Physica 20D (1986) 217–236
North-Holland, Amsterdam

EXTRACTING QUALITATIVE DYNAMICS FROM EXPERIMENTAL DATA

D.S. BROOMHEAD and Gregory P. KING*
Royal Signals and Radar Establishment, St Andrews Road, Great Malvern, Worcestershire, WR14 3PS, UK

Received 1 July 1985
Revised manuscript received 25 November 1985

We consider the notion of qualitative information and the practicalities of extracting it from experimental data. Our approach, based on a theorem of Takens, draws on ideas from the generalized theory of information known as singular system analysis due to Bertero, Pike and co-workers. We illustrate our technique with numerical data from the chaotic regime of the Lorenz model.

1. Introduction

In this paper we consider the notion of *qualitative* information† and how it may be extracted from experimental time series. That this type of information might be recovered from a time series was first suggested by Packard et al. [1]. These authors suggested that a phase portrait, equivalent in some sense to that of the underlying dynamical system, could be reconstructed from time derivatives formed from the data. Another method of phase portrait reconstruction was suggested independently by Takens [2]. This method, which we shall discuss below, is known as the "method of delays". In his paper, Takens provided both approaches with a firm theoretical foundation. As general experimental tools, however, they remain ill-defined since they do not take account of problems associated with the process of measurement.

In this paper we develop another method which draws on Takens' proof and on ideas from the generalized theory of information, known as singular system analysis, recently developed by Bertero, Pike and co-workers [3]. By casting the problem in an information theoretic context, a framework is established which allows us to address the problems associated with the noisy, finite precision, sampled data produced by an experimental measurement. We are further able to resolve in a self-consistent fashion the well-known ambiguities inherent in the application of the method of delays: the need for an ad hoc lag time and the choice of dimension for the space in which the data are plotted.

In section 2 of this paper we introduce some of the relevant language of dynamical systems theory, the definition of qualitative dynamics and the concept of equivalence relations, discuss Whitney's embedding theorem, and review the method of delays. Section 3 is primarily concerned with the singular system analysis and contains the main results of this paper. An example of the methodology introduced in this section is applied to a time series, obtained from the Lorenz model, in section 4. Certain details of the implementation of the methodology are relegated to an appendix. A brief conclusion is given in section 5.

*Also at: Imperial College of Science and Technology, Department of Mathematics, Huxley Building, Queens Gate, London, SW7 2BZ, UK.

†By "qualitative information" we mean that knowledge which may be obtained from a qualitative analysis of a dynamical system.

2. Dynamical systems and the method of delays

2.1. *Dynamical systems theory*

Consider a dynamical system formally as:

$$\frac{dy}{dt} = \boldsymbol{F}(y), \tag{2.1}$$

where each $y = (y_1, y_2, \ldots)$ represents a state of the system and may be thought of as a point in a suitably defined space – which we shall call *phase space*, S. The dimensionality of S, since it controls the number of possible states, will be associated with the number of a priori degrees of freedom of the system. The vector field, $\boldsymbol{F}(y)$, is in general a non-linear operator acting on points in S. Under well-known conditions on $\boldsymbol{F}(y)$ (i.e., $\boldsymbol{F}(y)$ is locally Lipschitz) equation (2.1) defines an initial value problem in the sense that a unique solution curve passes through each point y in the phase space. Formally we may write the solution at time t given an initial value y_0 as $y(t) = \varphi_t y_0$. φ_t represents a one parameter family of maps of the phase space into itself. We can conceive of solutions to all possible initial value problems for the system by writing them collectively as $\varphi_t S$. This may be thought of as a flow of points in the phase space.

Initially the dimension of the set $\varphi_t S$ will be that of S itself. As the system evolves, however, it is generally the case for so-called dissipative systems that the flow contracts onto sets of lower dimension. These are called *attractors*. For the purposes of the present work it will be assumed that the attractor of interest exists within a smooth manifold which we call M. This, too, will generally have a dimension less than that of S. On the attractor the system has fewer degrees of freedom and consequently requires less information to specify its state. Physically this corresponds to self-organization and is common in systems driven far from thermodynamic equilibrium. For example, in the Belousov–Zhabotinski chemistry experiment [4] there are about 30 chemical species participating in the reaction – thus dim $S = 30$. However, in the parameter regime where there exist oscillations with a single fundamental frequency, the attractor is a limit cycle in S and hence has a dimension of 1. A more extreme example is found in fluid dynamic experiments such as the Couette–Taylor system [5] where S is in fact a function space since equation (2.1) is actually a partial differential equation – thus dim $S = \infty$. However, when the system control parameters are adjusted so as to stabilize time-independent Taylor-vortex flow, the attractor is a fixed point in S and hence has a dimension of 0.

2.2. *Qualitative dynamics*

The complete solution of eq. (2.1) is equivalent to a complete knowledge of the family φ_t. However, in most problems of interest obtaining this knowledge is not a practical proposition. It was Poincaré who observed that a great deal of qualitative information about the dynamics could nevertheless be obtained [6]. A qualitative study of eq.

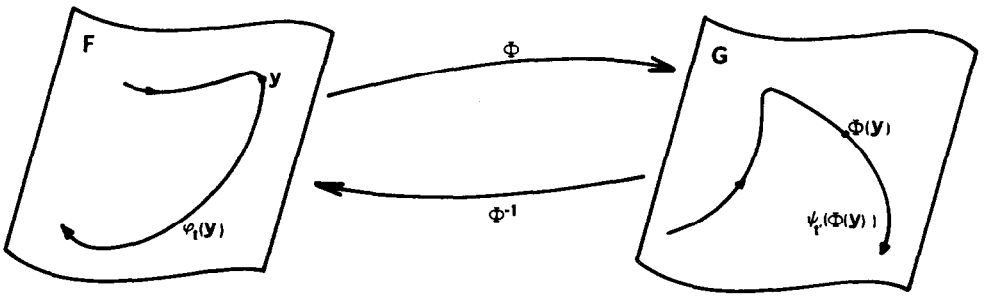

Fig. 1. Schematic representation of an equivalence relation between two vector fields $\boldsymbol{F}$ and $\boldsymbol{G}$. The invertible map, Φ, takes orbits $\varphi_t(y)$ of $\boldsymbol{F}$ into orbits $\psi_{t'}(\Phi(y))$ of $\boldsymbol{G}$.

(2.1) results in a geometric description of its orbits. This will be referred to as the *phase portrait*. In order to give meaning to the idea of qualitative information and to be able to compare and classify the phase portraits of different systems, an equivalence relation between differential equations must be introduced [7, 8].

In general, two C^r vector fields, $\boldsymbol{F}$ and $\boldsymbol{G}$, are said to be C^k equivalent ($k \leq r$) if there exists a C^k diffeomorphism, Φ, which takes orbits $\varphi_t(y)$ of $\boldsymbol{F}$ to orbits $\psi_{t'}(\Phi(y))$ of $\boldsymbol{G}$ in such a way as to preserve their orientation. Intuitively, one can think of this as meaning that Φ is an invertible, possibly non-linear, change of coordinates, which, though distorting the flow, will do so smoothly and will not confuse the order in which the points on the trajectory are visited. When $k = 0$ Φ is a homeomorphism – that is, continuous and one-to-one in both directions. This is known as *topological*, or C^0, *equivalence*. If, in addition, one has the above smoothness conditions on Φ ($k \geq 1$), then one has the stronger *differentiable equivalence*. These ideas are represented pictorially in fig. 1.

The following are useful consequences of topological equivalence:

(1) $y \in M$ is a singularity of $\boldsymbol{F}$ iff $\Phi(y)$ is a singularity of $\boldsymbol{G}$.
(2) The orbit of y for the vector field $\boldsymbol{F}$, is closed iff the orbit of $\Phi(y)$ for $\boldsymbol{G}$ is closed.
(3) The image of the ω-limit set of the orbit of y for $\boldsymbol{F}$ under Φ is the ω-limit set of the orbit of $\Phi(y)$ for $\boldsymbol{G}$ and similarly for the α-limit set.

This means that important topological objects defined by the flow are preserved by the equivalence relation. Furthermore, topological equivalence preserves the stability properties of fixed points, but does not, however, distinguish between nodes, improper nodes, and foci. For this the stronger condition of differentiable equivalence is required. It follows, therefore, that the classification of solutions to differential equations into qualitatively distinct types may be made on the basis of their topological or differentiable equivalence. Thus members of the same equivalence class will be said to have the same qualitative dynamics.

2.3. *Embeddings of manifolds*

Dynamical systems in phase spaces of widely differing dimensions may belong to the same equivalence class provided that their asymptotic dynamics are confined to attracting manifolds of the same dimensionality. This forms the basis of a technique which introduces the ideas of qualitative dynamics into the experimental domain. As has been observed, it is a common phenomenon in physical systems that self-organization gives rise to system evolution on low dimensional manifolds. This leads to the possibility that physically disparate systems may give rise to qualitatively equivalent dynamics, and, moreover, that all members of a given equivalence class may be represented by a canonical model equation. The apparent problem with pursuing this idea is that, in general, it is not at all clear what needs to be measured in order to relate to the dynamics on the underlying attractor. For example, in the Couette–Taylor flow experiment laser-doppler velocimetry enables the measurement of the evolution of one component of the velocity at a point in the fluid – apparently a very low-level description of an evolution in a function space. If it were possible to abstract in some way the low-dimensional manifold from its high-dimensional phase space, this difficulty might be avoided. Such an abstraction process may be achieved by an embedding of the manifold in a lower dimensional space.

An *embedding* is a smooth map, say Φ, from the manifold M to a space U such that its image $\Phi(M) \subset U$ is a smooth submanifold of U and that Φ is a diffeomorphism between M and $\Phi(M)$. In other words, the embedding of M in U is a "realization" of M as a submanifold within U. In particular, the fact that the embedding gives a diffeomorphism between two manifolds means that we have an important prerequisite with which to set up a differentiable equivalence relation. A general existence theorem for embeddings in Euclidean

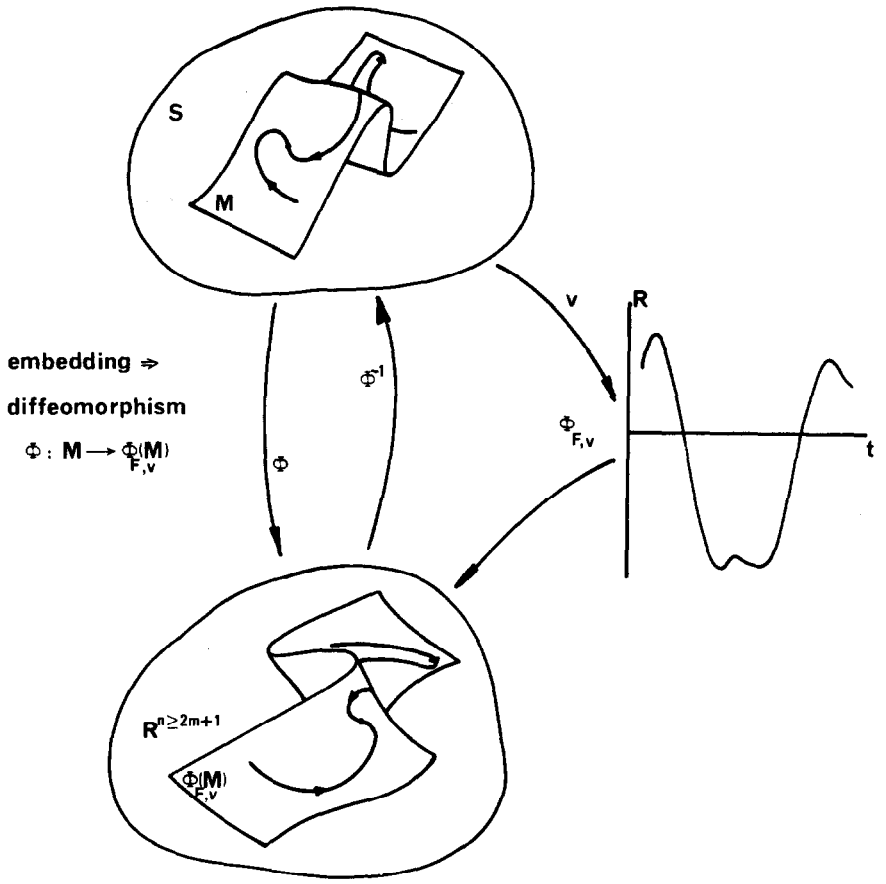

Fig. 2. A schematic representation of the method of delays. The asymptotic dynamics of an experimental system is assumed to correspond to an evolution on an m-dimensional submanifold of the state space S. A sequence of real-valued measurements, v, is used to construct a map $\Phi_{F,v}$ into a Euclidean n-space, $\mathbb{R}^n$. The image of M, $\Phi_{F,v}(M)$, is a submanifold of $\mathbb{R}^n$ under the hypotheses of Takens' theorem. Moreover, the evolution on M is C^1-equivalent to that on $\Phi_{F,v}(M)$.

spaces was given by Whitney [9] who proved that a smooth (C^2) m-dimensional manifold (which is compact and Hausdorff) may be embedded in $\mathbb{R}^{2m+1}$. This theorem is the basis of reconstruction techniques for phase portraits from time series measurements proposed by Packard et al. [1] and by Takens [2].

The present paper is concerned with systems for which the underlying dynamics may be associated with a flow corresponding to a physical process continuous in time. The relevant theorem for flows was proved by Takens and is the basis for the work to be described. In the present notation his theorem (theorem 2) states:

Let M be a compact manifold of dimension m. For pairs $(\boldsymbol{F}, v)$, $\boldsymbol{F}$ a smooth (i.e. C^2) vectorfield and v a smooth function on M, it is a generic property that $\Phi_{\boldsymbol{F},v}(y)$: $M \to \mathbb{R}^{2m+1}$, defined by

$$\Phi_{\boldsymbol{F},v}(y) = (v(y), v(\varphi_1(y)), \ldots, v(\varphi_{2m}(y)))^{\mathrm{T}}$$

is an embedding, where φ_t is the flow of $\boldsymbol{F}$.

Here $v(y)$ corresponds to the value of a measurement made on the system in a state given by $y \in M$.

The conceptual framework discussed here is illustrated in fig. 2, where it is shown how the above theorem provides an explicit construction of an embedding implied by Whitney's theorem. In practice it is necessary to relate the above to a time series of measurements made on the system:

$$v_1, v_2, \ldots, v_i, v_{i+1}, \ldots,$$

where $v_i \equiv v(\varphi_i(y))$. Clearly here we are dealing with a sampled time series for which the sampling interval need not correspond to the unspecified and arbitrary interval implied by the time one map, φ_1, utilized in the theorem. We shall call the practical implementation of this theorem the *method of delays*. The details of the method will now be discussed.

2.4. *Method of delays*

At this stage it is convenient to introduce some vocabulary. The space which contains the image of $\Phi_{F,v}$ will be called the *embedding space* and its dimension the *embedding dimension*. We will denote the embedding dimension by n to emphasize the fact that it will not, in general, equal $2m+1$ since the dimension of M is not known a priori. Nevertheless, it is supposed that $n \geq 2m+1$ to satisfy the Whitney embedding theorem.

In applying the method of delays a useful concept is an "(n, J)-window" which makes visible n elements of the time series. When $J=1$ the elements are consecutive, and when $J>1$ there is an interval of J sample times between each visible element. We shall refer to an $(n,1)$-window as an *n-window*. At any stage the elements visible in the (n, J)-window constitute the components of a vector in the embedding space, $\mathbf{R}^n$. As the time series is advanced step-wise through the window, a sequence of vectors in the embedding space is generated. These form a discrete trajectory. To represent this we use the notation

$$\begin{aligned} \boldsymbol{x}_i &= \Phi_{F,v}(\varphi_i(y)) \\ &= (v_i, v_{i+J}, \ldots, v_{i+(n-1)J})^{\mathrm{T}}. \end{aligned}$$

(a) A 5-window:

$$v_1\ v_2\ \ldots\ v_{i-1}\ \boxed{v_i\ v_{i+1}\ v_{i+2}\ v_{i+3}\ v_{i+4}}\ v_{i+5}\ \ldots$$

$$x_i = (v_i, v_{i+1}, v_{i+2}, v_{i+3}, v_{i+4})^{\mathrm{T}}$$

(b) A (5,3)-window:

$$v_1\ v_2\ \ldots\ \boxed{v_i}\ \ldots\ \boxed{v_{i+3}}\ \ldots\ \boxed{v_{i+6}}\ \ldots\ \boxed{v_{i+9}}\ \ldots\ \boxed{v_{i+12}}\ \ldots$$

$$x_i = (v_i, v_{i+3}, v_{i+6}, v_{i+9}, v_{i+12})^{\mathrm{T}}$$

(c) The trajectory matrix for a 5-window:

$$\boldsymbol{X} = N^{-\frac{1}{2}} \begin{bmatrix} x_1^{\mathrm{T}} \\ x_2^{\mathrm{T}} \\ \vdots \\ \vdots \\ \vdots \end{bmatrix} = N^{-\frac{1}{2}} \begin{bmatrix} v_1 & v_2 & v_3 & v_4 & v_5 \\ v_2 & v_3 & v_4 & v_5 & v_6 \\ \vdots & & \vdots & & \\ & & \vdots & & \\ & & \vdots & & \end{bmatrix}$$

Fig. 3. Illustrations of the use of the method of delays: (a) The construction of vectors in $\mathbf{R}^5$ from a sequence of measurements of v using a lag-time of one sample-time. (b) As in (a), but with a lag-time of three sample-times. (c) The construction of the trajectory matrix, $\boldsymbol{X}$, from vectors obtained by application of a 5-window (see (a)) to a time series.

The construction of vectors using an (n, J)-window is illustrated in fig. 3.

There are several difficulties in applying the method of delays in its present form. These can be traced to the fact that Takens' theorem makes no direct contact with the process of measurement. In particular, there are several time scales that are unspecified. The most obvious are the sampling time, τ_s, the "lag time", $\tau_L = J\tau_s$, and the window length, $\tau_w = n\tau_L$. In practice, short sampling times are employed to produce good approximations to smooth trajectories. However, this has the unfortunate effect of creating highly correlated samples within an n-window – thus causing the trajectory to lie close to the diagonal in the embedding space. To avoid this, the more general (n, J)-windows, with τ_L large enough to introduce a degree of statistical independence between the components, are used.

Takens' theorem appears to provide information about the choice of embedding dimension, stating that $n \geq 2m + 1$. This, however, is of little practical relevance since $m = \dim M$ is not generally known a priori. The approach taken in published work has been to increase n systematically, until trajectories no longer appear to intersect [10]. This is at best a rather subjective criterion, becoming rapidly unworkable in higher dimensions or in the presence of noise. A further difficulty is that plotting data in this way introduces an artificial symmetry into the phase portrait. A consequence of this is that time-averaged moments of trajectories projected onto the coordinate axes become independent of coordinate for long time series. More confusing, from the point of view of interpreting the multi-dimensional structure of attractors, is the fact that many projections onto orthogonal planes are identical. For example, the projection onto the (i, j)-plane is the same as that onto all the $(i + k, j + k)$-planes for all k such that $i + k, j + k \leq n$. Indeed, as one increases the embedding dimension, in each new embedding space the artificial symmetry is increased.

This undesirable property results from choosing a basis for the embedding space in an arbitrary manner. Intuitively one could hope for a choice of basis such that as n increases beyond $2m + 1$ the attractor will be found, with invariant geometry, confined to a subspace of fixed dimension. In the next section we show how an analysis of the information content of the time series can be used to derive such a basis. This will have the additional advantage of being able to deal with experimental noise in a systematic fashion, thereby remedying a deficiency in the implementation of the method of delays as described above.

3. A statistical approach to the method of delays

In this section we develop a *singular system* approach to the method of delays which deals with the ambiguities and limitations described in the previous section. A consequence of this analysis will be that we eliminate the need to introduce statistical independence through the use of a general (n, J)-window. Thus, we set $J = 1$ and dispense with arbitrary lag times.

3.1. *The singular system*

The application of an n-window to a time series of N_T data points results in a sequence of $N = N_T - (n - 1)$ vectors, $\{x_i \in \mathbb{R}^n | i = 1, 2, \ldots, N\}$ in the embedding space. Such a sequence can be used to construct a *trajectory matrix*, $\boldsymbol{X}$, which contains the complete record of patterns which have occurred within the window:

$$\boldsymbol{X} = N^{-1/2} \begin{bmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_N^T \end{bmatrix}, \tag{3.1}$$

where $N^{-1/2}$ has been introduced as a convenient normalization. The trajectory matrix and its transpose may be thought of as linear maps between the spaces $\mathbb{R}^n$ and $\mathbb{R}^N$. The embedding space, $\mathbb{R}^n$, is the space of all n-element patterns and is the

natural object of interest. A similar interpretation of $\mathbb{R}^N$ is clearly possible. Indeed, the columns of $\boldsymbol{X}$ bear the same relationship to one another as do the rows. However, for the present purposes we shall use $\mathbb{R}^N$ in a simple way which utilizes the obvious property that the standard basis vectors $\{\boldsymbol{e}_i \in \mathbb{R}^N\}$, can be used as an indexing system for points on the trajectory in $\mathbb{R}^n$ (where $\boldsymbol{e}_i$ is the ith column of the $N \times N$ unit matrix). That is, formally we can extract the ith vector, $\boldsymbol{x}_i$, from the trajectory matrix by operating from the left with $N^{1/2}\boldsymbol{e}_i^{\mathrm{T}}$:

$$\boldsymbol{x}_i^{\mathrm{T}} = N^{1/2}\boldsymbol{e}_i^{\mathrm{T}}\boldsymbol{X}. \tag{3.2}$$

It follows directly that operating with a general vector, $\boldsymbol{w}^{\mathrm{T}} = N^{1/2}\sum_{i=1}^{N} w_i \boldsymbol{e}_i^{\mathrm{T}}$, will produce a linear combination of vectors on the trajectory in $\mathbb{R}^n$. Then $\boldsymbol{w}^{\mathrm{T}}\boldsymbol{X}$ is the mean position in the embedding space relative to a measure induced on the sampled attractor by the choice of $\boldsymbol{w}$. Alternatively, it corresponds to a weighted time-average over the trajectory. We note that in this context the concepts of space- and time-average are interchangeable: Just as the standard basis of $\mathbb{R}^N$ indexes the points in the embedding space so also does it represent the labelling of a time-sequence.

The triple $(\boldsymbol{X}, \mathbb{R}^N, \mathbb{R}^n)$ can be analysed using a singular system as considered in the generalized theory of information developed by Bertero, Pike, and their co-workers [3]. In the following we shall develop a singular system analysis based on the method of delays.

3.2. *Independence and orthogonality*

A central concept in distinguishing chaotic systems from stochastic, Brownian-like systems is that of *number of degrees of freedom*. Unfortunately, there are numerous definitions of number of degrees of freedom depending on the context in which it arises. For example, in signal processing it is the number of modes used to describe a signal which contain significant power. In dynamical systems it is usually used to specify the dimensionality of an attracting manifold within the phase space, the dimensionality of the attractor itself or even the dimension of the whole phase space. We shall show that the signal processing definition when related to a dynamical system interpretation becomes the dimensionality of the subspace in which the embedded manifold is to be found. This has shortcomings since, unlike $m = \dim M$, it is not an invariant of the embedding process. However, it will be shown to give a reasonable upper bound to m.

We begin, therefore, by calculating the dimensionality of the subspace which contains the embedded manifold. To do this we need to know the number of linearly independent vectors that can be constructed from the trajectory in the embedding space by forming linear combinations of the $\boldsymbol{x}_i$. It has already been established that vectors in $\mathbb{R}^N$ give rise to such linear combinations when they act on the trajectory matrix. Consider the set of vectors $\{\boldsymbol{s}_i \in \mathbb{R}^N\}$, which we assume give a set of linearly independent vectors in $\mathbb{R}^n$ by their action on $\boldsymbol{X}$. We shall also assume, with no loss of generality, that the latter have been orthonormalized. Hence, in general, they constitute part of a complete orthonormal basis, $\{\boldsymbol{c}_i | i = 1, \ldots, n\}$, for the embedding space. By construction, the following relationship holds:

$$\boldsymbol{s}_i^{\mathrm{T}}\boldsymbol{X} = \sigma_i \boldsymbol{c}_i^{\mathrm{T}}, \tag{3.3}$$

where the $\{\sigma_i\}$ are a set of real constants which will be used to fix the normalization of both sets of vectors.

The orthonormality of the $\{\boldsymbol{c}_i\}$ imposes the following condition:

$$\boldsymbol{s}_i^{\mathrm{T}}\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{s}_j = \sigma_i \sigma_j \delta_{ij}, \tag{3.4}$$

where δ_{ij} is the Kronecker delta. The $N \times N$ matrix $\boldsymbol{\Theta} = \boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}$ is real symmetric, and hence its eigenvectors form a complete orthonormal basis for $\mathbb{R}^N$. In particular, the above equation is solved by the eigenvectors of $\boldsymbol{\Theta}$:

$$\boldsymbol{\Theta}\boldsymbol{s}_i = \sigma_i^2 \boldsymbol{s}_i \tag{3.5}$$

provided the $\{\sigma_i^2\}$ are interpreted as the corresponding eigenvalues. It should be noted that real matrices of the form $\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}$ are non-negative definite. Therefore, the $\{\sigma_i\}$ are real constants which may themselves be taken to be non-negative without loss of generality.

Eq. (3.1) may be used to show that $\boldsymbol{\Theta}$ is actually an array of scalar products of all pairs of points on the trajectory in the embedding space:

$$\boldsymbol{\Theta} = N^{-1}\begin{bmatrix} \boldsymbol{x}_1^{\mathrm{T}}\boldsymbol{x}_1 & \boldsymbol{x}_1^{\mathrm{T}}\boldsymbol{x}_2 \cdots \boldsymbol{x}_1^{\mathrm{T}}\boldsymbol{x}_N \\ \boldsymbol{x}_2^{\mathrm{T}}\boldsymbol{x}_1 & \boldsymbol{x}_2^{\mathrm{T}}\boldsymbol{x}_2 \cdots \boldsymbol{x}_2^{\mathrm{T}}\boldsymbol{x}_N \\ \vdots & \vdots \quad\quad \vdots \end{bmatrix}, \tag{3.6}$$

and for this reason it will be called the *structure matrix* of the trajectory. Equally, it may be interpreted as being composed of the correlations between all pairs of patterns to have appeared in the n-window. The considerable redundancy in specifying the correlations between all pairs of patterns results in $\boldsymbol{\Theta}$ having low rank. This is borne out by eq. (3.3) which shows that there are, at most, n of the $\{\sigma_i\}$ which are non-zero. Because of this the difficulty of diagonalizing $\boldsymbol{\Theta}$ when N is large can be avoided.

To explore this point further, we look for an inverse relationship to eq. (3.3) whereby an expression for the vector $\boldsymbol{s}_i$ which yields a particular $\boldsymbol{c}_i$ is obtained. This is:

$$\boldsymbol{X}\boldsymbol{c}_i = \sigma_i \boldsymbol{s}_i \tag{3.7}$$

which, when $\sigma_i \neq 0$, may be derived by taking the transpose of eq. (3.3), operating from the left with $\boldsymbol{X}$, and using eq. (3.5). The $\{\boldsymbol{s}_i\}$ have been shown to be an orthogonal set, therefore, the following eigenvalue equation can be derived analogously to eq. (3.5):

$$\boldsymbol{\Xi}\boldsymbol{c}_i = \sigma_i^2 \boldsymbol{c}_i. \tag{3.8}$$

Here $\boldsymbol{\Xi} = \boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}$ is a real, symmetric $n \times n$ matrix which may be written as:

$$\boldsymbol{\Xi} = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{x}_i \boldsymbol{x}_i^{\mathrm{T}} \tag{3.9}$$

using the definition of the trajectory matrix. This is the time-average of the dyadic product $\boldsymbol{x}_i\boldsymbol{x}_i^{\mathrm{T}}$. Thus $\boldsymbol{\Xi}$ is the *covariance matrix* of the components of the $\{\boldsymbol{x}_i\}$, averaged over the entire trajectory; that is, the time-averaged correlation between all pairs of elements in the n-window. Expressed in terms of the original time series, this has the form:

$$\boldsymbol{\Xi} = \frac{1}{N} \times \begin{bmatrix} \sum_{i=1}^{N} v_i v_i & \sum_{i=1}^{N} v_i v_{i+1} & \cdots & \sum_{i=1}^{N} v_i v_{i+n-1} \\ \vdots & \vdots & & \vdots \\ \sum_{i=1}^{N} v_{i+n-1} v_i & \sum_{i=1}^{N} v_{i+n-1} v_{i+1} & \cdots & \sum_{i=1}^{N} v_{i+n-1} v_{i+n-1} \end{bmatrix}.$$

Equation (3.8) is far more tractable than eq. (3.5) since it is implicit to the approach that the embedding dimension is small.

Returning to the question of the rank of $\boldsymbol{\Theta}$, we note that the derivation of eq. (3.8) from eq. (3.5) shows that the non-zero eigenvalues of the structure matrix equal the non-zero eigenvalues of the covariance matrix. That is, rank $\boldsymbol{\Theta}$ = rank $\boldsymbol{\Xi} = n' \leq n$. One is thus led to the observation that $\mathbb{R}^N$ can be decomposed into a subspace of dimension n' and its orthogonal complement. The n'-dimensional subspace is spanned by a set $\{\boldsymbol{s}_i | i = 1, \ldots, n'\}$ which is such that each corresponding average over the $\boldsymbol{x}_i$ gives rise uniquely to a basis vector $\boldsymbol{c}_i \in \mathbb{R}^n$ according to eq. (3.3). The complementary subspace, spanned by the set $\{\boldsymbol{s}_i | i = n' + 1, \ldots, N\}$, is the kernal of $\boldsymbol{X}$ mapping onto the origin of the embedding space through eq. (3.3).

The complete set $\{\boldsymbol{s}_i\}$ is a basis for the construction of all possible averages over the trajectory. The significance of the decomposition is that only the averages associated with the n'-dimensional subspace give a non-trivial vector in $\mathbb{R}^n$. There are at most n' linearly independent vectors in the embedding space that may be constructed from the trajectory. Therefore, it might be supposed that the number n', the rank of $\boldsymbol{\Xi}$, is the dimensionality of the subspace containing the embedded manifold. However, account must be taken of the fact that $\boldsymbol{X}$ has contributions from sources of experimental noise. This point is addressed in the next section.

3.3. *Singular value decomposition and noise*

To study the effect of noise it is necessary to discuss the eigenvectors and their associated spectrum. Consider the orthogonal $n \times n$ matrix $\boldsymbol{C}$ which has columns consisting of the vectors $\{\boldsymbol{c}_i\}$, $\boldsymbol{C} = (\boldsymbol{c}_1, \boldsymbol{c}_2, \ldots, \boldsymbol{c}_n)$, and the diagonal matrix $\boldsymbol{\Sigma} = \mathrm{diag}(\sigma_1, \sigma_2, \ldots, \sigma_n)$, where the ordering $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0$ is assumed. Using this notation, eq. (3.8) is:

$$\boldsymbol{\Xi C} = \boldsymbol{C\Sigma}^2. \tag{3.10}$$

Using the definition of $\boldsymbol{\Xi}$ it follows that:

$$(\boldsymbol{XC})^{\mathrm{T}}(\boldsymbol{XC}) = \boldsymbol{\Sigma}^2. \tag{3.11}$$

The matrix $\boldsymbol{XC}$ is the trajectory matrix projected onto the basis $\{\boldsymbol{c}_i\}$. This result expresses the fact that in the basis $\{\boldsymbol{c}_i\}$ the components of the trajectory are uncorrelated since the $\{\boldsymbol{c}_i\}$ are obtained from the diagonalization of the covariance matrix. It was in anticipation of this result that we omitted consideration of general (n, J)-windows. This result also shows that each σ_i^2 is the mean square projection of the trajectory onto the corresponding $\boldsymbol{c}_i$. Therefore, the spectrum $\{\sigma_i^2\}$ has information about the extent to which the trajectory explores the embedding space. One may think of the trajectory as exploring on average, an n-dimensional ellipsoid. The $\{\boldsymbol{c}_i\}$ then give the directions and the $\{\sigma_i\}$ the lengths of the principal axes of the ellipsoid. In the previous section the trajectory was found to be confined to a subspace of dimension equal to the rank of $\boldsymbol{\Xi}$. In a noisy environment this needs qualification since the presence of noise will tend to smear out the deterministic behaviour, and, in the directions associated with small or vanishing σ_i, the noise will dominate. Therefore, the rank of $\boldsymbol{\Xi}$ is an upper bound to the dimensionality of the subspace explored by the deterministic component of the trajectory.

The above formalism, as is well known from linear algebra, can be used to express the singular value decomposition of a singular, linear map [11]. The significance of this in the context of information theory has been discussed recently in a different application [3]. Here, we are interested in the singular value decomposition of the trajectory matrix:

$$\boldsymbol{X} = \boldsymbol{S\Sigma C}^{\mathrm{T}}, \tag{3.12}$$

where $\boldsymbol{S}$ is the $N \times n$ matrix of eigenvectors of $\boldsymbol{\Theta}$, n' of which have non-zero eigenvalues. The vectors of $\boldsymbol{C}$ and $\boldsymbol{S}$ will henceforth be referred to as the *singular vectors* of $\boldsymbol{X}$, while the elements of the diagonal matrix, $\boldsymbol{\Sigma}$, will be called the associated *singular values*.

To exemplify the effect of noise consider the following simple model of a time series with a noise component ξ_j: $v_j = \bar{v}_j + \xi_j$. Here, and below, an overbar indicates a quantity associated with the deterministic component. For white noise uncorrelated with $\bar{v}_j$ the autocorrelation function of the time series can be written $\langle v_0 v_j \rangle = \langle \bar{v}_0 \bar{v}_j \rangle + \langle \xi^2 \rangle \delta_{0j}$, where the angle brackets refer to a time average. Generally, in the limit of an infinite time series of statistically stationary data, the covariance matrix is expected to have the Toeplitz structure: $\boldsymbol{\Xi}_{ij} = g([i-j]\tau_s)$, where $g(\tau)$ is the autocorrelation function of the continuous time series. In this limit the covariance matrix for the above time series may be decomposed into two parts: $\boldsymbol{\Xi} = \bar{\boldsymbol{\Xi}} + \langle \xi^2 \rangle \boldsymbol{1}_n$, where $\boldsymbol{1}_n$ is the $n \times n$ unit matrix. In this case, the singular values of $\boldsymbol{X}$ are shifted uniformly:

$$\sigma_i^2 = \bar{\sigma}_i^2 + \langle \xi^2 \rangle, \quad i = 1, 2, \ldots, n, \tag{3.13}$$

where $\bar{\sigma}_i^2$ is an eigenvalue of $\bar{\boldsymbol{\Xi}}$. Thus the noise causes *all* the singular values of the trajectory matrix to be non-zero. Hence, the trajectory appears to explore all dimensions of the embedding space. We need, therefore, a method whereby these two types of contribution may be distinguished.

In the simple case of white noise, the existence of a non-zero constant background or noise floor is a noteable characteristic which can be used to distinguish the deterministic component. More

generally, the independent measurement of a time series consisting only of the experimental noise will enable the calculation of its root mean square projections onto the $\{\boldsymbol{c}_i\}$. By comparing these with the corresponding singular values for the time series containing the deterministic signal we can define a signal-to-noise ratio which may be associated with each singular vector. The technique will allow the identification of those singular values which are noise dominated even when the noise is not white. It is not always possible to measure the noise separately and for this reason there is interest in the signal processing community in developing so-called *cross-validation* techniques whereby the internal consistency of the data itself may be used to estimate the noise level [12].

Given that a suitable method has been used to partition the singular value spectrum, we move on to consider the implied partitioning of the embedding space. Consider the matrices $\boldsymbol{p}^{(i)}$: $\boldsymbol{p}^{(i)}_{jk} = \delta_{ij}\delta_{jk}$, which are representations of projection operators onto the basis functions $\{\boldsymbol{c}_i\}$. These may be used to construct projection operators onto the corresponding subspaces of the embedding space:

$$\boldsymbol{Q} = \sum_{\sigma_i \approx \text{noise}} \boldsymbol{p}^{(i)}, \tag{3.14}$$

$$\boldsymbol{P} = \sum_{\sigma_i > \text{noise}} \boldsymbol{p}^{(i)}. \tag{3.15}$$

Inserting the identity $\boldsymbol{P} + \boldsymbol{Q} = \boldsymbol{1}_n$ into eq. (3.12) gives the following:

$$\boldsymbol{X} = \bar{\boldsymbol{X}} + \Delta\boldsymbol{X}, \tag{3.16}$$

where

$$\bar{\boldsymbol{X}} = \boldsymbol{S}\boldsymbol{P}\boldsymbol{\Sigma}\boldsymbol{C}^{\mathrm{T}} \tag{3.17}$$

is the deterministic part of the trajectory matrix, and

$$\Delta\boldsymbol{X} = \boldsymbol{S}\boldsymbol{Q}\boldsymbol{\Sigma}\boldsymbol{C}^{\mathrm{T}} \tag{3.18}$$

is the noise-dominated part. This separation allows the rejection of a portion of the inherent noise of the experiment. In a signal processing context this is known as rejection of *out-of-band* noise. Here it amounts to oversampling the data in order to average the noise over more points in the window.

From now on our attention will shift to the reduced trajectory matrix $\bar{\boldsymbol{X}}$ since this contains all the information about the deterministic trajectory that we can sensibly extract from the experiment. A useful form for $\bar{\boldsymbol{X}}$ may be obtained by substituting eq. (3.7) into eq. (3.17):

$$\bar{\boldsymbol{X}} = \sum_{\sigma_i > \text{noise}} (\boldsymbol{X}\boldsymbol{c}_i)\boldsymbol{c}_i^{\mathrm{T}}. \tag{3.19}$$

This expression uses only quantities obtainable from the diagonalization of the covariance matrix. In this form there is an obvious interpretation of $\bar{\boldsymbol{X}}$ which relates directly to the methodology of plotting phase portraits. Here $\boldsymbol{c}_i^{\mathrm{T}}$ represents a coordinate axis of the embedding space referred to the standard basis, while $(\boldsymbol{X}\boldsymbol{c}_i)$ is a column vector containing a time series of the ith component, in the basis $\{\boldsymbol{c}_i\}$, of the vectors in the trajectory.

Clearly $\bar{\boldsymbol{X}}$ is an $N \times n$ matrix, however, by construction it consists of a trajectory confined to the deterministic subspace of $\mathbb{R}^n$ having dimension $d \leq n$ (where d is the number of the $\{\sigma_i\}$ above the noise floor). It is straightforward to define vectors restricted to this subspace. We shall distinguish these by a square bracket notation. Thus, $\bar{\boldsymbol{X}}$ restricted to the deterministic subspace is an $N \times d$ rectangular matrix the jth row of which is $[\boldsymbol{x}_j^{\mathrm{T}}\boldsymbol{c}_1, \boldsymbol{x}_j^{\mathrm{T}}\boldsymbol{c}_2, \ldots, \boldsymbol{x}_j^{\mathrm{T}}\boldsymbol{c}_d]$. In the case that the experimental noise is not white it may be necessary to relabel the $\{\boldsymbol{c}_i\}$ such that the first d vectors span the deterministic subspace.

3.4. *On the choice of time scales*

In the following paragraphs we outline an approach to the choice of sampling time, τ_s, and the window length, τ_w, so that the formalism developed above can be applied to time series data. It is more usual in the method of delays to consider the

choice of embedding dimension rather than the window length. However, the window length has particular significance in the singular system analysis since it determines the *form* of the singular spectrum.

Before we can consider the effect of the window length on the form of the singular spectrum, we must first establish a sampling criterion. Observe that for fixed τ_w increasing n by decreasing τ_s generates additional singular vectors capable of describing more rapid variations within the window. The corresponding singular values will, as we have said earlier, represent the "power" in the time series which corresponds to such variations. On physical grounds one can assume the existence of an inner time scale. Over times less than this inner scale the data does not vary significantly to within the precision of the measurement. In this case decreasing τ_s further results in additional singular vectors with singular values in the noise floor defined by the precision of the experiment. At this point we say the spectrum has converged. Therefore, a natural choice for τ_s for a given τ_w is that necessary to achieve convergence of the singular spectrum.

Having determined a sampling time that ensures a convergent spectrum, we now consider how τ_w affects the form of the spectrum. Clearly, as τ_w is decreased we approach the inner scale for which there is no measureable variation of the data within the window. The singular spectrum is thus reduced to a triviality. On the other hand as τ_w is increased the information to be represented within the window increases. This has the corresponding effect of increasing the number of significant singular values. Indeed in the limit $\tau_w \to \infty$ it can be shown that our method becomes a discrete Fourier transform [13]. Clearly, it is important to have a criterion for the choice of τ_w.

Takens, in his proof of the theorem given in section 2, required on generic grounds the exclusion of data with interger periods less than τ_w. It can be shown that for realistic measurements the generic argument is too weak. However, it is sufficient, for band-limited data, to replace this with a constraint on τ_w:

$$\tau_w \leq \tau^*, \tag{3.20}$$

where $\tau^* = 2\pi/\omega^*$ and ω^* is the band-limiting frequency. By band-limited data we mean here that the Fourier spectrum of the time series contains no frequencies with significant power greater than a cutoff frequency known as the band-limit. Furthermore, there is an obvious lower bound on τ_w:

$$\tau_w \geq (2m+1)\tau_s. \tag{3.21}$$

Since m is unknown, however, the only consistent a priori estimate is $\tau_w = \tau^*$.

The above arguments provide a self-consistent approach to the choice of a sampling time and window length which will satisfy the postulates of Takens' theorem. In our limited experience the use of $\tau_w = \tau^*$ has proved satisfactory. However, it must be emphasized that the derivation of this estimate for τ_w used a sufficient condition which in some circumstances may prove to be too strong. A complete answer to this problem must circumvent a fundamental limitation of Takens' theorem: the implicit assumption of data with infinite precision. It is clear from the above analysis that it is necessary to specify two quantities – the embedding dimension and a time scale (τ_w or τ_s). It is equally clear that Takens' theorem makes no mention of a time scale. This is due to an assumption that successive measurements contain new information whatever the time interval between them. For finite precision measurements this is manifestly untrue. Thus it is insufficient merely to require $2m+1$ measurements to specify an embedding – a time scale is also required. It should be emphasized that this problem exists for all data analysis techniques that rely on the construction of an embedding (e.g., entropy and dimension calculations using time series data). Theoretically the problem is still open; in practice the effects of sampling and window length on the results should be investigated.

3.5. *Concluding comments*

The derivation of a coordinate system by diagonalizing a covariance matrix is known as the Karhunen–Loeve method [13] and is widely known in the signal processing and pattern recognition fields. One feature of a basis obtained in this way is that it produces an optimum compression of information. By this we mean that in order to distinguish points in a set of interest (in this case the trajectory in $\mathbb{R}^n$) to within a given accuracy the Karhunen–Loeve basis requires the fewest components to be specified. Thus, for a fixed embedding dimension the error produced by projecting onto the first ν basis vectors is, when averaged over the set, minimized if the first ν singular vectors are used. This suggests a systematic sequence of coarse grainings – the small scale limit of which is the grain size set by the noise as measured by the noise floor of the singular value spectrum. In contrast the standard basis for $\mathbb{R}^n$ – the one implicit in a naive implementation of the method of delays – gives the worst possible information compression since for long times the rms projections of the trajectory onto the basis vectors are all equal.

An additional consequence of deriving a basis by diagonalization of the covariance matrix is that orthogonality in the embedding space is related to the statistical properties of the time series. An outcome of this is that a trajectory, when plotted against the $\{\boldsymbol{c}_i\}$, will appear coherent to the extent that it deviates from a gaussian random process. This is an important point since it is known that deterministic chaos is non-gaussian. The possibility of using higher multi-point correlation properties of the time series to distinguish chaos from stochasticity has not to our knowledge been systematically investigated [14], although these concepts do underlie attempts to define dimensions and entropies for attracting sets [15].

4. Application to the Lorenz model

In this section we illustrate the singular system approach with data obtained from the Lorenz model [16]. Applications to experimental data will appear elsewhere [17].

The Lorenz model is defined by the equations

$$\frac{\mathrm{d}X}{\mathrm{d}t} = \sigma(Y - X),$$

$$\frac{\mathrm{d}Y}{\mathrm{d}t} = rX - Y - XZ,$$

$$\frac{\mathrm{d}Z}{\mathrm{d}t} = -bZ + XY.$$

In this work the equations were solved with a fourth-order Runge–Kutta routine using a step size of either 0.003 or 0.009. When the parameters take on the values $\sigma = 10$, $b = 8/3$, and $r = 28$, the system produces turbulent dynamics. The time-evolution is organized by two unstable foci and an intervening saddle point. Fig. 4 shows a projection of the phase-space orbit of the system onto the XY-plane. The study of the continuous system can be replaced by the study of a discrete dynamical system known as the Poincaré map. This is obtained by recording the intersections of the trajectory with a surface of section oriented transverse to the flow. The surface of section shown in fig. 5a

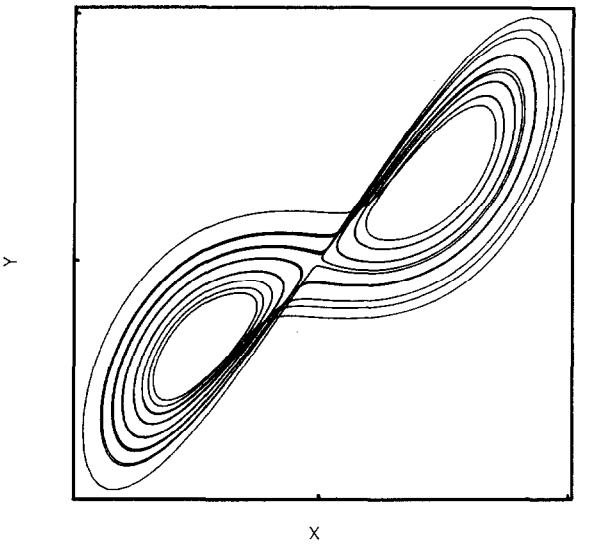

Fig. 4. The attractor of the full Lorenz model ($\sigma = 10$, $b = 8/3$, $r = 28$) projected onto the XY-plane.

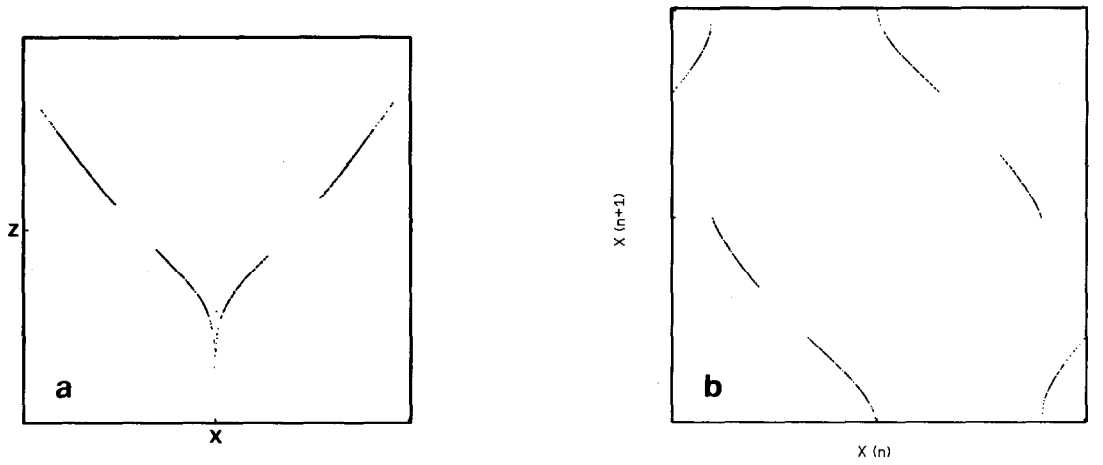

Fig. 5. (a) The intersection of the attractor of the full Lorenz model (see fig. 4) with a planar surface of section containing the Z-axis and passing through the two unstable foci. (b) The corresponding first-return map using the X-coordinates of the points of intersection.

was chosen to contain the two unstable foci and to be transverse to their unstable manifolds (it in fact contains the Z-axis). An approximately one-dimensional first-return map constructed from the Poincaré map is shown in fig. 5b. This map has been calculated to within a suitable approximation [14] and forms a connection between the full system and analytic theory.

We now turn our attention to the time series of the X variable shown in fig. 6a. Using the approach outlined in section 3.4, we choose a window length by estimating the band-limiting frequency of the power spectrum of $X(t)$ shown in fig. 6b. This gives a window length of approximately a tenth of the period of oscillation about the unstable foci [18]. In Lorenz units we have $\tau_w = 0.063$.

Let us now consider the choice of sampling time. Fig. 7a illustrates the effect of changing the sampling time on the singular value spectrum while maintaining the window length at the above value. This was done by using a time series, sampled with a step size of 0.003, and generating more sparsely

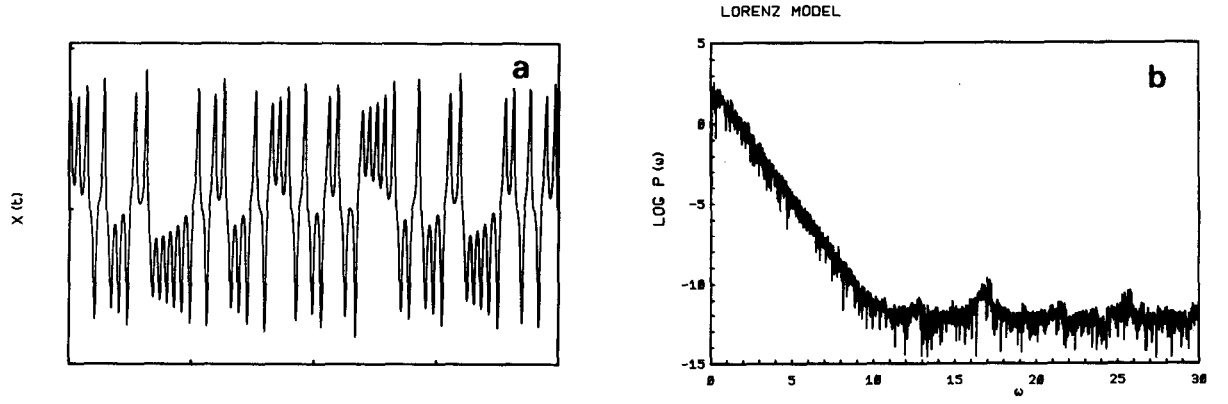

Fig. 6. (a) A sample of the time series $X(t)$ corresponding to motion on the attractor of the full Lorenz model (see fig. 4). (b) The corresponding power spectrum constructed from $X(t)$. Frequencies are scaled with the frequency associated with the unstable foci [18].

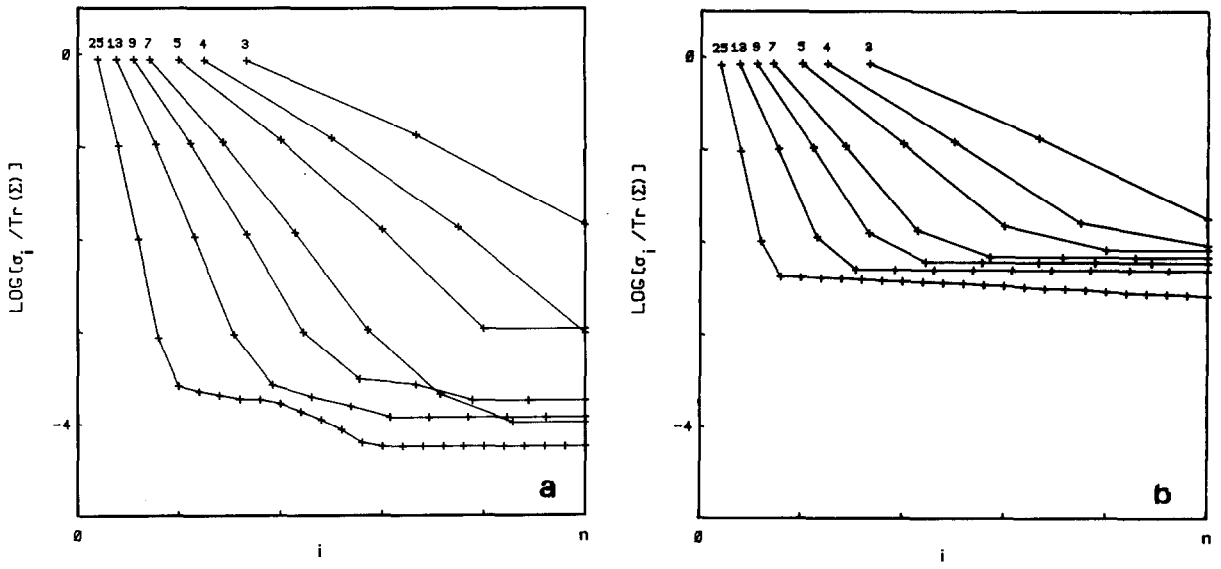

Fig. 7. (a) The normalized singular spectra $\{\log_{10}(\sigma_i/\Sigma\sigma_i)\}$ obtained from $X(t)$ data of the Lorenz model (see fig. 6a). The figure illustrates the effect of changing the sampling time, τ_s, for fixed window length, $\tau_w = n\tau_s$. Each curve involves the diagonalization of an $n \times n$ symmetric matrix. The matrix elements are obtained from the autocorrelation function of the time series which was calculated using $\sim 2 \times 10^4$ data points. The noise floors are a result of numerical noise in the calculations. (b) As with (a), the data were, however, passed through a simulated 6-bit A/D converter. Quantization error dominates the noise in this case.

sampled time series from it. The example shown in fig. 7a began with 25 samples within τ_w†. Omitting every other data point gave a time series sampled at half the rate and having 13 samples in the window. Similarly, taking points at $\frac{1}{3}$, $\frac{1}{4}$, $\frac{1}{6}$, $\frac{1}{8}$, and $\frac{1}{12}$ the original rate gave time series with 9, 7, 5, 4, and 3 samples in the window. For sampling rates yielding 7 or more points in the window, the spectrum has two distinct parts – one part which can be associated with the noise floor, and the other part which can be associated with the deterministic components of the data. The distinguishing features of the noise floor are its magnitude and flatness. The magnitude represents rounding errors within the computer as measured by the magnitude of spurious negative eigenvalues generated by the diagonalization routine. This should be contrasted with fig. 7b for which the data were passed through a simulated 6-bit analog-to-digital converter. In this case the noise floor is higher since it is dominated by quantization noise.

It is clear from figs. 7a and b that the form of the singular spectrum is insensitive to the range of sampling times used. In particular, the effect of decreasing τ_s is essentially to increase the number of singular values in the noise floor. On the basis of this analysis we choose to work with data sampled at intervals $\tau_s = 0.009$ which implies an embedding dimension of 7. This is largely a matter of computational ease since our data is very clean. With noisy data it is often better to increase the sampling rate in order to be able to average over more data points within the window.

The singular value spectrum shown in fig. 8 results from the above choices of τ_s and τ_w. It is clear from figs. 7a and 8 that the important dynamics will be confined to a 4-dimensional subspace of the embedding space. In the absence of any prior knowledge about the system it would be necessary to consider the dynamics in this 4-

†Note that this implies $\tau_w = 0.075$. This choice was made for convenience in producing the figure and does not have a significant effect on the spectra obtained, cf. figs. 7a and 8.

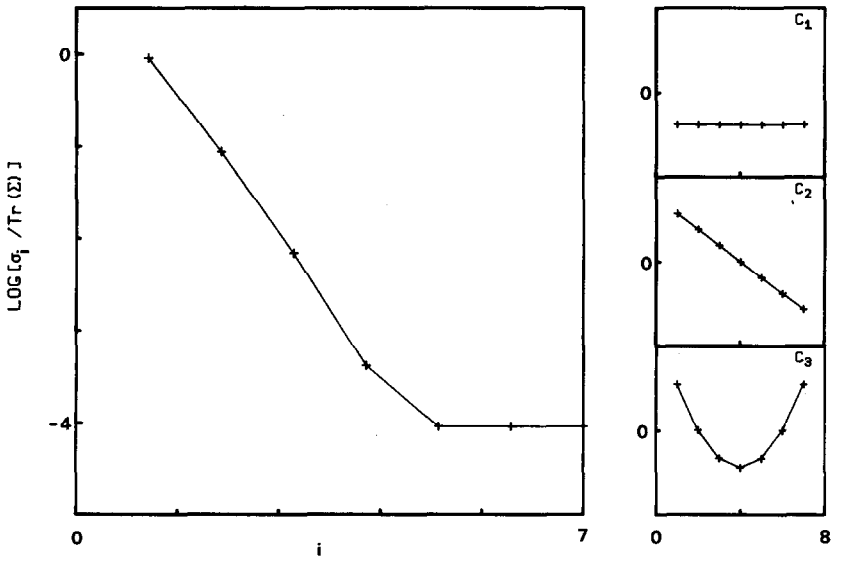

Fig. 8. The singular spectrum and the first three singular vectors obtained from the unquantized $X(t)$ Lorenz data using $\tau_s = 0.009$ and $\tau_w = 7\tau_s$. (The ordinates of each singular vector plot cover the range $[-0.5, +0.5]$).

dimensional space. However, since we know that the Lorenz attractor can be embedded in a 3-dimensional Euclidean space, we shall only consider the projection onto the subspace spanned by the first three singular vectors. These vectors, which are shown in fig. 8, have a form reminiscent of orthogonal polynomials. In fig. 9 we show the projections of the trajectory onto the planes spanned by (c_1, c_2), (c_1, c_3), and (c_2, c_3). The details of this have been given in section 3.3. We interpret the polynomial appearance of the $\{c_i\}$ to mean that the components of $c_i^T x_j$ are averaged "time derivatives" of the time series. This recalls the approach of Packard et al. [1]. An important difference, however, is that, unlike the sequences of components generated here, the time derivatives of the time series are not statistically independent. Moreover, time derivatives estimated by finite differences are sensitive to noise since no averaging process is involved.

The projections shown in fig. 9 do not exhibit the spurious symmetries inherent in the basic method of delays and readily suggest the 3-dimensional form of the reconstructed attractor. Indeed the use of a suitable orthogonal transform enables one to generate stereo pairs as shown in fig. 10. We believe these to be a valuable tool for the development of geometric intuition about attractors extracted from experimental data [19].

Figures 9 and 10 demonstrate the clear qualitative relationship between the Lorenz attractor and its reconstruction from the $X(t)$ time series. The major features which are obviously preserved are the number of fixed points, their stability properties, and the disposition of the flow about them. The extent of the relationship can be further demonstrated by constructing a surface of section in the embedding space. The surface chosen in fig. 11a is analogous to that used in fig. 5a – viz., it contains the two unstable foci and is transverse to their unstable manifolds (and in fact contains the c_3-axis). The pseudo one-dimensional first-return map is shown in fig. 11b. We note that the return map of fig. 11b is indistinguishable from the return map in fig. 5b obtained from the full Lorenz model. While this result is gratifying, it is unexpected since in general one expects the two maps to be related by a diffeomorphism.

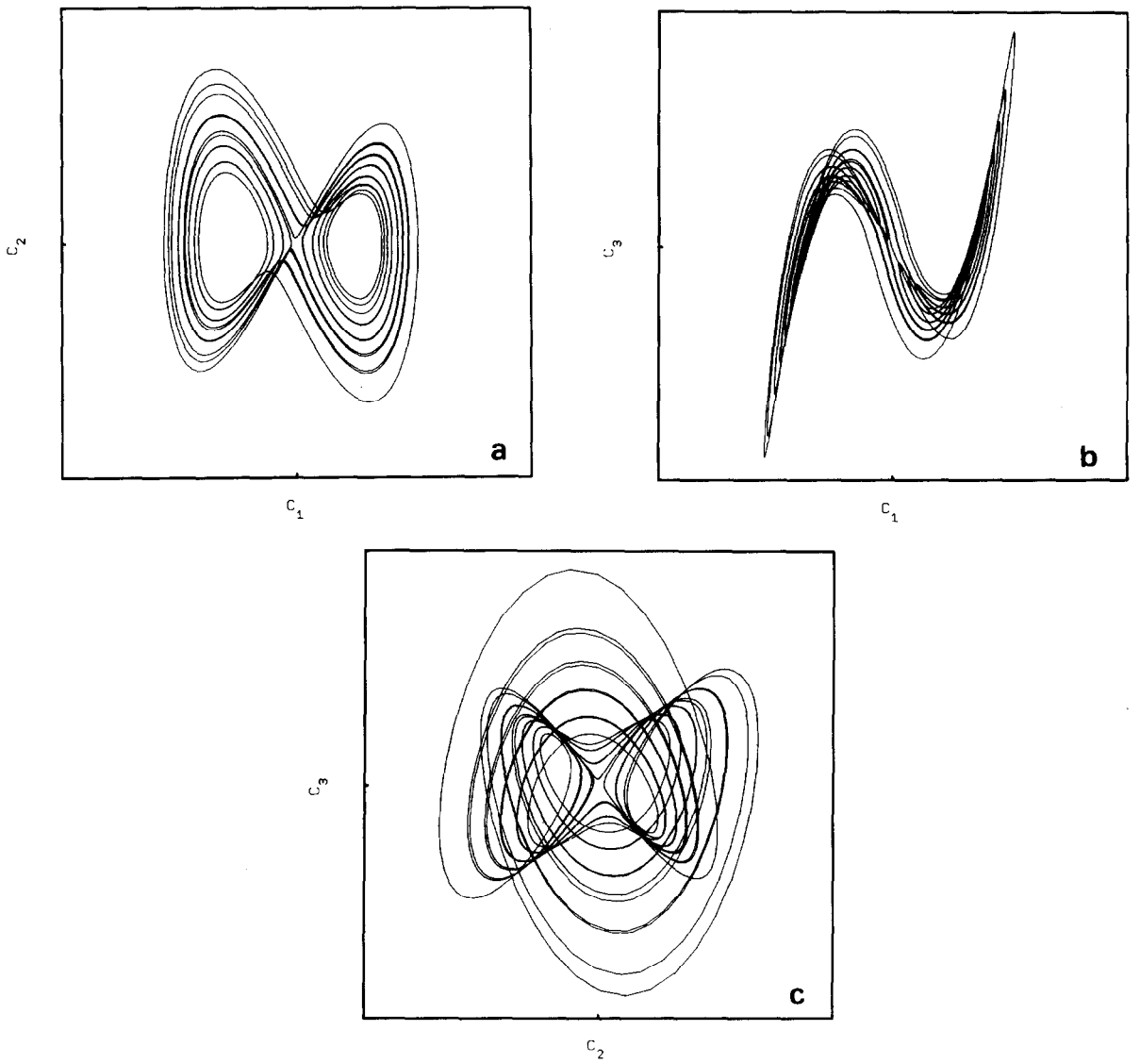

Fig. 9. Plots of the trajectory, $\boldsymbol{X}$, projected onto the three mutually orthogonal planes spanned by the singular vectors $\{\boldsymbol{c}_1, \boldsymbol{c}_2, \boldsymbol{c}_3\}$ shown in fig. 8. The ith point on the $(\boldsymbol{c}_j, \boldsymbol{c}_k)$-plane is given by $(\boldsymbol{c}_j^{\mathrm{T}}\boldsymbol{x}_i, \boldsymbol{c}_k^{\mathrm{T}}\boldsymbol{x}_i)$.

The closeness of the two maps can be understood by considering the form of the singular vectors $\boldsymbol{c}_1$ and $\boldsymbol{c}_2$ (see fig. 8). Recall that the vector $\boldsymbol{x}$ is constructed from a sequence of (seven) consecutive samples of the time series $X(t)$. Therefore, the component $\boldsymbol{c}_1^{\mathrm{T}}\boldsymbol{x}$ corresponds to the average of $X(t)$ in the window, and the component $\boldsymbol{c}_2^{\mathrm{T}}\boldsymbol{x}$ corresponds to an averaged central difference approximation to $\mathrm{d}X(t)/\mathrm{d}t$. The surface of section used in fig. 11b is the plane $\boldsymbol{c}_2^{\mathrm{T}}\boldsymbol{x} = 0$ and the coordinate used for the return map is $\boldsymbol{c}_1^{\mathrm{T}}\boldsymbol{x}$. Thus the surface of section picks out turning points in $X(t)$ and the return map is constructed from the window-averaged value of $X(t)$. On the other hand, by inspection of fig. 4 one can see that the surface of section chosen for the full Lorenz model will intersect the trajectory near its turning points in the XY-plane. Since the coordinate used for the

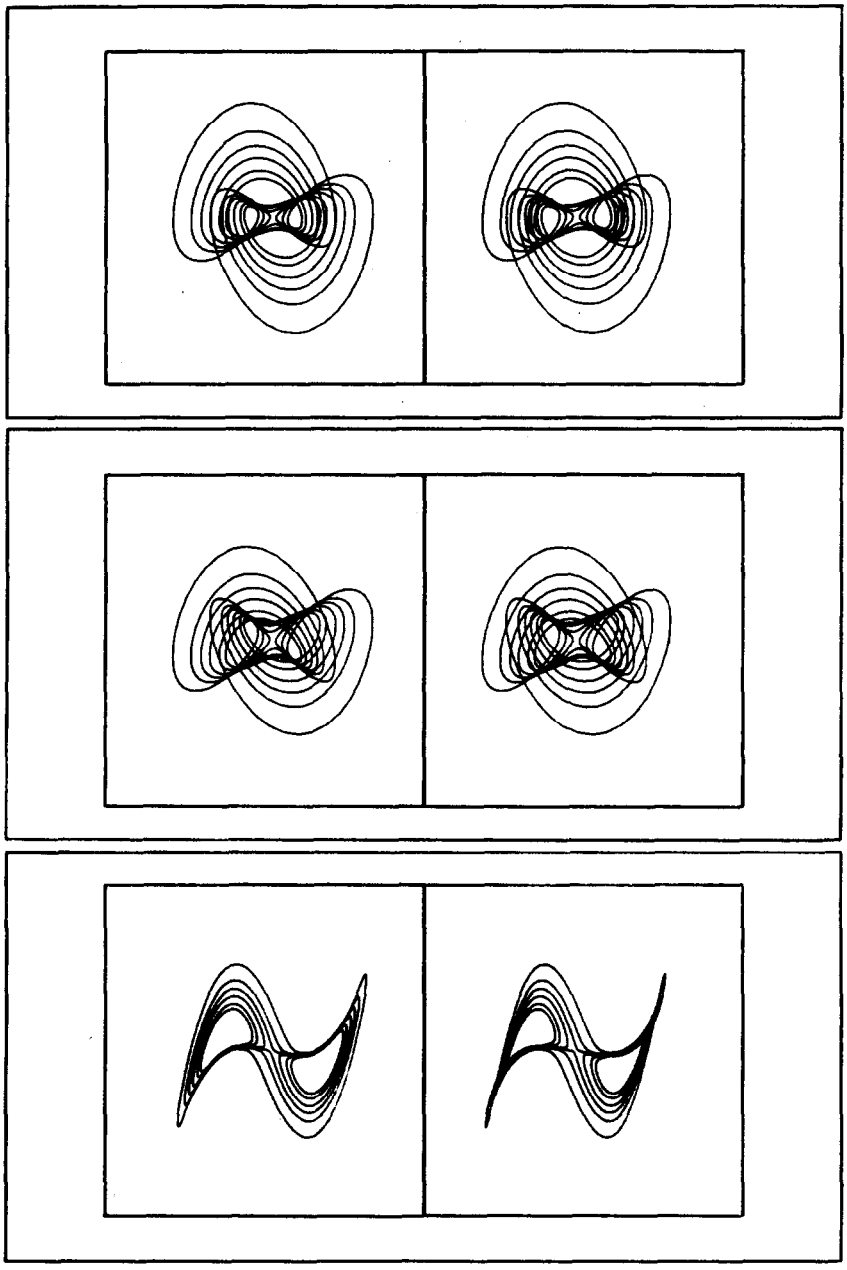

Fig. 10. Stereo pairs of $\boldsymbol{X}$ in the subspace spanned by $\{c_1, c_2, c_3\}$ corresponding to the trajectory shown in fig. 9.

return map in fig. 5b was chosen to be $X(t)$, this map is close to that shown in fig. 11b. Thus the apparent congruence of the two maps is a feature of the simple example we have taken and should not be expected in a more general setting.

Finally, we illustrate the advantages of the singular system approach in extracting qualitative dynamics from data corrupted by noise. For this we use $X(t)$ data obtained with $\tau_s = 0.009$ and quantized to 6-bits of precision. Fig. 12a shows the

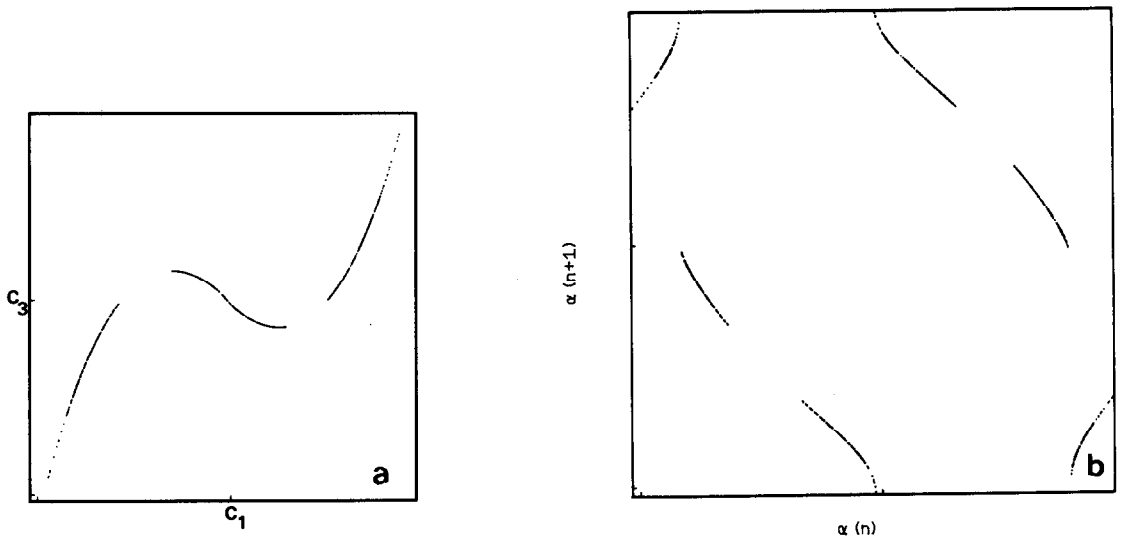

Fig. 11. (a) The intersection of the attractor reconstructed from $X(t)$ data with a planar surface of section containing the c_3-axis and passing through the two unstable foci. (b) The corresponding first-return map using the c_1-coordinate of the points of intersection. ($\alpha = c_1^T x$ where x is a point on the surface of section.)

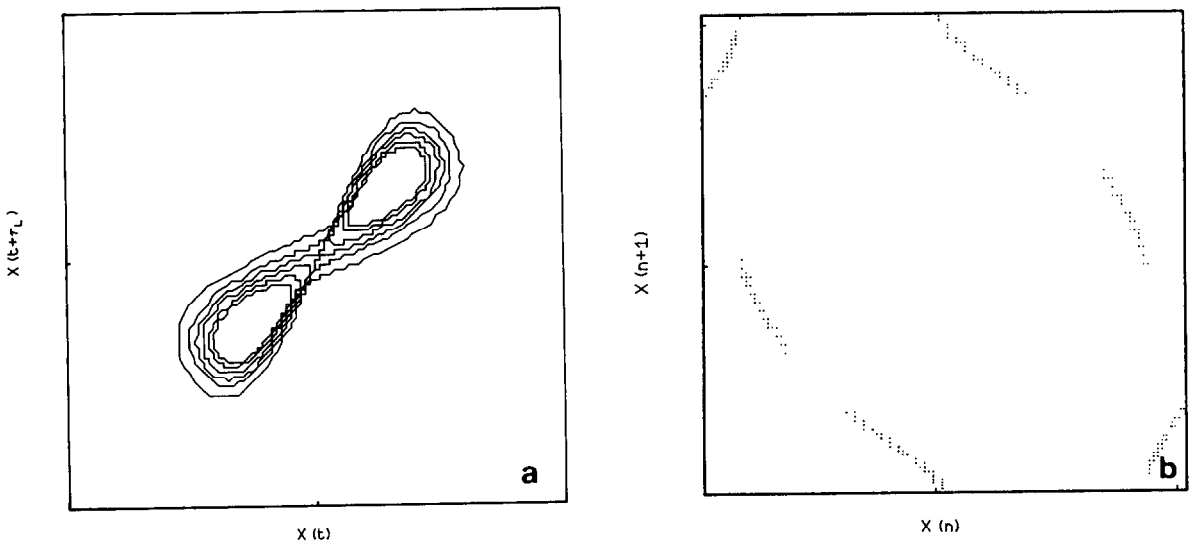

Fig. 12. (a) A phase portrait constructed by the method of delays with a lag time, $\tau_L = 10\tau_s$, using $X(t)$ data obtained with $\tau_s = 0.009$ and quantized to 6-bits of precision. (b) The corresponding first-return map obtained using a surface of section perpendicular to the $(X(t), X(t+\tau_L))$-plane and containing the two unstable foci.

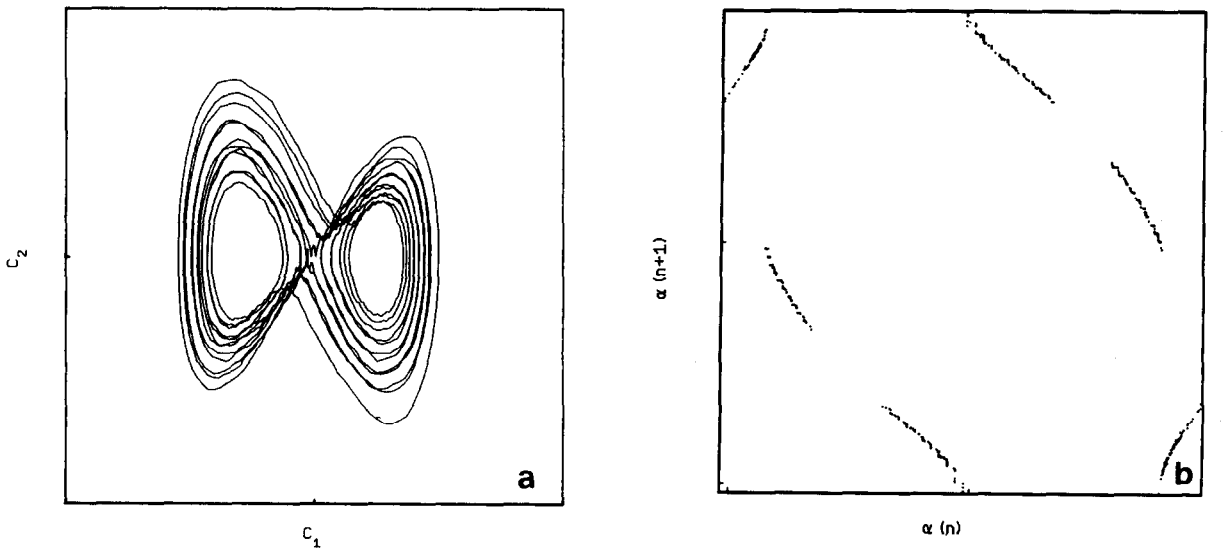

Fig. 13. (a) A phase portrait constructed by the singular system method with a window length $\tau_w = 13\tau_s$ using $X(t)$ data obtained with $\tau_s = 0.009$ and quantized to 6-bits of precision. Compare with fig. 12a. (b) The corresponding first-return map obtained using a surface of section perpendicular to the (c_1, c_2)-plane and containing the two unstable foci. Compare with fig. 12b.

phase portrait constructed by the method of delays with $\tau_L = 10\tau_s$, and fig. 12b shows the corresponding first-return map (which was obtained on a surface of section analogous to those used previously). The effect of the noise is obvious. These figures should be compared with figs. 13a and b which have been obtained from the same data using the singular system method. Following the prescription described in section 3.4, we require a window length larger than that obtained previously. This is due to the increased level of noise which results in a smaller value of ω^*. For the present data the prescription suggests a window length of $13\tau_s$. The singular spectrum for this window length and noise level has only three singular values above the noise floor, but the forms of their corresponding singular vectors are unchanged from those shown in fig. 8. The improved quality of the phase portrait and first-return map is due to the fact that the singular system approach generates averages of data within the window. In contrast the method of delays uses raw (unaveraged) data.

5. Conclusions

Dynamical systems theory has provided a language and a point of view for the study of nonlinear phenomena in physical systems. However, it is a geometric theory and hence requires that experimental results be analysed for geometric information. The work reported here and developed elsewhere [20] is intended ultimately to provide the experimentalist with statistical tools for qualitative analysis.

Acknowledgements

We gratefully acknowledge useful discussions with E.R. Pike, E. Jakeman, R. Jones, M. Johnson, I. Clarke, and J. Mather.

Appendix

In this appendix we give some practical details and timings for the implementation of the methodology presented in this paper.

As in section 3, consider a discrete time series of an observable, v, of length $N_T = N + (n-1)$:

$$v_1, v_2, \ldots, v_{N_T}$$

The (k,l)th element of the $n \times n$ covariance matrix, Ξ, is:

$$\Xi_{kl} = \frac{1}{N} \sum_{i=1}^{N} v_{i+k-1} v_{i+l-1}. \qquad (A.1)$$

It is easy to show that:

$$\Xi_{k+1,l+1} = \Xi_{kl} + \frac{1}{N} \{ v_{N+k} v_{N+l} - v_k v_l \}. \qquad (A.2)$$

Therefore, the time involved in calculating Ξ is essentially the time required to calculate the first row of this matrix. A typical time for the calculation of a 25×25 covariance matrix for a data set with $N_T = 32{,}768$ on a 16-bit laboratory minicomputer possessing a floating point hardware unit is about 3 minutes. The diagonalization of a matrix of this size takes only about 10 seconds.

The amount of data required for the calculation of Ξ should follow the usual rules for ensuring the convergence of the autocorrelation function of a stationary process. Since the purpose at this stage is to extract the second-order statistics from the time series for the calculation of the Karhunen–Loeve basis, the requirement on the amount of data is far less formidable than would be needed for the calculation of higher-order quantities such as Lyapunov exponents, the dimension of the manifold, etc.

References

[1] N.H. Packard, J.P. Crutchfield, J.D. Farmer and R.S. Shaw, "Geometry from a time series", Phys. Rev. Lett. 45 (1980) 712.

[2] F. Takens, "Detecting strange attractors in turbulence", Lecture Notes in Mathematics, D.A. Rand and L.-S. Young, eds. (Springer, Berlin, 1981) p. 366.

[3] (a) M. Bertero and E.R. Pike, "Resolution in diffraction-limited imaging, a singular value analysis. I: The case of coherent illumination", Opt. Acta. 29 (1982) p. 727.
(b) E.R. Pike, J.G. McWhirter, M. Bertero and C. de Mol, "Generalised information theory for inverse problems in signal processing", IEE Proceedings, 131, pt. F, No. 6 (Oct. 1984) p. 660. Also see the references therein.

[4] D. Edelson, R.J. Field, R.M. Noyes, "Mechanistic details of the Belousov–Zhabotinskii oscillations", Int. J. Chem. Kinet. 7 (1975) 417.

[5] R.C. Di Prima and H.L. Swinney, "Instabilities and transition in flow between concentric rotating cylinders", in: Hydrodynamic Instabilities and the Transition to Turbulence, H.L. Swinney and J.P. Gollub, eds. (Springer, Berlin, 1981) p. 139.

[6] H. Poincaré, "Mémoire sur les courbes définies par les équations différentielles I–VI", Oeuvre I. Gauthier-Villar: Paris (1880–1890).
H. Poincaré, "Sur les équations de la dynamique et le Problème de trois corps", Acta Math. 13 (1890) 1–270.
H. Poincaré, "Les Méthodes Nouvelles de la Mécanique Céleste", 3 vols. (Gauthier-Villars, Paris 1899).

[7] J. Guckenheimer and P. Holmes, Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields (Springer, Berlin, 1983).

[8] J. Palis Jr. and W. de Melo, Geometric Theory of Dynamical Systems: An Introduction (Springer, New York, 1982).

[9] (a) H. Whitney, "Differentiable Manifolds", Ann. Math. 37 (1936) 645.
(b) M.W. Hirsch, Differential Topology (Springer, New York, 1976).
(c) for a good introduction to embeddings:
D.R.J. Chillingworth, Differential Topology with a View to Applications, Research Notes in Mathematics 9 (Pitman, London 1976).

[10] J.C. Roux, R.H. Simoyi and H.L. Swinney, "Observation of a strange attractor", Physica 8D (1983) 257–266.

[11] G.W. Stewart, Introduction to Matrix Computations (Academic Press, New York, 1973).

[12] J.G. McWhirter, private communication.

[13] P.A. Devijver and J. Kittler, Pattern Recognition: A Statistical Approach (Prentice-Hall, New York, 1982).

[14] D.S. Broomhead, J.N. Elgin, E. Jakeman, S. Sarkar, S.C. Hawkins and P. Drazin, "Statistical properties in the chaotic regime of the Maxwell–Bloch equations", Optics Comms. 50 (1984) 56.

[15] P. Grassberger and I. Procaccia, "Dimensions and entropies of strange attractors from a fluctuating dynamics approach", Physica 13D (1984) 34.

[16] E.N. Lorenz, "Deterministic nonperiodic flow", J. Atmospheric Sci. 20 (1963) 130.

[17] D.S. Broomhead, R. Jones and G.P. King, "Qualitative dynamics of an electronic oscillator", in preparation.

[18] G. Rowlands, "An approximate analytic solution of the Lorenz equations", J. Phys. A16 (1983) 585.

[19] O. Rössler, "Different types of chaos in two simple differential equations", Z. Naturforsch 31a (1976) 1661.

[20] D.S. Broomhead and G.P. King, "On the qualitative analysis of experimental dynamical systems", to appear in: Nonlinear Phenomena and Chaos, S. Sarkar, ed. (Adam Hilger, Bristol, 1986).