This is a repository copy of *On Overparameterization of Nonlinear Discrete Systems*.

White Rose Research Online URL for this paper:
http://eprints.whiterose.ac.uk/81019/

Monograph:

Mendes, Eduardo M.A.M. and Billings, S.A. (1997) On Overparameterization of Nonlinear Discrete Systems. Research Report. ACSE Research Report 663 . Department of Automatic Control and Systems Engineering
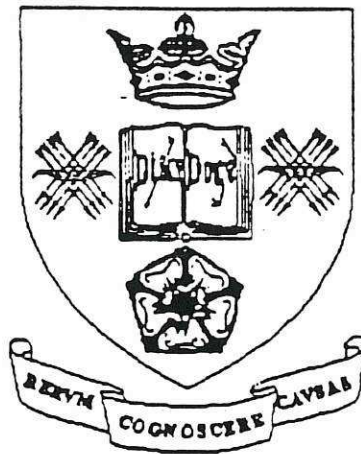
On Overparametrization of Nonlinear Discrete Systems

by

Eduardo M. A. M. Mendes and S. A. Billings

Department of Automatic Control and Systems Engineering
University of Sheffield
Mappin Street
Sheffield S1 3DJ
U.K.

On Overparametrization of Nonlinear Discrete Systems

EDUARDO M. A. M. MENDES

Departamento de Eletricidade

Fundação de Ensino Superior de São João Del Rei

Praça Frei Orlando 170 - Centro

São João Del Rei. MG - 36300.000 - Brazil

Tel.: +55(0)32 3792358, FAX: +55(0)32 3792306,

E-mail: mazoni@cpdee.ufmg.br

S. A. BILLINGS

Department of Automatic Control and Systems Engineering

University of Sheffield

P.O. Box 600. Mappin Street — Sheffield S1 3DJ - UK

June 1997

Abstract

One of the subjects which has received a great deal of attention is the overparametrization problem. It is known that the dynamical performance of the model representations deteriorates if the respective model structure is too complex. This paper investigates the problem of model overparametrization. Two new types of overparametrization, *fixed-point* and *dimension* overparametrization, are introduced and based upon this a new procedure for improving structure detection of nonlinear models is developed. This procedure uses all the information from the cluster cancellation and the location of the fixed points. Numerous examples are given to illustrate the ideas.

1 Introduction

The system identification problem typically starts with the construction of a model form or type followed by the estimation of the model parameters. Each model can be viewed as a means of expressing the properties within the data. The aim in identification is to search for a model or a class of models which best captures the properties of the data.

The single most important step in the identification process using empirical or black-box modelling is to decide upon a model structure. that is. the mathematical form with which the unknown system can be described. But first, it is necessary to define which model representation to employ, linear or nonlinear models.

Linear models have been widely used in system identification for two major reasons. First. the principle of superposition holds for such models and as a result. the effects that different and

combined input signals have on the output are easily determined. Second, linear systems are homogeneous. However, for nonlinear systems neither of these principles are valid. For instance, if an input containing different frequencies is used to excite a nonlinear system, the resultant output can exhibit not only these frequencies but also new frequencies due to the presence of nonlinearities. Harmonic generation, intermodulation, chaos and other complex effects can be generated by nonlinear systems.

Since many control systems encountered in practice are nonlinear, it seems reasonable to represent such systems using nonlinear models. The disvantage of using nonlinear models arises from the inherent complexity compared with linear models which are supported by a well established theory. One feature of this complexity is the possibility of a large number of model structures within a chosen representation. Contrary to the linear case, model structure selection for nonlinear systems is not restricted to the determination of the model order (dimension), but involves other factors. For this reason, nonlinear model structure selection is not a simple subject.

The importance of structure detection can be exemplified by reviewing the use of polynomial models in the identification of systems exhibiting chaotic motions. Katdke *et al.* (1993), and Aguirre and Billings (1994) have demonstrated that polynomial models can successfully and fairly fully reproduce chaotic behaviour provided the terms in the model or the model structures are appropriately selected. Model structure which are too complex or overparametrized were shown to introduce spurious dynamical regimes and bifurcations and ultimately may become unstable (Aguirre and Billings, 1994).

In this paper two new types of overparametrization, called *fixed-point* and *dimension* overparametrization, are introduced. In (Aguirre and Mendes, 1996) it was shown that the number and location of the fixed points of global nonlinear polynomials can be specified in relation to *term clusters* and *cluster coefficients* (Aguirre and Billings, 1995) of the respective models. Aguirre and Mendes showed that if the structure, that is the model basis, of a nonlinear polynomial model of degree ℓ includes all possible terms then such a model will always have ℓ nontrivial fixed points. In modeling problems, where the estimated polynomials are to reproduce fundamental invariants of the original system, this situation will not always be welcome and this therefore suggests that the structure of polynomial models should be chosen carefully.

Polynomial models will be considered because they are far more amenable to analysis than other model types. The structure of polynomial models can also easily be quantified and fundamental dynamical properties such as the number, location and stability of the fixed points can be readily obtained analytically when the degree of nonlinearity ℓ does not exceed four (Aguirre and Mendes, 1996). This is true for single input and single output polynomial NARMAX models but not for their multivariable counterparts. In this case, numerical methods tend to be the more natural choice. However an alternative method, called Gröbner basis (GB), has recently been introduced in the engineering field which allows fixed points to be calculated analytically in many cases (Forsman, 1991). Forsman also made use of the Gröbner basis to solve some cases of the nonlinear realization

problem. In this paper the Gröbner basis will be used merely as tool for calculating fixed points for multivariable models.

Instead of inferring the importance of the clusters as suggested in (Aguirre and Billings, 1995) from variations of the *cluster coefficients* a more physical quantity will be used, that is, the *fixed point* itself. The fixed point contains information which originates from the *cluster coefficients* and provides an indication of how the system will behave. It is shown that by monitoring the values of the fixed points, it is possible to determine which clusters are important to the model *even* for discrete systems. Numerous examples will be used to demonstrate this new approach.

The question of which minimal dimension should be used is also addressed. Recently Parlitz (1992) devised a method of determining the spurious *Lyapunov Exponents* originated from embedding the data with a dimension greater than necessary. Despite the inherent usefulness, the method proposed by Parlitz cannot be employed for NARMAX models. Instead, an alternative approach based upon early ideas discussed in (Eckmann and Ruelle, 1985) will be used in this paper. In this manner, the minimal dimension required to describe the original system is selected, hence avoiding the problem of dimension overparametrization.

The paper is organized as follows. Section (2) includes the background concepts needed to understand the paper. In this section the number, location and symmetry of fixed points are reviewed in the light of term clusters and cluster coefficients. The concept of Lyapunov exponents is also reviewed. In section (3) the different types of overparametrization of nonlinear discrete systems are defined and analyzed. It is shown how the number and location of fixed points usually alter when terms from different clusters are selected to compose a model. Section (3) also suggests some ways in which the results presented in this work can be used to aid in selecting the structure of polynomial models. In section (4) a preliminary study of overparametrization in the context of multivariable nonlinear models is made. The main points of the paper are summarized in section (5).

2   Important Concepts

The aim of this section is to review the basic concepts of *term clusters* and *cluster coefficients* proposed in (Aguirre and Billings, 1995). Recent results involving fixed points and clusters (Aguirre and Mendes, 1996) are briefly discussed. Finally a short review of Lyapunov exponents is given.

Polynomial models will be used throughout the present study. A thorough discussion of the advantages of using polynomial models in the context of chaos is given in (Aguirre and Mendes, 1996). Such models are a subset of the so-called NARMAX (*Non-Linear AutoRegressive Moving Average with eXogenous inputs* model (Leontaritis and Billings, 1985). Because polynomial NARMAX models are linear in the parameters the structure of these models can be effectively detected using available algorithms (Billings *et al.*, 1988; Mees, 1993a; Katdke *et al.*, 1993; Judd and Mees, 1994).

3

2.1 Identification of Polynomial models and Structure Detection

It has been shown that the NARMAX model provides a unified representation for a wide class of non-linear systems (Leontaritis and Billings, 1985). Several well-known models such as the Hammerstein, Wiener and bilinear models are known to be a special case of the NARMAX model (Billings and Leontaritis, 1981, 1982). This model can represented as follows

$$y(k) = F(y(k-1),\ldots,y(k-n_y),u(k-d),\ldots,u(k-d-n_u),$$
$$e(k-1),\ldots,e(k-n_e)) + e(k) \tag{1}$$

where $y(k)$, $u(k)$ and $e(k)$ represent the output, input and noise, respectively. n_y, n_u and n_e are the corresponding maximum lags and $\{e(t)\}$ accounts for possible noise, uncertainties, unmodelled dynamics, etc. $d \in \mathbb{Z}^+$ is the delay. F is some non-linear function, the form of which is usually unknown.

It is interesting to note that the NARMAX model came as a natural extension of the well-known ARMAX. An enormous amount of literature exists on estimation of ARMAX models from data. For instance, see (Box and Jenkins, 1976; Söderström and Stoica, 1989).

For many real sampled nonlinear systems the exact NARMAX models, described by the function F in equation (1), are very difficult to determine. Therefore it is often necessary to approximate F by some known function. Polynomial NARMAX models have been shown to be a good choice (Billings and Fadzil, 1985; Billings et al., 1989) and their use was studied in detail in (Chen and Billings, 1989).

When the purpose is to identify polynomial nonlinear discrete models from data equation (1) can be rewritten as

$$y(k) = \Psi_{yu}^T(k-1)\Theta_{yu} + \Psi_{yue}^T(k-1)\Theta_{yue} + e(k) , \tag{2}$$

where $\Psi_{yu}^T(k-1)$ includes a constant and all the output and input terms as well as all combinations of these up to degree ℓ and time $k-1$. These terms will henceforth be referred to as *process terms*. The vector Θ_{yu} contains the parameters associated with these terms. The matrix $\Psi_{yue}^T(k-1)$ and the vector Θ_{yue} are defined likewise. $\Psi_{yue}^T(k-1)$ will be referred to as *noise terms*.

Unfortunately equation (2) is not suitable for estimating the parameters of a polynomial NAR-MAX model because the noise terms are not known. However the noise sequence $e(k)$ can be estimated interactively as

$$\xi(k) = y(k) - \hat{y}(k|\hat{\Theta}) \tag{3}$$

where $\xi(k)$ is the residual at time k and $\hat{y}(k|\hat{\Theta})$, the prediction of $y(k)$, can be written as

$$\hat{y}(k|\hat{\Theta}) = \Psi_{yu}^T(k-1)\hat{\Theta}_{yu} + \Psi_{yu\xi}^T(k-1)\hat{\Theta}_{yu\xi} . \tag{4}$$

Finally, substituting equation (4) into equation (3) and rearranging yields

$$y(k) = \Psi^T(k-1)\hat{\Theta} + \xi(k) \tag{5}$$

where $\Psi^T(k-1) = [\Psi_{yu}^T(k-1) \; \Psi_{yu\xi}^T(k-1)]$ and $\hat{\Theta}^T = [\hat{\Theta}_{yu}^T \; \hat{\Theta}_{yu\xi}^T]^T$. Equation (5) clearly belongs to the linear regression model

$$y(k) = \sum_{i=1}^{n_\theta} \mathsf{p}_i(k)\theta_i + \xi(k), \qquad k = 1, \ldots, N \tag{6}$$

where N is the data length, $p_i(t)$ are column-vectors which represent *process* and *noise terms*, n_θ is the number of distinct such column-vectors, θ_i are unknown parameters to be estimated and n_θ is the summation of n_p process terms and n_n noise terms. In the orthogonal estimator the parameter estimation is performed for a linear-in-the-parameters model which is closely related to (6) and which can be represented as

$$y(k) = \sum_{i=1}^{n_\theta} \mathsf{w}_i(k)g_i + \xi(k), \qquad k = 1, \ldots \tag{7}$$

where the orthogonal vectors w_i and the parameters g_i are constructed from equation (6). The original parameters θ_i of the model in equation (6) can be calculated from the $\{g_i\}_{i=1}^{n_\theta}$.

The possibility of selecting the relevant vectors (terms) as a by-product is the great advantage of the orthogonal estimator. To demonstrate this, consider again the orthogonal regression equation (7). Assuming that the orthogonal property $\mathsf{w}_i^T\mathsf{w}_j = 0$ for $i \neq j$ holds. Therefore, multiply equation (7) by itself and take the time average to give

$$\frac{1}{N}\mathsf{y}^T\mathsf{y} = \frac{1}{N}\sum_{i=1}^{n_\theta} g_i^2 \mathsf{w}_i^T\mathsf{w}_i + \frac{1}{N}\xi^T\xi \tag{8}$$

The output variance $\mathsf{y}^T\mathsf{y}/N$ consists of two terms. The first term $\sum_{i=1}^{n_\theta} g_i^2\mathsf{w}_i^T\mathsf{w}_i/N$ is the part of the output variance explained by the regressors whereas the second term $\xi^T\xi/N$ accounts for the unexplained variance. Because of the properties of the orthogonal estimator the increment towards the overall output variance of each regressor (term or vector) can be computed independently as

5

$g_i^2 \mathbf{w}_i^T \mathbf{w}_i$. Expressing this quantity as a fraction of the overall output variance yields the *Error Reduction Error* (ERR)

$$[ERR]_i = \frac{g_i^2 \mathbf{w}_i^T \mathbf{w}_i}{\mathbf{y}^T \mathbf{y}}, \qquad 1 \leq i \leq n_\theta \tag{9}$$

ERR can be used as a simple and effective means of selecting the most relevant regressors in a forward-regression manner. Therefore ERR imposes a hierarchy of terms according to their contribution towards the overall output variance.

The orthogonal estimator will be used throughout this work to select the terms and estimate their coefficients for polynomial models.

2.2 Term clustering

In order to review the concepts of term clustering it is necessary to expand the deterministic part of the NARMAX model, that is, a NARX model, as the summation of terms with degrees of nonlinearity in the range $1 \leq m \leq \ell$

$$y(k) = \sum_{m=0}^{\ell} \sum_{p=0}^{m} \sum_{n_1, n_m}^{n_y, n_u} c_{p,m-p}(n_1, \ldots, n_m) \prod_{i=1}^{p} y(k-n_i) \prod_{i=p+1}^{m} u(k-n_i) \ , \tag{10}$$

where

$$\sum_{n_1, n_m}^{n_y, n_u} \equiv \sum_{n_1=1}^{n_y} \cdots \sum_{n_m=1}^{n_u}$$

and the upper limit is the maximum lag n_y if the summation refers to factors in $y(k - n_i)$ or the maximum lag n_u for factors in $u(k - n_i)$. For $m = 0$ equation (10) is reduced to $y(k) = c_0$, that is, a constant term. It can be noticed from equation (10) that there are many possible terms in a polynomial model. Although there is no term repetition it seems natural to consider groups of similar terms which describe the same type of nonlinearity. For instance, terms such as $y(k-1)u(k-2)$, $y(k-2)u(k-1)$ and $u(k-3)y(k-1)$ can be considered as members of the group which includes all the terms of the type $y(k - n_1)u(k - n_2)$ ($n_1 = 1, \ldots, n_y$ and $n_2 = 1, \ldots, n_u$). These groups of similar terms have been named as *term clusters* in (Aguirre and Billings, 1995). In a mathematical notation, the set of terms represented by $\Omega_{y^p u^{m-p}}$ contains terms of the form $y(k - n_i)^p u(k - n_j)^{m-p}$ for $m = 0, \ldots, \ell$ and $p = 0, \ldots, m$.

The summation of the coefficients of all the terms which pertain to a certain cluster is the *cluster coefficient* (Aguirre and Billings, 1995) denoted by $\Sigma_{y^p u^{m-p}}$. Generically, for equation (10) the cluster coefficients are $\sum_{n_1, n_m}^{n_y, n_u} c_{p,m-p}(n_1, \ldots, n_m)$. For example, if the maximum degree of nonlinearity in equation (10) is $\ell = 2$ all clusters can readily be determined as Ω_0 (that is, the

6

constant term), Ω_y, Ω_u, Ω_{yu}, Ω_{y^2} and Ω_{u^2}. Setting $n_y = 1$ and $n_u = 2$ the cluster coefficients can also be determined: $\sum_0 = c_{0,0}$, $\sum_y = c_{1,0}(1)$, $\sum_u = c_{0,1}(1) + c_{0,1}(2)$, $\sum_{yu} = c_{1,1}(1,1) + c_{1,1}(1,2)$, $\sum_{y^2} = c_{2,0}(1)$ and $\sum_{u^2} = c_{0,2}(1) + c_{0,2}(2)$.

From the definitions above it can be concluded that the set of candidate terms for a NARX model is the union of all possible clusters up to degree ℓ. This can be represented as follows

$$
\begin{aligned}
\{\text{all possible terms}\} \quad &= \bigcup_{\substack{p=0\ldots m \\ m=0\ldots \ell}} \Omega_{y^p u^{m-p}} \\
&= \text{constant} \cup \Omega_y \cup \Omega_u \cup \Omega_{y^2} \cup \Omega_{yu} \cup \Omega_{u^2} \cup \ldots \\
&\quad \ldots \cup \text{ all possible combinations up to degree } \ell \; .
\end{aligned}
\tag{11}
$$

In (Aguirre and Billings, 1995) it is stated that it is highly desirable to eliminate certain clusters prior to term selection because in this case the number of candidate terms would be considerably reduced. If a certain cluster is not required to reproduce the underlying dynamics such a cluster is said to be *spurious* as opposed to *effective* clusters which are required to obtain a dynamically valid model.

It has been shown that if a certain term cluster is spurious, the respective coefficient will gradually become small or will oscillate around zero as the number of terms in the model is increased Aguirre and Billings (1995). This procedure called cluster cancellation is simple, quite robust and can be used in structure selection problems.

When the main interest is to estimate global nonlinear discrete models directly from data Mendes and Billings (1996a) have established that the aforementioned cancellation is not always obvious. In fact, in the cases where the data do not contain information at all the fixed points (reviewed in the next section) cluster cancellation does not indicate the spurious terms. In these cases Mendes and Billings showed that several different model structures can only reproduce the local dynamical characteristics of the system under study.

2.3 Fixed Points

The objective of this section is to investigate some relationships between the structure of nonlinear polynomials and their respective fixed points. The concept of using cluster coefficients for characterizing the number of fixed points of a map and the location of such fixed points will be briefly reviewed.

Based upon the definition of fixed points, that is, $y(k) = y(k + i)$, $i \in \mathbb{Z}$ and using the cluster coefficients, the fixed points of an autonomous polynomial with degree of nonlinearity ℓ can be calculated by finding the roots of the following "clustered polynomial" (Aguirre and Mendes, 1996)

$$
\Sigma_{y^\ell}\, y(k)^\ell + \ldots + \Sigma_{y^2}\, y(k)^2 + (\Sigma_y - 1)\, y(k) + \Sigma_0 = 0 \; .
\tag{12}
$$

where $\Sigma_0 = c_{0,0}$ is a constant. From equation (12) it can be seen that an autonomous polynomial with degree of nonlinearity ℓ will have ℓ fixed points if $\Sigma_{y^\ell} \neq 0$.

One important aspect in the study of nonlinear systems is to verify if the fixed points are symmetric. Aguirre and Mendes (1996) investigated the simplest nontrivial symmetry of fixed points in $\mathbb{R}$, that is $\mathbb{Z}_2$, for nonlinear polynomial models. These authors pointed out that the data used for identification purposes will always be real and only real fixed points should be considered. This conclusion and the procedure of determing the fixed points directly from the data (described in the sequel) can be used as an aid to structure detection. This will be explored in the last sections of this work.

2.3.1 Determining Fixed Points directly from the data set

Recently Glover and Mees (1993) devised a very simple method of estimating the position of the fixed points directly from the data. For the sake of clarity, their method is now briefly described.

The method is based on the fact that a fixed point corresponds to a measurement which remains constant. Trajectories, which come closer to a fixed point, mimic this behaviour by remaining relatively constant for a short period of time. Just by looking at a time series, the influence of one or more fixed points can often be readily seen. This constitutes a strong result since it allows the important clusters to be defined *before* estimating a model.

By fitting local $AR(m)$[1] models to sections or windows $\{y(1), \ldots, y(2 * m + 1)\}$ of the time series, the nonlinear dynamics can be locally represented. The fitted AR models can then be used for estimating the fixed point p. This procedure is performed repeatedly so that values of the fixed points are calculated over the *entire* time series using overlapping windows of size $2 * m + 1$. The next step is to calculate the distance r_k between the m-dimensional point $v_k = (y(k), \ldots, y(k+m-1))$ and the candidate m-dimensional fixed point $q_k = (p_k, \ldots, p_k)$. If they are far apart the value of p_k is rejected. However if v_k is near the fixed point so is its estimate. A plot of p_k against r_k will show the location of the fixed points.

2.4 Lyapunov Exponents

Lyapunov exponents and Lyapunov dimension are two of the most studied diffeomorphic invariants of autonomous dynamical systems. The dimension of the attractor describes approximately the number of degrees of freedom whereas the Lyapunov exponents quantify the sensitivity of the system to initial conditions. Algorithms for estimating Lyapunov exponents have been described in, for instance, (Wolf *et al.*, 1985; Sano and Sawada, 1985; Eckmann *et al.*, 1986; Sato *et al.*, 1987) and constitute an active field (Briggs, 1990; Brown *et al.*, 1991; Bryant *et al.*, 1990; Gencay and Dechert, 1992; Kruel *et al.*, 1993) etc. In these references the main concern is to calculate the whole

---

[1] m is the dimension of an AutoRegressive Model.

Lyapunov spectrum. When only the largest exponent needs to be estimated, simpler algorithms can be used (Wolf *et al.*, 1985; Kantz, 1994).

It is normally accepted that any system containing at least one positive Lyapunov exponent is considered to be chaotic, with the magnitude of the exponent reflecting the time scale on which the system dynamics become unpredictable. The magnitudes of Lyapunov Exponents quantify an attractor's dynamics in information theoretic terms. The exponents also measure the rate at which system processes create or destroy information. For systems whose equations of motion are explicitly known there is a straightforward technique for computing a complete Lyapunov spectrum.

A great problem when Lyapunov exponents are calculated is how to define the dimension = maximum lag n_y. Recently Parlitz (1992) rediscovered and explored an earlier idea described by Eckmann and Ruelle (1985) that under time reversal of the data, the true Lyapunov exponents will reverse in sign, that is, positive exponents become negative and vice-versa. Parlitz argues that the spurious exponents do not behave in the same way, that is, the sign remains unchanged when the time is reversed. Abarbanel and Sushchik (1993) pointed out that such behaviour would be difficult to see when the exponent is close to zero. Although Parlitz has been quite successful in determining the true Lyapunov exponents from experimental data (Parlitz, 1993), Abarbanel and Sushchik also claimed that such a method is not robust to noisy data. They advocate the use of the method of false neighbours instead (Kennel *et al.*, 1992; Abarbanel and Kennel, 1993; Kennel and Abarbanel, 1994).

It seems quite reasonable to think that even with the pitfalls of Parlitz's method, this could be extended to polynomial NARMAX models. In a case in which the identified models are invertible (usually not many), it could be argued that if the time reversal procedure were applied, the *spurious* exponents would not change signs. Unfortunately when the inverted model is free run the signs of the spurious exponents change as occurs to the true Lyapunov exponents. As a result Parlitz's method can only be applied when local models are directly fitted to the data. Fortunately, Eckmann and Ruelle (1985) pointed out that the spurious Lyapunov exponents tend to be multiples of the effective exponents. This simple procedure will be used in this work to detect the spurious exponents hence avoiding the deleterious effects of overparametrization.

3 Overparametrization of Nonlinear Systems

In the context of analysis of both linear and nonlinear systems a model where the structure is more complex than necessary is said to be **overparametrized**. In other words the *Principle of Parsimony* should ideally be adhered to. This principle states that the complexity of the model should be penalized so as to produce the simplest model possible. There are sound and technical reasons for always penalizing overparametrization (See Tong. 1992, for details and references). Such parasitic effects occur very commonly in nonlinear modelling. For some discussion on overparametrization of radial functions, see (Mees, 1993b; Judd and Mees. 1994; Mees and Judd, 1995).

9

The objective of this section is to define the different types of overparametrization according to the spurious dynamics introduced into the model. For instance, it will be shown how over-parametrization of nonlinear models affects the number and location of the respective fixed points and how it affects the number of Lyapunov exponents. Numerous examples will be given to illustrate the various types of overparametrization.

Some aspects of structure selection will also be addressed and analyzed in this section. The results developed in the previous section will be used as a framework both to illustrate the effects of overparametrization in nonlinear models and also to aid the structure selection.

3.1 Different types of Overparametrization: Definitions

When overparametrization is mentioned in the literature, it is often associated with linear system modelling. The inclusion of more terms than necessary in a linear system model implies that the model order is overestimated. In the nonlinear case such terms do not always lead to an increase in dimension as will be elaborated shortly. Owing to the variety of terms that are possible in nonlinear models, overparametrization may be classified into three different types:

Definition 3.1 *Term Overparametrization*
This kind of overparametrization occurs when a model has more terms than necessary.

This is the commonly held and understood definition of overparametrization. In the case of linear systems, this would imply an increase in model order (dimension). A possible consequence of such an overparametrization is the loss of predictability of the model.

Definition 3.2 *Dimension Overparametrization*
Overparametrization occurs when an n-dimensional system is represented by a model whose dimension is $n + i$ $(i > 0)$.

In the identification of linear systems this kind of overparametrization can be detected by cancelling pole-zero pairs especially when the data are noise-free. It is well known that exact cancellations are not possible in real data, but nevertheless the principle still applies. In the nonlinear case, such cancellations are often not obvious, although they can occur when the linear part is considered. This will be illustrated in the following example.

Example 3.1
The system considered in this example is the Duffing-Ueda oscillator (Ueda, 1985) described by the following differential equations:

$$\ddot{y} + k\dot{y} + y^3 = u \tag{13}$$

10

where $k = 0.1$. This equation exhibits chaotic behaviour when the the input $u(t)$ is a cosine of amplitude 11 and frequency $w = 1.0$ rad/s. To simulate the system, a square wave with increasing amplitude with a superimposed Gaussian sequence was applied as input. Only 200 data points were used for the identification although over 1500 points were available. Insofar as dimension (lag) is concerned, such a system can be represented fairly well by models of dimension 2 or greater. As the data are noise-free, the great majority of models were able to reproduce the system dynamics. The means of verifying cancellation amongst the linear terms in the models was achieved by studying the linear transfer functions (TF_l) (which are obtained by keeping only the linear terms from an estimated nonlinear model). The TF_l can easily be obtained via inspection of the linear clusters Ω_y and Ω_u. For this example, three different models were considered. Each one represents a specific dimension:

Model of dimension 2

The linear transfer function of the model estimated from the Duffing-Ueda with dimension $n_y = 2$ is

$$TF_{l_2} = \frac{+0.1169 \times 10^{-2} z^1 + 0.1554 \times 10^{-2}}{z^2 - 0.1995 \times 10^{-1} z^1 + 0.9953 \times 10^{+0}} \tag{14}$$

from equation (14), the poles and zeros can easily be calculated as

$$\text{Poles} = \begin{cases} 0.9976 + 0.003323i \\ 0.9976 - 0.003323i \end{cases} \quad \text{Zeros} = \begin{cases} -1.329 \end{cases} \tag{15}$$

Model of dimension 3

The linear transfer function of the model estimated from the Duffing-Ueda with dimension $n_y = 3$ is

$$TF_{i_3} = \frac{+0.1165 \times 10^{-2} z^2 + 0.1121 \times 10^{-2} z^1 - 0.543 \times 10^{-3}}{z^3 - 0.2356 \times 10^{+1} z^2 + 0.1716 \times 10^{+1} z^1 - 0.3598 \times 10^{+0}} \tag{16}$$

from equation (16), the following poles and zeros can be calculated.

$$\text{Poles} = \begin{cases} 1.002 \\ 0.9929 \\ 0.3617 \end{cases} \quad \text{Zeros} = \begin{cases} -1.316 \\ 0.3541 \end{cases} \tag{17}$$

Model of dimension 4

The linear transfer function of the model estimated from the Duffing-Ueda with $n_y = 4$ is

$$TF_{l_4} = \frac{+0.1168 \times 10^{-2} z^3 + 0.1101 \times 10^{-2} z^2 - 0.3263 \times 10^{-3} z^1 + 0.2957 \times 10^{-3}}{z^4 - 0.2369 \times 10^{+1} z^3 + 0.1935 \times 10^{+1} z^2 - 0.759 \times 10^{+0} z^1 + 0.1928 \times 10^{+0}} \quad (18)$$

with poles and zeros

$$\text{Poles} = \begin{cases} 1.002 \\ 0.9931 \\ 0.1871 + 0.3985i \\ 0.1871 - 0.3985i \end{cases} \quad \text{Zeros} = \begin{cases} -1.305 \\ 0.1813 + 0.4014i \\ 0.1813 - 0.4014i \end{cases} \quad (19)$$

The pole-zero cancellation is evident from equations (17) and (19). It is worth pointing out that TF_{l_3} and TF_{l_4} are not stable since one of the poles is placed outside the unit circle. This seems a characteristic of chaotic systems especially when time series (NARMA models) are considered. The cancellation is however not so easy to detect when noise is added to the data. This is a common situation when real data are analyzed.

$\square$

Definition 3.3 *Fixed-Point Overparametrization*

Overparametrization occurs when spurious term clusters are included in a model. As an immediate consequence the model will have a spurious fixed point.

Both *dimension overparametrization* and *fixed-point overparametrization* can be thought of as special cases of *term overparametrization*. However the latter definition will only be used when the model has more terms than necessary but does not introduce spurious clusters and higher order terms.

A diagram of overparametrization of nonlinear systems is depicted in Figure (1). The dashed lines indicate that both *dimension and fixed-point overparametrization* can possibly occur as a consequence of *term overparametrization*. However their effects are different as will be demonstrated in the following example.
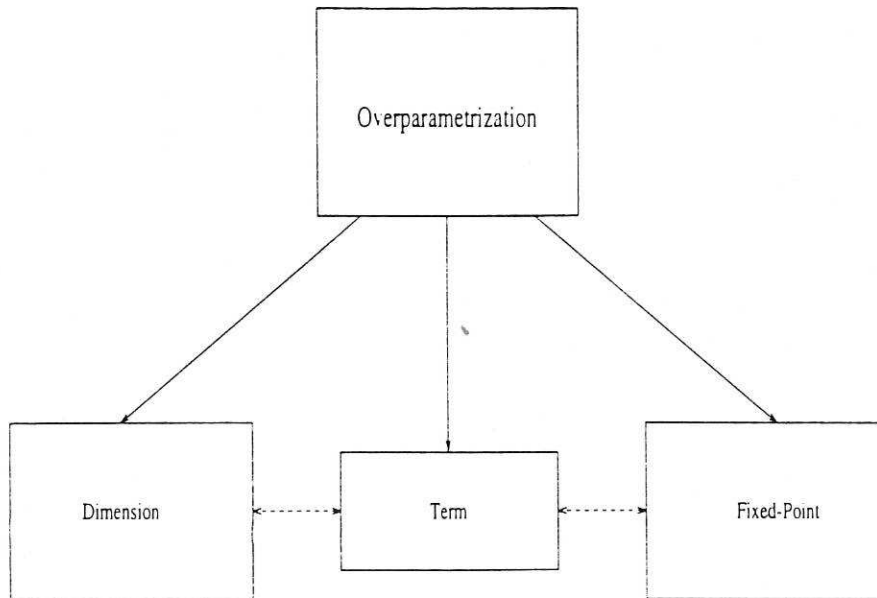
Figure 1: Overparametrization of nonlinear discrete systems.

Example 3.2

Consider the Hénon map (Hénon, 1976):

$$x(k) = 1 + \alpha x(k-2) + \beta x(k-1)^2 \tag{20}$$

where $\alpha = 0.3$ and $\beta = -1.4$, the values often used in the literature. For the Hénon map, the fixed points are placed at (-1.131,0.6314), the Lyapunov exponents are (0.425,-1.629) bits/s and the Lyapunov dimension is $D_L = 1.2609$. The poles of the linear part are (0.5477,-0.5477). To illustrate the types of overparametrization, two variants of equation (20) are analyzed:

Model A

$$x(k) = 1 + \alpha x(k-2) + \beta x(k-1)^2 + \gamma x(k-3) \tag{21}$$

and

Model B

$$x(k) = 1 + \alpha x(k-2) + \beta x(k-1)^2 + \gamma x(k-1)^2 x(k-2) \tag{22}$$

It can readily be seen that, compared to equation (20), Model A is dimension overparametrized and Model B is fixed-point overparametrized. The coefficient γ is then varied ($\gamma < \alpha; \gamma < \beta$) so as to assess the influence of the spurious term.

(a)

(b)

(c)
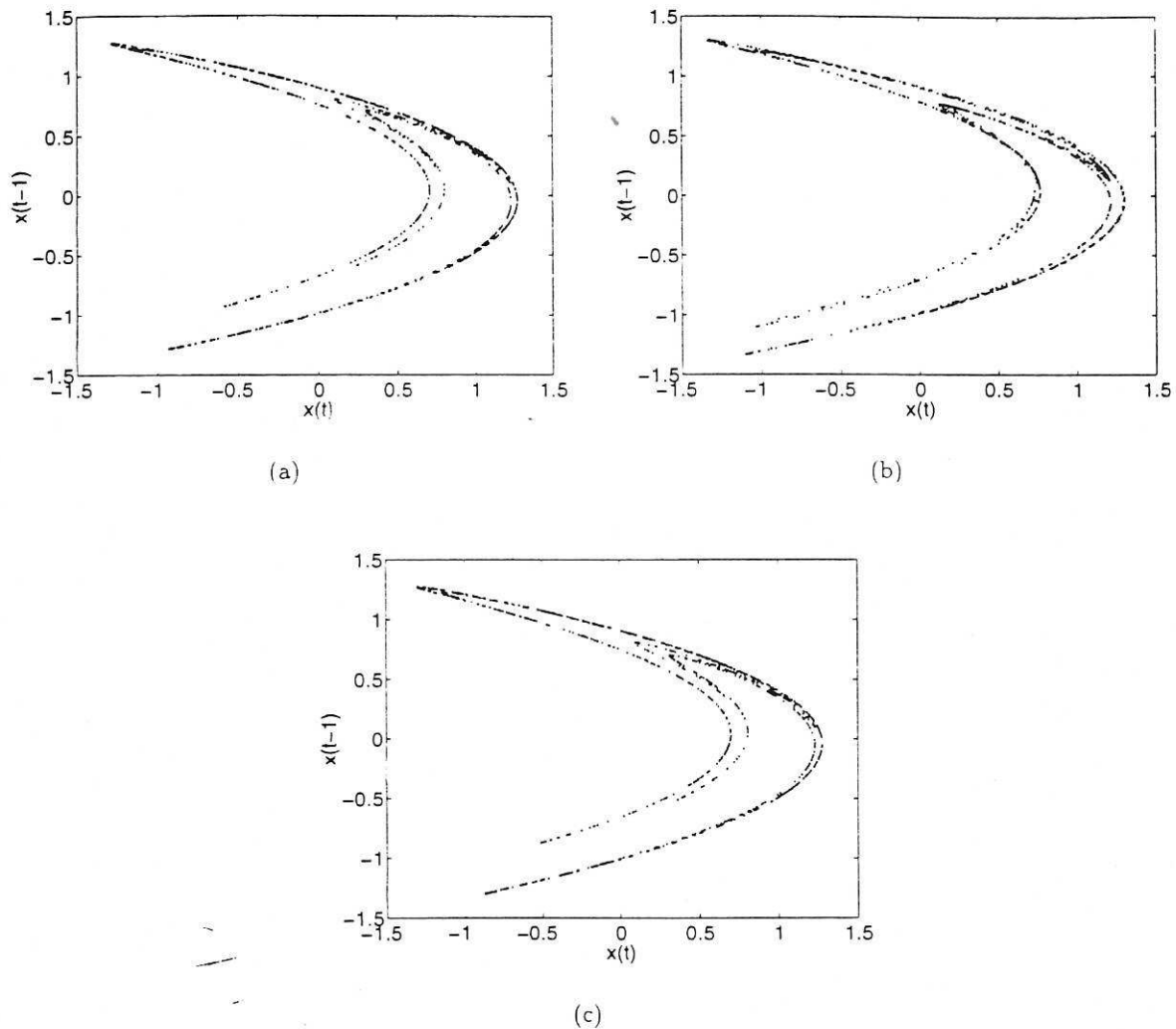
Figure 2: First return map of (a) Hénon Map, equation (20), (b) Model A, equation (21), with $\gamma = 0.05$ and (c) Model B, equation (22), with $\gamma = 0.05$.

Figure (2) depicts the first return map of the Hénon map and its variants. When figures (2-b) and (2-c) are compared to (2-a) no apparent differences can be noted although they do exist. The bifurcation diagrams shown in Figures (3), (4) and (5) reveal quite clearly the differences between the models which are, of course, due to overparametrization. They also show that the two overparametrized models have distinct behaviour. In Model A which is dimension overparametrized

14

the distortions are accentuated. The model invariants are shown in Table (1) for $\alpha = 0.3$, $\beta = -1.4$ and $\gamma = 0.05$. Note that the fixed points of **Model A** are similar to the original fixed points. Whereas **Model B** has a spurious fixed point owing to the cluster Ω_{x^3}. It is important to stress that the spurious fixed point is placed far from the original fixed points which tends to affect the bifurcation diagram only mildly. Once the parameter γ is increased beyond a certain value so that spurious fixed point is moved to be near to the original fixed points, the effects of overparametrization are more devastating.

(a)

(b)

(c)

Figure 3: Bifurcation diagram of the Hénon Map (equation (20)): (a) the parameter β is varied from -1.5 to -0.3, (b) Zoom over the region between -1.4 and -1.1 and (c) Zoom over the region between -1.1 and -0.9.

The same analysis can be done for **Model A**. In this case there is no spurious fixed point. However the spurious Lyapunov exponents and the pole shown in Table (1) can be accounted for by the effects of overparametrization. Firstly, the Lyapunov spectrum of **Model A** is not as good as that of **Model B**. Consequently the Lyapunov Dimension is higher for the previous model. Also owing to the inclusion of the extra dimension, the poles of the linear part have lost the

original symmetry. Secondly, when the parameter γ is increased the model becomes unstable. After numerous simulations it was conjectured that the increase of dimension is largely responsible for the identification of unstable models and should therefore be avoided.

	Hénon Map	Model A	Model B
Fixed Points	(-1.131,0.6314)	(-1.109,0.6443)	(28.47,-1.104,0.6366)
Lyapunov Exponents	(0.425,-1.629)	(0.4138,-1.327,-2.082)	(0.406,-1.519)
Lyapunov Dimension	1.2609	1.3118	1.2672
Poles	(0.5477,-0.5477)	(0.6173,-0.428,-0.1893)	(0.5477,-0.5477)

Table 1: Fixed Points, Lyapunov exponents, Lyapunov dimension and poles for Model A and Model B, equations (21) and (22) respectively.

□

3.2 A Fixed-Point approach to structure selection

The objective of this section is to introduce a new approach to structure selection based upon term clusters and cluster coefficients. The central idea of using cluster methods in model structure selection is that if terms of an unnecessary cluster are selected (this will often be the case for short sampling times or when the data are noisy) the parameter estimation algorithm will usually counteract this situation by estimating parameters in such a way that the effects of the terms within the inappropriate cluster cancel out. This cancelling effect is indicated by a small absolute value of the respective cluster coefficient (Aguirre and Billings, 1995). According to the authors, in situations when the coefficient of a spurious cluster is not convincingly small, it is often helpful to monitor the behavior of the cluster coefficients as the number of terms in the model is gradually increased. By following this procedure spurious clusters are usually revealed either by *decreasing* or *oscillating* cluster coefficients. In (Mendes and Billings, 1996a) it is shown that such a procedure is only useful when the data contain information about all the fixed points. Moreover the procedure is valid only for discrete models identified from data generated from a continuous system.

Another alternative, the concept of *zeroing-and-refitting* has been investigated in the context of model structure selection (Kadtke *et al.*, 1993). Briefly, in this approach the terms which have coefficients smaller than a certain threshold are eliminated from the original structure. Although this method is quite effective in some situations, it was pointed out that noise in the data and also the sampling time usually affect the term coefficients in such a way that the absolute values of spurious terms increase and therefore it becomes very difficult to decide which terms should be removed from the original structure (Aguirre and Billings, 1995).
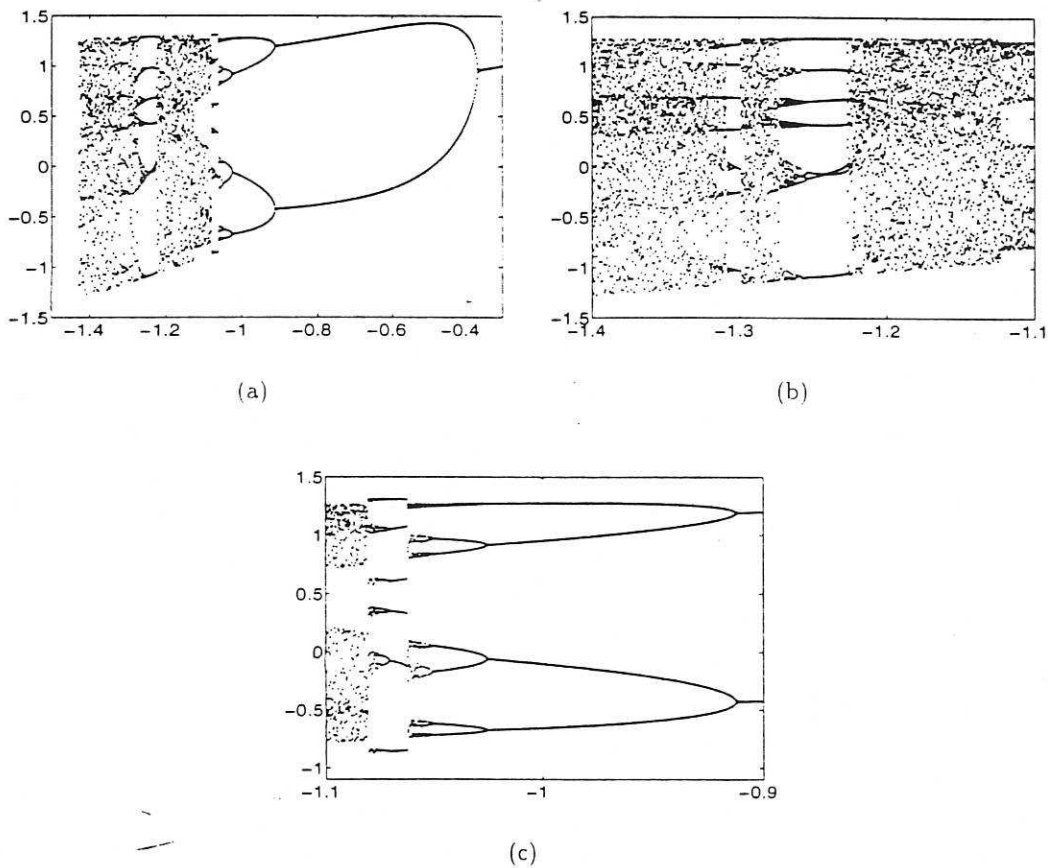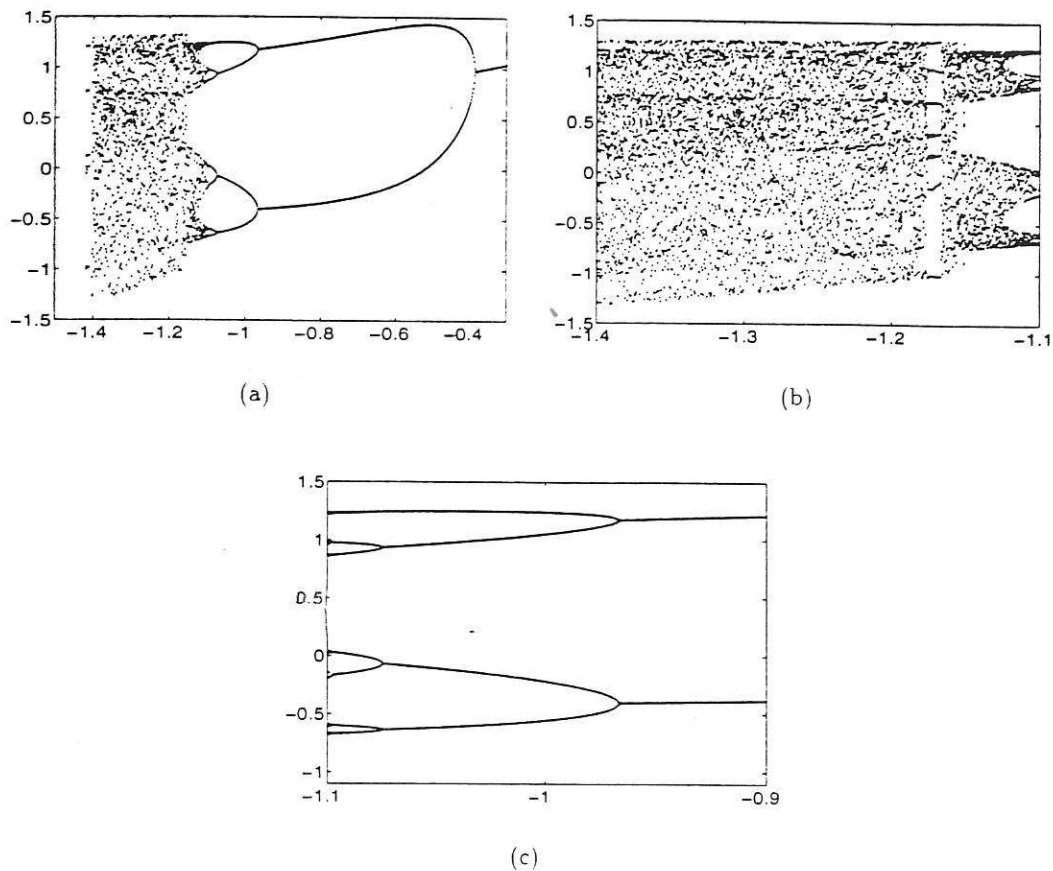
Figure 4: Bifurcation diagram of Model A. equation (21): (a) the parameter β is varied from -1.5 to -0.3, (b) Zoom over the region between -1.4 and -1.1 and (c) Zoom over the region between -1.1 and -0.9.

The new procedure named the fixed-point approach uses all the information from the cluster cancellation together with the location of the fixed points. It will be shown that the degree of nonlinearity ℓ can be determined when the location and values of the fixed points are extracted from the data. Consider an example to illustrate the approach.

Example 3.3

The objective of this example is to illustrate how the fixed point approach can often be used to determine not only the *effective* clusters but also the degree of nonlinearity ℓ. The system considered in this example is Chua's circuit (Chua and Hasler, 1993) represented as follows
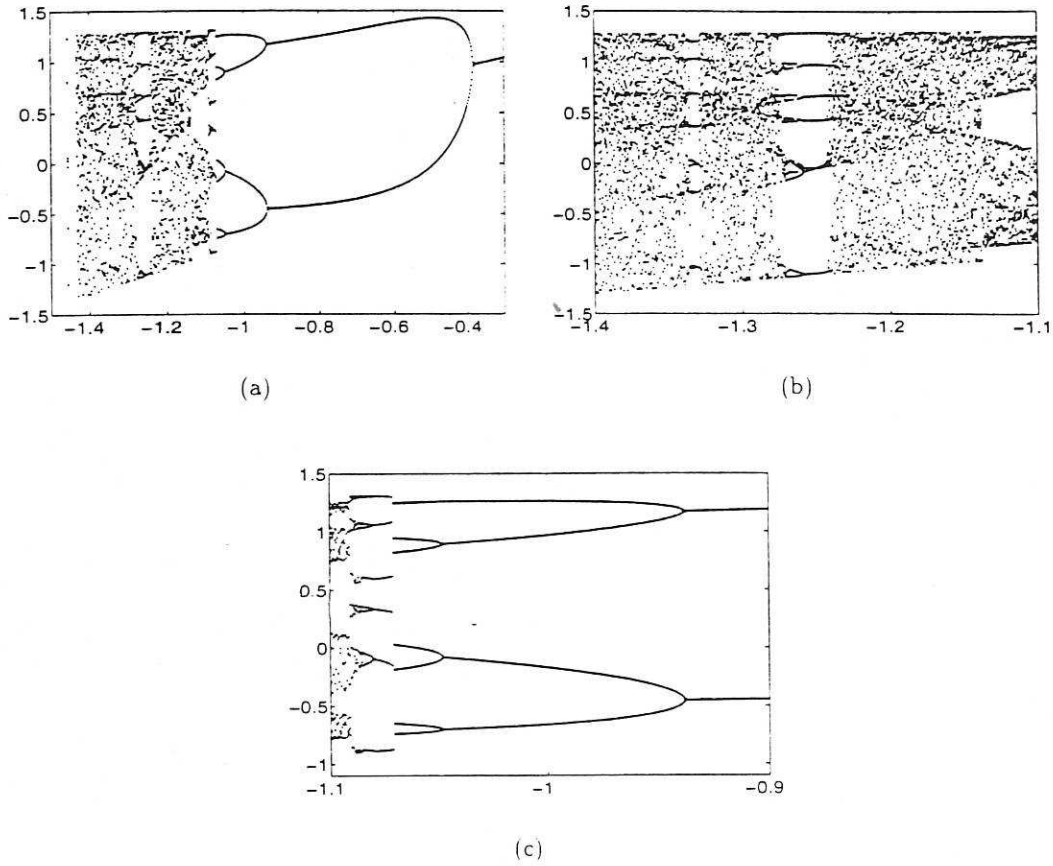
17

Figure 5: Bifurcation diagram of Model B. equation (22): (a) the parameter β is varied from -1.5 to -0.3. (b) Zoom over the region between -1.4 and -1.1 and (c) Zoom over the region between -1.1 and -0.9.

$$\begin{cases} \dot{x} = \alpha(y - h(x)) \\ \dot{y} = x - y + z \\ \dot{z} = -\beta y \end{cases} , \quad h(x) = \begin{cases} m_1 x + (m_0 - m_1) & x \geq 1 \\ m_0 x & |x| \leq 1 \\ m_1 x - (m_0 - m_1) & x \leq -1, \end{cases} \tag{23}$$

where $\alpha = 9.0$, $\beta = 100/7$, $m_0 = -1/7$ and $m_1 = 2/7$. For these values of parameters the system settles to the so-called double scroll attractor. Equation (23) was simulated using a Runge-Kutta of 4th-order with step size 0.001. The data from the z-coordinate were then sampled at $T_s = 0.15$. It has been shown that when the attractor settles in the well-known double scroll. it is possible to determine the *effective* clusters precisely (Mendes and Billings, 1996a). It was also shown that the fixed points can easily be identified directly from data by, for instance, using the procedure of Glover and Mees (1993). The same results could be achieved when NARMAX models are fitted to

18

the data. This is the basis of the fixed point approach.

To determine ℓ for the double scroll, polynomial NARMAX models with increasing number of terms and degrees were identified from 1801 noise-free data points. In this example the structure of such models was automatically selected by the ERR criterion. All the models identified were of dimension $n_y = 3$ which is the minimum dimension to describe Chua's circuit. The results will remain qualitatively the same even when higher dimensions are used. Tables (2), (3) and (4) show the values of the fixed points for ℓ varying from 2 to 4. At this point, it is worth mentioning that the true fixed points, despite being easily calculated from equation (23), were calculated using the procedure of Gloves and Mees described in section (2.3.1). For the polynomial models the fixed points can readily be calculated using equation (12). For $\ell = 2$, the fixed points diverge indicating that the OLS algorithm could not determine the real fixed points. This situation does not occur for the fixed points of Table (3) and (4). Clearly two fixed points hardly vary even with an increasing number of terms suggesting that they pertain to the original set of fixed points. To detect the trivial fixed point, only the cluster Ω_0 needs be inferred (Aguirre and Mendes, 1996; Mendes and Billings, 1996a). Verifying the cancelling effect is obviously not a valid procedure for such a cluster since it is a singleton. A solution to this problem is now suggested. From the two fixed points already detected the symmetry of the double scroll can be reviewed. A necessary condition to have such symmetry is that the coefficient of cluster Ω_0 equals zero. To further confirm this, Tables (3) and (4) show that the constant term is included only after 16 terms have been selected and has a small coefficient.

n_θ	Fixed Points		
10	-0.03913	4.635	-
9	-0.2796	3.588	-
8	-0.3809	2.703	-
7	-0.08267	5.458	-
6	-0.1589	53.92	-
TRUE	0	1.5	-1.5

Table 2: Fixed Points for models estimated from the z coordinate of Chua's circuit, equation (23), with the degree of nonlinearity $\ell = 2$. n_θ is the number of model terms. The true fixed points were calculated using the procedure of Glover and Mees described in section (2.3.1). For the polynomial models the fixed points can easily be calculated using equation (12).

The last column of Table (4) shows that a spurious fixed point is introduced in all models of degree of nonlinearity $\ell = 4$ due to the the presence of the cluster $\Omega_{,3}$. It can readily be noticed that such fixed points are placed beyond the range of the data (-3.18,3.1761) and change considerably with the number of terms indicating that this fixed point is indeed spurious. Therefore $\ell = 3$ appears to be the most appropriate degree of nonlinearity for the models. Following this procedure *fixed-point overparametrization* is therefore avoided.

n_θ	Fixed Points		
20	-0.004343	-1.542	1.542
19	-0.001639	-1.542	1.543
18	-0.001513	-1.542	1.542
17	-0.001406	-1.542	1.542
16	0	-1.542	1.543
15	0	-1.542	1.543
14	0	-1.542	1.543
13	0	-1.542	1.542
12	0	-1.537	1.537
11	0	-1.53	1.53
10	0	-1.53	1.53
9	0	-1.527	1.527
8	0	-1.582	1.582
7	0	-1.524	1.524
6	0	-1.547	1.547
TRUE	0	-1.5	1.5

Table 3: Fixed Points for models estimated from the z coordinate of Chua's circuit, equation (23), with the degree of nonlinearity $\ell = 3$. The true fixed points were calculated using the procedure of Glover and Mees described in section (2.3.1). For the polynomial models the fixed points can easily be calculated using equation (12).

The properties used above to distinguish *true* fixed points from *spurious* fixed points are lost when noise is added to the data. Table (5) shows the fixed points of models identified from noise contaminated data with a signal to noise ratio of 42 dB ($\approx 20\log_{10}(\sigma_z^2/\sigma_e^2)$). Since the only fixed point whose value varies considerably is the fixed point listed in the first column, it seems reasonable to suppose that the cluster Ω_0 is *spurious* (This is in perfect agreement with the results presented in (Aguirre and Mendes, 1996)). All the values of the remaining fixed points are largely unchanged. However it can be noticed that the fixed point shown in the fourth column of Table (5) is placed outside the range of the data and must therefore be *spurious*. This is in perfect agreement with the results presented in (Mendes and Billings, 1996a). The reason for this is that the data do not have information about fixed points located outside of the range. It is worth stressing that the noise added to the data has brought the *spurious* fixed points closer to the *true* ones. This effect has been observed in numerous simulations and causes distortions to bifurcation diagrams.

The symmetry can be detected in table (5) from the fixed points of columns two and three. The model terms automatically selected by the ERR criterion indicates that, when n_θ is equal to 6 or 7, the cluster Ω_{z^2} is not present in the model and as a result the fixed points are symmetrical around the origin. As soon as more terms are included in the model, the symmetry is lost. However the values of the fixed points remain almost constant which indicates that the cluster Ω_{z^2} is not

n_θ	Fixed Points			
21	-0.002234	-1.542	1.542	-1130
20	-0.002073	-1.542	1.542	-1044
19	-0.001364	-1.542	1.543	-5098
18	-0.001513	-1.542	1.542	-
17	-0.001406	-1.542	1.542	-
16	0	-1.542	1.543	-
15	0	-1.542	1.543	-
14	0	-1.542	1.543	-
13	0	-1.542	1.542	-
12	0	-1.537	1.537	-
11	0	-1.53	1.53	-
10	0	-1.53	1.53	-
9	0	-1.527	1.527	-
8	0	-1.582	1.582	-
7	0	-1.524	1.524	-
6	0	-1.547	1.547	-
TRUE	0 -	-1.5	1.5	-

Table 4: Fixed Points for models estimated from the z coordinate of Chua's circuit, equation (23), with the degree of nonlinearity $\ell = 4$. The true fixed points were calculated using the procedure of Glover and Mees described in section (2.3.1). For the polynomial models the fixed points can easily be calculated using equation (12).

necessary and can be eliminated.

Finally it is worth mentioning that the results presented above show that the orthogonal estimator can be taken as a procedure to determine the location and number of the fixed points. Simulations have shown that even when the data are corrupted by noise the orthogonal estimator is able to identify the location of the fixed points directly from the data.

□

Example 3.4

This example demonstrates that the fixed-point approach is similar to cluster cancellation. Though some differences should be pointed out. Consider the following model (Aguirre and Mendes, 1996)

$$
\begin{aligned}
y(k) = \ & 1.4269\, y(k-1) - 0.41549\, y(k-3) + 0.012\, y(k-2) \\
& + 0.11736\, u(k-3) - 0.04904\, y(k-1)^3 + 1.2007\, y(k-1)^2 u(k-3) \\
& + 0.252\, y(k-3)^2 u(k-2) - 0.078346\, u(k-2) - 0.47759\, y(k-2)y(k-3)u(k-3) \\
& - 0.030695\, y(k-3)^3 + 0.05843\, y(k-2)^3 - 0.39072\, y(k-2)^2 u(k-3)
\end{aligned}
$$

n_θ	Fixed Points			
21	-0.552	1.59	-1.627	12.08
20	-0.5474	1.581	-1.641	12.03
19	-0.5228	1.582	-1.637	12.36
18	-0.4501	-1.588	1.6	15.35
17	-0.394	1.579	-1.619	15.16
16	-0.3499	1.548	-1.68	13.76
15	-0.3391	1.544	-1.684	13.79
14	-0.3461	1.525	-1.662	13.75
13	-0.3376	1.542	-1.684	13.59
12	-0.3323	1.543	-1.681	13.61
11	-0.4019	1.561	-1.664	15.35
10	-0.4261	1.58	-1.63	15.34
9	-0.3298	-1.509	1.653	-
8	-0.1554	-1.506	1.661	-
7	0	1.601	-1.601	-
6	0	1.615	-1.615	-

Table 5: Fixed Points for models estimated from the z coordinate of Chua's circuit, equation (23), with the degree of nonlinearity $\ell = 4$. Noise contaminated data were used for the identification.

$$-1.0272\, y(k-1)^2 u(k-2) + 0.44085\, y(k-2)y(k-3)u(k-1) - 0.20771 \times 10^{-2}$$
$$+0.032643\, y(k-1)^2 - 0.054208\, y(k-2)^2 + 0.023113\, y(k-3)^2 \tag{24}$$

which was estimated based on data generated by the Duffing-Holmes oscillator (Holmes, 1979)

$$\dot{y} = x,$$
$$\dot{x} = -0.15\, y + x - x^3 + u(t) , \tag{25}$$

with a signal to noise ratio of 54dB ($\approx 20 \log_{10}(\sigma_y^2/\sigma_e^2)$) and an input $u(t) = 0.3\, cos(t)$.

In (Aguirre and Mendes, 1996) it was shown that the cluster coefficients of model (24) are $\Sigma_y = 1.0234$, $\Sigma_{y^3} = -2.1305 \times 10^{-2}$, $\Sigma_u = 3.9019 \times 10^{-2}$, $\Sigma_{y^2 u} = -1.9184 \times 10^{-3}$, $\Sigma_0 = -2.0771 \times 10^{-3}$ and $\Sigma_{y^2} = 1.54771 \times 10^{-3}$. From the analysis of these cluster coefficients it seems reasonable to suppose that the *spurious* clusters are Ω_0, $\Omega_{y^2 u}$, and Ω_{y^2}. Indeed, the fact that Ω_y, Ω_u and Ω_{y^3} seems to be the effective clusters[2] can be verified from extensive simulations and calculations of bifurcation diagrams and also from higher-order spectral analysis (Chandran *et al.*, 1993). although the higher-order spectral analysis only detects properties exhibited in the data. The fixed-point approach leads

---

[2]Note that the coefficients of the clusters are ten times smaller than the coefficients of the true clusters.

to similar results but nothing can be said about the cluster $\Omega_{y^2 u}$. Such a cluster vanishes completely when fixed points are calculated. However it has been shown in (Mendes and Billings, 1996b) that this cluster results from the process of discretization. Models with this cluster can also reproduce the bifurcation diagram quite well.

$\square$

Example 3.5

The following polynomial model was estimated from a set of 300 data points with a signal to noise ratio of 50 dB generated by the Hénon Map (Hénon, 1976) (See equation (20))

$$
\begin{aligned}
x(k) =\ & -0.22005 \times 10^{-1} x(k-1)x(k-1)x(k-3) + 0.71496 \times 10^{+0} \\
& -0.74714 \times 10^{+0} x(k-1)x(k-1) + 0.31391 \times 10^{+0} x(k-2) \\
& -0.23093 \times 10^{-1} x(k-2)x(k-2)x(k-2) + 0.37603 \times 10^{+0} x(k-2)x(k-2) \\
& +0.84738 \times 10^{+0} x(k-1)x(k-2)x(k-2) - 0.81355 \times 10^{-1} x(k-3) \\
& -0.17482 \times 10^{+0} x(k-1)x(k-3) - 0.3471 \times 10^{+0} x(k-1) \\
& +\Psi_\xi^T(k-1)\hat{\Theta}_\xi + \xi(k)
\end{aligned}
\tag{26}
$$

This map has fixed points at (1.22,0.6262,-1.166) whereas the Hénon Map (equation (20)) has fixed points at (-1.131,0.631). The fact that the map in equation (26) has one fixed point in excess can be justified by the appearance of the cluster Ω_{x^3} in the model structure.

Figures (6a-b) show the first return map and bifurcation diagrams of the map in equation (26). When compared to Figure (2), the first return map is quite similar, whereas the bifurcation diagrams show significant differences which are a consequence of overparametrization. Equation (26) is overparametrized in two different ways. Firstly, the degree of nonlinearity is $\ell = 3$ instead of $\ell = 2$ and secondly, the order of the system is $n_x = 3$ instead of $n_x = 2$. As a consequence of these two aspects of overparametrization, the total number of terms n_θ (parameters) of the estimated model in equation (26) is more than three times larger than the number of terms in the Hénon map.

It should be noted that the two worst effects of overparametrization are present in equation (26): *i*) a spurious fixed point within the data range and *ii*) a spurious Lyapunov exponents less negative than the actual Lyapunov exponent which leads to a larger Lyapunov dimension ($D_L = 1.3778$). Models with such effects can be avoided by using the fixed point procedure.

In order to illustrate this, consider models with an increasing number of terms identified from the noisy Hénon data. The order and degree of nonlinearity are both 3. The Lyapunov spectrum is shown in Table (6). It can be noticed that the spurious Lyapunov exponent (third column) is roughly twice the true one (second column) in all models. This is a symptom of *dimension*
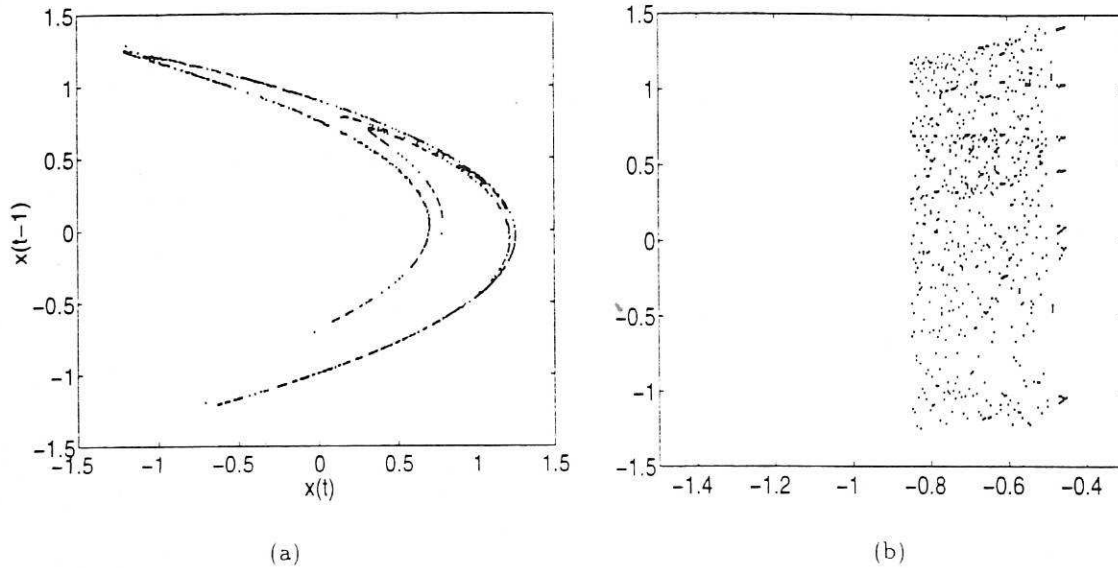
23

Figure 6: (a) first return map of equation (26). and (b) bifurcation diagram of equation (26) with minus the coefficient of the term $x(k-1)^2$ taken as the bifurcation parameter.

overparametrization. The spurious fixed point can easily be detected by its values when the number of terms varies. Table (7) shows this variation. The conclusion is therefore that $\ell = 2$ is the correct degree of nonlinearity. Finally Table (8) shows the behaviour of the poles of the linear part. Clearly after 8 terms the poles are placed at a completely different location which might explain why the bifurcation diagram of Figure (6-b) is so contrasting.

□

4   A glance at overparametrization of MIMO nonlinear systems

The objective of this section is to show briefly that overparametrization as defined in previous sections also occurs in MIMO discrete nonlinear models.

Instead of formally introducing the cluster ideas for multivariable systems. the following simple model is provided in order to illustrate how to define the clusters for such systems.

$$
\begin{aligned}
y_1(k) \;=\; & -0.37862\times10^{+1}\, y_1(k-1) + 0.40749\times10^{+0}\, y_2(k-1) + 0.4786\times10^{+1}\, y_1(k-2) \\
& -0.27366\times10^{-3}\, y_1(k-1)y_1(k-1)y_2(k-1) + 0.19833\times10^{+0}\, y_2(k-2) \\
& +0.10322\times10^{-3}\, y_1(k-2)y_1(k-2)y_2(k-2) + 0.29222\times10^{-6}\, y_2(k-1)y_2(k-1)y_2(k-2) \\
& +0.53665\times10^{-5}\, y_1(k-1)y_2(k-2)y_2(k-2) - 0.40865\times10^{-4}\, y_1(k-1)y_2(k-1)y_2(k-1)
\end{aligned}
$$

24

n_θ	Lyapunov Exponents		
10	0.3525	-0.8868	-1.618
9	0.3647	-0.5777	-1.495
8	0.3866	-0.8512	-1.583
7	0.3404	-1.606	-2.63
6	0.3561	-1.884	-3.207
5	0.3276	-1.854	-3.161
4	0.34	-1.83	-3.354

Table 6: Lyapunov Exponents calculated by iterating models estimated from noise contaminated data of the Hénon map (equation (20)). All models have $n_y = 3$, $\ell = 3$.

n_θ	Fixed Points		
10	0.6262	-1.166	1.22
9	0.6301	-1.182	0.895
8	0.6354	-1.309	3.789
7	0.6352	-1.23	-76.37
6	0.6293	-1.191	-20.09
5	0.6301	-1.185	-20.79
4	0.6271	-1.166	-57.96

Table 7: Fixed Points for models estimated from noise contaminated data of the Hénon map (equation (20)). All models have $n_y = 3$, $\ell = 3$

$$+0.17269 \times 10^{-3} y_1(k-1)y_1(k-1)y_2(k-2) \tag{27}$$

$$
\begin{aligned}
y_2(k) = \ & +0.76566 \times 10^{+1} y_2(k-1) + 0.4659 \times 10^{+1} y_2(k-2) \\
& -0.14696 \times 10^{-1} y_1(k-1)y_1(k-1)y_2(k-1) - 0.60014 \times 10^{-2} y_1(k-2)y_1(k-2)y_2(k-2) \\
& -0.11867 \times 10^{-2} y_1(k-1)y_2(k-1)y_2(k-1) + 0.68294 \times 10^{-2} y_1(k-1)y_1(k-2)y_2(k-1) \\
& -0.1081 \times 10^{+3} y_1(k-1) + 0.1081 \times 10^{+3} y_1(k-2) \\
& +0.14137 \times 10^{-1} y_1(k-1)y_1(k-1)y_2(k-2) \\
& -0.9748 \times 10^{-3} y_1(k-2)y_2(k-2)y_2(k-2)
\end{aligned}
\tag{28}
$$

where y_1 and y_2 are the outputs of the system. The cluster terms for model (27), which generates the output y_1, are defined as follows:

n_θ	Poles		
10	-0.8377	0.2453-0.1922i	0.2453+0.1922i
9	-0.7351	0.3676-0.2903i	0.3676+0.2903i
8	-0.7001	0.35-0.1416i	0.35+0.1416i
7	0	0.5715	-0.5715
6	0	0.5704	-0.5704
5	0	0.5679	-0.5679
4	0	0.5432	-0.5432

Table 8: Poles of the linear transfer function of models estimated from noise contaminated data of the Hénon map (equation (20)). All models have $n_y = 3$, $\ell = 3$

$$
\begin{aligned}
\Omega_{y_1} &= \left(y_1(k-1).y_1(k-2) \right), \\
\Omega_{y_1} &= \left(y_2(k-1).y_2(k-2) \right), \\
\Omega_{y_1^2 y_2} &= \left(y_1(k-1)y_1(k-1)y_2(k-1).y_1(k-2)y_1(k-2)y_2(k-2).y_1(k-1)y_1(k-1)y_2(k-2) \right), \\
\Omega_{y_2^3} &= \left(y_2(k-1)y_2(k-1)y_2(k-2) \right). \\
\Omega_{y_1 y_2^2} &= \left(y_1(k-1)y_2(k-2)y_2(k-2).y_1(k-1)y_2(k-1)y_2(k-1) \right),
\end{aligned}
\tag{29}
$$

The cluster coefficients can be obtained by summing the coefficients of terms within the cluster defined in equation (29). The cluster terms and cluster coefficients for model (28) are defined likewise. It is interesting to note that if the maximum lag of both model (27) and model (28) were $n_y = n_{y_1} = n_{y_2} = 1$, each term would represent a cluster.

Although the cluster ideas can easily be extended to multivariable systems, there is a lack of simple methods which relate cluster and fixed point. For this reason no attempt will be made to discuss how the cluster affects the presence of *spurious* fixed points.

Calculating the fixed points for SISO models has been shown to be a relatively easy task. Unfortunately MIMO models do not present such nice properties. In this particular case, a set of nonlinear equations must be solved to obtain the fixed points. Although numerical methods can be used the convergence depends upon how good the initial guess is. Furthermore, finding all fixed points may be troublesome. Gröbner basis (Cox *et al.*, 1992) however constitute an alternative without the setback of numerical methods. Briefly, Gröbner basis transform the initial set of nonlinear equations into a new set where typically one of the nonlinear equations is a function of a single variable and can therefore easily be solved using the same approach applied for SISO models. This explanation is rather simplistic, but nevertheless it is enough to understand the

26

results presented in this section. For a thorough explanation of Gröbner basis refer to the excellent book by Cox *et al.* (1992). Although Gröbner basis can provide an analytical solution in many cases, the requirement of computer memory can be very demanding even for simple models.

Example 4.1

Consider the set of equations governing the dynamics of Lorenz's equations (Lorenz, 1963):

$$\begin{cases} \dot{x} &= \pi(y - x) \\ \dot{y} &= \rho x - y - xz \\ \dot{z} &= xy - 3z \end{cases} \tag{30}$$

where τ, ρ and 3 are constants. Setting $\dot{x}$, $\dot{y}$ and $\dot{z}$ equal zero, the fixed points can be calculated by solving a system of nonlinear equations. Fortunately this system can easily be solved by just computing the Gröbner basis which are:

$$\begin{cases} 0 &= -y + x \\ 0 &= y^2 - 3z \\ 0 &= yz + (1 - \rho)y \\ 0 &= (1 - \rho)z + z^2 \end{cases} \tag{31}$$

Note that the last equation is a function of just one variable, i. e., $f(z)$. The fixed points are then:

$$\bar{z} = \left\{ \begin{array}{c} 0 \\ \rho - 1 \end{array} \right. \quad \rightarrow \quad \bar{y} = \left\{ \begin{array}{c} 0 \\ \pm\sqrt{3(\rho - 1)} \end{array} \right. \quad \rightarrow \quad \bar{x} = \left\{ \begin{array}{c} 0 \\ \mp\sqrt{3(\rho - 1)} \end{array} \right. \tag{32}$$

Setting $\tau = 10$, $\rho = 28$ and $3 = 8/3$, the fixed points become:

$$\bar{z} = \left\{ \begin{array}{c} 0 \\ 27 \end{array} \right. \quad \rightarrow \quad \bar{y} = \left\{ \begin{array}{c} 0 \\ \pm 6\sqrt{2} \end{array} \right. \quad \rightarrow \quad \bar{x} = \left\{ \begin{array}{c} 0 \\ \mp 6\sqrt{2} \end{array} \right. \tag{33}$$

Equation (30) is now used to generate data for the identification. From the resultant data sampled at $T_s = 0.01$, 6001 points were used to identify the model

$$
\begin{aligned}
x(k) &= +0.91802 \times 10^{+0}x(k-1) + 0.95132 \times 10^{-1}y(k-1) \\
&\quad -0.46439 \times 10^{-3}x(k-1)z(k-1) \\
&\quad -0.26686 \times 10^{-4}y(k-1)z(k-1) \\
y(k) &= +0.10041 \times 10^{+1}y(k-1) - 0.82126 \times 10^{-3}y(k-1)z(k-1) \\
&\quad -0.93900 \times 10^{-2}x(k-1)z(k-1) + 0.26837 \times 10^{+0}x(k-1) \\
z(k) &= +0.97605 \times 10^{+0}z(k-1) + 0.84948 \times 10^{-2}x(k-1)y(k-1) \\
&\quad +0.10934 \times 10^{-2}y(k-1)y(k-1) \\
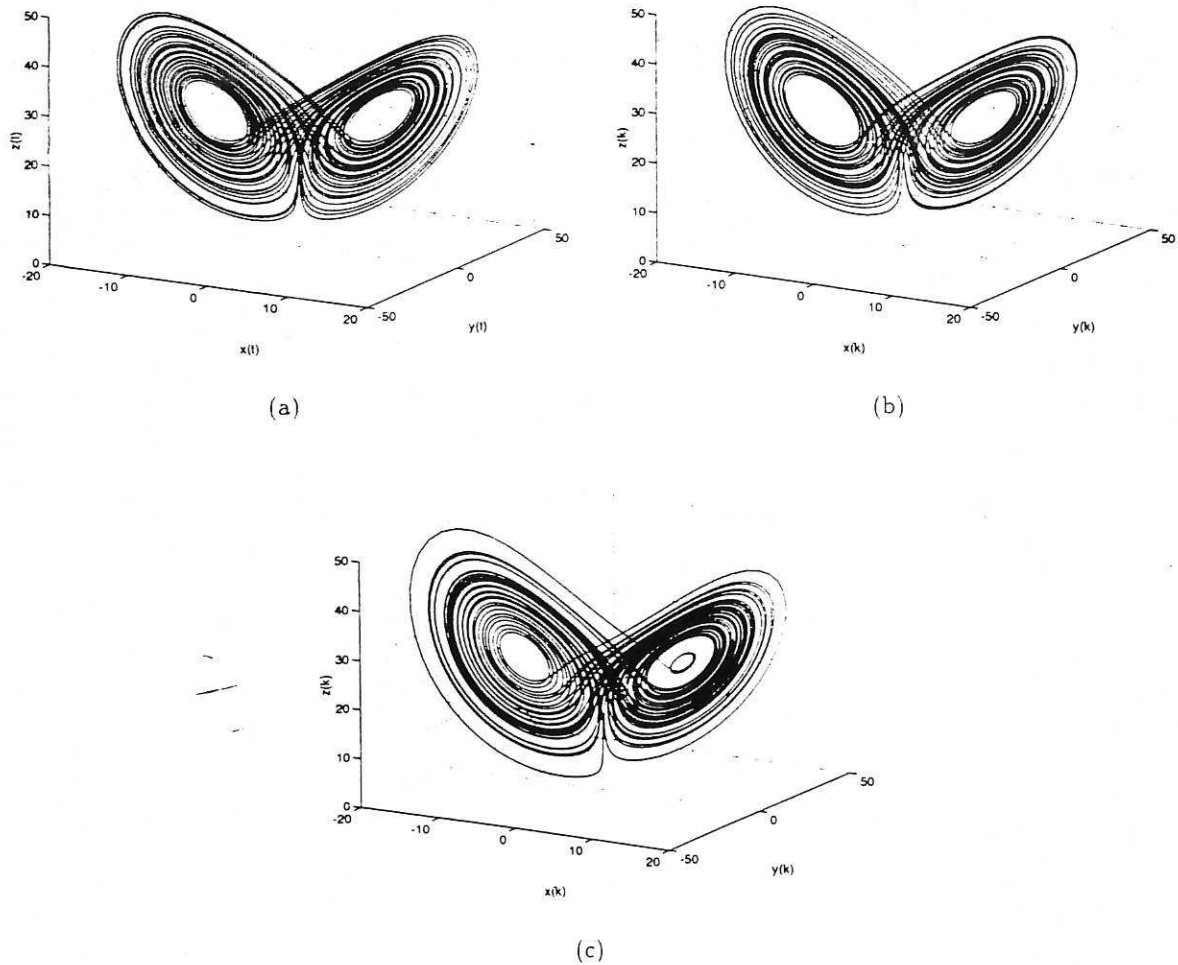&\quad -0.89652 \times 10^{-4}z(k-1)z(k-1)
\end{aligned}
\tag{34}
$$

(a)

(b)

(c)

Figure 7: The Lorenz attractor (a) original. equation 30). (b) reconstructed from the model identi-fied from the noise-free data. equation (34). (c) reconstructed from the model identified from the noisy data. equation (35).

The model obtained can be iterated to obtain time series for the three components x, y and z. Figure (7) shows the comparison between the original attractor and the reconstructed one. Clearly the identified model reproduces the chaotic behaviour very well.

From equation (34) and by using the Gröbner basis all the fixed points of the discrete model were found. Table (9) shows that the fixed points are in good agreement with the system fixed points, but the model has four *spurious* fixed points which indicates *fixed-point overparametrization*. For MIMO models such overparametrization cannot be avoided so easily as for SISO models.

Considering a more realistic case, white noise was added to the data. The resultant data had a signal noise ratio of $\approx$ 37 dB and were used to identify the model

$$
\begin{aligned}
x(k) = & +0.90934 \times 10^{+0}x(k-1) + 0.10310 \times 10^{+0}y(k-1) \\
& -0.19078 \times 10^{-3}x(k-1)z(k-1) - 0.31794 \times 10^{-3}y(k-1)z(k-1) \\
& +\Psi^T_{\xi_x\xi_y\xi_z}(k-1)\hat{\Theta}_{\xi_x\xi_y\xi_z} + \xi_x(k) \\
y(k) = & +0.10121 \times 10^{+1}y(k-1) - 0.99824 \times 10^{-3}y(k-1)z(k-1) \\
& -0.89927 \times 10^{-2}x(k-1)z(k-1) + 0.25357 \times 10^{+0}x(k-1) \\
& +\Psi^T_{\xi_x\xi_y\xi_z}(k-1)\hat{\Theta}_{\xi_x\xi_y\xi_z} + \xi_y(k) \\
z(k) = & +0.97139 \times 10^{+0}z(k-1) + 0.11904 \times 10^{-1}x(k-1)y(k-1) \\
& +0.18035 \times 10^{-3}z(k-1)z(k-1) - 0.30014 \times 10^{-2}x(k-1)x(k-1) \\
& +\Psi^T_{\xi_x\xi_y\xi_z}(k-1)\hat{\Theta}_{\xi_x\xi_y\xi_z} + \xi_z(k)
\end{aligned}
\tag{35}
$$

Lorenz equations (30)	Model (34)
(0,0,0)	(0,0,0)
(8.4853,8.4853,27)	(8.5647,8.5601,26.6802)
(−8.4853,−8.4853,27)	(−8.5647,−8.5601,26.6802)
*	(−319.7713,3667.7845,−7411.0984)
*	(319.7713,−3667.7845,−7411.0984)
*	(0.1263,0.0132,−267.1929)
*	(−0.1263,−0.0132,−267.1929)

Table 9: Comparison between the original fixed points and the fixed points of model (34). Noise-free data of Lorenz equations (30) were used for the identification.

The model of equation (35) also exhibits dynamical invariants close to the original (See Figure (7) for the attractor and Table (10) for the comparison between the original fixed points and the fixed points of model (35). The spurious values were not calculated since they are similar to the

noise-free case. However it is worth stressing that the models identified from noisy data tend to have the spurious fixed points closer to the *real* ones which agrees with the examples shown for monovariable systems.

Lorenz equations (30)	Model (35)
(0.0.0)	(0,0.0)
(8.4853,8.4853.27)	(8.3679,8.4638,26.5759)
(−8.4853,−8.4853.27)	(8.3679,-8.4638,26.5759)

Table 10: Comparison between the original fixed points and the fixed points of model (35). Noise contaminated data of Lorenz equations (30) were used for the identification.

Finally, even though the relation between clusters and fixed points has not been established for MIMO models, the significance of the specific cluster Ω_o, that is, the constant term, can be determined directly from the data. The procedure of Glover and Mees described in section (2.3.1) can be used for this purpose. For the Lorenz data the results showed that a constant term (cluster Ω_o) is not necessary and can be eliminated from the model. The same conclusion can be drawn by noticing the symmetry inherent in the fixed points. An introduction of the cluster Ω_o would break this symmetry. $\square$

5 Conclusions

This paper has introduced two new types of overparametrization of nonlinear discrete systems. Fixed point and dimension overparametrization were defined according to the spurious dynamics introduced into the model. It has been shown how overparametrization of nonlinear models affects the number and location of the respective fixed points and how it affects the number of Lyapunov exponents. Furthermore whenever there are spurious term clusters in the model, the fixed points of the model will probably be misplaced. Similar results have also been observed when only terms from correct clusters are included in the model but the number of terms is excessive especially when the model is *dimension overparametrized*.

A new procedure referred to as the fixed-point approach for improving structure detection has been introduced to detect and avoid overparametrization. The fixed-point approach uses all the information from cluster cancellation and location of the fixed points. The approach often indicates the degree of nonlinearity ℓ if the location and values of the fixed points can be extracted from the data. Although the fixed-point approach is based upon the cluster ideas it has been shown that this approach can be used in the structure selection of discrete nonlinear systems identified from both sampled continuous systems and from a "pure" discrete systems. Numerous examples were given to illustrate the various types of overparametrization and the new procedures.

Finally the concept of fixed-point overparametrization has been extended to multivariable systems. Although no clear connection between the cluster ideas and fixed points has been established for these models it has been demonstrated that it is still possible to detect some spurious clusters.

Acknowledgements

References

Abarbanel, H. D. I. and Kennel, M. B. (1993). Local false nearest neighbors and dynamical dimensions from observer chaotic data. *Phys. Rev. E*, **47**(5), 3057–3068.

Abarbanel, H. D. I. and Sushchik, M. M. (1993). Local or dynamical dimension of nonlinear systems inferred from observations. *Int. J. Bif. Chaos*, **3**(3), 543–550.

Aguirre, L. A. and Billings, S. A. (1994). Discrete reconstruction of strange attractors in Chua's circuit. *Int. J. Bif. Chaos*, **4**(4), 853–864.

Aguirre, L. A. and Billings, S. A. (1995). Improved structure selection for nonlinear models based on term clustering. *Int. J. Control*, **62**, 569–587.

Aguirre, L. A. and Mendes, E. M. A. M. (1996). Global nonlinear polynomial models: Structure, Term cluster and Fixed points. *Int. J. Bif. Chaos*, **6**(2), 279–294.

Billings, S. A. and Fadzil, M. B. (1985). The practical identification of systems with non-linearities. In *7-th IFAC Symposium on Identification and System Parameter Estimation*, pages 155–160.

Billings, S. A. and Leontaritis, I. J. (1981). Identification of nonlinear systems using parameter estimation techniques. In *Proc. IEE Conf. Control and Its Applications*, pages 183–187, Warwick, U.K.

Billings, S. A. and Leontaritis, I. J. (1982). Parameter estimation techniques for nonlinear systems. In *Proc. 6th IFAC Symposium on Identification and System Parameter Estimation*, pages 505–510, Washington, DC.

Billings, S. A., Korenberg, M., and Chen, S. (1988). Identification of non-linear output-affine systems using an orthogonal least-squares algorithm. *Int. J. Systems Sci.*, **19**(8), 1559–1568.

Billings, S. A., Chen, S., and Backhouse, R. J. (1989). Identification of linear and non-linear models of a turbocharged automotive diesel engine. *Mechanical Systems and Signal Processing*, **3**(2), 123–142.

Box, G. E. P. and Jenkins, G. M. (1976). *Time Series analysis forescating and control*. Holden-Day, San Francisco.

Briggs, K. (1990). An improved method for estimating Liapunov exponents of chaotic time series. *Physics Letters A*, **151**(1/2), 27–32.

Brown, R., Bryant, P., and Abarbanel. H. D. I. (1991). Computing the Lyapunov spectrum of a dynamical system from an observed time series. *Phys. Rev. A*, **43**, 2787–2806.

Bryant, P., Brown, R., and Abarbanel. H. D. I. (1990). Lyapunov exponents from observed time series. *Phys. Rev. Lett.*, **65**(13), 1523–1526.

Chandran. V., Elgar. S., and Pezeshki, C. (1993). Bispectral and trispectral characterization of transition to chaos in the Duffing oscillator. *Int. J. Bif. Chaos*, **3**(3), 551–557.

Chen, S. and Billings. S. A. (1989). Representations of non-linear systems: the NARMAX model. *Int. J. Control*, **49**(3), 1013–1032.

Chua, L. O. and Hasler. M. (1993). (Guest Editors). Speciall issue on Chaos in nonlinear electronic circuits. *IEEE Trans. Circuits Syst.*, **40**(10–11).

Cox, D., Little. J., and O'Shea, D. (1992). *Ideals. varieties, and algorithms, an introduction to computational algebraic geometry and commutative algebra*. Undergraduate texts in mathematics. Springer Verlag, New York, London.

Eckmann. J. P. and Ruelle. D. (1985). Ergodic theory of chaos and strange attractors. *Rev. Mod. Phys.*, **57**, 617–656.

Eckmann. J. P., Kamphorst. S. O., Ruelle, D., and Ciliberto. S. (1986). Liapunov exponents from time series. *Phys. Rev. A*, **34**(6), 4971–4979.

Forsman, K. (1991). *Constructive Commutative Algebra in Nonlinear Control Theory*. Ph.D. thesis, Department of Electrical Engineering, Linköping University, S-581 83 Linköping, Sweden.

Gencay, R. and Dechert, W. D. (1992). An algorithm for the n Lyapunov exponents of an n-dimensional unknown dynamical system. *Physica D*, **59**, 142–157.

Glover, J. and Mees, A. I. (1993). Reconstructing the dynamics of chua's circuit. *Journal of Circuits, Systems, and Computers*, **3**(1), 201–214.

Hénon, M. (1976). A two-dimensional mapping with a strange attractor. *Commun. Math. Phys.*, **50**, 69–77.

Holmes, P. J. (1979). A nonlinear oscillator with a strange attractor. *Philos. Trans. Royal Soc. London A*, **292**, 419–448.

Judd, K. and Mees, A. I. (1994). On selecting models for nonlinear time series. Technical report, Department of Mathematics, University of Western Australia.

Kadtke, J. B., Brush, J., and Holzfuss, J. (1993). Global dynamical equations and Lyapunov exponents from noisy chaotic time series. *Int. J. Bif. Chaos*, **3**(3), 607–616.

Kantz, H. (1994). A robust method to estimate the maximal Lyapunov exponent of a time series. *Phys. Rev. A*, **185**, 77–87.

Katdke, J. B., Brush, J., and Holzfuss, J. (1993). Global dynamical equations and Lyapunov exponents from chaotic time series. *Int. J. Bif. Chaos*, **3**(3), 607–616.

Kennel, M. B. and Abarbanel, H. D. I. (1994). False neighbors and false strands: A reliable mininum embedding dimension algorithm. Technical report, Institute of Nonlinear Science, University of California, San Diego, La Jolla, CA 92093-0402.

Kennel, M. B., Brown, R., and Abarbanel, H. D. I. (1992). Determining embedding dimension for phase space reconstruction using the method of false nearest neighbors. Technical report, Institute of Nonlinear Science, University of California, San Diego, La Jolla, CA 92093-0402.

Kruel, T. M., Eiswirth, M., and Schneider, F. W. (1993). Computation of Lyapunov spectra: Effect of interactive noise and application to a chemical oscillator. *Physica D*, **63**, 117–137.

Leontaritis, I. J. and Billings, S. A. (1985). Input-output parametric models for non-linear systems – part i: deterministic non-linear systems. *Int. J. Control*, **41**(2), 303–328.

Lorenz, E. N. (1963). Deterministic nonperiodic flow. *Journal of The Atmospheric Sciences*, **20**, 131–141.

Mees, A. (1993a). Parsimonious dynamical reconstruction. *Int. J. Bif. Chaos*, **3**(3), 669–675.

Mees, A. I. (1993b). Parsimonious dynamical reconstruction. *Int. J. Bifurcation and Chaos*, **3**(3), 669–675.

Mees, A. I. and Judd, K. (1995). Parsimony in dynamical systems. Technical report, Department of Mathematics, University of Western Australia.

Mendes, E. M. A. M. and Billings, S. A. (1996a). On identifying global nonlinear discrete models from chaotic data. *Int. J. Bif. Chaos*.

Mendes, E. M. A. M. and Billings, S. A. (1996b). Discretizing a nonlinear analytic system: Consequences on the model structure problem. *(In preparation)*.

Parlitz, U. (1992). Identification of true and spurious Lyapunov exponents. *Int. J. Bif. Chaos*, **2**, 155–165.

Parlitz, U. (1993). Lyapunov exponents from chua's circuit. *Journal of Circuits, Systems, and Computers*, **3**(2), 507–523.

Sano, M. and Sawada, Y. (1985). Measurement of the Lyapunov spectrum from a chaotic time series. *Phys. Rev. Lett.*, **55**, 1082–1085.

Sato, S., Sano, M., and Sawada, Y. (1987). Practical methods of measuring the generalized dimension and largest Lyapunov exponent in high dimensional chaotic systems. *Progress of Theorical Physics*, **77**(1), 1–5.

Söderström, T. and Stoica, P. (1989). *System Identification*. Prentice Hall, London.

Tong, H. (1992). Some comments on a bridge between nonlinear dynamicists and statisticians. *Physica D*, **58**, 299–303.

Ueda, Y. (1985). Randomly phenomena resulting from nonlinearity in the system described by Duffing's equation. *Int. J. Non-Linear Mech.*, **20**(5/6), 481–491.

Wolf, A., Swift, J. B., Swinney, H. L., and Vastano, J. A. (1985). Determining Lyapunov exponents from a time series. *Physica D*, **16**, 285–317.