorem can then be proved as follows. Let $\theta = \{\alpha_{ijk}, h_{ij}(k)\}$ and $\theta^* = \{\alpha^*_{ijk}, h^*_{ij}(k)\}$. The corresponding information matrices M_θ and M_{θ^*} are related by †:

$$M_{\theta^*} = E_{Y|\theta^*}\{(\frac{\partial \log p(Y|\theta^*)}{\partial\theta^*}^T)(\frac{\partial \log p(Y|\theta)}{\partial\theta^*})\}$$

$$=(\frac{\partial\theta}{\partial\theta^*})^T E_{Y|\theta}\{(\frac{\partial \log p(Y|\theta)}{\partial\theta}^T)(\frac{\partial \log p(Y|\theta)}{\partial\theta})\}(\frac{\partial\theta}{\partial\theta^*})$$

$$= (\frac{\partial\theta}{\partial\theta^*})^T M_\theta (\frac{\partial\theta}{\partial\theta^*})$$

It follows from the bijective relationship between the intrinsic invariants $\alpha_{ijk}, h_{ij}(k)$ and the corresponding overlapping parametrizations H,F,K or A(z), C(z) that the theorem also holds when two overlapping (state-space or ARMA) parametrizations are compared.

Corollary : Given two overlapping parametrizations F,K,H and F^*, K^*, H^* in the form (3.7) (resp. A(z),B(z) and $A^*(z), B^*(z)$ in the form (3.12)) for the same process, corresponding to two different sets of structure indices $\{n_1, \ldots, n_p\}$ and $\{n^*_1, \ldots, n^*_p\}$ then the determinants of the information matrices corresponding to these two parametrizations are identical.

If the parameters are estimated using a maximum likelihood or a prediction error method, then the covariance matrix of the estimation errors is asymptotically equal to the inverse of the Fisher information matrix M_θ.

Therefore all overlapping parametrizations are asymptotically equivalent, as far as the accuracy of the parameter estimates is concerned, when this accuracy is measured by the determinant of the covariance matrix of the estimation errors. Of course other criteria could be used that might be able to discriminate, even asymptotically, between different structures, see e.g. [7]. Some structures might also be better than others when only a finite data record is available. In [4] and [6] some heuristic selection procedures have been proposed to handle the finite data situation.

CONCLUSIONS.

The problem of specifying identifiable parametric structures for multivariable systems can be solved by a factorization of the probability map in such a way as to define a finite set of invariants which completely characterize the process. Proceeding in this way we have constructed a family of overlapping parametrizations which completely cover the set of finite-dimensional minimal-order systems. Since a given process can in general be represented by different overlapping parametrizations, the question then arises as to whether some parametrizations might yield more accurate parameter estimates than others. Our main result is that all overlapping parametrizations yield asymptotically the same value to the determinant of the information matrix. Therefore, when a prediction error identification method is used for the estimation of the parameters, all overlapping parametrizations will give the same value to the determinant of the asymptotic error covariance matrix.

REFERENCES.

[1] M.J. DENHAM, "Canonical Forms for the Identification of Multivariable Linear Systems", IEEE Trans. A.C., vol. 19, pp. 646-656, 1974.

[2] R.P. GUIDORZI, "Invariants and Canonical Forms for Systems Structural and Parametric Identification", Automatica, vol. 17, pp. 117-133, 1981.

[3] K. GLOVER, J.C. WILLEMS, "Parametrization of Linear Dynamical Systems, Canonical Forms and Identifiability", IEEE Trans. A.C., vol. 19, pp. 640-646, 1974.

[4] L. LJUNG, J. RISSANEN, "On Canonical Forms, Parameter Identifiability and the Concept of Complexity", 4th IFAC Symp. on Identification and System Parameter Estimation, Tbilisi, URSS, 1976.

[5] A.J.M. VAN OVERBEEK, L. LJUNG, "On Line Structure Selection for Multivariable State Space Models", 5th IFAC Symp. on Identification and System Parameter Estimation, Darmstadt, FRG, 1979.

[6] V. WERTZ, M. GEVERS, E. HANNAN, "The Determination of Optimum Structures for the State Space Representation of Multivariate Stochastic Processes", to appear in IEEE Trans. Autom. Control, Oct. 1982.

[7] J. RISSANEN, "Estimations of Structure by Minimum Description Length", Proc. Intern. Workshop on Rational Approximations for Systems. Leuven, Belgium, 1981.

[8] G. PICCI, "Some Numerical Aspects of Multivariable Systems Identification", Proc. of the Workshop on "Numerical Methods for System Engineering Problems", Lexington, Kentucky, June 1980.

---

† If x is a scalar and θ a column k-vector, then $\frac{\partial x}{\partial\theta}$ denotes the row vector $\frac{\partial x}{\partial\theta_1}, \ldots, \frac{\partial x}{\partial\theta_k}$.