IFAC World Congress, Munich, 1987

C29

PARAMETRIZATION ISSUES IN SYSTEM IDENTIFICATION

M. Gevers and V. Wertz (Chargé de recherches FNRS)

Laboratoire d'Automatique, Dynamique et Analyse des Systèmes
Université Catholique de Louvain, Bâtiment Maxwell
Place du Levant, 3, B-1348 LOUVAIN-LA-NEUVE (BELGIUM)

Abstract. We present a brief introduction to the problems of parametrization and identifiability. We introduce a distinction, which we believe useful, between identifiability of specific parameters and identifiability of a model structure. For the latter we present the structure, and some possible parametrizations, for the set of all linear multivariable systems of given McMillan degree.

Keywords. System identification; multivariable systems; identifiability.

1. INTRODUCTION

At the earliest stage of any identification experiment the would be identifier has to consider the question of identifiability. This question contains two aspects :

- an input signal aspect : is the input signal rich enough (or informative enough) to distinguish between different models ? For example, if the input signal is constant, it is impossible to distinguish between the models

$$Y(s) = K\,U(s) \text{ and } Y(s) = \frac{K}{1+sT}\,U(s)$$

from input-output measurements.

- a parametrization aspect : loosely speaking, the problem is whether the model set is parametrized in such a way that every input-output map is represented by a unique value of the parameter vector.

In this paper we shall not discuss the problem of sufficiently rich or persistently exciting input signals. We shall concentrate on the parametrization aspects of identifiability. We shall distinguish between two concepts : parameter identifiability and identifiability of the model structure (or structural identifiability). The two concepts arise from two different problems.

Parameter identifiability deals with the question of whether a set of specific and physically meaningful parameters that appear in the description of a physical system can be identified from input-output experiments. The outcome of the investigation might be that some parameters can be identified but others not.

The question of identifiability of a model structure (or structural identifiability) arises when a parametrized black box model set is chosen to identify or approximate a system. The parametrization must then be chosen so that every input-output map (or transfer function) corresponds to a unique value of the parameter vector. This notion is a property of the parametrization without any reference to a true system.

The main thrust of this paper will be concerned with the identifiability of model sets, and in particular with the choice of identifiable parametrizations for multivariable linear systems. This is because parameter identifiability can only be checked on a case by case basis, while a comprehensive theory is now available for the parametrization of multivariable black box models.

To introduce these two concepts of identifiability in a rigourous way, we first present some definitions on models, model sets and model structures in section 2. In section 3 we shall very briefly deal with the concept of identifiability, the motivation being mainly to distinguish parameter identifiability from the identifiability of a model structure. In section 4 we describe the structure of all linear multivariable systems of given McMillan degree. This then leads directly to a presentation of identifiable parametrizations for such systems, which is the subject of section 5. A much more detailed presentation of these concepts, together with a brief survey of results on structure estimation, can be found in Gevers and Wertz (1987).

2. MODELS, MODEL SETS AND MODEL STRUCTURES

Because the problems of parametrization are essentially algebraic, we shall for simplicity assume that we deal with deterministic input-output (I/O) models. We shall also limit ourselves to linear discrete time models. Our development is inspired by but not identical to Ljung (1987).

Definition 2.1 : A model M is a stable algebraic operator that transforms a given input process u(t) into a unique output process y(t).

Therefore the following are all models by our definition, provided specific values (rational functions, or real-valued parameters, or polynomials) are put into the operators G(z), A, B, C, etc.

Transfer function model (TF) :

$$y(t) = G(z)\,u(t) \qquad (2.1)$$

where $y(t) \in \mathbb{R}^p$, $u(t) \in \mathbb{R}^m$ and $G(z)$ is a strictly causal rational stable transfer function matrix, $G(z) = \sum_1^{\infty} G_i z^{-i}$ and $z^{-i}\{u(t)\} = \{u(t-i)\}$.

State space model (SS) :

$$\begin{cases} x(t+1) = Ax(t) + Bu(t) \\ y(t) = Cx(t) \end{cases} \qquad (2.2)$$

where $x(t) \in \mathbb{R}^n$, $u(t)$ and $y(t)$ are as before, A, B, C are real matrices of appropriate dimensions and dim $x(t)$ is minimal.

Matrix fraction description (MFD) :

$$P(z)y(t) = Q(z)u(t) \qquad (2.3)$$

where $u(t)$ and $y(t)$ are as before, and $P(z)$, $Q(z)$ are left coprime polynomial matrices in z, with $z^i\{u(t)\} = \{u(t+i)\}$.

One can think of other model descriptions such as an impulse response representation or an ARMA model. Whatever the description chosen for our model (e.g. SS, MFD, ARMA), one can always compute the TF $G(z)$ from it, e.g.

$$G(z) = C(zI-A)^{-1}B \quad = \quad P^{-1}(z)Q(z)$$

One can therefore introduce the following definition.

Definition 2.2 : Two models M^1 and M^2 are **equivalent** (we write $M^1 = M^2$) if and only if their transfer functions are identical : $G^1(z) = G^2(z)$ for all z.

In identification a search will typically be conducted over a **model set**, which is obtained as the range of a **smoothly** parametrized algebraic operator where the parameter vector θ, of dimension d, is allowed to cover a subset D of $\mathbb{R}^d$. The operator is called a **model structure**.

Definition 2.3 : A model set M^* is a set of models :

$$M^* = \{M(\theta) | \theta \in D \subset \mathbb{R}^d\}$$

Definition 2.4 : A model structure $\tilde{M}$ is a **differentiable** mapping from a subset D_M of $\mathbb{R}^d$ to a model set :

$$\tilde{M} : \theta \in D_M \rightarrow M(\theta) \in M^*$$

Example :

1) M : $y(t) = 0.85\ y(t-1) + 1.5\ u(t-1)$

2) M^* : $\{y(t) = \alpha y(t-1) + \beta u(t-1) | (\alpha,\beta) \in D_M \subset \mathbb{R}^2\}$
with $D_M = \{\alpha,\beta|\ |\alpha|<1,\ |\beta| \leq 100\}$

3) $\tilde{M}$: $(\alpha,\beta) \in D_M \rightarrow M^* = \{y(t) = \alpha y(t-1) + \beta u(t-1) | (\alpha,\beta) \in D_M\}$

3. IDENTIFIABILTY

Roughly speaking, a model structure is called identifiable if the mapping of Definition 2.4 is invertible. We now give a precise definition of structural identifiability.

Definition 3.1 : A model structure $\tilde{M}$ is **globally identifiable at** θ^* if

$$M(\theta) = M(\theta^*),\ \theta \in D_M \rightarrow \theta = \theta^*$$

It is **strictly globally identifiable** if it is globally identifiable $\forall\ \theta^* \in D_M$, and **globally identifiable** if it is globally identifiable at almost all $\theta^* \in D_M$.

Comment 3.1 : We stress that this definition covers only one aspect of identifiability : it is a property of the model structure (i.e. the parametrization) only and is totally independent of a possible "true system". In addition it does not say anything about the richness of the data set.

To stress the relevance of these definitions to our problem of constructing identifiable model structures, we point out the following facts :

1) The set of all SISO models of order up to n cannot be described as the range of a strictly globally identifiable model structure because of pole-zero cancellations. It can, however, be described as the range of a globally identifiable model structure.

2) The set of all multivariable systems of order up to n (or even exactly n) cannot be described as the range of a unique globally identifiable model structure, but it can be described as a union of ranges of model structures :

$$M^* = \bigcup_{i=1}^{\ell} R(\tilde{M}_i)$$

The problem of **structural identifiability** arises when, in a blackbox identification context, it is desired to select a model structure whose range covers a desired model set, e.g. the set of all linear multivariable systems of order n with p outputs and m inputs. If a structure is globally identifiable and if the data are informative enough, the parameter estimation problem will be well posed. Notice that this does not guarantee convergence of the estimated model to the true system since that would require, in addition, that the true system be contained in the model set.

The problem of **parameter identifiability** arises in a different context, namely when a model of a true system has been obtained from physical principles. The question of parameter identifiability is then whether some specific physically meaningful parameters within that model can be identified from input-output experiments. Parameter identifiability is also dependent on the invertibility of the model structure, but with two essential differences :

1) if a set of parameters is called identifiable, it means that these parameters converge to their true values, while structural identifiability is independent of any notion of true values.

2) some parameters of interest may be identifiable whilst the structure itself may not be globally identifiable. The following example, from Godfrey and Distefano (1985), illustrates this.

Example 1 : Suppose that a physical system gives rise to the following model :

$$\begin{cases} \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} b \\ 0 \end{bmatrix} u \\ y = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \end{cases}$$

From I/O experiments, the following parameters can be identified : a_{11}, a_{22}, b and the product $a_{21}a_{12}$. Therefore the parameters a_{11}, a_{22} and b are identifiable, but the structure is not identifiable, unless a_{12} or a_{21} is known from the physics of the problem.

The problem of checking the parameter identifiability of a physical model is essentially a case by case study. A number of methods have been developed. See Walter (1982) and Godfrey and Distefano (1985) for recent surveys.

4. THE STRUCTURE OF MULTIVARIABLE LINEAR SYSTEMS

Definition 4.1 : We call S(n) the set of all strictly proper stable rational transfer function matrices G(z) of order (i.e. Mc Millan degree) n, with m inputs and p outputs.

From now on we look at the problem of parametrizing S(n). The reason for looking at this problem is twofold :

- the set S(n) cannot be described as the range of a single model structure ; hence the parametrization problem is non trivial ;
- the structure of S(n) is by now well understood.

We consider the system (2.1) with a strictly proper transfer function matrix (TFM) G(z) and we introduce the following Hankel matrix :

$$\mathcal{H}_{1,N}[G] \triangleq \begin{bmatrix} G_1 & G_2 & \cdots\cdots & G_N \\ G_2 & G_3 & \cdots\cdots & G_{N+1} \\ \cdot & & & \\ \cdot & & & \\ \cdot & & & \\ G_N & G_{N+1} & \cdots\cdots & G_{2N-1} \end{bmatrix} \qquad (4.1)$$

$\mathcal{H}_{1,N}[G]$ is made up of blocks of p rows. The rank of $\mathcal{H}_{1,\infty}[G]$ is called order or McMillan degree of G(z). We assume that the rows of the first block of $\mathcal{H}_{1,\infty}[G]$ are linearly independent. (This only eliminates degenerate transfer functions whose rows would be linearly dependent). We call r_{ij} the i-th row of the j-th block of $\mathcal{H}_{1,\infty}[G]$.

Our objective now is to find a parametrization of the elements of S(n), i.e. we would like to be able to represent an arbitrary element of S(n) by a model (SS or MFD or ARMA, ...) that depends on a finite parameter vector θ, as opposed to the infinite set of matrices G_i. To solve this problem we need to state a few results about the structure of S(n).

Result 4.1 : If G(z) $\in$ S(n), then there exists at least one partition $n = n_1 + n_2 + \ldots + n_p$ of n such that the set of rows

$$\{r_{ij} : i=1, \ldots, p \; ; \; j=1, \ldots, n_i\} \triangleq R(n_1, \ldots, n_p) \quad (4.2)$$

form a basis of $\mathcal{H}_{1,\infty}[G]$.

This follows immediately from the Hankel structure. Note that the rows in (4.2) are chosen in such a way that if $r_{ij} \in R(n_1, \ldots, n_p)$ for j>1, then $r_{i,j-1} \in R(n_1, \ldots, n_p)$. The set of the rows selected in $R(n_1, \ldots, n_p)$ is completely determined by the partition $\mu \triangleq (n_1, \ldots, n_p)$. The indices $n_1, \ldots, n_p$ are called structure indices. We also denote $|\mu| = \sum_1^p n_i = n$. It is easy to see that there are $\begin{bmatrix} n+p-1 \\ p-1 \end{bmatrix}$ such partitions.

Definition 4.2 : We call U_μ the set of all elements in S(n) for which the corresponding set of rows (4.2) forms a basis for the rows of $\mathcal{H}_{1,\infty}[G]$.

We now show that U_μ can be completely coordinatized by n(p+m) coordinates.
Row r_{i,n_i+1} is a unique linear combination of the basis rows .

$$r_{i,n_i+1} = \sum_{j=1}^{p} \sum_{\ell=1}^{n_j} \alpha_{ij\ell} \, r_{j\ell} \quad i = 1, \ldots, p \qquad (4.3)$$

It follows again from the Hankel structure that knowing the first block of elements (i.e. the first m elements) of each basis row (4.2) and the coefficients $(\alpha_{ij\ell}$; $i,j = 1, \ldots, p$; $\ell = 1, \ldots, n_j)$ allows one to compute any other row of $\mathcal{H}_{1,\infty}[G]$ and therefore to specify G(z) completely. Therefore any element G(z) $\in U_\mu$ can be mapped into a vector $\tau_\mu \in \mathbb{R}^d$ by the following $d \triangleq n(p+m)$ coordinates :

$$\tau_\mu = \begin{Bmatrix} \alpha_{ij\ell} \; ; \; i,j = 1, \ldots, p \; ; \; \ell = 1, \ldots, n_j \\ g_\ell(i,j); \; i=1,\ldots,p; \; j=1,\ldots,m; \; \ell=1,\ldots,n_i \end{Bmatrix} \qquad (4.4)$$

where $g_\ell(i,j)$ is the (i,j)-th element of the matrix G_ℓ.

With this mapping, which we denote by Φ_μ, every G(z) $\in U_\mu$ can be mapped into a point in a subspace θ_μ of $\mathbb{R}^{d}$:

$$\Phi_\mu : G(z) \in U_\mu \to \tau_\mu = \Phi_\mu(G(z)) \in \theta_\mu \subset \mathbb{R}^d \quad (4.5)$$

We now have the following important result.

Result 4.2 : see Clark (1976)

1) S(n) is a real analytic manifold of dimension n(p + m).
2) S(n) is the union of all U_μ for which $|\mu| = n$. Each U_μ is open and dense in S(n) ; θ_μ is open and dense in $\mathbb{R}^d$.

Comment 4.1 : Since U_μ with $|\mu| = n$ is open and dense in S(n) it follows that almost all points of S(n) are in U_μ for any such μ. The choice of μ specifies a local coordinate system and, as we shall see in Section 5, a parametrization (in SS or MFD form, say) called pseudocanonical form. The message therefore is that a given system of order n can be represented almost surely by any one of the $\begin{pmatrix} n+p-1 \\ p-1 \end{pmatrix}$ pseudocanonical forms corresponding to the $\begin{pmatrix} n+p-1 \\ p-1 \end{pmatrix}$ partitions μ of n, i.e. the U_μ overlap. However, no unique set U_μ can cover all of S(n). It means that no unique identifiable paramatrization can represent all systems in S(n).

Having described the structure of S(n), one can now think of the identification problem in the following terms. Estimate the order n, then take any partition μ of n such that $|\mu| = n$ and compute the maximum likelihood estimate $\hat{\tau}_\mu$ of the corresponding vector τ_μ that completely specifies the system. Almost surely the system can be represented in that coordinate system. The computation of the parameter vector by, say, a maximum likelihood method necessitates an I/O description of the point τ_μ. In section 5 we shall present pseudocanonical SS or MFD models which can be written directly as functions of τ_μ.

An alternative is to cover S(n) by disjoint subsets V_μ which can be coordinatized by non-overlapping coordinate systems. These will give rise to the somewhat simpler canonical SS or MFD forms.

Definition 4.3 : We call V_μ the subset of U_μ for which the rows (4.2) specified by μ are the first n linearly independent rows of $\overline{\mathcal{H}_{1,\infty}[G]}$. Since the row r_{i,n_i+1} is now a linear combination of the basis rows above it, (4.3) is replaced by

$$r_{i,n_i+1} = \sum_{j=1}^{p} \sum_{\ell=1}^{n_{ij}} \alpha_{ij\ell}\, r_{j\ell} \qquad i = 1, \ldots, p \quad (4.6)$$

where

$$n_{ij} = \min (n_i, n_j) \quad \text{if } i \leqslant j \qquad (4.7)$$
$$= \min (n_i+1, n_j) \text{ if } i > j$$

It follows, by the same argument as before, that any element $G(z) \in V_\mu$ can be mapped into a vector ρ_μ in $\mathbb{R}^{d(\mu)}$ defined by the following coordinates:

$$\rho_\mu = \begin{cases} \alpha_{ij\ell};i,j=1, \ldots, p;\ell=1,\ldots, n_{ij} & (4.8) \\ g_\ell(i,j);\ i=1,\ldots,p;j=1,\ldots, m;\ell=1,\ldots, n_i & \end{cases}$$

where

$$d(\mu) = n(s+1) + \sum_{i<j} (\min (n_i,n_j) + \min (n_i,n_j+1)) \qquad (4.9)$$

We call this mapping ψ_μ. It maps every $G(z) \in V_\mu$ into a point in a subspace X_μ of $\mathbb{R}^{d(\mu)}$:

$$\psi_\mu : G(z) \in V_\mu \rightarrow \rho_\mu = \psi_\mu(G(z)) \in X_\mu \subset \mathbb{R}^{d(\mu)} \quad (4.10)$$

Result 4.3 : see Hazewinkel and Kalman (1976)

1) The V_μ are disjoint, $V_\mu \subset U_\mu$ and

$$\bigcup_{|\mu|=n} V_\mu = S(n)$$

2) X_μ is an open and dense subset of $\mathbb{R}^{d(\mu)}$

Comment 4.2 : For p > 1, S(n) is partitioned into the $\binom{n+p-1}{p-1}$ disjoint sets V_μ, $|\mu| = n$, which are of different dimensions $d(\mu)$. Hence every system $\sigma \in S(n)$ belongs to one and only one of the V_μ and has therefore a set of structure indices $\mu = (n_1, \ldots, n_p)$ attached to it. Those structure indices are usually called observability indices or left Kronecker indices ; they determine the local coordinate system in which that system is described by the vector ρ_μ. They in turn define canonical SS or MFD forms as we shall see in the next section. Note that in all cases $d(\mu) \leqslant d = n(p+m)$. The canonical forms therefore will generally have fewer parameters than the corresponding pseudocanonical ones. This is one advantage of canonical forms.

Recall now that a model structure was defined (see Definition 2.4) as a mapping from a parameter vector θ to a particular model $M(\theta)$, and that a structure was called globally identifiable if that mapping was injective for allmost all θ in a subset D_M. Now we have shown that

(i) given a system $G(z) \in S(n)$, there exists a uniquely defined set of structure indices $\mu = (n_1, \ldots, n_p)$ and a uniquely defined mapping ψ_μ from G(z) to a parameter vector ρ_μ. This in turn yields a unique canonical form as we shall see next.

(ii) given a system $G(z) \in S(n)$, then for almost any arbitrary set of structure indices

$\mu = (n_1, \ldots, n_p)$ such that $\sum_1^p n_i = n$

there exists a uniquely defined mapping Φ_μ from G(z) to a parameter vector τ_μ. This parameter vector in turn defines a pseudocanonical form.

In the next section we define model structures (in SS or MFD form) that are entirely specified by the integer-valued structure indices and the real-valued parameter vectors (ρ_μ or τ_μ); hence these model structures will be identifiable since the sought after inverse mappings are precisely ψ_μ and Φ_μ.

5. CANONICAL AND PSEUDOCANONICAL FORMS

We now describe canonical and pseudocanonical SS and MFD forms for a strictly proper p x m TFM $G(z) \in S(n)$.

5.1. Canonical forms

It follows from section 4 that if $G(z) \in S(n)$, then $G(z) \in V_\mu$ for a unique μ and that it is represented by a coordinate vector ρ_μ : see (4.8). From ρ_μ the following canonical SS model A, B, C can be obtained (see (2.2)) :

$$\underset{(nxn)}{A =} \left[\begin{array}{c|c|c|c}
\begin{matrix} 0 & & \\ \vdots & I_{n_1-1} & \\ 0 & & \\ \alpha_{111} & \cdots & \alpha_{11n_1} \end{matrix} &
\begin{matrix} 0 & \cdots & & 0 \\ \vdots & & & \vdots \\ 0 & & & 0 \\ \alpha_{121} & \cdots\ \alpha_{12n_{12}} & 0 & \cdots\ 0 \end{matrix} &
\cdots &
\begin{matrix} 0 & \cdots & & 0 \\ \vdots & & & \vdots \\ 0 & & & 0 \\ \alpha_{1p1} & \cdots\ \alpha_{1pn_{1p}} & 0 & \cdots\ 0 \end{matrix} \\
\hline
\vdots & \vdots & \cdots & \vdots \\
\hline
\begin{matrix} 0 & \cdots & & 0 \\ \vdots & & & \\ 0 & & & \\ \alpha_{p11} & \cdots\ \alpha_{p1n_{p1}} & 0 & \cdots\ 0 \end{matrix} &
\begin{matrix} 0 & \cdots & & 0 \\ \vdots & & & \\ 0 & & & \\ \alpha_{p21} & \cdots\ \alpha_{p2n_{p2}} & 0 & \cdots\ 0 \end{matrix} &
&
\begin{matrix} 0 & & \\ \vdots & I_{n_p-1} & \\ 0 & & \\ \alpha_{pp1} & \cdots & \alpha_{ppn_p} \end{matrix}
\end{array}\right] \qquad (5.1b)$$

$$\underset{(p\times n)}{C} = \left[\begin{array}{ccc|ccc|c|ccc}
1 & 0 \ldots & 0 & 0 & \ldots & 0 & & 0 & \ldots & 0 \\
0 & & \cdot & 1 & 0 \ldots & 0 & & \cdot & & \\
\cdot & & & 0 & \ddots & \cdot & & \cdot & & \\
\cdot & & \cdot & \cdot & & \cdot & \cdots & 0 & \ldots & 0 \\
0 & \ldots & 0 & 0 & & 0 & & 1 & 0 \ldots & 0
\end{array}\right] \qquad (5.1a)$$

$$\underset{(n\times m)}{B} = \left[\begin{array}{ccc}
g_1(1,1) & \ldots & g_1(1,m) \\
\vdots & & \\
g_{n_1}(1,1) & \ldots & g_{n_1}(1,m) \\
\hline
 & \vdots & \\
\hline
g_1(p,1) & \ldots & g_1(p,m) \\
\vdots & & \vdots \\
g_{n_p}(p,1) & \ldots & g_{n_p}(p,m)
\end{array}\right] \qquad (5.1c)$$

Note that the parameters appearing in A, B, C are precisely the elements of ρ_μ. A and C have a specific structure, specified by the structure indices n_i, while B is fully parametrized. Any state space representation of a n-th order system with Kronecker indices $(n_1, \ldots, n_p)$ can be transformed to this canonical form by a similarity transformation : see e.g. Guidorzi (1981).

A canonical MFD form P(z), Q(z) such that K(z) = $P^{-1}(z)Q(z)$ is obtained as follows from ρ_μ.

Let $\underset{p\times p}{P(z)} \triangleq [p_{ij}(z)]$ and $\underset{p\times m}{Q(z)} \triangleq [q_{ij}(z)]$. Then

$$p_{ii}(z) = z^{n_i} - \alpha_{iin_i} z^{n_i - 1} - \ldots - \alpha_{ii1} \qquad (5.2a)$$

$$p_{ij}(z) = -\alpha_{ijn_{ij}} z^{n_{ij} - 1} - \ldots - \alpha_{ij1} \quad \text{for } i \neq j \qquad (5.2b)$$

$$q_{ij}(z) = \beta_{ijn_i} z^{n_i - 1} + \beta_{ij1} \qquad (5.2c)$$

where the coefficients β_{ijk} are bilinear functions of the coefficients α_{ijk} and $g_k(i,j)$ obtained as follows. Let

$$\underset{(n\times m)}{G} \triangleq \begin{bmatrix} G_1 \\ \vdots \\ G_p \end{bmatrix} \qquad \underset{(n_i\times m)}{G_i} \triangleq \begin{bmatrix} \beta_{i11} & \ldots & \beta_{im1} \\ \vdots & & \\ \beta_{i1n_i} & \ldots & \beta_{imn_i} \end{bmatrix} \qquad (5.3)$$

$$\underset{(n\times n)}{M} \triangleq [M_{ij}] \quad i,j = 1, \ldots, p \qquad (5.4)$$

with

$$\underset{(n_i\times n_i)}{M_{ii}} \triangleq \begin{bmatrix}
-\alpha_{ii2} & -\alpha_{ii3} & \ldots & -\alpha_{iin_i} & 1 \\
-\alpha_{ii3} & \ldots & \ldots & 1 & 0 \\
\vdots & & ⋰ & & \vdots \\
-\alpha_{iin_i} & 1 & & & \vdots \\
1 & 0 & \ldots & \ldots & 0
\end{bmatrix} \qquad (5.5a)$$

$$\underset{(n_i\times n_j)}{M_{ij}} \triangleq \begin{bmatrix}
-\alpha_{ij2} & -\alpha_{ij3} \ldots & -\alpha_{ijn_{ij}} & 0 \ldots & 0 \\
-\alpha_{ij3} & & ⋰ & & \vdots \\
\vdots & ⋰ & & & \vdots \\
-\alpha_{ijn_{ij}} & & & & \vdots \\
0 & & & & \vdots \\
\vdots & & & & \vdots \\
0 & \ldots & \ldots & \ldots & 0
\end{bmatrix} \qquad (5.5b)$$

Then :

$$G = MB \qquad (5.6)$$

See Guidorzi (1981) for a proof and for further properties of this form. In an identification context, once the Kronecker indices have been estimated, the structure of P(z) and Q(z) is completely specified by (5.2) ; the α_{ijk} and β_{ijk} are free parameters, whose number is again $d(\mu)$.

The main disadvantage of canonical forms is that one has to estimate the Kronecker indices. Hannan and Kavalieris (1984) have proposed a consistent procedure for the estimation of the n_i. An alternative is to use pseudocanonical forms : this requires the estimation of only one integer-valued parameter, the order n.

5.2. Pseudocanonical forms

If $G(z) \in S(n)$, then $G(z) \in U_\mu$ for almost any μ such that $|\mu| = n$ and for each such μ it is represented uniquely by a coordinate vector τ_μ : see (4.4). A pseudocanonical SS form for G(z) is obtained from τ_μ by taking C as in (5.1a), B as in (5.1c) and

$$\underset{(n\times n)}{A} \triangleq [A_{ij}] \text{ with } \dim A_{ij} = n_i \times n_j \qquad (5.7)$$

where the A_{ii} are as in (5.1b) and where

$$\underset{(n_i\times n_j)}{A_{ij}} \begin{bmatrix}
0 & \ldots & 0 \\
\vdots & & \vdots \\
0 & \ldots & 0 \\
\alpha_{ij1} & \ldots & \alpha_{ijn_j}
\end{bmatrix} \qquad (5.8)$$

See e.g. Gevers and Wertz (1984) for a derivation of this form. The free parameters are exactly the coordinates of τ_μ, which uniquely describe the system in the coordinate system defined by μ. Hence this form is identifiable.

A pseudocanonical MFD form for P(z), Q(z) is obtained from the coordinates of τ_μ (Compare with (5.2)-(5.6)).

$$p_{ii}(z) = z^{n_i} - \alpha_{iin_i} z^{n_{i-1}} - \ldots - \alpha_{ii1} \qquad (5.9a)$$

$$p_{ij}(z) = -\alpha_{ijn_j} z^{n_{j-1}} - \ldots - \alpha_{ij1} \quad \text{for } i \neq j \qquad (5.9b)$$

$$q_{ij}(z) = \beta_{ij\rho_i} z^{\rho_{i-1}} + \ldots + \beta_{ij1} \qquad (5.9c)$$

where $\rho_i \triangleq$ i-th row degree of $P(z) = \max\,(n_i, \max_j \{n_j\} - 1)$. The bilinear relations (5.6) between the β_{ijl} and $g_l(i,j)$ still hold, with B as before, but with

$$\underset{(\rho_i x m)}{G_i} \triangleq \begin{bmatrix} \beta_{i11} & \cdots & \beta_{im1} \\ \vdots & & \vdots \\ \beta_{i1\rho_i} & \cdots & \beta_{im\rho_i} \end{bmatrix} \qquad (5.10)$$

$$\underset{(\rho_i x n_i)}{M_{ii}} \triangleq \begin{array}{r} \\ \\ \\ \\ \\ \rho_i - n_i \Bigg\{ \\ \\ \end{array} \left[\begin{array}{cccc} -\alpha_{ii2} & \cdots & -\alpha_{iin_i} & 1 \\ \vdots & & \iddots & 0 \\ -\alpha_{iin_i} & & & \\ 1 & & & \\ 0 & & & \\ \vdots & & & \\ 0 & \cdots & \cdots & 0 \end{array}\right]$$

$$\underset{(\rho_i x n_j)}{M_{ij}} \triangleq \begin{array}{r} \\ \\ \\ \\ \rho_i - n_j + 1 \Bigg\{ \\ \\ \end{array} \left[\begin{array}{cccc} -\alpha_{ij2} & \cdots & -\alpha_{ijn_j} & 0 \\ \vdots & & \iddots & \\ -\alpha_{ijn_j} & & & \\ 0 & & & \\ \vdots & & & \\ 0 & \cdots & \cdots & 0 \end{array}\right] \qquad (5.11)$$

The parameters β_{ijl} are not all free here, and the number of elements in G is larger than in B : see e.g. Correa and Glover (1984) for details.

The choice between using either the canonical forms or the pseudocanonical forms described in this section is a fairly subjective one. Canonical forms may have slightly fewer parameters leading to more efficient estimates ; on the other hand the structure estimation step of the identification requires the estimation of p structure indices while, with pseudocanonical forms, only the order must be estimated. In practice, however, one starts off the identification procedure by assuming a generic set of structure indices, i.e. μ is such that $n_1 = n_2 = n_q = n_{q+1} + 1 = \ldots = n_p + 1$ for some q. With this generic μ, $U_\mu = V_\mu$. If the parameter estimation algorithm exhibits numerical ill-conditioning problems, one can then perform a coordinate transformation to a better-conditioned pseudo-canonical form without loosing the information already gained : see Van Overbeek and Ljung (1982).

6. COMMENTS AND CONCLUSIONS

We have given a very brief overview of the issues involved in the parametrization of linear systems. The problem of parametrization is intimately related to that of identifiability. We have attempted to make a clear distinction between the case where specific and physically meaningful parameters in an exact model description must be identified (parameter identifiability) and the case where an identifiable parametrized black box model must be selected to represent or approximate a system (structural identifiability).

Most efforts so far have been spent on solving the latter problem, which is by no means trivial because the structure of S(n), the set of systems of order n, is fairly complicated : the set of systems in S(n) cannot be described by a single parametrization. In this brief tutorial introduction we have tried to explain the structure of S(n) in simple terms. We have illustrated the canonical and pseudocanonical forms, and their connections with the structure of S(n), by two models : SS and MFD models. Other parametrizations are possible : in particular, much recent work has been done on the degree properties (Gevers, 1986) and the parametrization (Deistler and Gevers, 1987) of monic ARMA models. Many other problems have not been touched on in this brief tutorial paper, the most important being structure and order estimation procedures. We refer to Gevers and Wertz (1987) for a comprehensive survey and to Hannan and Kavalieris (1984) and Deistler (1986) for technical details.

REFERENCES

Clark J.M.C. (1976), "The consistent selection of parametrizations in system identification", JACC, Purdue University.

Correa G.O. and K. Glover (1984), "Pseudo-canonical forms, identifiable parametrizations and simple parameter estimation for linear multivariable systems", Automatica, vol.20, pp. 443-452.

Deistler M. (1986), "Multivariate time series and linear dynamic systems", Advances Stat. Analysis and Stat. Computing, vol.1, pp. 51-85.

Deistler M. and Gevers (1987), "Properties of the parametrization of monic ARMA systems", submitted for publication.

Gevers M. (1986), "ARMA models, their Kronecker indices and their McMillan degree", Int. J. Control, vol. 43, N°6, pp. 1745-1761.

Gevers M. and V. Wertz (1984), "Uniquely identifiable state-space and ARMA parametrizations for multivariable linear systems", Automatica, vol.20, pp. 333-348.

Gevers M. and V. Wertz (1987), "Techniques for the selection of identifiable parametrizations for multivariable linear systems", to appear in 'Control and Dynamic Systems, vol.XXIV', C.T. Leondes Ed., Academic Press.

Godfrey K.R. and J.J. Distefano III (1985), "Identifiability of model parameters", Proc. 7th IFAC Symp. on Identification and System Parameter Estimation, York (U.K.), pp. 89-114.

Guidorzi R.P. (1981), "Invariants and canonical forms for systems structural and parametric identification", Automatica, vol.17, pp. 117-133.

Hannan E.J. and L. Kavalieris (1984), "Multivariate linear time series models", Adv. Applied Probability, vol.16, pp. 492-561.

Hazewinkel M. and R.E. Kalman (1976), "On invariants, canonical forms and moduli for linear constant finite-dimensional dynamical systems", in Lecture Notes Econ.-Math. System Theory, vol. 131, pp. 48-60, Springer-Verlag.

Ljung L. (1987), "System Identification : theory for the user", Prentice Hall.

Van Overbeek A.J.M. and L. Ljung (1982), "On line structure selection for multivariable state-space models", Automatica, vol.18, pp. 529-544.

Walter E. (1982), "Identifiability of state space models", Lecture Notes in Biomathematics N°46, Springer Verlag.