# Time Series Analysis

A Part III Essay

Ulf Martin Engel

Darwin College
Cambridge CB3 9EU
United Kingdom

1 May 1991

(second, improved version: 15 November 1991)

**Abstract**

It is possible to reconstruct from a one-dimensional time series (obtained by an experiment or numerically) the attractor of the corresponding dynamical system. The technique used for this purpose, the method of delays, is described, a theoretical justification (using an embedding trick) is given and the problems one encounters when implementing this technique in practice are discussed. The importance of the method of delays lies in the fact that, in principle, it enables us to retrieve "all" the characteristic quantities of a (strange) attractor from the time series, no matter how complicated the structure of the attractor is, despite of the simple form of the data.

# Contents

# 1 Introduction

In dynamical systems theory it is a common problem to analyze a discrete set of observed quantities, a *time series*.

Consider for example the dynamical system defined by the differential equation

$$\frac{dy}{dt} = F(y) \quad , \tag{1}$$

where $y = y(t) = (y_1, y_2, \ldots, y_s)^T \in S$ is a vector representing the state of the system at time $t$ in some $s$-dimensional phase space $S$. The vector field $F : S \to TS$ ($TS$ is the tangent space of $S$.) induces the flow $\varphi_t : S \to S$ :

$$y(t) = \varphi_t(y_0) \quad ; \tag{2}$$

here, $y_0$ comes from the initial condition $y(t = 0) = y_0$. Often, when one analyzes a system, the analytic solution $y(t)$ is not known. Instead, only a *time series* is available, a series of values of one single quantity sampled at regular intervals:

$$v(y_0), v(\varphi_\tau(y_0)), v(\varphi_{2\tau}(y_0)), \ldots \equiv v_0, v_1, v_2, \ldots \quad . \tag{3}$$

$v(\varphi_t(y_0)) = v(y(t))$ is the value of some observable $v$ of the system at time t. The time interval $\tau$ between two successive measurements of $v$ is the *sampling time* or *delay time*. $\tau$ need not be fixed; it is possible to consider the sequence

$$v(\varphi_{t_0}(y_0)), v(\varphi_{t_1}(y_0)), v(\varphi_{t_2}(y_0)), \ldots \equiv v_0, v_1, v_2, \ldots \quad , \tag{4}$$

as well, where the times $t_0 < t_1 < t_2 < \cdots$ are not equidistant. One can get such a time series as the result of an experiment (where it is often convenient to measure only one single observable) or from the numerical integration of eq. (1). It is clear that having only a time series, instead of the complete solution, is a strong restriction of our knowledge about the system.

Alternatively, one might have a discrete dynamical system, defined by the mapping

$$y_{n+1} = \varphi(y_n), \qquad n = 0, 1, 2, \ldots \quad , \tag{5}$$

where $y_n \in S$, $\varphi : S \to S$, and get the time series $\{v_n \equiv v(y_n)\}_{n \in \{0,1,2,\ldots\}}$.

As the system evolves, the trajectory in the phase space $S$ approaches an *attractor*[1] $A$ which lies within some submanifold $M$ of $S$:

$$A \subseteq M \subseteq S \quad , \tag{6}$$

[1] We assume that our dynamical system is *dissipative*, i.e. the phase flow $\varphi_t$ (or the map $\varphi$, respectively) contracts volume in (some part of) the phase space. Otherwise the discussion of attractors would not be very meaningful.

where

$$\dim A \leq \dim M \leq \dim S \quad . \tag{7}$$

The general aim of the analysis of the time series is to describe the behaviour of the system on this attractor $A$. One would like to reconstruct a geometric picture of $A$ in phase space. Also one is interested in characterizing $A$ by calculating its Hausdorff-Besikovich (or fractal) dimension, information dimension, correlation dimension, Lyapunov exponents, etc. The problems one encounters when computing these quantities from a time series will be discussed in this essay.

In section 2, we will show how an embedding trick can be used to get a simple technique of reconstruction of phase space pictures. Section 3 deals with the application of this technique. In particular, we will present several strategies how to choose some of the important parameters of the reconstruction. Additionally, we will consider the problems arising from noisy data, and we will discuss the requirements the data (i.e. the time series) has to fulfill to allow for reasonable results. The important news is that, in fact, it is possible to get those reasonable results, despite of the very restricted form of the data.

# 2 Geometry from an Ideal Time Series

In this section only an ideal time series is considered which is to be thought of as consisting of *infinitely* many measurements $\{v_i\}_{i=0}^{\infty}$. What is more, this data is assumed to be totally error-free, i.e. we do not consider the effect of noise.

## 2.1 The Principle of Reconstruction – Method of Delays

The first group to propose a solution for the problem of extracting geometric information from time series was Packard et al. [3]. The basic idea of their approach is that the state of an $m$-dimensional dynamical system can be uniquely characterized by $m$ independent quantities. One such set of independent quantities are, of course, the phase space coordinates (more exact: the coordinates in some $m$-dimensional basis which spans $M$)

$$y(t) = (y_1(t), \ldots, y_m(t))^T \quad , \tag{8}$$

but these are not available, since the only data one has is the one-dimensional time series. Based on the conjecture that *any* $m$-tupel of numbers should give

equivalent results (in the sense that, if one reconstructs several phase portraits in accordance with this idea, then for any two of these phase portraits there should be a diffeomorphism which maps one onto the other) Packard et al. proposed two different possibilities to construct $m$-vectors which in some sense contain the same information as the original state vectors in eq. (8):

One can work with the *method of delays* which simply takes $m$ consecutive elements of the time series directly as coordinates in phase space:

$$x_i = (v_i, v_{i+1}, \ldots, v_{i+m-1})^T, \quad i = 0, 1, 2, \ldots \quad . \tag{9}$$

For example when analyzing a dynamical system it is often convenient just to take one coordinate of phase space as the observable $v$:

$$x_i = \Big(y_1(i\tau), y_1((i+1)\tau), \ldots, y_1((i+m-1)\tau)\Big)^T, \quad i = 0, 1, 2, \ldots \quad . \tag{10}$$

The other possibility suggested by Packard et al. is to use the time series to construct time derivatives of the measured observable[2] and use the vectors

$$\left(v(i\tau), \dot{v}(i\tau), \ddot{v}(i\tau), \ldots, \frac{d^{m-1}}{dt^{m-1}} v(i\tau)\right)^T, \quad i = 0, 1, 2, \ldots \quad . \tag{11}$$

In this way the one-dimensional time series is used to build a sequence of $m$-dimensional vectors which form a trajectory in some $m$-dimensional space. Numerical experiments done by Packard et al. show that this procedure gives reasonable results: the phase-space pictures one gets seem closely related to the corresponding pictures of the original dynamical system. In particular, topological properties and the geometrical form of the attractor seem to be conserved under this procedure.

---

[2] One can compute an approximation to the derivatives from the time series for example by using the formula

$$\frac{d^k}{dt^k} v(i\tau) = \frac{k!}{\tau^k} \sum_{j=-p/2}^{p/2} \alpha_j v((i+j)\tau) \; + \; \mathcal{O}\left(\tau^{\mathrm{p+1-k}}\right) \quad ,$$

where the $\alpha_j$ are determined by

$$\sum_{j=-p/2}^{p/2} j^l \alpha_j = \begin{cases} 1 & \text{if} \quad l = k \\ 0 & \text{if} \quad l \neq k \end{cases} \quad .$$

This formula results from differentiating the interpolating polynomial. See [10], §§ 27 and 28, for details of this method of numerical differentiation. See [11], chapter 3.4, for a more sophisticated extrapolation method.

As an illustration of this method of phase space reconstruction, we present a result of Broomhead and King [7]: They considered the *Lorenz system*:

$$\begin{aligned} \frac{dx}{dt} &= \sigma(y-x) \\ \frac{dy}{dt} &= \rho x - y - xz \\ \frac{dz}{dt} &= -\beta z + xy \end{aligned} \tag{12}$$

(with the parameter values $\sigma = 10$, $\beta = \frac{8}{3}$, $\rho = 28$). First they integrated this set of ordinary differential equations to get the "complete solution"; the projection of this solution onto the $xy$-plane is shown in Fig. 1.a. They chose the $x$-coordinate to be their "observable" $v$ and thus got the one-dimensional data stream as shown in Fig. 1.b. Then, after choosing some delay time $\tau$, they obtained a discrete time series from the data stream and used this time series to construct a phase portrait, using the method of delays: See Fig. 1.c. Comparison of the original Lorenz attractor (Fig. 1.a) and the reconstructed one justifies Packard's approach: Although the pictures are not exactly the same and although Fig. 1.a is much coarser than Fig. 1.c, both of them seem to show an object with the same geometry.

A shortcoming of Packard's method is that it does not give a general strategy how to choose $m$. If one knows that the dynamical system is evolving in $s$-dimensional phase space then one can use the idea mentioned above and take $m = s$. But if one is analyzing a system about the dimensionality of which nothing is known a priori then how to proceed? This problem will be addressed in section 3.

It is also not quite clear what $m$ should be when we try to reconstruct phase portrait of a system with infinitely many degrees of freedom, such as the Mackey-Glass equation ([5, 12] and [14], chapter 5.3):

$$\frac{dx(t)}{dt} = \frac{ax(t-\Delta)}{1+(x(t-\Delta))^{10}} \quad . \tag{13}$$

Figure 1.a. The attractor of the Lorenz system ($\sigma = 10$, $\beta = \frac{8}{3}$, $\rho = 28$), projected onto the $xy$-plane.

Figure 1.b. $x(t)$, the one-dimensional data stream obtained from the integration of eq. (12). From this data one gets the time series $x(i\tau)$, $i = 0, 1, 2, \ldots$.

Figure 1.c. A phase portrait of the Lorenz system, constructed by the method of delays.

## 2.2 Theoretical Foundation – Takens' Embedding Theorems

Apparently without knowing about the work of Packard et al., Floris Takens gave their approach a safe theoretical foundation [4]. Although different in details such as the dimension of the reconstructed phase space, the spirit of his work is much the same.

Perhabs inspired by *Whitney's embedding theorem*, which states that every $m$-dimensional manifold (of class $\mathcal{C}^r$, $2 \leq r \leq \infty$) can be embedded (via a $\mathcal{C}^r$ diffeomorphism) in $(2m+1)$-dimensional Euclidean space[3], Takens proposed to embed the attractor-manifold $M$ (i.e. the $m$-dimensional manifold which contains the attractor $A$) in $\mathbf{R}^{2m+1}$.

A smooth map $\Phi : M_1 \to M_2$, where $M_1$ and $M_2$ are smooth manifolds, embeds $M_1$ in $M_2$ ("is an *embedding*") if $\Phi$ is a diffeomorphism from $M_1$ to a smooth submanifold of $M_2$. $M_2$ is called the embedding space; the embedding dimension is $\dim(M_2)$. Notice that, in general, we have $\dim(M_1) \neq \dim(M_2)$[4]. The notion of embeddings comes into the game here, since one can think of $\Phi(M_1)$ as being the realization of $M_1$ as a submanifold of $M_2$: The topological structures of $M_1$ and $\Phi(M_1) \subset M_2$ are diffeomorphically equivalent. See Fig. 2.

Figure 2. Example of a manifold $M_1$ which is embedded in the manifold $M_2$ via the diffeomorphism $\Phi : M_1 \to M_2$.

This means that if we can find, using Takens' method, the embedding $\Phi$ which maps from $M$ to $\mathbf{R}^{2m+1}$ then we can analyze the structure of the

[3] For details about Whitney's theorem see e.g. [1].

[4] Refer to [2], for example, for a thorough treatment of embeddings.

trajectory of the dynamical system in $\mathbf{R}^{2m+1}$ and then easily from this infer properties of the actual trajectory on the attractor in $M$. Consider, for example, the dynamical system in eq. (1); the embedding $\Phi$ would tell us that there is a dynamical system

$$\begin{aligned} G : \mathbf{R}^{2m+1} &\rightarrow \mathbf{R}^{2m+1} \\ \frac{dz}{dt} &= G(z) \\ z &= \Phi(y) \quad , \end{aligned} \tag{14}$$

and it follows already from "topological equivalence" (for which we even only need to have a homeomorphism $\Phi$ instead of a diffeomorphism) that [7]:

- having a singularity of $F$ in $y \in M$ is equivalent to having a singularity of $G$ in $\Phi(y) \in \mathbf{R}^{2m+1}$;
- closed orbits of $y$ correspond one-to-one to closed orbits of $\Phi(y)$;
- the forward and backward attractors of $\Phi(y)$ with respect to $G$ are the $\Phi$-pictures of the respective attractors of $y$ under action of $F$;
- in more general terms, the flow $\phi_t$ of the dynamical system (1) and the flow $\psi_t$ of (14) are topologically equivalent:

$$\begin{array}{ccc} y_0 & \xrightarrow{\phi_t} & y(t) \\ \Phi\downarrow & & \downarrow\Phi \\ z_0 & \xrightarrow{\psi_t} & z(t) \end{array} \quad . \tag{15}$$

The dynamical systems of eqs. (1) and (14) are said to have the same qualitative dynamics. This situation is represented pictorially in Fig. 3.

Figure 3. Two dynamical systems with the same qualitative dynamics, connected via the diffeomorphism $\Phi : M \rightarrow \mathbf{R}^{2m+1}$.

It is especially nice to have the embedding $\Phi$ if one wants to characterize the dynamics of the system quantitatively: In this case the dimensions

(Hausdorff, topological, correlation, ...) of the attractor in $M$ and in $\mathbf{R}^{2m+1}$ are the same [8].

Takens stated the following **theorem** (theorem 2 in [4]):
*Given a compact $m$-dimensional manifold $M$, with $F : M \to TM$ a $\mathcal{C}^2$-vector field ($F$ being the generating vector field of the flow $\varphi_t$) and $v : M \to \mathbf{R}$ a $\mathcal{C}^2$-function, define*

$$\begin{aligned} \Phi_{(F,v)} &: M \to \mathbf{R}^{2m+1} \\ \Phi_{(F,v)}(y) &= \Big(v(y), v(\varphi_1(y)), \ldots, v(\varphi_{2m}(y))\Big)^T \quad . \end{aligned} \tag{16}$$

*Then, genericallly, $\Phi_{(F,v)}$ is an embedding.*

Here, the term "generically" is of central importance: The proof of this theorem is based on the idea that, if $\Phi_{(F,v)}$ is an embedding, then for all points $y \in M$ the co-vectors

$$\Big(dv\Big)(y), \ \Big(d(v\varphi_1)\Big)(y), \ \ \ldots, \ \Big(d(v\varphi_{2m})\Big)(y) \tag{17}$$

must span the cotangent space $T_y^*(M)$. This is ensured if one requires $F$ to fulfill the following conditions:
$(i)$ for all points $y \in M$ with $F(y) = 0$ all eigenvalues of $(d\varphi_1))\,(y)$ must be different and not equal to one;
$(ii)$ no periodic integral curve of $F$ (i.e. solution $y(t)$ of eq. (1)) must have an integer period $\leq 2m + 1$.
Takens argues that these conditions are generically met by $F$; i.e. practically all functions $F$ meet these conditions[5], because the cases excluded by $(i)$ and $(ii)$ are structually unstable: If one adds only a very small perturbation to $F$ then the very special situation of degenerate eigenvalues will be destroyed and one will get two different eigenvalues instead. This is comparable to the non-generic case that a smooth function has a double zero: one can get two different zeroes by changing the function "a little". Similarly it is non-generic that one of the eigenvalues is 1, since a nearby function $\tilde{F}$ will have a corresponding eigenvalue $1+\varepsilon$ instead. A situation as in $(ii)$ can be changed by adding a small perturbation as well[6]. So we can say that usually the situations $(i)$ and $(ii)$ do not occur; hence it makes sense to speak of generic $F$ giving rise to $\Phi_{(F,v)}$ being an embedding.

[5] One can interpret the term "generically" in this case as follows: Consider the function space $W$ of all $\mathcal{C}^2$-functions $F$ which map from $M$ into $TM$; then every subset of $W$ consisting only of functions which do not meet condition $(i)$ and $(ii)$ has zero measure in $W$.

[6] As hinted by Broomhead and King [7] it is not perfectly clear on generic grounds that one can exclude solutions with integer periods $\leq 2m+1$: In general one cannot argue that a perturbation of the generating function $F$ will automatically change the period of the flow. Broomhead and King circumvent this problem by not considering $\tau$ to be fixed; instead they make it small enough so that Takens' period-condition is met. See section 3.4.3.

What are the conclusions to be drawn from this theorem? The theorem considers a time series which is sampled in regular time intervals as measurements of the observable $v$ at times $t = 0, 1, 2, \ldots$. According to Takens' theorem one can construct a $(2m+1)$-dimensional vector $\Phi_{(F,v)}(y)$ from the data and this vector is equivalent to the vector $y$ representing the system on the manifold $M$ which contains the attractor (the attractor is assumed to be simple enough such that it can be contained in a compact manifold). This equivalence is mathematically described by the diffeomorphism $\Phi_{(F,v)}$. So we have

$$\begin{aligned} y \in M \quad &\overset{\Phi_{(F,v)}}{\Longleftrightarrow} \quad (v_0, v_1, v_2, \ldots, v_{2m})^T \in \mathbf{R}^{2m+1} \\ \varphi_\tau(y) \in M \quad &\overset{\Phi_{(F,v)}}{\Longleftrightarrow} \quad (v_1, v_2, v_3, \ldots, v_{2m+1})^T \in \mathbf{R}^{2m+1} \\ \varphi_{2\tau}(y) \in M \quad &\overset{\Phi_{(F,v)}}{\Longleftrightarrow} \quad (v_2, v_3, v_4, \ldots, v_{2m+2})^T \in \mathbf{R}^{2m+1} \\ &\ldots \end{aligned} \tag{18}$$

(At this stage, we are still restricted to $\tau = 1$, but below we will show that, in fact, nearly every $\tau \in \mathbf{R}_+$ can be chosen.) The difference between Packard's conjecture und Takens' approach is that Takens requires the embedding space to have a higher dimension $(2m+1)$ than one would ad hoc expect $(m)$. This requirement *ensures* that the embedding exists[7], but of course it still may be possible to get reasonable results with a smaller embedding dimension, as one can see for example from Packard's numerical results.

Takens proves two further theorems which give similar results: One of these theorems justifies the method of delays for maps (systems defined by eq. (5)) instead of flows (systems like eq. (1), as considered in the above theorem); the other one works with embeddings reconstructed using time derivatives of the observable and corresponds to Packard's second proposed method (eq. (11)): In both cases it is again possible, under genericity assumptions for $\varphi$ and $F$, to embed $M$ in $(2m+1)$-dimensional Euclidean space[8]. We have for the derivative method:

$$\begin{aligned} y \in M \quad &\overset{\Phi_{(F,v)}}{\Longleftrightarrow} \quad \left(v_0, \dot{v}_0, \ddot{v}_0, \ldots, \frac{d^{2m}}{dt^{2m}} v_0\right)^T \in \mathbf{R}^{2m+1} \\ \varphi_\tau(y) \in M \quad &\overset{\Phi_{(F,v)}}{\Longleftrightarrow} \quad \left(v_1, \dot{v}_1, \ddot{v}_1, \ldots, \frac{d^{2m}}{dt^{2m}} v_1\right)^T \in \mathbf{R}^{2m+1} \\ \varphi_{2\tau}(y) \in M \quad &\overset{\Phi_{(F,v)}}{\Longleftrightarrow} \quad \left(v_2, \dot{v}_2, \ddot{v}_2, \ldots, \frac{d^{2m}}{dt^{2m}} v_2\right)^T \in \mathbf{R}^{2m+1} \end{aligned} \tag{19}$$

[7] For attractors with a simple geometric structure a smaller embedding dimension may be sufficient, but the more complicated the structure of $A$ is (e.g. if the attractor is Cantor set-like or if there are many "backfoldings") the higher the embedding dimension must be [5]. The importance of Takens' result is that, no matter how complex the structure of $A$, $2m+1$ dimensions always suffice.

[8] There is one peculiarity for the method using time derivatives: Here, $F$ and $v$ must be at least $\mathcal{C}^{2m+1}$-functions, and this stricter requirement may become a problem if the system or the observable are not that "well-behaved".

...

So both methods suggested by Packard rather intuitively are hereby justified, although slightly altered.

With all these theorems one knows how to construct meaningful phase space vectors from one-dimensional data which has been measured with the sampling time $\tau = 1$. This result is interesting but not exhaustive, because one would like to have the possibility of adjusting $\tau$ to each individual situation rather than having to fix it at some given value; also there is obviously no physical reason for giving this very special role to the unit time interval. So we would like to generalize the above result to arbitrary time steps $\tau$. Additionally, the important question if one can really reconstruct the *attractor* of the dynamical system still remains unanswered. It is not clear at all that the reconstructed phase space vectors tend to the same attractor as the picture of the actual flow $\varphi_t(y)$ of the dynamical system does. For example it could be that the reconstructed vectors visit only a part of the attractor's equivalent in $\mathbf{R}^{2m+1}$. The reason for these doubts is that one is not using measurements which are made at random times (This would give rise to the assumption that all these measurements together actually give a true picture of all parts of the system's trajectory.) but at equidistant times. So one has to be aware of the possibility that this very special selection of data points could result in non-equivalence of the original and the reconstructed attractors. This uncertainty would mark a fundamental flaw of the attempt to get a geometrical picture of the attractor, but, fortunately, Takens provides us here with a theorem, too, which solves both problems.

This **theorem** (theorem 4 in [4]) says that, *for a compact manifold $M$, a vector field $F$ on $M$ with the flow $\varphi_t$ and for $y \in M$, the attractors for the point $p$ of the flow $\varphi_t$ and of the mapping $\varphi_{\tau \cdot i}, \tau \in \mathbf{R}_+, i = 0, 1, 2, \ldots$ are the same, generically.*

The term "generically" refers in this case to the number $\tau$ and means that the theorem is true for "almost all" positive real numbers $\tau$. Only for a small subset of $\mathbf{R}_+$ the theorem does not hold, and the probability of choosing "accidentally" one of the elements of this subset is zero.

Thus both of the above problems are solved hereby; the theorem tells us that we actually can use a time series with a sampling time which we are free to choose, and despite of the discretization of the original continuous flow the limit of $\varphi_{\tau \cdot i}$ is really equivalent to the original attractor. So eq. (18) and (19) hold for nearly all $\tau$.

Summarizing the results of this section we have seen that, given a time series of infinite length, *one can construct (using e.g. the method of delays)*

*in practically all cases (i.e. "generically") an infinite series of vectors the limit set of which is diffeomorphically equivalent to the attractor of the original dynamical system.* The embedding process which gives this result is summarized in Fig. 4.

Figure 4. Embedding process after Packard et al. and Takens.

One has to stress that this result is somewhat *theoretical*, since talking about limit sets and attractors requires an infinitely long time of observation and thus an infinitely long time series: The last theorem does *not* give a hint how many data points or reconstructed phase space vectors one needs to get an approximation which is "good" enough for the "diffeomorphical equivalence" to be true at least approximately. What is more, it is implicitely assumed that transient initial behaviour has died away and that the measured time series really corresponds only to phase states *on* the attractor. Obviously this can be only an approximation to any real experimental situation where one will always have trajectories which are near to the attractor (whatever that means in each individual case) but not on it. No general information can be given how long we must wait to be sure that the the trajectory is near enough to the attractor, so it is necessary to investigate this problem in each case individually.

Also, the above treatment implicitely assumes error-free measurements; the accuracy of the time series is not being questioned but taken for granted. In the next section, we will deal with these problems in more detail.

# 3 Analysis of Real-World Data

So far, the method of delays remains ill-defined since it does not take into account the practical obstacles which arise: generally, the dimension $m$ of $M$ is not known a priori, so the choice of a proper embedding dimension $n$ is a problem.
Also, instead of having the infinitely long time series considered by Takens one works in practice with a time series of finite length (from which we construct a series of $N$ vectors with the method of delays), so we need to find a criterion for the smallest number of data points required to get "good" results.
Takens' theory does not say anything about the sampling time $\tau$ between two successive measurements. (It has not been necessary so far to say something about it, because for $N \to \infty$ generically any value of $\tau$ can be taken.) For a real data set $N$ is finite, so we cannot expect any more that every sampling time will give good results. We must assign a suitable value to $\tau$.
Additionally, all data, no matter if it is taken from numerical or physical experiments, will be noisy (i.e. it will be measured with finite precision only) and we must consider this problem when building our strategy of data analysis.

## 3.1 Basic Remarks about the Choice of Embedding Dimension and Sampling Time

The first idea one might have about the choice of the sampling time $\tau$ is to make it as small as possible, such that one can reconstruct a "smooth" trajectory in the embedding space. However, this approach is rather short-sighted [5]: If we choose $\tau$ too small then consecutive measurements of $v$ will give nearly the same results,

$$v_i \approx v_{i+1} \approx v_{i+2} \approx \ldots \approx v_{i+n-1} \quad . \tag{20}$$

This means that the vectors

$$x_i = (v_i, v_{i+1}, v_{i+2}, \ldots, v_{i+n-1})^T \quad , \tag{21}$$

constructed via the method of delays, will be stretched along the diagonal in the $n$-dimensional embedding space and thus the analysis of the picture of

the attractor will be very difficult. To get an intuitive picture of what is happening in this case one can think of the phase space picture being artificially compressed towards the diagonal and this decreases the dimensionality of the attractor although there is no physical reason for this. In fact, numerical experiments show that for small sampling times one gets spuriously low results of dimension calculations, for example the correlation dimension tends to zero as $\tau$ approaches zero [8].

On the other hand $\tau$ must not be chosen too large, because in this case the $v_i$ become totally uncorrelated (since one of the features of a "chaotic" system is the exponential separation of nearby trajectories, and thus the noise which is present in every real system gives rise to total non-correlation of measurements which are made in sufficiently large time intervals). This means that the vectors $x_i$ fill (the relevant part of) the embedding space more or less homogeneously and extraction of any information from this phase space picture becomes impossible [5]. It is possible to regard this problem of $\tau$ being too large from another point of view [8]: We have seen in section 2.2 that one can also use (instead of the method of delays) a vector the components of which are the first $n$ derivatives of the observable to construct a phase space picture. Now, since we cannot measure the derivatives themselves we compute approximations to them using the time series. It is typical for such approximation formulae that the error of the approximant is $\mathcal{O}(\tau^k)$ with some integer $k$. (See for example footnote 2 in section 2.2.) This means that to get a sensible result for this approximation, and thus for $x_i$, we want to have $\tau$ small!

We illustrate the effect of a "bad" choice of $\tau$ with a numerical experiment: We consider a derivate of the Lorenz system, the *Rössler system* (see e.g. [14], chapter 5.3), which is defined by

$$\begin{aligned} \frac{dx}{dt} &= -z - y \\ \frac{dy}{dt} &= x + ay \\ \frac{dz}{dt} &= b + z(x - c) \quad . \end{aligned} \tag{22}$$

For the parameter values $a = 0.15$, $b = 0.20$, $c = 10$ the flow of this system becomes attracted to the Rössler attractor, the projections of which onto the $xy$- and to the $xz$-planes are shown in Fig. 5.a and Fig. 5.b, respectively. Numerical integration (with some step size $\Delta$) of eq. (22) provides us, for example, with the time series $\{x(i\Delta)\}_{i=0}^{\infty}$, depicted in Fig. 5.c. Using this data and several different delay times $\tau$, we can reconstruct (in 2-dimensional embedding space) the phase space pictures of the Rössler attractor which are shown in Fig. 5.d, ..., Fig. 5.h[9]. The effects we described in the above

[9] Fig. 5.d, ..., Fig. 5.h consist of clouds of points, in contrast to the continuous orbit

Figure 5. (a,b) Projections of the Rössler attractor onto the $xy$- and $xz$-planes, respectively. (c) Time series $x(i\Delta)$, $i = 0, 1, 2, \ldots$, obtained by numerical integration of the Rössler system. (d,e,f) Results of the method of delays, with $\tau = 10\Delta$, $50\Delta$, $140\Delta$, respectively.

of the system shown in Fig. 5.a and Fig. 5.b. This is not an inherent problem of the method of delays, but due to the limited computing power we had available for this little experiment: By choosing a smaller step size $\Delta$ while keeping $\tau$ constant, we would have got pictures similar to Fig. 5.d, ..., Fig. 5.h, but consisting of "quasi-continuous" orbits.

Figure 5. (g,h) Results of the method of delays, with $\tau = 235\Delta$, $2000\Delta$, respectively.

paragraph for $\tau$ too small (stretching along the diagonal) and $\tau$ too large (increasing non-correlation of the data) are clearly visible in Fig. 5.d and Fig. 5.h, respectively.

So it is clear that a *proper* choice for the sampling time is a necessary requirement for the method of delays to give sensible results. A similar statement holds for the embedding dimension $n$:

Obviously $n$ must be large enough to allow $M$ to be embedded within $\mathbf{R}^n$. The more complex the attractor is the higher $n$ must be (see footnote 7 in section 2.2). What is more, to make sure that the choice of $n$ allows Takens' theory to be applied one might be tempted to make $n$ *very* large in order to guarantee that $n \geq 2m + 1$. If $n$ is chosen too small then we change the distribution of the points on the attractor artificially and the point densitiy

$$p(x) = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \delta(x_i - x) \tag{23}$$

develops singularities which are absent in an embedding space with $n$ large enough [5]. (As a simple example consider the surface of a sphere in 3 dimensions which is "embedded" in 2 dimensions. Even if the point density on the sphere in 3-space is constant the "embedded" sphere, which is essentially a projection of the sphere into the plane, will have singularities on the boundary.) If it is known a priori that the attractor is visited uniformly then one can use this observation to determine the embedding dimension: Take the smallest $n$ such that the densitity on the reconstructed attractor does not have any singularities. However, this method requires some pre-knowledge about the system and cannot be taken as a general approach to the choice of $n$.

There are good arguments to keep $n$ small as well: It is easier to work with a low-dimensional embedding space, especially if one aims at building up some geometrical intuition about the attractor's shape. Also, working in higher dimensions results in larger numerical errors when one wants to use the reconstructed attractor to calculate the usual quantities which characterize the system, such as the Hausdorff or correlation dimension [5].

But the best argument not to make $n$ too large is perhabs the problem which we pointed to in section 2.2: in contrast to Takens' statement that systems with periods smaller than or equal to $(2m + 1)\tau$ can be excluded on generic grounds we saw that these periods actually *can* appear, generally speaking. So we have to take additional precautions to prevent such periodic flows with small periods to cause trouble, since otherwise Takens' theory can not be applied! An easy solution for this problem is to make $n\tau$ small [7].

## 3.2 Topological Considerations

Schuster [15] proposed to base the choice of $\tau$ and $n$ on the idea that an embedding using delay coordinates is a topological mapping which preserves neighbourhood relations. This means that points on the attractor in $M$ which are near to each other should also be near in the embedding space $\mathbf{R}^n$.

The distance of any two points $x_i,\ x_j \in \mathbf{R}^n$ cannot decrease but only increase when one increases the embedding dimension $n$. But if this distance increases under a change from $n$ to $n + 1$ then it is clear that $n$ is not sufficiently large, as Fig. 6 shows.

Figure 6. Illustration of an embedding dimension $n$ which is too small. The reconstructed picture of the attractor is shown schematically in $n$- and $(n + 1)$-dimensional embedding space.

$n$ being too small means that the attractor is projected onto a space of lower dimensionality $n$ and this projection possibly destroys neighbourhood relations, resulting in some points appearing nearer to each other in the embedding space than they actually are. (For example, $x_j$ may be the nearest neighbour of $x_i$ in $\mathbf{R}^n$ although this is not true in the proper embedding space $\mathbf{R}^{n+1}$. See, again, Fig. 6 for an illustration of this observation.) If, on the other hand, $n$ is sufficiently large then the distance of any two points of the attractor in embedding space should stay the same when one changes $n$ into $n+1$.

Applying this geometrical point of view, one can find the proper embedding dimension by choosing initially a small value of $n$ and then increasing it systematically. One knows that the proper value of $n$ is found when all distances between any two points $x_i$ and $x_j$ do not grow any more when increasing $n$. Practically, one constructs the quantity

$$Q(i,k,n) = \frac{d_{n+1}\left(x_i(n+1), x(i,k,n+1)\right)}{d_n\left(x_i(n), x(i,k,n)\right)} \quad , \tag{24}$$

where $x_i(n)$ is the $i$-th reconstructed vector in $n$-dimensional embedding space, $x_i(n) = (v_i, v_{i+1}, \ldots, v_{i+n-1})^T \in \mathbf{R}^n$, and

$$x(i,k,n) = \left\{ \begin{array}{c} \text{the } k\text{-th nearest neighbour } x_{j\neq i}(n) \text{ of } x_i(n) \\ \text{in } n\text{-dimensional embedding space} \end{array} \right\} \quad . \tag{25}$$

$Q(i,k,n)$ measures the increase of the distance between $x_i$ and its $k$-th nearest neighbour, as $n$ increases. ($d_n(.,.)$ is some appropriate, fixed metric in $\mathbf{R}^n$.) According to the observations stated above $Q$ should be greater than or equal to one. To get a notion what happens not only to the single point $x_i$ and its neighbours but to all the $x_i$ the next step is to calculate

$$\overline{W}(n) = \frac{1}{\tau N(N-1)} \sum_{i=1}^{N} \sum_{k=1}^{N-1} \log Q(i,k,n) \tag{26}$$

which considers all $N$ reconstructed points in the embedding space and all the $N-1$ "neighbours" of these[10] and adds up the logarithms of the ratios of the respective distances. (The number of the $x_i$ increases linearly with $\tau$, such that there would be a trivial linear $\tau$-dependence in $\overline{W}(n)$. This is removed by dividing by $\tau$.)

Clearly, for $n$ equal to the proper embedding dimension and for the right sampling time $\tau$, $\overline{W}(n)$ should approach zero (within the experimental and numerical errors). Thus systematic variation of $n$ and $\tau$ seems to enable us to find sensible values for these quantities. In fact, numerical experiments done by Schuster [15] show that one can get reasonable results when using this method.

[10] To save computing time it is also possible to consider not all the $N-1$ neighbours of each $x_i$ but only the $p$ nearest ones.

## 3.3 An Algorithm based on the Calculation of the Correlation Dimension

One of the many quantities which are usually used to characterize a strange attractor is the correlation dimension. It was introduced by Grassberger and Procaccia as a measure of strangeness which is easier to handle (especially numerically) than the measures used until then [5, 6]. For instance, if the dynamical laws of a system are not known then the Hausdorff dimension is usually computed using box-counting algorithms (see for example [14], chapter 3.3). These are very time-consuming, especially for higher-dimensional systems, because the number of calculations grows exponentially with the dimension [5, 6, 8]. The problems we encounter for our special case of analyzing a dynamical system where the only information about the system is given in the form of a time series are even larger, since we have to go through the embedding process first. Grassberger and Procaccia suggested to use, instead of the Hausdorff dimension, the correlation dimension $\nu$, which can be computed directly from the time series without greater difficulties. We will see that calculation of $\nu$ will provide us with a method to find a proper embedding dimension ([8] and [14], chapter 5.3), as well.

To determine the correlation dimension we first calculate the *correlation integral*:

$$C_N(r) = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} H\left(r - \|x_i - x_j\|\right) \quad , \tag{27}$$

where $r$ is the correlation length and $H$ the Heaviside function. For $N$ sufficiently large and $r$ sufficiently small the logarithm of $C_N$ as a function of the logarithm of $r$ will have a linear region, the scaling region, and the slope in this region is the *correlation dimension* $\nu$:

$$\log\left(C_N(r)\right) = \nu \log r + \cdots \quad . \tag{28}$$

Because $\nu$ is closely related to the Hausdorff-dimension[11] which itself is a generalization of the intuitive concept of dimensionality [17], it seems sensible to apply a procedure similar to the one presented in the previous section [8]: We calculate $C_N$ for a sequence of embedding dimensions $n$ and infer the respective $\nu(n)$'s. We expect that if the embedding dimension is too small then $\nu(n)$ equals $n$. (In this case $\nu(n)$ is not the true correlation dimension of the attractor but only an artefact, due to the embedding dimension being

[11] For the relationship of $\nu$ to the Hausdorff and information dimensions see [5, 6]; Schuster [14], chapter 5.3, shows that these three types of dimensions can be organized within the framework of *generalized dimensions* $D_q$: $D_{q=2} = \nu$, for example. It seems to be true that in most cases all three dimensions do not differ very much and often they are even the same, e.g. when the attractor is covered uniformly by the $x_i$ [6].

too small.) So increasing $n$ yields an increasing $\nu(n)$. But for $n$ sufficiently large $\nu(n)$ will be smaller than $n$ and equal to the true correlation dimension. A further increase should not change the value of $\nu$ any more.

Thus we have a method to compute the proper embedding dimension: we simply use the correlation dimension as the embeding dimension.

## 3.4 Singular System Analysis

The two techniques presented above solved only partly the problem of determining the delay time $\tau$ and none of it addressed the virulent problem of noise. Therefore we now turn to the *singular system analysis* [7] which shows a way how to determine the embedding dimension and the delay time reliably even in the presence of noise.

We proceed in three steps: In the first step (in section 3.4.1), we discuss a method to derive a suitable embedding dimension from a finite data set which is measured with infinite precision. Then, in section 3.4.2, we generalize the result of the first step to noisy data. The third step (in the second part of section 3.4.2) gives a slightly more sophisticated method of construction of the vectors $x_i$. In section 3.4.3, we again consider the problem of choosing $\tau$ and the embedding dimension.

### 3.4.1 Choice of a Proper Embedding Dimension for a Noise-Free Time Series of Finite Length

As a first step we choose some $n$. We want to embed $M$ in $\mathbf{R}^n$. In general $m$ will be unknown so that it is not obvious how to meet Takens' condition $n = 2m + 1$. But if we choose $n$ large enough to ensure that $n \geq 2m + 1$ then we *can* embed $M$ in $\mathbf{R}^n$, according to Takens' theorem 2, because it is a well known fact that every Euclidean space $\mathbf{R}^{2m+1}$ can again be embedded in any higher-dimensional Euclidean space $\mathbf{R}^n$ without any problems. So we start by making an *ansatz* for $n$ and our concern for the next few paragraphs will be to find a more appropriate $n' \leq n$ such that $\mathbf{R}^{n'}$ is a space containing Takens' embedding space. Thus $n'$ will be a better embedding dimension than the $n$ we started with, since it gives a lower-dimensional embedding space.

We want to use the method of delays to construct vectors in $\mathbf{R}^n$ from the $v_i$ of the time series. Doing this one realizes that there is even one more quantity which is not yet specified: One could, for example, take the series

of $n$-vectors

$$x_i = \left(v_i, v_{i+J}, \ldots, v_{i+(n-1)J}\right)^T, \quad i = 0, 1, 2, \ldots \quad , \tag{29}$$

which we are obviously allowed to use in accordance with Takens' statements. So we also have to choose the "lag time" $\tau_L = J\tau$, where $J \in \mathbf{N}_+$. We will see in section 3.4.2 that we do not have to spend much effort on choosing $J$: when we use the singular systems technique the influence of the lag time becomes insignificant, hence we will choose it from now on to equal $\tau$ (i.e. $J = 1$).

Consider a sequence of $N$ vectors $x_i \in \mathbf{R}^n$, $i = 1, 2, \ldots, N$ (i.e. we take a time series containing $N + n - 1$ data points)[12]. There seems to be no analytical way to compute the proper (i.e. minimal) embedding dimension from the time series. However, it is possible to determine a reasonable estimate for it: For some given $n$ the $n$-vectors $x_i$ usually do not explore the whole space $\mathbf{R}^n$. Rather than that they are restricted to some subspace $T$ of $\mathbf{R}^n$; T contains the embedded manifold $M$ which contains the picture of the attractor:

$$T = \mathrm{span}\{x_i \mid i = 1, 2, \ldots, N\} \quad . \tag{30}$$

When we assume that the $x_i$ really visit the whole attractor in the embedding space (more or less) uniformly and we bear in mind that $N$ usually is much larger than $n$ then $n' \equiv \dim(T)$ is a sensible upper bound for the minimal embedding dimension. In order to determine $n'$ we compute the maximum number of linearly independent vectors that can be constructed as linear combinations of the $x_i$. To do this, we define the $(N, n)$-*trajectory matrix* $X$:

$$X \equiv \frac{1}{\sqrt{N}} \begin{pmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_N^T \end{pmatrix} \quad , \tag{31}$$

which is built out of all vectors we want to use to reconstruct the attractor. Notice that when operating with $X^T$ on some $N$-vector we get an $n$-vector:

$$X^T s = c, \quad \forall s \in \mathbf{R}^N : c \in \mathbf{R}^n \quad . \tag{32}$$

Since we are interested in linearly independent $n$-vectors, we choose a set of vectors $\tilde{s}_i \in \mathbf{R}^N$ such that the $n$-vectors

$$X^T \tilde{s}_i = c_i \tag{33}$$

are orthonormal. We introduce some real constants $\sigma_i \in \mathbf{R}$ into this equation, in order to normalize the $\tilde{s}_i$:

$$X^T s_i = \sigma_i c_i \quad . \tag{34}$$

[12] We will generally assume $N \gg n$, which is obviously true in most cases.

The important point about this equation is that, after transposing, it can be rewritten as

$$s_{i1}x_1 + s_{i2}x_2 + \cdots + s_{iN}x_N = s_i^T X = \sigma_i c_i^T \quad , \tag{35}$$

i.e. as a linear combination of the reconstructed trajectory vectors. This tells us, when we keep in mind the definition of $n'$, that we can get $n'$ linearly independent vectors $c_i$, using eq. (34); so we have $n'$ vectors $s_i$ and $n'$ numbers $\sigma_i \neq 0$, too. The $c_1, c_2, \ldots, c_{n'}$ are elements of an orthonormal basis $\{c_i\}_{i=1}^n$ of $\mathbf{R}^n$. Thus we are left to determine $n'$ as the number of those $\sigma_i$ which are non-zero.

Define the *structure matrix* $\Theta \equiv XX^T$; then it follows from eq. (34) that the $\sigma_i^2$ are the eigenvalues of this matrix:

$$\Theta s_i = \sigma_i^2 s_i, \quad i = 1, 2, \ldots, n' \quad . \tag{36}$$

We could determine $n'$ as the number of the non-zero eigenvalues of $\Theta$. But $\Theta$ is a huge $(N, N)$-matrix, and singular, and its diagonalization is in practice impossible. Instead, we notice that the *covariance matrix* $\Xi \equiv X^T X$ has the same non-zero eigenvalues as $\Theta$, and $\Xi$ is much easier to diagonalize, because it is only an $(n, n)$-matrix. So all one has to do in order to calculate $n'$ which, cum grano salis, estimates the minimal embedding dimension $(2m+1)$, is to determine the number of the non-zero eigenvalues of

$$\Xi = \frac{1}{N} \begin{pmatrix} \sum_{i=1}^N v_i v_i & \sum_{i=1}^N v_i v_{i+1} & \cdots & \sum_{i=1}^N v_i v_{i+n-1} \\ \vdots & \vdots & & \vdots \\ \sum_{i=1}^N v_{i+n-1} v_i & \sum_{i=1}^N v_{i+n-1} v_{i+1} & \cdots & \sum_{i=1}^N v_{i+n-1} v_{i+n-1} \end{pmatrix} \quad . \tag{37}$$

Now we know that the trajectory is confined to an $n'$-dimensional subspace of $\mathbf{R}^n$, and we can use $\mathbf{R}^{n'}$ as the embedding space.

However, this treatment only makes sense in the case that we have noise-free data.

### 3.4.2 Singular Value Decomposition as a Noise Reduction Scheme

We will see that in the presence of noise one has to expect that all zero eigenvalues $\sigma_i$ (of the noise-free case) will be changed into non-zero numbers, small but non-zero! So the simple method of the above section breaks down and the result of the computation will be in almost all cases $n' = n$, no matter how large $n$ is.

To deal with this problem we introduce two new matrices

$$C = (c_1, c_2, \ldots, c_n) \quad , \tag{38}$$
$$\Sigma = \mathrm{diag}\,(\sigma_1, \sigma_2, \ldots, \sigma_n) \quad . \tag{39}$$

The elements $\sigma_i$ of $\Sigma$ are called the *singular values* of $X$. Using these definitions, the definition of $\Xi$ and eq. (34) we get the important relation

$$(XC)^T (XC) = \Sigma^2 \quad . \tag{40}$$

Recall that $\{c_i\}_{i=1}^n$ is by definition an orthonormal basis of $\mathbf{R}^n$. So the $(N, n)$-matrix $\tilde{X} \equiv XC$ can be interpreted as the projection of the trajectory matrix onto this new basis. Then $\tilde{X}$ is the trajectory matrix expressed in the basis $c_i$, and $\tilde{X}^T\tilde{X}$ in the above equation is the covariance matrix in the same new basis. If one considers the general form of a covariance matrix, $\Xi = \frac{1}{N}\sum_{i=1}^N x_i x_i^T$, then it is clear that $\Xi$ measures the correlation of all the vectors $x_i$, averaged over the entire trajectory. Thus the fact that the product $\tilde{X}^T\tilde{X}$ gives a diagonal matrix (eq. (40)) shows that *in the basis* $\{c_i\}$ *the vectors of the trajectory* ($x_i^T C$) *are uncorrelated* [13].

What is more, we can see from eq. (40) that $\sigma_i^2$ is proportional to the extent to which the trajectory varies in the $c_i$-direction. One can think of the trajectory as exploring on the average an ellipsoid in $\mathbf{R}^n$, the directions and lengths of the axes of which are given by the $c_i$ and the $\sigma_i^2$, respectively. This picture is very valuable for the discussion of the effect of noise on the trajectory: If there is, for example, a floor of white noise (i.e. we have $v_i = \overline{v}_i + \xi$, where $\overline{v}_i$ is the deterministic contribution to the observable and $\xi$ stands for the noise) then we get for the covariance matrix an expression like

$$\Xi = \overline{\Xi} + \left\langle \xi^2 \right\rangle \cdot I_n \quad , \tag{41}$$

where $\langle . \rangle$ denotes the time average, $I_n$ is the $(n, n)$-identity matrix and $\overline{\Xi}$ is the deterministic contribution we have discussed until now. Diagonalization of $\Xi$ for this "white noise case" yields the eigenvalues

$$\sigma_i^2 = \overline{\sigma}_i^2 + \left\langle \xi^2 \right\rangle, \quad i = 1, 2, \ldots, n \quad . \tag{42}$$

Notice that in the presence of noise all eigenvalues are non-zero (since $\langle \xi^2 \rangle > 0$), even for those directions $c_i$ which are not explored by the deterministic movement of the system. Thus the influence of noise on our analysis is to make us think that the (deterministic) trajectory explores *all* directions of $\mathbf{R}^n$, instead of only $n'$!

[13] This result justifies the choice $J = 1$ in section 3.1: since in a well-chosen basis ($\{c_i\}$) the vectors which form the trajectory are uncorrelated, the lag-time does not influence our results and we can choose it at our convenience.

The solution of this problem is found by considering the *singular value decomposition* of $X$:

$$X = S\Sigma C^T \quad . \tag{43}$$

($S$ is the $(N,n)$-matrix formed of the eigenvectors of $\Theta$.) We will see below that this decomposition is useful, because $\Sigma$ appears as one of the factors constituting $X$. We want to find those entries of $\Sigma$ which are obviously non-zero only due to the effects of noise. One possibility to do this is to measure, in addition to the true time series, a time series which consists only of the noise and then to compute the corresponding mean square projections onto the $c_i$-directions (which we found for the true time series). These quantities are to be compared with the respective $\sigma_i$ (of the true time series) and if both are found to be of the same order of magnitude then we know that this particular $\sigma_i$ is only non-zero due to the noise; in other words, the corresponding direction $c_i$ is noise-dominated. A straightforward strategy (which uses the special form of eq. (43)) to get rid of this most significant influence of noise is to set those selected entries equal to zero[14]. Thus we arrive at the following corrected equation for the approximate deterministic part $\overline{X}'$ of the trajectory matrix:

$$\overline{X}' = S\overline{\Sigma}'C^T \tag{44}$$

$$\overline{\Sigma}' = \mathrm{diag}\,(\overline{\sigma}'_1, \overline{\sigma}'_2, \ldots, \overline{\sigma}'_n) \tag{45}$$

$$\overline{\sigma}'_i = \begin{cases} \sigma_i & \text{if} \quad \sigma_i > \text{noise} \\ 0 & \text{if} \quad \sigma_i \approx \text{noise} \end{cases} \tag{46}$$

(Notice that the $\overline{\sigma}'_i$ are *not* the deterministic component $\overline{\sigma_i}$ of the $\sigma_i$; but, by construction, $\overline{\sigma_i}'$ and $\overline{\sigma_i}$ should be *approximately* the same.), which can be simplified to get a representation of the corrected trajectory matrix $\overline{X}'$ which is most easy to work with:

$$\overline{X}' = \sum_{\substack{\{i \mid i \in \{1,\ldots,n\}, \\ \sigma_i > \text{noise floor}\}}} (Xc_i)\, c_i^T \quad . \tag{47}$$

This equation is nice to work with since the $c_i$ and $\sigma_i$ are easy to compute by diagonalization of the covariance matrix.

[14] One should be aware of the fact that this strategy does not remove *all* effects of the noise on the trajectory: The noise contribution to those $\overline{\sigma}_i$ which are non-zero according to eq. (46) remains unchanged. It is not totally clear to me why Broomhead and King in [7] do not propose to subtract the noise floor from *all* eigenvalues $\sigma_i$ by replacing $\Sigma$ in eq. (43) with $\left(\Sigma - diag\left(\langle\xi_1^2\rangle, \ldots, \langle\xi_n^2\rangle\right)\right)$ (or with $\left(\Sigma - \langle\xi^2\rangle \cdot I_n\right)$ in the case of white noise); this would (on the average!) remove the noise floor from all eigenvalues. Maybe the reason for not doing this are:

- we would not get eq. (47) in this nice form which is easy to handle numerically;
- the effect of noise on the first $n'$ components is assumed to be not very large;
- for the calculation of $n'$ the numerical values of the $\sigma_i$ do not matter anyhow, apart from being above or in the noise floor.

The sum in eq. (47) will run over $d$ summands, and of course we expect $d \approx n'$. We relabel the $\{c_i\}_{i=1}^{n}$ such that the first $d$ of them correspond to those eigenvalues $\sigma_i$ which are not noise-dominated. Then we know, after having eliminated the effect of noise as far as possible, that the trajectory is confined to a $d$-dimensional subspace of $\mathbf{R}^n$ which is spanned by $\{c_i\}_{i=1}^{d}$. So we can take $\mathbf{R}^d$ as the embedding space instead of $\mathbf{R}^n$. Finally we get the following vectors on the trajectory in $\mathbf{R}^d$:

$$\left(x_i^T c_1, x_i^T c_2, \ldots, x_i^T c_d\right)^T, \quad i = 0, 1, 2, \ldots, N-1 \quad , \tag{48}$$

and it is these vectors which we can now plot in $d$ dimensions (taking e.g. two- or three-dimensional cross-sections) to get the geometric picture of the attractor which we have been aiming at.

### 3.4.3 Choice of First Embedding Dimension and Sampling Time

By "first embedding dimension" we mean here $n$, the dimension of the space with which we begin the procedure described in the last section. We did not specify $n$ yet, but simply said that it should be greater than or equal to $2m + 1$, in order to guarantee that embedding is possible.

Broomhead and King in [7] suggested to compute the power spectrum, which shows to which degree the frequencies contribute. Typically, the power spectrum consists of a noise floor (i.e. all frequencies contribute equally in the case of white noise) to which the amplitudes due to the deterministic contribution are added (see Fig. 7).

Figure 7. Example for a band-limited power spectrum which allows to determine the band-limit frequency $\omega^*$.

If the deterministic contribution is only significant for frequencies which are smaller than some "band-limit" frequency $\omega^*$ (which corresponds to the time-

interval $\tau^* = \frac{2\pi}{\omega^*}$) then one can choose $n\tau = \tau^*$. This can be justified as follows: On the one hand one wants to make $n\tau$ large, in order to have it $\geq (2m+1)\tau$ (see section 3.1), on the other hand choosing $n$ so large that $n\tau > \tau^*$ seems not to be sensible, since it is obviously easier to work with a lower-dimensional embedding space. So the only consistent a priori estimate seems to be $n\tau = \tau^*$.

Although this justification for the choice of $n$ is rather handwaving and only valid for band-limited data, numerical experiments in [7] show that in many cases it gives good results. One reason for this is that most dynamical systems which have been investigated until now usually have rather low-dimensional attractors due to the dissipative properties of the systems. This can be true even if the system is moving in a phase space as high-dimensional as in the case of the Belousov-Zhabotinski reaction (see chapter 1.1 in [14]).

According to Broomhead and King the choice of the sampling time can be based on physical considerations as well. Many systems have a characteristic time-scale: The observables of the system do not change significantly in times smaller than this. If one decreases $\tau$ while keeping the "window length" $n\tau$ constant then one gets vectors with more and more components and thus more and more singular values $\sigma_i$. Doing this one will reach a point where the number $d$ of those singular values which are not noise-dominated does not go up further: decreasing $\tau$ then results essentially in increasing the number of singular values in the noise floor. This means that the corresponding value of $\tau$ is small enough to match the characteristic time of the system and we can take $\tau$ as the sampling time.

A similar approach to the choice of sampling time is described by Schuster (chapter 5.3 in [14]): He considers some fixed $n$ and determines $\tau$ as the decay time of the autocorrelation function $C(t)$ of the time series which can be computed as follows:

$$C(t) = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N-1} v_n v_{n+1} \quad , \tag{49}$$

where $t$ is the time between two successive measurements of $v$. Then we get $\tau$ from

$$C(\tau) \approx \frac{1}{2} C(0) \quad . \tag{50}$$

Since the power spectrum is proportional to the Fourier transform of the autocorrelation function[15] both approaches ([7] and [14], chapter 5.3) should give comparable estimates for $\tau$.

[15] This is the Wiener-Khinchine theorem.

### 3.5 On the Required Number of Data Points

The question of how many reconstructed phase space vectors ($N$) we need to secure that the reconstructed picture of the attractor is "accurate" is seldom addressed in the literature. Often it is simply implicitely assumed that $N$ is large enough to make the error small. It is clear (when keeping Takens' results in mind) that for $N \to \infty$ our picture converges to an object which is (diffeomorphically) equivalent to the attractor. We will take a brief look on what is known about the requirements on $N$ to come as near to this ideal as possible.

The first thing to mention is that the needed $N$ depends on the "fractality" of the attractor, that is the more complex the attractor is (i.e. the larger the modulus of the difference between the attractor's dimension and the nearest integer is) the more points we need to reproduce it correctly. Secondly the attractor-picture is not only constructed as an end in itself but to derive characteristic properties from it. From this it follows that one should choose $N$ according to the quantity to be computed. For example, calculations of Lyapunov exponents usually require more data points than calculations of the correlation dimension. Thirdly, an inhomogeneous distribution of the $x_i$ on the reconstructed attractor obviously results in the need for more data points, since we want to resolve all parts of the attractor accurately enough.

Considering the above hints which seem to suggest that usually a large $N$ is required it is good news that one can conclude from systematic studies surveyed by Albano et al. [8] that often a surprisingly small number of data points is sufficient, especially when one computes numerically robust quantities such as the correlation dimension. Unfortunately this observation seems to be correct only if the required accuracy is small. If one wants to compute for instance dimensions up to several decimal places then it can become necessary to use millions of data points. This observation is illustrated by the numerical experiment with the Rössler system, mentioned above: Although the geometrical structures of the objects in Figs. 5.a/5.b and e.g. in Fig. 5.f/5.g are very similar, one cannot expect to derive very accurate numerical invariants from the latter ones, since they are quite coarse-grained. Much more effort must be spent on slightly more exact computations.

## 4 Conclusion

We have described and to some extent justified a method to retrieve geometrical information from a one-dimensional time series by reconstructing

a phase space picture of the attractor. Modifications and extensions of this method which take into account the practical problems of time series analysis have been discussed. It is important to notice that there are many different approaches to these problems and until now we do not have a general technique which can be applied to all time series. Rather than that one has to choose the technique which seems to be most appropriate for each individual data set.

Further improvements of the reconstructed attractor-picture are possible, e.g. by using the tesselation method [13]. Having gone through the embedding process one can use the well-known methods to quantify the system by calculating the usual dimensions, Lyapunov exponents, etc. An interesting application of the embedding technique is the prediction of a dynamical system, using the information retrieved from a time series [9, 12, 16].

# References

[1] V. Guillemin and A. Pollack, *Differential Topology* (Prentice Hall, New Jersey, 1974) 48–56.

[2] D.R.J. Chillingworth, *Differential Topology with a View to Applications* (Pitman Publishing, London, 1976) 146–152.

[3] N.H. Packard, J.P. Crutchfield, J.D. Farmer and R.S. Shaw, *Geometry from a Time Series*, Phys. Rev. Lett. **45** (1980) 712–716

[4] F. Takens, *Detecting Strange Attractors in Turbulence*, in *Proceedings of the Symposion on Dynamical Systems and Turbulence, University of Warwick, 1979–1980*, edited by D.A. Rand and L.S. Young (Springer, Berlin, 1981).

[5] P. Grassberger and I. Procaccia, *Measuring the Strangeness of Strange Attractors*, Physica **9D** (1983) 189–208.

[6] P. Grassberger and I. Procaccia, *Characterization of Strange Attractors*, Phys. Rev. Lett. **50** (1983) 346–349.

[7] D.S. Broomhead and G.P. King, *Extracting Qualitative Dynamics from Experimental Data*, Physica **20D** (1986) 217–236.

[8] A.M. Albano, A.I. Mees, G.C. de Guzman and P.E. Rapp, *Data Requirements for Reliable Estimation of Correlation Dimensions*, in *Chaos in Biological Systems*, edited by H. Degn, A.V. Holden and L.F. Olsen (Plenum Press, New York, 1987).

[9] J.D. Farmer and J.J. Sidorowich, *Predicting Chaotic Time Series*, Phys. Rev. Lett. **59** (1987) 845–848.

[10] F. Natterer, *Numerische Mathematik 1 und 2 (Vorlesungsskript)*, 1987, unpublished.

[11] J. Stoer, *Numerische Mathematik 1*, fünfte verb. Auflage (Springer, 1989).

[12] M. Casdagli, *Nonlinear Prediction of Chaotic Time Series*, Physica D **35** (1989) 335–356.

[13] A. Mees, *Modelling Dynamical Systems from Real-World Data*, in *Measures of Complexity and Chaos*, edited by N.B. Abraham, A.M. Albano, A. Passamante, P.E. Rapp (Plenum Press, New York, 1989).

[14] H.G. Schuster, *Deterministic Chaos. An Introduction*, 2nd rev. ed. (VCH Verlagsgesellschaft mbH, Weinheim, 1989).

[15] H.G. Schuster, *Extraction of Models from Complex Data*, in *Measures of Complexity and Chaos*, edited by N.B. Abraham, A.M. Albano, A. Passamante, P.E. Rapp (Plenum Press, New York, 1989).

[16] L.A. Smith, *Quantifying Chaos with Predictive Flows and Maps: Locating Unstable Periodic Orbits*, in *Measures of Complexity and Chaos*, edited by N.B. Abraham, A.M. Albano, A. Passamante, P.E. Rapp (Plenum Press, New York, 1989).

[17] J. Palis and F. Takens, *Homoclinic Bifurcations and Sensitive-Chaotic Dynamics*, Proceedings of the *School on Dynamical Systems*, Trieste (1991)

# Declaration

In accordance with the Ordinances of the University of Cambridge, pages 394–6 (1988 edition), I hereby declare that this essay *Time Series Analysis*, submitted as one part of the examination for Part III of the Mathematical Tripos, represents solely my own unaided work.

Cambridge, 1 May 1991