Monograph:
Leontaritis, I.J. and Billings, S.A. (1984) Input-Output Parametric Models for Nonlinear Systems. Part 1- Deterministic Nonlinear Systems. Research Report. Acse Report 251 . Dept of Automatic Control and System Engineering. University of Sheffield

[redacted]

INPUT - OUTPUT PARAMETRIC

MODELS FOR NONLINEAR

SYSTEMS

PART I - DETERMINISTIC NONLINEAR SYSTEMS

by

I.J.Leontaritis B.Sc., M.Eng., Ph.D.

and

S.A.Billings B.Eng., Ph.D., C.Eng. MIEE,
AFIMA, M.InstMC

Department of Control Engineering

Abstract

Recursive input-output models for nonlinear multivariable discrete-time systems are derived and sufficient conditions for their existence are defined. The paper is divided into two parts. The first part introduces and defines concepts such as Nerode realization, multistructural forms and results from differential geometry which are then used to derive a recursive input-output model for multivariable deterministic nonlinear systems. The second part introduces several examples, compares the derived model with other representations and extends the results to create prediction error or innovation input-output models for nonlinear stochastic systems. These latter models are the generalization of the multivariable ARMAX models for linear systems and are referred to as NARMAX or Nonlinear AutoRegressive Moving Average models with eXogenous inputs.

1. Introduction

The identifications and digital control of linear systems is largely based on the linear difference equation model which relates sampled output signals to sampled inputs. Numerous parameter estimation routines and controller synthesis procedures [Goodwin and Payne (1977), Astrom and Eykhoff (1970).] have been developed based on this description which provides a concise representation of both the process and the feedback controller. When the system is nonlinear however the traditional system descriptions are based on functional series such as the Volterra [Volterra (1930)] or Wiener [Wiener(1982)] series [Billings (1980).] Whilst these provide an adequate representation for a wide class of nonlinear systems several hundred parameters are often required to characterise even simple nonlinear systems. The excessive computational effort required to estimate the unknown parameters, the difficulty of interpreting the results and the necessity to use special input signals are further disadvantages of functional series methods. The usefulness of these system descriptions for identification and control purposes is therefore limited and alternative representations are required.

Recently several authors [Brockett (1976), Herman and Krener (1977) Sussman (1977), Crouch (1979), (1981), Jakubczyki (1980), Sontag (1979a), Fliess and Normand-Cyrot (1972)] have studied the realization problem for nonlinear systems and numerous interesting results have been obtained. However little work has been directed towards deriving input-output models for nonlinear discrete systems. The wide application of linear difference equations makes it natural to search for nonlinear difference equation models that can be used to represent general nonlinear systems. The development of such models could provide a class of models which could form the basis for the development of identification and digital controller design techniques for nonlinear systems. With the exception of Sontag's work [Sontag (1970), (1979a), (1979b)] in this field little has been achieved.

In the present study recursive input-output models for both deterministic and stochastic nonlinear multivariable discrete time systems are derived. Conditions for their existence are provided which can be given a simple physical interpretation. The recursive input-output models are valid only in a restricted region of operation around the equilibrium point. The determination of such a region of operation is also given. The single-input single-output recursive nonlinear input-output models were first derived heuristically and recursive estimation methods were developed for them in Billings and Leontaritis [1981, 1982]. Alternative least squares algorithms based on these models are derived in Billings and Voon [1984] and structure detection and model validation techniques have been developed in Billings and Voon [1983]. Several examples which illustrate the derivation of the models are included and a comparison is made with other models for nonlinear systems especially with the globally valid input-output models developed by Sontag.

The deterministic models are used to create prediction error or innovation input-output models for nonlinear stochastic systems. These models are the generalisation of the multivariable ARMAX or controlled ARMA models of linear stockastic systems [Goodwin and Payne (1977)]. Consequently the nonlinear models are referred to as the NARMAX or Nonlinear AutoRegressive Moving Average model with eXogenous inputs.

The paper is split into two parts. The first part introduces results from system theory such as Nerode realization, multistructural forms and concepts from differential geometry which are used to derive linear multivariable input-output models which are then generalized to provide a recursive input-output model for multivariable deterministic nonlinear systems. The second part [Leontaritis and Billings, 1984] introduces several examples, compares the derived model with other representations and extends the results to create prediction error models for nonlinear stochastic systems.

2. Basic Concepts

Consider the discrete-time, time invariant system S

$$x(t+1)=g[x(t),u(t)]$$

$$y(t)=h[x(t),u(t)] \tag{2.1}$$

where $t \in Z$ is the set of integers,

$x(t+1), x(t) \in X$ is the state set,

$u(t) \in U$ is the input set, of dimension r,

$y(t) \in Y$ is the output set of dimension m

g: $ZxXxU \to X$ is the one step ahead state-transition function and

h: $ZxXxU \to Y$ is the output function

The many step ahead state-transition function Φ can be found by repeated applications of the function g.

$$x(t+2)=g[x(t+1),u(t+1)] = g[g[x(t), u(t)],u(t+1)]$$

$$=\Phi[x(t),u(t+1),u(t)]$$

$$x(t+3)=g[x(t+2),u(t+2)]=g[g[g[x(t),u(t)],u(t+1)],u(t+2)]$$

$$=\Phi[x(t),u(t+2),u(t+1),u(t)]$$

$$x(t+4)= \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \tag{2.2}$$

Let U* be the set of all sequences of members of the set U

$$U^* = \{u_{k-1} \ldots u_1 u_0 | k \geq 0 \text{ and } u_j \in U,\ j=0,1,2,\ldots ,k\} \tag{2.3}$$

The empty sequence denoted by e is included in U* and corresponds to k=0. The input sequences will be allowed to be concatenated one after the other so that if $w_1, w_2 \in U^*$, their concatenation will then be denoted $w_1 w_2 \in U^*$. The set of all sequences minus the empty sequence e is the set $U+ = U^* - \{e\}$

The length of a sequence w will be denoted by $|w|$.

The state-transfer function can now be defined as

$$\Phi: \quad XxU^* \to X \tag{2.4}$$

so that if $u(t+j)=u_j$, $j=0,1,2,\dots,k-1$ and $w=u_{k-1}\dots u_1u_0$ then

$$x(t+k)=\Phi[x(t),w] \quad \text{for } k>0 \text{ and } x(t)=\Phi[x(t),e] \text{ for } k=0$$

Note that an input sequence $w=u_{k-1}\dots u_1u_0$ is written contrary to the usual way from right to left, so that the input latest in time. u_{k-1}, appears first in the sequence and the most remote in the past, u_0, appears last in the sequence. This symbolism is preferred because it is more convenient for later use.

The function of primary importance in system theory is the function that describes the input-output behaviour of the system since this is all an external observer can see. When the system is at the state x_o the behaviour of the system from that state can be described by a function f_{x_o} called the input-output map or the response function of the system and it is defined as

$$f_{x_o} : U^+ \to Y \tag{2.5}$$

where

$$f_{x_o}(w)=h[\Phi[x_o,u_{k-1}\dots u_1u_0],u_k] \tag{2.6}$$

and $\quad w=u_ku_{k-1}\dots u_1u_0 \in U+$

A state x_o is an equilibrium state if there exists an input u_o such that $g[x_o,u_o]=x_o$. A constant input sequence w such that $w=u_ou_o\dots u_o$ will leave the state x_o unchanged i.e. $\Phi[x_o,w]=x_o$. When a system is at its equilibrium state x_o and the constant input sequence w is applied, the output takes the constant value y_o where $y_o=h[x_o,u_o]$. The equilibrium state x_o is a state where the system is at rest for an input sequence that has a constant value of u_o. It is assumed that every system has at least one equilibrium state. When the sets U and Y have the structure of a vector space, a change of the origin of the co-ordinate system in U and Y can transfer the input u_o to the origin of U and the output y_o to the origin of Y. It will henceforth be assumed that such a transformation has been done and the origin of U will be called a zero input and the origin of Y will be called a zero output. A sequence W that consists of k zero inputs will be called a zero input sequence and it will be denoted 0^k.

A state x is called reachable from another state x_o if there exists an input sequence $w \in U^*$ such that $x=\Phi[x_o,w]$. A system S is called reachable from a state x_o if every state $x \in X$ is reachable from the state x_o.

Two states x_1 and x_2 are called indistinguishable if for any input sequence $w \in U+$ it holds $f_{x_1}(w)=f_{x_2}(w)$

Thus when the system is at state x_1, the output from the system when an input sequence w is applied, is the same as the output when the system is at state x_2 and the same input sequence is applied. Consequently no input-output experiment can determine which of the two states the system is at.

A system S is called observable when there are no indistinguishable states in the state set X of the system.

The theoretical basis of identification theory is realization theory.

The function that summarizes all the future input-output behaviour of a system in state x_o is the response function f_{x_o}. This is the maximum information an observer can get from measurements on the system. The realization problem then is the following:
Given a function F: U+→Y find a system S such that it has a particular state x_o for which the response function of the system f_{x_o} is equal to the function F.

This system S is a model that from an input-output point of view behaves identically to the function F. There might be many model systems S but out of all of these one must be simpler than the others or minimal in some particular way. The sense in which a system is considered minimal is quite important in the theory.

The most general solution of the realization problem is a method called Nerode realization [Arbib (1968)], [Padulo and Arbib (1974)]. The criterion for minimality is that the model system should be reachable from x_o and observable.

The basic idea behind Nerode realization is that all input sequences that bring a system S from the state x_o to the same state x can be considered to belong to an equivalence class and any one of them can represent the state x. If the system S is observable the way to check that two input sequences w_1 and w_2 are equivalent in the above way is to make sure that the behaviour of the system after the two sequences have been applied is identical since this proves that the system has been driven to the same state. The sequences w_1 and w_2 are thus equivalent if

$$f_{x_o}(ww_1)=f_{x_o}(ww_2) \qquad \text{for all} \quad w \in U+$$

Since the aim of Nerode realization is to construct a reachable and observable system, this idea can work backwards.

A set equivalence relationship E can be defined on the set U* according to the function F. Two equivalent sequences w_1 and w_2 will be symbolized as w_1Ew_2. Two sequences w_1 and w_2 are then equivalent if and only if

$$w_1Ew_2 \quad \leftrightarrow \quad F(ww_1)=F(ww_2) \quad \text{for all } w \in U+$$

The equivalence class a sequence w belongs to will be symbolized as $[w]$, thus

$$[w] = \{w' \in U^* \mid w'Ew\}$$

The state space of the Nerode realized system S is the quotient set $x=U^*/E$

The one step ahead state-transition function g and the output function h are given as $g: \; X \times U \rightarrow X \qquad h: \; X \times U \rightarrow Y$

where

$$g([w],u)=[uw]$$

$$h([w],u)=F(uw) \tag{2.7}$$

where w can be any sequence of the class $[w] \in X$. It can easily be seen that these two functions are properly defined, that is, their definition is independent of the choice of the sequence w in $[w]$. The state-transition function is

$$\Phi[[w_1],w_2]=[w_2w_1] \tag{2.8}$$

The state $[e]$ which is the equivalence class of the empty sequence e will be the state x_o. The Nerode realized system S has as an input-output map from the state x_o, the given function F. In fact let w be any non-empty sequence that belongs to U+. The sequence w can be written as

$$w=uw_1 \qquad \text{where } u \in U,\ w_1 \in U^*$$

Then

$$
\begin{aligned}
f_{x_o}(w) &= f_{x_o}(uw_1) \\
&= h(\Phi(x_o,w_1),u) && \text{from the definition of the response function (2.6)} \\
&= h(\Phi([e],w_1),u) && \text{since } x_o=[e] \\
&= h([w_1e],u) && \text{from (2.8)} \\
&= h([w_1],u) && \text{since e is the empty sequence} \\
&= F(uw_1) && \text{from (2.7)} \\
&= F(w) && (2.9)
\end{aligned}
$$

The system S has therefore response function from the state x_o equal to the given function F. It is completely reachable from x_o because any state x is an equivalence class of input sequences and any member of this class drives the system from x_o to the state x. It is observable because any two distinct states always have different responses from these states. In fact let two distinct states x_1 and x_2 be $x_1=[w_1]$ and $x_2=[w_2]$. Assume that they are indistinguishable, that is $f_{x_1}(w)=f_{x_2}(w)$ for all $w \in U+$. Then

$$
f_{x_1}(w)=f_{x_2}(w) \rightarrow f_{x_o}(ww_1)=f_{x_o}(ww_2) \rightarrow F(ww_1)=F(ww_2) \rightarrow [w_1]=[w_2] \rightarrow x_1=x_2 \qquad (2.10)
$$

a contradiction which proves that the states x_1 and x_2 are distinguishable.

3. Realization and input-output models of linear systems

In this section several results from the theory of linear systems will be reviewed. These results will be needed as background for generalisation purposes in the field of non-linear system theory.

A discrete-time, time-invariant linear system can be defined as

$$
\begin{aligned}
x(t+1) &= g[x(t),u(t)]=Ax(t)+Bu(t) \\
y(t) &= h[x(t),u(t)]=Cx(t)+Du(t)
\end{aligned} \qquad (3.1)
$$

Where U, Y and X are vector spaces and $A: X \rightarrow X$, $B: U \rightarrow X$, $C: X \rightarrow Y$, $D: U \rightarrow Y$. Let $x_i=x(t+i)$ $u_i=u(t+i)$ and $y_i=y(t+i)$ for $i=0,1,2,\ldots$. The many step ahead state-transition function then is

$$
x_k=\Phi[x_o,u_{k-1}\ldots u_1u_0]=A^kx_o+\sum_{\ell=0}^{k-1}A^{k-\ell-1}Bu_\ell \quad \text{for } k>0 \qquad (3.2)
$$

and the response function is

$$y_k=f_{x_0}(u_ku_{k-1}\dots u_1u_0)=CA^kx_0+\sum_{\ell=0}^{k-1}CA^{k-\ell-1}Bu_\ell+Du_k \qquad \text{for } k\geq 0 \tag{3.3}$$

The origin of the state space is an obvious equilibrium state of a linear system. The zero input sequence leaves the system at the zero state with output the zero of the output space.

Only finite-dimensional systems will be considered where

$\dim(U)=r \quad \dim(Y)=m \quad \dim(X)=n$

The specialization of Nerode realization to linear systems will be discussed next [Arbib (1968)], [Padulo and Arbib (1974)]. The realization of response functions from the zero state only will be studied since it can be assumed that the system was at the zero state at some point in time. However, the realization from a non-zero state can be done using similar arguments. The zero state response function will be denoted by f for notational convenience.

The zero state response function of a linear system is given by (3.3) for $x_0=0$

$$y_k=f(u_ku_{k-1}\dots u_1u_0)=H_0u_K+H_1u_{k-1}+\dots+H_ku_0=\sum_{\ell=0}^{k}H_\ell u_{k-\ell} \tag{3.4}$$

where $\quad H_0=D \quad$ and $\quad H_\ell=CA^{\ell-1}B \quad$ for $\quad \ell\geq 1$

The linear functions H characterize the zero state response function and are called impulse functions since for an impulse input sequence $0\dots 0u_0$ the output of the system is $y_k=H_ku_0$. For some ordered basis of the vector spaces U and Y the impulse functions H_ℓ are represented by matrices of dimension mxr called impulse matrices.

The specific problem of realization of a linear zero state response function is the following:

Given the impulse functions H_ℓ: $U\to Y$ $\ell\geq 0$ where U and Y are finite dimensional vector spaces and H_ℓ are linear functions, find a vector space X and four linear functions A, B, C, D where A: $X\to X$ B: $U\to X$ C: $X\to Y$ D: $U\to Y$

such that $H_0=D$ and $H_\ell=CA^{\ell-1}B$ for $\ell\geq 1$.

In other words find a linear system that has as impulse functions the given functions H_ℓ. The Nerode realization gives a general solution to the problem but the special form of the response function has to be taken into account to prove that the Nerode realized system is a linear system.

The zero state response function of a linear system can be intuitively seen to be a linear function but to be precise the space U^* must first be given the structure of a vector space. In order to do so addition between members of U^* and multiplication with scalars have to be defined first. Component-wise multiplication is the obvious choice for the multiplication. For addition sequence w can be post-loaded with zeros without altering the imate it has through the response function i.e. it holds that $f(w)=f(w0^k)$ for $k\geq 0$

This obviously happens just because the response function is a zero-state response function and any zero input sequence leaves it at the zero state. An equivalence relationship P can thus be defined in U^* such that two sequences w_1 and w_2 are equivalent if and only if there exist two integers k_1 and k_2 such that $w_1 0^{k_1}=w_2 0^{k_2}$. Obviously all the members of one such equivalent class of sequences have the same image through the response function f and thus the quotient space $U^o=U^*/P$ can be substituted as the domain of the response function. Addition in U^o can be easily defined since two sequences w_1 and w_2 can be made to have the same length by post-loading the shorter one with zeros and can thus be added together component-wise. It can be readily proved that the space U^o is then a vector space and that f is a linear function.

The first step of Nerode realization is to group together in equivalent sets all the sequences with the same response and define the quotient space as the state-space. The equivalence relationship E among the members of U^o is thus defined as

$$w_1 E w_2 \leftrightarrow f(ww_1)=f(ww_2) \quad \text{for all } W \in U+ \tag{3.5}$$

where the juxtapositions ww_1 and ww_2 for $w_1, w_2 \in U^o$, $w \in U+$ are also members of U^o. As a consequence of the linearity of f

$$\begin{aligned} f(ww_1)&=f(w)+f(0^{|w|}w_1) \\ f(ww_2)&=f(w)+F(0^{|w|}w_2) \end{aligned} \tag{3.6}$$

and thus

$$f(ww_1)=f(ww_2) \leftrightarrow f(0^{|w|}w_1)=f(0^{|w|}w_2) \tag{3.7}$$

Two sequences are then equivalent when

$$
\begin{aligned}
w_1 E w_2 &\leftrightarrow f(ww_1)=f(ww_2) && \text{for all } w \in U+ \\
&\leftrightarrow f(0^{|w|}w_1)=f(0^{|w|}w_2) && \text{for all } |w| \geq 1 \\
&\leftrightarrow f(0^k w_1)=f(0^k w_2) && \text{for all } k \geq 1
\end{aligned}
$$

Thus in order to check that two sequences are equivalent an infinite zero sequence should be applied after the applications of the two sequences and the response of the system should be checked to be identical. Any arbitrary sequence of infinite length can actually be used as a test sequence but the zero sequence is a common and obvious choice.

A series of equivalence relationships that require only finite length test sequences can now be introduced. Two sequences w_1 and w_2 are k-equivalent if and only if

$$w_1 E_k w_2 \leftrightarrow f(0^{\ell} w_1)=f(0^{\ell} w_2) \qquad \text{for all } 1 \leq \ell \leq k \tag{3.9}$$

It is obvious that the Nerode equivalence relationship E is actually

$$E = E_{\infty}$$

The quotient spaces $Q_k = U^o/E_k$, $k=1,2,\ldots$

can be seen to be linear spaces. In fact let the vector $\langle w\rangle_k$ be defined as

$$\langle w\rangle_k = \begin{vmatrix} f(0w) \\ f(00w) \\ \cdot \\ \cdot \\ \cdot \\ f(0^k w) \end{vmatrix} = L_k(w) \qquad \text{for} \qquad k=1,2,\ \ldots \tag{3.10}$$

The vector $\langle w\rangle_k$ is a member of the vector space Y^k. The function $L_k : U^o \rightarrow Y^k$ that takes the sequence w to $\langle w\rangle_k$ is obviously a linear function since the function f is linear. The equivalence relationship E_k can be alternatively defined as

$$w_1 E_k w_2 \leftrightarrow L_k(w_1)=L_k(w_2) \qquad k=1,2,\ldots \tag{3.11}$$

The quotient space $Q_k = U^o/E_k$ is then nothing else but the factor space $U^o/\mathrm{Ker}(L_k)$ which is a vector space in its own right. Thus

$$Q_k = U^o/\mathrm{Ker}(L_k) \qquad k=1,2,\dots \tag{3.12}$$

The state-space X of the Nerode realization is then the vector space Q that is $X = U^o/\mathrm{Ker}(L_\infty)$

The one step ahead state-transition function and the output function or equivalently the functions A, B, C, D can now be found. Let a member of the state space X be $x=[w]$, where w is some sequence of the equivalence class $[w]$. The one step ahead state-transition function of the Nerode realized system is

$$g(x,u)=g([w],u)=[uw]= [0w+u]=[0w]+[u] \tag{3.13}$$

The functions A and B are defined as

$$\begin{aligned} A &: X\to X : [w] \mapsto [0w] \\ B &: U\to X : u \mapsto [u] \end{aligned} \tag{3.14}$$

It can be seen that the functions are linear and well defined (independent of the choice of the sequence w that represents the state x). The output function of the Nerode realization is

$$h(x,u)=h([w],u) = f(uw) = f(0w+u) = f(0w) + f(u) \tag{3.15}$$

The functions C and D are defined as

$$\begin{aligned} C &: X\to Y : [w] \mapsto f(0w) \\ D &: U\to Y : u \mapsto f(u) \end{aligned} \tag{3.16}$$

The realization of the zero state linear response function has thus been theoretically solved but a more numerical interpretation will be given so that the realization can also be done in practice.

The set U^o can be given a different interpretation as the set of all right-infinite sequences with only a finite number of non-zero elements. All sequences can be made right-infinite by post-loading them with zeros. A sequence $w=u_k u_{k-1}\dots \in U^o$ will be written as an infinite vector

$$w=\left|\begin{matrix} u_k \\ u_{k-1} \\ \cdot \\ \cdot \\ \cdot \end{matrix}\right.$$

The response function can then be written as

$$y(t)=H_0u(t)+H_1u(t-1)+H_2u(t-2)+ \ldots = \sum_{\ell=0}^{\infty} H_\ell u(t-\ell)=$$

$$= \begin{bmatrix} H_0 & H_1 & H_2 & \ldots \end{bmatrix} \begin{vmatrix} u(t) \\ u(t-1) \\ u(t-2) \\ \cdot \\ \cdot \\ \cdot \end{vmatrix} \qquad (3.17)$$

where the sum is well defined since only a finite number of terms are non-zero.

The elements $f(0w)$, $f(00w)$, ..., $f(0^k w)$ of the vector $<w>_k$ are actually the output of the system after the application of the input sequence w, followed by the application of a number of zero inputs. For the sequence $w=u(t-1)u(t-2)u(t-3)\ldots$

the output $f(0w)$ will be denoted by $y(t|t-1)$ and it is

$$y(t|t-1)=H_1u(t-1)+ \ldots = \sum_{\ell=1}^{\infty} H_\ell u(t-\ell) \qquad (3.18)$$

In general

$$y(t|t-1)=H_k u(t-k)+H_{k+1}u(t-k-1)+ \ldots = \sum_{\ell=k}^{\infty} H_\ell u(t-\ell) \quad k \geq 0 \qquad (3.19)$$

When the realization problem is interpreted as the realization of a stochastic system, the vector $y(t|t-k)$ can be considered to be the best prediction of $y(t)$ given all the inputs up to time t-k [Akaike (1974a)]. This is the reason for the present symbolism. The stochastic interpretation of the realization problem seems to be more natural to people not familiar with automata theory and the Nerode realization. Such a view of the realization problem does not actually lead to any new results and it has no other advantage apart from providing a familiar mental picture for people with a statistical background.

The functions $L_k : U^0 \to Y^k$ are given by

$$L_k(w) = \begin{bmatrix} y(t|t-1) \\ y(t+1|t-1) \\ \cdot \\ \cdot \\ \cdot \\ y(t+k-1|t-1) \end{bmatrix} = \begin{bmatrix} H_1 & H_2 & H_3 & \cdots \\ H_2 & H_3 & H_4 & \cdots \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ H_k & H_{k+1} & H_{k+2} & \cdots \end{bmatrix} \begin{bmatrix} u(t-1) \\ u(t-2) \\ u(t-3) \\ \\ \cdot \\ \cdot \end{bmatrix} \qquad (3.20)$$

Given ordered bases for the input and output vector spaces, the impulse functions H_ℓ are represented by the impulse matrices and the linear functions L_k are represented by semi-infinite matrices, the Hankel matrices $\mathcal{H}_{k,\infty}$

$$\mathcal{H}_{k,\infty} = \begin{bmatrix} H_1 & H_2 & H_3 & \cdots \\ H_2 & H_3 & H_4 & \cdots \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ H_k & H_{k+1} & H_{k+2} & \cdots \end{bmatrix} \qquad (3.21)$$

Finite Hankel matrices $\mathcal{H}_{k,l}$ can be defined similarly as

$$\mathcal{H}_{k,l} = \begin{bmatrix} H_1 & H_2 & \cdots & H_l \\ H_2 & H_3 & \cdots & H_{l+1} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ H_k & H_{k+1} & \cdots & H_{k+l-1} \end{bmatrix} \qquad (3.22)$$

The Hankel matrix, because of its special structure, has several useful properties

If for some k

rank $(\mathcal{H}_{k,\infty})$=rank $(\mathcal{H}_{k+1,\infty})$

then

(1) rank $(\mathcal{H}_{k,\infty})$=rank$(\mathcal{H}_{k',\infty})$ for all $k' \geq k$

(2) rank $(\mathcal{H}_{k,k})$=rank $(\mathcal{H}_{k,\infty})$

Thus for such a k the matrix $\mathcal{H}_{k,k}$ has the highest possible rank any Hankel matrix $\mathcal{H}_{k',k,"}$ can have.

A response function that is realized by a linear system with finite dimensional state-space is called finitely realizable. The dimension of the state-space is called the order of the linear system. The properties of the Hankel matrix can be exploited in the realization of the finitely realizable response functions.

The factor spaces Q_k in (3.12) have dimension equal to rank (L_k)

$$\dim(Q_k)=\mathrm{rank}(L_k)=\mathrm{rank}(\mathcal{H}_{k,\infty}) \qquad k=1,2,\ldots \tag{3.24}$$

If a response function is finitely realizable the state-space $X=Q_\infty$ has finite dimension equal to n. Then from (3.24) it holds

$$\dim(Q_\infty)=\mathrm{rank}(\mathcal{H}_{\infty,\infty})=n$$

The condition then for a response function to be finitely realizable is

$$\max(\mathrm{rank}\ \mathcal{H}_{k,1})=n \qquad \text{for any} \quad \begin{array}{l} k=1,2,\ldots \\ 1=1,2,\ldots \end{array}$$

Property (1) of (3.23) can be exploited to prove that for a finitely realizable response function of order n the state-space is also given by

$$X=Q_n=U^o/(L_n) \tag{3.25}$$

From (3.25) and (3.24)

$$\dim(\mathcal{H}_{n,\infty})=n \tag{3.26}$$

Property (2) of (3.23) can be used so that (3.26) also gives $\dim(\mathcal{H}_{n,n})=n$

The state-space of the Nerode realization is given by the factor space $U^o/\mathrm{Ker}(L_n)$. A basis for this space has to be found so that the state can be represented by a column vector. If a rule is provided for choosing a unique basis for the state-space, the realization is called canonical, since for such a basis, the matrices A,B,C,D of the linear system of the Nerode realization are unique. A method that can be used to determine a basis for a vector space is to find another vector space, isomorphic to the original one, that has a naturally occuring basis. The basis chosen for the original space is the basis of the isomorphic space mapped by the isomorphism back to the original space. The advantage of this choice is that two isomorphic vectors have exactly the same column vector representation for the isomorphic bases. In essence, the original space can be forgotten and everything can be done in relation to the isomorphic space.

The factor space $U^o/\mathrm{Ker}(L_n)$ is isomorphic to the space image (L_n) [Wonham (1974)]. This space can thus be thought os as the state-space of the Nerode realization. There are two types of bases that can be easily defined for the space image (L_n). The first is any set of independent column vectors of the matrix representation of the linear function L_n, $\mathcal{H}_{n,\infty}$. The state-space models in the controllable canonical form can be created by this type of bases. The description of such realization is given in [Padulo and Arbib (1974)]. Another type of bases will be considered here. The state-space models in the observable canonical form can be created by this type of bases. The reason for such a choice is that they also lead to input-output models that can be easily used in identification.

The rank of the linear function L_n or equivalentnly the rank of the Hankel matrix $\mathcal{H}_{n,\infty}$ is equal to n. Thus n independent rows of this matrix can be selected and the rest of them will be linearly dependent on them. If the linearly dependent rows are deleted, the linear function represented by the matrix that consists of the linearly independent rows only is called L_n^{*}. The space Image (L_n^*) is obviously isomorphic to Image (L_n), so the state-space can be identified with Image (L_n^*). Choosing the standard basis for this space, a column state vector consists of the elements of the column vector.

$$\begin{bmatrix} y(t|t-1) \\ y(t+1|t-1) \\ \cdot \\ \cdot \\ \cdot \\ y(t+n-1|t-1) \end{bmatrix} \qquad (3.27)$$

that correspond to the independent rows of $\mathcal{H}_{n,\infty}$. The state vector x(t) is thus given by

$$x(t)=R \begin{bmatrix} u(t-1) \\ u(t-2) \\ \cdot \\ \cdot \\ \cdot \end{bmatrix} \tag{3.28}$$

Where R is the $n \times \infty$ matrix of the function L^*_n, i.e. the matrix that consists of the chosen independent rows of $\mathcal{H}_{n,\infty}$. The vector x(t+1) is then equal to

$$x(t+1)=R \begin{bmatrix} u(t) \\ u(t-1) \\ \cdot \\ \cdot \\ \cdot \end{bmatrix} = \begin{bmatrix} B & R' \end{bmatrix} \begin{bmatrix} u(t) \\ u(t-1) \\ \cdot \\ \cdot \\ \cdot \end{bmatrix}$$

$$=R' \begin{bmatrix} u(t-1) \\ u(t-2) \\ \cdot \\ \cdot \\ \cdot \end{bmatrix} + Bu(t) \tag{3.29}$$

Where B is an nxr dimensional matrix that consists of the first r columns of the matrix R.

Every row of the matrix R' is a row of the matrix R with the first r elements removed. Every row of the matrix R is actually some row of the matrix $\mathcal{H}_{\infty,\infty}$. Every row of the matrix R' is thus some row of the matrix $\mathcal{H}_{\infty,\infty}$ with the first r elements removed. Let the i^{th} row of $\mathcal{H}_{\infty,\infty}$ have the first r elements removed. This derived row is, because of the structure of the Hankel matrix, the $(i+m)^{th}$ row of the matrix $\mathcal{H}_{\infty,\infty}$. Consequently every row of R' is a row of the matrix $\mathcal{H}_{\infty,\infty}$. Since every row of $\mathcal{H}_{\infty,\infty}$ is linearly dependent on the rows of matrix R, every row of R' is a linear combination of the rows of the matrix R. It then holds that

$$R' = A\,R \tag{3.30}$$

where A is an nxn matrix. From (3.29) and (3.30)

$$x(t+1) = A\,R \begin{bmatrix} u(t-1) \\ u(t-2) \\ \cdot \\ \cdot \\ \cdot \end{bmatrix} + B\,u(t) \tag{3.31}$$

and using (3.28)

$$x(t+1) = A\,x(t) + B\,u(t) \tag{3.32}$$

From (3.17) it holds

$$y(t) = \begin{bmatrix} H_0 & H_1 & \cdots \end{bmatrix} \begin{bmatrix} u(t) \\ u(t-1) \\ \cdot \\ \cdot \\ \cdot \end{bmatrix} = \begin{bmatrix} H_1 & H_2 & \cdots \end{bmatrix} \begin{bmatrix} u(t-1) \\ u(t-2) \\ \cdot \\ \cdot \\ \cdot \end{bmatrix} + H_0 u(t) \tag{3.33}$$

But the matrix $[H_1\ H_2\ \ldots]$ consists of the first r rows of the Hankel matrix $\mathcal{H}_{\infty,\infty}$ which are some linear combination of the rows of matrix R. Thus

$$\begin{bmatrix} H_1 & H_2 & \cdots \end{bmatrix} = C\,R \tag{3.34}$$

and (3.33) becomes

$$y(t) = C\,R \begin{bmatrix} u(t-1) \\ u(t-2) \\ \cdot \\ \cdot \\ \cdot \end{bmatrix} + H_0\,u(t) \tag{3.35}$$

and using the definition of x(t) in (3.28)

$$y(t) = C\,x(t) + D\,u(t) \tag{3.36}$$

where $D=H_0$

The state-space representation is thus complete.

The independent rows of the matrix $\mathcal{H}_{n,\infty}$ which constitute a row basis were chosen in an arbitrary way. If however a particular rule is followed in the selection of the row basis, the realized state-space model has some nice properties. The rule is the following:

If the i^{th} row of the matrix $\mathcal{H}_{n,\infty}$ is chosen tobelong to the row basis, the $(i-m)^{th}$ row must also belong to the row basis provided that i-m is positive.

This rule is easy to understand if the rows of the matrix $\mathcal{H}_{n,\infty}$ are grouped together in groups of m rows starting from the first row. The selection rule then becomes:

If the i^{th} row of some m-group of rows belongs to the basis, then all the i^{th} rows of the above m-groups must also belong to the basis.

The selection of a basis when such a rule is followed produces a state-space with a smaller number of parameters than when the rule is not followed. In addition, all the parameters are independent so that different sets of different parameters correspond to different systems.

Let the chosen row basis have n_1 rows from the set of rows which are first in every m-group of rows, n_2 rows from the set of rows which are second in every m-group of rows, etc., n_m rows from the set of rows which are m^{th} in every m-group. Obviously

$$n_1+n_2+\dots+n_m=n$$

The integers $n_1, n_2, \dots, n_m$, will be called observability indices. For every specific choice of observability indices, the state-space model takes a specific form which is called an observable multi-structural form. The state-vector can be arranged to be equal to the column vector

$$x(t)=\begin{bmatrix} y_1(t|t-1) \\ \cdot \\ \cdot \\ \cdot \\ y_1(t+n_1-1|t-1) \\ \cdot \\ \cdot \\ \cdot \\ y_m(t|t-1) \\ \cdot \\ \cdot \\ \cdot \\ y_m(t+n_m-1|t-1) \end{bmatrix} \tag{3.37}$$

It can easily be seen that for such a choice of a row basis, the state transition matrix A takes a very specialized form. For the state vector (3.37), the matrix A has the multi-companion form. For instance, for m=3 and $n_1=3$, $n_2=2$, $n_3=4$, it has the form

$$\begin{array}{cl} \left[\begin{array}{ccccccccc} 0&1&0&0&0&0&0&0&0\\ 0&0&1&0&0&0&0&0&0\\ x&x&x&x&x&x&x&x&x\\ \hline 0&0&0&0&1&0&0&0&0\\ x&x&x&x&x&x&x&x&x\\ \hline 0&0&0&0&0&0&1&0&0\\ 0&0&0&0&0&0&0&1&0\\ 0&0&0&0&0&0&0&0&1\\ x&x&x&x&x&x&x&x&x \end{array}\right] & \begin{array}{l} \rightarrow n_1 \text{ rows } (=3) \\ \\ \rightarrow n_2 \text{ rows } (=2) \\ \\ \rightarrow n_3 \text{ rows } (=4) \end{array} \end{array} \tag{3.38}$$

where x signifies a free parameter. Let the row basis have m_1 rows from the first m-group of rows of $\mathcal{H}_{n,\infty}$ or equivalently let m_1 be the number of non-zero observability indices. Very often it is $m_1=m$ since the outputs are usually linearly independent and thus all the first m rows can be taken into the basis. The matrix A then has $n-m_1$ rows with all elements zero except one element equal to 1. The rest of the m_1 rows have elements that depend on the system, i.e. on the impulse matrices. The matrix B has all its elements dependent on the system. The matrix C also has a special form. For instance for m=3 and $n_1=3$, $n_2=2$, $n_3=4$, it is

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix} \tag{3.39}$$

The matrix C has m_1 rows with all elements zero except one element equal to 1 and $(m-m_1)$ rows with all elements dependent on the system. The matrix D has all its elements dependent on the system. The total number of non-constant parameters of matrix A is nm_1, of matrix B is nr. of matrix C is $n(m-m_1)$ and of matrix D, mr. The total number of parameters then is

$$n_d=nm+nr+mr \tag{3.40}$$

Thus whatever the choice of the observability indices, or equivalently, whatever observable multi-structural form is used for the state-space model, the total number of parameters is always equal to n_d. A system that can be represented in two different observable multi-structural forms has two sets of n_d parameters that can describe the system. The one set of parameters can be uniquely derived from the other and the functions that give the one set of parameters with respect to the other are rational functions, as described in the appendix of [Wertz, Gevers and Hannan (1982)].

The observable multi-structural forms are generalizations of the observable canonical forms described in [Guidorzi (1975),(1981)]. The observable canonical forms can be derived in exactly the same way the observable multi-structural forms were derived with the extra condition that the rows of the row basis must be independent from all the rows above them. This extra rule makes the chosen row basis unique and thus the state-space models are canonical. The multi-structural forms are obviously not canonical since a given system can have several different sets of observability indices and can thus be represented by several different state-space models. The observable multi-structural forms were proposed in [Ljung and Rissanen(1976)] to overcome the difficulty of a possible almost linearly dependent row basis of the observable canonical form. They were called overlapping parametrizations and they were found superior since a change of state-space model could be done to avoid numerical problems when the row basis was found almost linearly dependent. The observability indices of observable canonical forms are invariants of the system and they are also called Kronecker indices. The observability indices of the multi-structural forms are not invariants of the system and if they are to be distinguished from the observability indices of the canonical form, they should be called multi-structural observability indices. In [Gevers and Wertz (1982)], they are called structural indices. This name is not very appropriate since it cannot distinguish them from the multi-structural controllability indices. Here canonical forms are not considered and thus the observability indices refer only to the multi-structural forms and they are not invariants of the system.

The one important consequence of the realization described above is that it also leads to an input-output model.

Let the maximum of the observability indices be equal to p

$$p=\max(n_1,n_2,\ldots,n_m) \tag{3.41}$$

The vector $y(t+p|t-1)$ is then linearly dependent on the state vector $x(t)$ in (3.37). It then holds

$$y(t+p|t-1)=A_1y(t+p-1|t-1)+A_2y(t+p-2|t-1)+ \ldots + A_py(t|t-1) \tag{3.42}$$

The matrices A_i $i=1,\ldots,p$ are of dimension mxm. If the element $y_i(t+p-j|t-1)$, $i=1,\ldots,r$ $j=1,\ldots,p$, does not belong to the state vector, then the i^{th} column of the matrix A_j is zero. Since the vector has only n elements, the matrices $A_1, A_2,\ldots, A_p$ have in total only n non-zero columns. It holds in general that

$$y(t+k)=y(t+k|t-1)+H_0\,u(t+k)+H_1\,u(t+k-1)+ \ldots +H_k\,u(t) \tag{3.43}$$

$$\text{for } K=0,1,2,\ldots$$

thus equation (3.42) becomes

$$\begin{aligned} y(t+p)-H_0u(t+p)\ldots H_pu(t)=\;&A_1\left[y(t+p-1)-H_0u(t+p-1)- \ldots -H_{p-1}u(t)\right]+\\ &A_2\left[y(t+p-2)-H_0u(t+p-2)- \ldots -H_{p-2}u(t)\right]+\\ &\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot\\ &A_p\left[y(t)-H_0u(t)\right] \end{aligned} \tag{3.44}$$

and thus

$$y(t+p)=A_1y(t+p-1)+\ \dots\ +A_py(t)+B_0u(t+p)+B_1u(t+p-1)+\ \dots\ +B_pu(t) \tag{3.45}$$

where

$$\begin{aligned} &B_0=H_0\\ &B_1=H_1-A_1H_0\\ &B_2=H_2-A_1H_1-A_2H_0\\ &\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\\ &B_p=H_p-A_1H_{p-1}-\dots\ -A_pH_0 \end{aligned} \tag{3.46}$$

The model (3.45) can be put in a form that does not include the zero columns present in the matrices $A_1, A_2, \dots, A_p$. Let the matrix A* consist of all the non-zero columns of the matrices $A_1, A_2, \dots, A_p$. The matrix A* has dimension mxn. Equation (3.45) then becomes

$$y(t+p)=A^*\begin{bmatrix} y_1(t+n_1-1)\\ \cdot\\ \cdot\\ \cdot\\ y_1(t)\\ \cdot\\ \cdot\\ \cdot\\ y_m(t+n_m-1)\\ \cdot\\ \cdot\\ \cdot\\ y_m(t) \end{bmatrix} + B_0u(t+p)+(t+p-1)+\ \dots +B_pu(t)$$

Equation (3.47) written individually for every component of y(t+p) is

$$y_i(t+p)=\begin{bmatrix} a_{i1} & a_{i2} & \dots & a_{in}\end{bmatrix}\begin{bmatrix} y_1(t+n_1-1)\\ \cdot\\ \cdot\\ \cdot\\ y_1(t)\\ \cdot\\ \cdot\\ \cdot\\ y_m(t+n_m-1)\\ \cdot\\ \cdot\\ \cdot\\ y_m(t) \end{bmatrix} +$$

$$\begin{bmatrix} b_{i1} & b_{i2} & \cdots & b_{is} \end{bmatrix} \begin{bmatrix} u_1(t+p) \\ \cdot \\ \cdot \\ \cdot \\ u_1(t) \\ \cdot \\ \cdot \\ \cdot \\ u_r(t+p) \\ \cdot \\ \cdot \\ \cdot \\ u_r(t) \end{bmatrix} \tag{3.48}$$

where $\begin{bmatrix} b_{i1} & b_{i2} & \cdots & b_{is} \end{bmatrix}$ is a row matrix that consists of the i^{th} rows of the matrices $B_0, B_1, \ldots, B_p$ and $s=(p+1)r$. The model has in total n_d^* parameters where $n_d^*=mn+ms$. Thus

$$n_d^*=nm+mpr+mr \tag{3.49}$$

The integer $p=\max(n_1,n_2,\ldots,n_m)$ obviously satisfies

$$p \geq n_1,\ p \geq n_2,\ \ldots\ ,\ p \geq n_m \tag{3.50}$$

and adding them together

$$mp \geq n \tag{3.51}$$

with the equality valid only if $n_1=n_2=\ldots=n_m$. Comparing the number of coefficients of the canonical state-space model n_d and the number of coefficients of the input-output model n_d^* and using inequality (3.51), it holds

$$n_d^* \geq n_d \tag{3.52}$$

with equality valid only when all the observability indices are equal.

The input-output models derived from the observable canonical forms have the same number of parameters as the state-space model in contrast to the input-output model (3.48). There are two reasons why canonical input-output models are not considered here. First, in the event of a badly conditioned row basis, the identification based on the canonical input-output models present numerical problems. The multi-structural input-output model avoids the problem by choosing a different, better conditioned, row basis. Second and more important for this study, the canonical input-output models are very difficult to extend to the non-linear case. However, the generalization of the multi-structural input-output model (3.48) to a non-linear model is presented in section 4.

The input-output model model (3.48) is different from the input-output model presented in [Gevers and Wertz (1982)] and [Guidorzi (1982)]. The model in these references can be anticipating and certain relationships between the parameters of the model must be valid for the model to be non-anticipating. It is thus much more difficult to use for identification purposes. Its only advantage is that it can be more easily related to the multi-structural state-space model.

The input-output model (3.48) may have more parameters than the state-space model but it is still an identifiable parametrization [Gevers and Wertz (1982)]. This means that there are no two different sets of parameters that can represent the same response function and thus the parameters of the input-output model can be found by performing input-output experiments.

4. Nonlinear Theory

A few results and notions from modern analysis and differential geometry will be needed for the development of the recursive non-linear input-output model. The key theorems that will be used later are summarised next.

An important theorem in modern analysis is the Rank Theorem [Dieudonne (1969)] [Brocker and Lander (1975)]. It is the generalization of the Inverse Function Theorem that provides the conditions under which a differentiable map can be locally transformed into a linear map by some coordinate change in the domain and co-domain of the differentiable map.

The Rank Theorem

Let E and F be finite dimensional vector spaces and let W be an open subset of E that contains the point x_o. Let f be a continuously differentiable map from W to F i.e. $f: W \to F$. Let the derivative of f at a point $x \in W$ be $Df(x)$.

If

the derivative $Df(x)$ has constant rank for all $x \in W$

then

(a) There exists a set V, an open subset of E, such that $V \subset W$ and another set V^*, an open subset of F, such that $f(V) \subset V^*$. There also exist two diffeomorphisms d_1 and d_2 where $d_1: V \to E$ and $d_2: V^* \to F$.

(b) The restriction of $f|V$ of the mapping f to the set V is equal to

$$f|V = d_2^{-1} \circ Df(x_o) \circ d_1 \qquad (4.1)$$

The essence of the Rank Theorem is that if the rank of the derivative of a function is constant around a point, local changes of coordinates (created by the diffeomorphisms d_1 and d_2) transform the function into its own derivative at that point. Actually any linear function from E to F that has rank equal to the rank of the derivative of f could be used in the place of the derivative of f in (4.1).

A function that has constant rank in the whole domain of its definition, behaves locally like a linear function of the same rank. Globally however, the image of the function in some region of the domain may interfere with the image in another region of the domain and create problems like in the function f from R^2 to R^2 illustrated in figure 1.

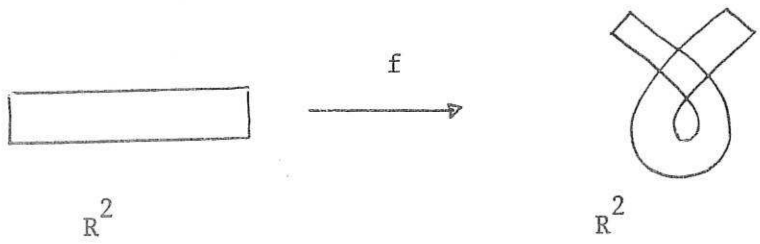

Figure 1

In order to prevent such problems an extra condition has to be imposed.

Let the derivative of the function $f: W \to F$ have constant rank in all the open sets $W \subset E$ where E and F are finite dimensional vector spaces. The set of points of W that map to the same point in the co-domain F have the structure of a submanifold in the vector space E and they are called level submanifolds [Chillingworth (1975)]. The extra condition imposed on the function f is that level submanifolds should be connected. If this condition is imposed, functions like the one shown in figure 1 are not allowed.

The linear functions always have connected level submanifolds. Thus if a function has derivative with constant rank but fails to have connected level submanifolds, it can be restricted to an open set V such that (4.1) holds. Consequently, since it is diffeomorphically equivalent to a linear function in the set V it has connected level submanifolds.

The rank of a matrix A is equal to the dimension of a square minor matrix with non-zero determinant. The determinant is a continuous function and thus if the matrix A is changed slightly, the determinant of the minor matrix will remain non-zero. The rank of the matrix A will thus either increase or stay the same. This fact can be used if the derivative of a function f at the point x_o, $Df(x_o)$, is known to have the maximum possible rank that the derivative $Df(x)$ can achieve for any x. Since the rank of $Df(x)$ cannot increase for an x different from x_o, there will be an open set W containing x_o such that the derivative $Df(x)$ has constant rank equal to the maximum rank for every $x \in W$. The Rank Theorem can then be used for the function $f|W$.

Another concept needed later is the concept of a function dependent on another function.

Let the functions $f: V \to F$ and $g: V \to G$. The function g is dependent on the function f if there exists another function $h: \mathrm{Image}(f) \to G$ such that

$$g = h \circ f \tag{4.2}$$

or equivalently if the diagram Fig.2 commutes

$$\begin{tikzcd} V \arrow[r, "f"] \arrow[dr, "g"'] & F \arrow[d, "h"] \\ & G \end{tikzcd}$$

Figure 2

Equation (4.2) can alternatively be written as

$$g(x) = h(f(x)) \qquad \text{for any } x \in V \tag{4.3}$$

The image of a point in V through the function g is thus determined when its image through the function f is known.

If the function f has constant rank and it also has connected level submanifolds, it is very easy to check that another function g is dependent on f. The result is given in the following theorem.

Theorem 1

Let the continuously differentiabel functions $f: V \to F$ and $g: V \to G$, where V is an open subset of a finite dimensional vector space E and the sets F and G are finite dimensional vector spaces.

Assume

(a) rank $Df(x)=n$ for every $x \in V$

(b) the level submanifolds of the function f are connected

Let the function $\bar{f}=(f,g): V \to F \times G$.

If also

(c) rank $D\bar{f}(x)=n$ for all $x \in V$

then

the function g is dependent on the function f.

The proof of this theorem is given as Lemma 1 in [Jakubczyk (1980)]. In essence it says that if the rank of the derivative of the function f is not increased when the function f is extended to the function (f,g) the function g is dependent on f.

In matrix form the derivative of $\bar{f}$ is

$$D\bar{f}(x) = \begin{bmatrix} Df(x) \\ Dg(x) \end{bmatrix}$$

A slightly different version of Theorem 1 will be needed later. It is the following Corollary.

<u>Corollary 1</u>

Let E*, E, F and G be finite dimensional vector spaces. Let V be an open set of the space $E^* \times E$ and let the continuously differentiable functions f and g be $f: V \to F$, $g: V \to G$. The function $f(z,x)$ where $z \in E^*$ and $x \in E$ has derivative with respect to the variable x $D_x f(z,x)$ at every point $(z,x) \in V$. The function $g(z,x)$ also has derivative with respect to x $D_x g(z,x)$.

Assume

(a) rank $D_x f(z,x)=n$ for every $(z,x) \in V$

(b) for any fixed z, the submanifold that consists of all the points $(z,x) \in V$ that map to the same point of F through the function f, is connected.

(c) the derivative of the function $\bar{f}=(f,g)$ with respect to x $D_x \bar{f}(z,x)$ also has rank equal to n for every (z,x).

In matrix form

$$D_x\overline{f}(z,x)= \begin{bmatrix} D_x f(z,x) \\ D_x g(z,x) \end{bmatrix} \tag{4.5}$$

Let the function $f^*: V\to E^*xF: (z,x)\mapsto(z,f(z,x))$ have Image $(f^*)=W$

Then

there exists a function $h: W\to G$ such that

$$g=h\circ f^* \tag{4.6}$$

or equivalently

$$g(z,x)=h(z,f(z,x)) \qquad \text{for every } (z,x) \in V \tag{4.7}$$

In other words if the variable z is known and the image of the point (z,x) through f is also known, the function h provides the image of the point (z,x) through the function g.

The proof is quite trivial and consists of showing that the conditions for the function g to be dependent on the function f^* required by Theorem 1 are the same conditions required by Corollary 1.

If the function f has rank $D_x f(z,x)$=constant around a point (z_o,x_o), then the Rank Theorem assures that a set V, an open subset of E^*xE containing (z_o,x_o), can be found so that the level submanifolds of Corollary 1 are connected.

4.1 Recursive input-output models

The results of section 4 can be used for the creation of a non-linear input-output model. First of all it is necessary to assume that the input and the output spaces U and Y are normed vector spaces so that derivatives can be defined. It is also assumed that they are finite dimensional vector spaces of dimension r and m respectively and that ordered bases have been specified for them. The zero state response function can be written individually for input sequences of length 1,2,... . Assuming that the system is at the zero equilibrium point at time t=1, the zero state response for an input sequence of length t is given by

$$y(t)=f_t(u(t),u(t-1),\ldots ,u(1)) \tag{4.8}$$

where $f_t: U^t\to Y$

The function f_t is a different function for every t=1,2,... , since the domain of definition U^t is a different one. It is assumed that the response functions f_t are continuously differentiable functions so that the Rank Theorem and Corollary 1 can be used. The functions that correspond to the function L_k used in the realization of linear systems can be found for the non-linear systems. Let the vector

$$u^t = \begin{bmatrix} u(t) \\ u(t-1) \\ \cdot \\ \cdot \\ \cdot \\ u(1) \end{bmatrix} \tag{4.9}$$

then

$$\begin{aligned} &y(t)=f_t(u(t),u^{t-1}) \\ &y(t+1)=f_{t+1}(u(t+1),u(t),u^{t-1}) \\ &\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\ &y(t+k-1)=f_{t+k-1}(u(t+k-1),\ldots ,u(t),u^{t-1}) \end{aligned} \tag{4.10}$$

let

$$y_t^k = \begin{bmatrix} y(t) \\ y(t+1) \\ \cdot \\ \cdot \\ \cdot \\ y(t+k-1) \end{bmatrix} \tag{4.11}$$

and

$$u_t^k = \begin{bmatrix} u(t+k-1) \\ u(t+k-2) \\ \cdot \\ \cdot \\ \cdot \\ u(t) \end{bmatrix} \tag{4.12}$$

Then equations (4.10) can be written as

$$y_t^k = F_{k,t-1}(u_t^k, u^{t-1}) \tag{4.13}$$

where $\quad y_t^k \in Y^k,\ u_t^k \in U^k,\ u^{t-1} \in U^{t-1}$

The function

$$F_{k,t}: U^k \times U^t \to Y^k \tag{4.14}$$

can thus be defined from (4.11) for any $k=1,2,\ldots$, $t=1,2,\ldots$ when the response function f_t is known for $t=1,2,\ldots$. The function $F_{k,t}$ is continuously differentiable since it is the Cartesian product of k response functions which are continuously differentiable. For notational simplicity let

$$\begin{aligned} z &= u^k_{t+1} \in U^k \\ x &= u^t \in U^t \end{aligned} \tag{4.15}$$

The first assumption about the function $F_{k,t}$ is the following

<u>Assumption 1</u>

$$\max(\operatorname{rank} D_x F_{k,t}(z,x)) = n \qquad \text{for any } z \in U^k,\ x \in U^t \text{ and for any } t=1,2,\ldots,\ k=1,2,\ldots \tag{4.16}$$

Assumption 1 is exactly equivalent to the requirement in the linear case that a finitely realizable system have Hankel matrix $\mathcal{H}_{k,t}$ of maximum rank equal to n for any k or t. In fact for a linear system, the function $F_{k,t}$ is a linear function and also

$$D_x F_{k,t}(z,x) = \mathcal{H}_{k,t} \qquad \text{for any } z \in U^k \text{ and } x \in U^t \tag{4.17}$$

Assumption 1 in essence guarantees that the state-space of the Nerode realization does not have infinite 'dimensions', whatever interpretation the term dimension can be given in the particular structure the state-space may have.

Assumption 1 can be checked directly by construding the function $F_{k,t}$ from the response function, calculating the derivative and finding the rank of the derivative for any k and t. Another way of checking Assumption 1 is also possible if the response function is known to be derived from a state-space description

$$\begin{aligned} x(t+1) &= g(x(t),u(t)) \\ y(t) &= h(x(t),u(t)) \end{aligned} \tag{4.18}$$

where the state-space X is a vector space of finite dimension n.

The state-transition function can be written individually as

$$x(t+k)=\Phi_k(u_t^k, x(t)) \tag{4.19}$$

where

$$\Phi_k: U^k xX\to X \tag{4.20}$$

Equations (4.10) can be written as

$$y(t)=h[\Phi_0(x(t)),u(t)]$$
$$y(t+1)=h[\Phi_1(u(t),x(t)),u(t+1)]$$
$$\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot$$
$$y(t+k-1)=h[\Phi_{k-1}(u(t+k-2),\dots ,u(t),x(t)),u(t+k-1)] \tag{4.21}$$

Assuming as usual that the system is at the zero state at t=1, the state at time t, x(t), is given by

$$x(t)= \Phi_{t-1}(u^{t-1}, x(1))=\Phi_{t-1}(u^{t-1}, 0)=\Phi^*_{t-1}(u^{t-1}) \tag{4.22}$$

where the function Φ^*_t is

$$\Phi^*_t: U^t\to X \tag{4.23}$$

The functions (4.21) can be written as

$$y_t^k=G_k(u_t^k, x(t)) \tag{4.24}$$

where the function G_k is

$$G_k: U^k xX\to Y^k \tag{4.25}$$

The function $F_{k,t}$ can now be decomposed using (4.24) and (4.22) as

$$F_{k,t}(z,x)=G_k(z, \Phi^*_t(x)) \tag{4.26}$$

Using the chain rule of differentiation

$$D_x F_{k,t}(z,x)=D_x G_k(z, \Phi^*_t(x))^{\circ} D_x \Phi^*_t(x) \tag{4.27}$$

If the state-space X has finite dimension equal to n, the co-domain of the function Φ^*_t in (4.23) is of finite dimension n and thus the derivative has rank less than or equal to n. Thus

$$\text{rank}(D_x \, \Phi_t^*(x)) \leq n \tag{4.28}$$

and consequently from (4.27)

$$\text{rank}\,(D_x \, F_{k,t}(z,x)) \leq n \qquad \text{for any } z,x \text{ and any } k,t \tag{4.29}$$

Thus

$$\max(\text{rank}(D_x \, F_{k,t}(z,x))=\text{finite} \qquad \text{for any } z,x \text{ and } k,t \tag{4.30}$$

If also for a particular z,x and k,t it holds that rank $(D_x \, F_{k,t}(z,x))=n$ then the maximum value of the rank of the derivative is equal to the dimension of the state-space n. Otherwise the maximum value of the rank of the derivative has to be found directly. The particular values of z,x and k,t that may give the rank of the derivative the value of the dimension of the state-space is not known beforehand and such values might not even exist. The value of z and x that should be tried first is zero since this is the value needed for Assumption 2 discussed next. A state-space description of a physical system is sometimes known and thus this way of checking Assumption 1 is possible. It also provides a physical explanation for Assumption 1 as it will be discussed later.

The second assumption about the function $F_{k,t}$ is that there exist a k and a t such that the derivative $D_x F_{k,t}(0,0)$ at the origin of U^k and U^t has rank n, i.e.

Assumption 2

$$\text{rank } D_x F_{k,t}(0,0)=n \qquad \text{for some } t \text{ and some } k \tag{4.31}$$

Assumption 2 can be given a simple interpretation. Let the derivative of the function f_t at the origin of U^t be

$$Df_t(0)= \begin{bmatrix} H_0 & H_1 & H_2 & \cdots & H_{t-1} \end{bmatrix} \tag{4.32}$$

The linearized system around the zero state and the zero input has impulse matrices the matrices H_ℓ, $\ell=0,1,2,\ldots$, given by the derivative of the response function f_t. It can be seen from the definition of the function $F_{k,t}(z,x)$ that the linear function $D_x F_{k,t}(0,0)$ is equal to the Hankel matrix $\mathcal{H}_{k,t}$ of the linearized system around the origin, i.e.

$$D_x F_{k,t}(0,0)=\mathcal{H}_{k,t} \tag{4.33}$$

Assumption 2 then just requires that the linearized system around the origin have some Hankel matrix of rank the maximum possible one (=n), or equivalently that the state-space of the linearized system have dimension equal to n. Because of the properties of the Hankel matrix, the k and the t in condition (4.31) need only be equal to n.

The realization of the linearized system can be carried out in the way described in section 3. Let the observability indices of the linearized system be $n_1, n_2, \ldots, n_m$ and let $p=\max(n_1, n_2, \ldots, n_m)$. Let the vector y_t^* be

$$y_t^* = \begin{bmatrix} y_1(t+n_1-1) \\ \cdot \\ \cdot \\ \cdot \\ y_1(t) \\ \cdot \\ \cdot \\ \cdot \\ y_m(t+n_m-1) \\ \cdot \\ \cdot \\ \cdot \\ y_m(t) \end{bmatrix} \tag{4.34}$$

Theorem 2

Let a non-linear system satisfy Assumption 1 and Assumption 2. Then there exists a non-linear function q such that

$$y(t+p)=q(y_t^*, u(t+p), u_t^p) \tag{4.35}$$

for a restricted region of operation around the zero equilibrium point.

The equation (4.35) can be written in expanded form as

$$\begin{aligned} y_i(t+p)=q_i\big[&y_1(t+n_1-1),\ y_1(t+n_1-2),\ \dots,\ y_1(t),\\ &y_2(t+n_2-1),\ y_2(t+n_2-2),\dots,\ y_2(t),\\ &\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\\ &y_m(t+n_m-1),\ y_m(t+n_m-2),\dots,y_m(t),\\ &u_1(t+p),\ u_1(t+p-1),\dots,u_1(t),\\ &u_2(t+p),\ u_2(t+p-1),\dots,\ u_2(t),\\ &\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\\ &u_r(t+p),\ u_r(t+p-1),\dots,\ u_r(t)\big] \end{aligned} \tag{4.36}$$

for $i=1,2,\dots,m$

The model (4.36) is a recursive input-output model completely equivalent to the linear multi-structural input-output derived in section 3.

The proof of Theorem 2 is based on the Rank Theorem and Corollary 1 presented in section 4. The vector $y_t^* \in R^n$ is given by a function F_{t-1}^* such that

$$y_t^*=F_{t-1}^*(u_t^p,\ u^{t-1}) \tag{4.37}$$

where

$$F_t^*:\ U^p x U^t \to R^n \tag{4.38}$$

The function F_t^* is constructed in the same way the function $F_{k,t}$ was constructed. Let again

$$x=u^t \in U^t \tag{4.39}$$

$$z=u_{t+1}^p \in U^p$$

The function F_t^* (z,x) has partial derivative $D_x F_t^*(0,0)$, the matrix that consists of the rows of the Hankel matrix $\mathcal{H}_{n,t}$ of the linearized system that belong to the row basis determined by the observability indices $n_1,n_2,\dots,n_m$. This matrix has rank n for $t\geq n$. Thus the function $F_t^*(z,x)$ has derivative with respect to x at the origin of rank equal to the maximum possible rank (=n) for any $t\geq n$. From the Rank Theorem there exists, for every $t\geq n$, an open set W_t, a subset of $U^p x U^t$, such that the following requirements of Corollary 1 are satisfied.

(a) rank $D_x\ F^*_t(z,x)=n$ for every $(z,x) \in W_t$

and where $t \geq n$ (4.40)

(b) The restriction $F^*_t(z,x)|W_t$ has connected level submanifolds for every fixed z (4.41)

The determination of the sets W_t for $t=n,n+1,...$ will be discussed later. The response function that gives the output $y(t+p)$ is

$$y(t+p)=f_{t+p}(u^{t+p})=f_{t+p}(u(t+p),u^p_t,u^{t-1}) \tag{4.42}$$

It is now necessary to prove that the function f_{t+p} is dependent on the function F^*_{t-1} for given $u(t+p)$ and u^p_t. Corollary 1 provides sufficient conditions that need to be satisfied for this to be true. The conditions about the function F^*_{t-1} are provided by (4.40) and (4.41) for $(u^p_t,u^{t-1}) \in W_{t-1}$. The other condition required is that the rank of the partial derivative with respect to u^{t-1} of the function

$$(F^*_{t-1}(u^p_t,u^{t-1}),\ f_{t+p}(u(t+p),u^p_t,u^{t-1})) \tag{4.43}$$

have rank equal to n for $(u^p_t,u^{t-1}) \in W_{t-1}$. This is quite simple to see since function (4.43) can be extended to the function $F_{p+1,t-1}$, which has partial derivative of maximum rank equal to n from Assumption 1. Thus function (4.43) must have partial derivative of rank n. Since all the requirements of Corollary 1 are satisfied, there exists a function q such that

$$y(t+p)=q(y^*_t,u(t+p),u^p_t) \qquad \text{for } (u^p_t,u^{t-1}) \in W_{t-1} \tag{4.44}$$

The set W_{t-1} is the set of points $(u^p_t,u^{t-1}) \in U^p x U^{t-1}$ such that the function $F^*_{t-1}(u^p_t,u^{t-1})$ has derivative with respect to u^{t-1} of constant rank n and the level submanifolds for a given u^p_t are connected. The set W_t is a set of points

$$(u^p_t,u^t)=(u^p_t,u_2^{t-1},u(1)) = ((u^p_t,u_2^{t-1}),u(1)) \in (U^p x U^{t-1})xU \tag{4.45}$$

Using the fact that the response function is a zero response function it holds that the restriction of the function F^*_t in the space $U^p x U^{t-1}$ is equal to the function F^*_{t-1}. Thus the level submanifolds of F^*_t intersect the space $U^p x U^{t-1}$ at the level submanifolds of F^*_{t-1} and, since F^*_t and F^*_{t-1} have derivatives of the same rank, they also intersect it transversely [Jakubczyk (1980)]. The set W_t can thus be chosen such that its intersection with the space $U^p x U^{t-1}$ is equal to the set W_{t-1}. Consequently the sets W_n, W_{n+1} ... can be chosen so that each one contains the one that precedes it. For $t \leq n$ each one of the sets $W_1, W_2, \ldots, W_n$ in (4.44) can be trivially chosen as the restriction of the one that follows it. The restricted region of operation of the system in which the recursive input-output model (4.36) is valid is determined by the sets $W_1, W_2, \ldots$ chosen as described above.

Assumptions 1 and 2 that guarantee the existence of the recursive input-output model (4.36) are actually only sufficient conditions and a system that does not satisfy them may still satisfy a recursive input-output model. Such an example, discussed Part II [Leontaritis and Billings (1984)], is a non-linear system which is a cascade of a static non-linearity, with zero derivative at the point zero, followed by a linear system. Assumptions 1 and 2 are however non-restrictive and it is quite natural to expect a physical non-linear system to satisfy them. In fact Assumption 1, as discussed earlier, is satisfied by a system that can be described by state-space equations with a finite dimensional state-space. The physical laws that describe real life systems can always be put into state-space form and, for the majority of the physical systems, the state-space is finite dimensional. Assumption 2 is somehow more restrictive. In essence it requires that the system, when operated in a region very close to the equilibrium point, can be successfully approximated by the linearized system at that point of operation. The success of the method of linearizing non-linear systems when operated near the equilibrium point indicates that many physical non-linear systems will satisfy Assumption 2.

Conclusions

The first part of this paper has introduced the theory of Nerode realization and the numerical solution of the linear realization problem based on the creation of the Hankel matrix. Multistructural observable forms for linear multivariable systems were then derived by considering row bases of the Hankel matrix with a special property. These results were then used to develop input-output models that correspond to the observable multi-structural forms. The multistructural input-output models are similar to the models derived from observable canonical forms. The canonical input-output models however, cannot be generalized for nonlinear systems and this was the reason for developing multistructural observable forms. Although multistructural input-output models may have more parameters than canonical models they are still identifiable parameterizations and thus they can be used for identification purposes.

The linear input-output multi-structural models were then generalised to the nonlinear case. The recursive nonlinear input-output models were derived based on two assumptions about the response function of the nonlinear system. The first assumption was that the system is finitely realizable which in essence means that the state-space of the system cannot be infinite dimensional. The second assumption was that the linearized system around the equilibrium point has the maximum possible order. Such an assumption will hold provided the system when operated near the equilibrium point can be successfully represented by a linear model. The recursive nonlinear input-output models are valid only in some restricted region of operation around the equilibrium point. The mathematical results presented also provide a specific region in which the models are valid.

In the second part of this paper [Leontaritis and Billings (1984)] the non-linear model derived above is compared with other representations, several examples are given and the results are extended to create prediction error input-output models for nonlinear stochastic systems. These latter models are referred to as the NARMAX model. Details of the application of these models in the identification of nonlinear systems are reported elsewhere [Billing and Leontaritis (1981), (1982), Billings and Voon (1983), (1984)].

Acknowledgements

The author gratefully acknowledges that this work is supported by SERC under grant GR/B/31163.

REFERENCES

AKAIKE H. (1974a) "Stochastic theory of minimal realization," IEEE Trans. on Automat. Contr., vol. AC-19, pp 667-674.

AKAIKE H. (1974b) "A new look at the statistical mode identification," IEEE Trans. on Automat. Contr., vol. AC-19, pp 716-723.

ARBIB M.A. (1968) Theories of Abstract Automata. Prentice-Hall, Eaglewood Cliffs, NJ.

ASTROM K.J. and EYKHOFF P. (1970) "System identification, a survey," Preprints 2nd IFAC Symp. Identification and Process Parameter Estimation, Prague, pp 1-38.

BILLINGS S.A. (1980) "Identification of nonlinear systems - a survey" Proc. IEE, Part D, vol 127, pp 272-285.

BILLINGS S.A. and LEONTARITIS I.J. (1981) "Identification of non-linear systems using parameter estimation techniques," Proc. IEE Conference, Control and Its Applications, Warwick Univ. pp 183-187.

BILLINGS S.A. and LEONTARITIS I.J. (1982) "Parameter estimation techniques for non-linear systems," 6th IFAC Symp. Identification and System Parameter Estimation, Washington DC. pp 427-433.

BILLINGS S.A. VOON W.S.F. (1983) "Structure detection and model validity tests in the identification of nonlinear systems," Proc. IEE, Part D, vol 130, pp 193-199.

BILLINGS S.A. VOON W.S.F. (1984) "Least squares parameter estimation algorithms for nonlinear systems," Int. J. Syst. Sci. (to appear).

BROCKETT R.W. (1976) "Nonlinear systems and differential geometry," Proc. IEEE, vol. 64, pp 61-72.

BROCKER Th. and LANDER L. (1975) Differential Germs and Catastrophes. L.M.S. Lecture Notes Series 17, Cambridge University Press.

CHILLINGWORTH D.R.J. (1975) Differential Topology with a View to Applications. Research Notes in Mathematics 9, Pittman, London.

CROUCH P.E. (1979) "Realization theory for dynamical systems," Proc. IEE, Vol. 126, pp 605-615.

CROUCH P.E. (1981) "Dynamical realization of finite volterra series," Siam J. Control and Optimization, vol. 19, pp 177-202.

DIEUDONNE J. (1969) Foundations of Modern Analysis. Academic Press, New York.

FLIESS M., NORMAND-CYROT D. (1982) "On the approximation of nonlinear systems by some simple state space models," 6th IFAC Symp. Ident. & Syst.Par.Est.

GOODWIN G.C. and PAYNE R.L. (1977) Dynamic System Identification Experiment Design and Data Analysis. Academic Press, New York.

GUIDORZI R. (1975). "Canonical structures in the identification of Multivariable systems," Automatica, vol.11, pp 361-374.

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

GUIDORZI R. and BEGHELLI S. (1982). "Input-output multistructural models in multivariable system identification," Proc. IFAC Symp. Identification and parameter estimation, Washington DC.

HERMANN R. and KRENER A.J. (1977) "Nonlinear controllability and observability," IEEE Trans. on Automat. Contr., vol. AC-22, pp 728-740.

JAKUBCZYK B. (1980) "Existence and uniqueness of realization of non-linear systems," Siam J. Control and Optimization, vol.18, pp 455-471.

LEONTARITIS I.J., BILLINGS S.A., (1984): Input-output parametric models for nonlinear systems. Part II Stochastic nonlinear systems, Res, Report No.252 University of Sheffield 1984.

LJUNG L. and RISSANEN J. (1976) "On canonical forms, parameter identifiability and the concept of complexity," Proc. IFAC Symp. Identification and System Parameter Estimation, Tbilisi, USSR, pp 58-69.

PADULO L. and ARBIB M.A. (1974) System Theory, a unified state-space approach to continuous and discrete-time systems. W.B. Sonders, Philadelphia.

PAPOULIS A. (1965) Probability, Random Variables and Stochastic Processes. McGraw-Hill, Tokyo.

SONTAG E.D. (1976) "On the internal realization of polynomial response maps." Ph.D. Dissertation, University of Florida.

SONTAG E.D. (1979a) Polynomial Response Maps. Lecture notes in Control and Information Sciences No.13, Springer-Verlag.

SONTAG E.D. (1979b) "Realization theory of discrete-time nonlinear systems - the bounded case," IEEE Transactions on circuits and systems, vol. CAS-26, pp 342-356.

SUSSMAN H.J. (1977) "Existence and uniqueness of minimal realizations of nonlinear systems," Math. Systems Theory. vol. 10, pp 263-284.

VOLTERRA V. (1930) Theory of Functionals. Blackie and Sons.

WERTZ V., GEVERS M. and HANNAN E.J. (1982) "The discrete determination of optimum structures for the state space representation of multivariable stochastic processes," IEEE Trans. on Automat. Contr. vol.AC-27, pp 1200-1211.

WIENER N. (1958) Nonlinear Problems in Random Theory. Wiley.

WONHAM W.M. (1974) Linear Multivariable Control; a geometric approach, Lecture Notes in Economics and Mathematical Systems, Springer-Verlag, Berlin.