Chapter 3

Phase space methods

3.1 Determinism: uniqueness in phase space

The nonlinear time series methods discussed in this book are motivated and based on the theory of *dynamical systems*; that is, the time evolution is defined in some phase space. Since such nonlinear systems can exhibit deterministic chaos, this is a natural starting point when irregularity is present in a signal. Eventually, one might think of incorporating a stochastic component into the description as well. So far, however, we have to assume that this stochastic component is small and essentially does not change the nonlinear properties. Thus all the successful approaches we are aware of either assume the nonlinearity to be a small perturbation of an essentially linear stochastic process, or they regard the stochastic element as a small contamination of an essentially deterministic, nonlinear process. If a given data set is supposed to stem from a genuinely *non-linear stochastic* processes, time series analysis tools are still very limited and their discussion will be postponed to Section 12.1.

Consider for a moment a purely deterministic system. Once its present state is fixed, the states at all future times are determined as well. Thus it will be important to establish a vector space (called a *state space* or *phase space*) for the system such that specifying a point in this space specifies the state of the system, and vice versa. Then we can study the dynamics of the system by studying the dynamics of the corresponding phase space points. In theory, dynamical systems are usually defined by a set of first-order ordinary differential equations (see below) acting on a phase space. The mathematical theory of ordinary differential equations ensures the existence and uniqueness of the trajectories, if certain conditions are met. We will not hold up any academic distinction between the state and the phase space, but we remark that except for mathematical dynamical models with given equations of motion, there will not be a unique choice of what the phase space of a system can be.

The concept of the *state of a system* is powerful even for nondeterministic systems. A large class of systems can be described by a (possibly infinite) set of states and some kind of transition rules which specify how the system may proceed from one state to the other. Prominent members of this category are the stochastic *Markov processes* for which the transition rules are given in the form of a set of transition probabilities and the future state is selected randomly according to these probabilities. The essential feature of these processes is their strictly finite memory: the transition probabilities to the future states may only depend on the present state, not on the past.[1] If you like you can regard a purely deterministic system as a limiting case of a Markov process on a continuum of states. The transition to the state specified by the deterministic rule occurs with probability 1 and every other transition has probability 0. We mention this approach because it treats uncertainties in the transition rule – what we call *dynamical noise* – as the generic case of a peaked distribution of transition probabilities. The noise-free – deterministic – case is only obtained in the limit of a delta peak. This is not the most useful formulation for the purpose of this book. Instead, we formulate the particular case of a deterministic system in its own right as a starting point which allows us to establish some interesting ways of understanding signals and systems. Let us keep in mind the fact that we can treat dynamical noise not only as an additional complication in an otherwise clean situation, but we can also regard strict determinism as a limiting case of a very general class of models.

Let us now introduce some notation for deterministic dynamical systems in phase space. Modifications necessary for not exactly deterministic systems will be discussed later. For simplicity we will restrict ourselves to the case where the phase space is a finite dimensional vector space $\mathbb{R}^m$ (partial differential equations such as the Navier–Stokes equation for hydrodynamic flow form highly interesting dynamical systems as well, living in infinite dimensional phase spaces). A state is specified by a vector $\mathbf{x} \in \mathbb{R}^m$. Then we can describe the dynamics either by an m dimensional map or by an explicit system of m first-order ordinary differential equations. In the first case, the time is a discrete variable:

$$\mathbf{x}_{n+1} = \mathbf{F}(\mathbf{x}_n), \qquad n \in \mathbb{Z}, \tag{3.1}$$

and in the second case it is a continuous one:

$$\frac{d}{dt}\mathbf{x}(t) = \mathbf{f}(\mathbf{x}(t)), \qquad t \in \mathbb{R}. \tag{3.2}$$

The second situation is usually referred to as a *flow*. The vector field $\mathbf{f}$ in Eq. (3.2) is defined not to depend explicitly on time, and thus is called *autonomous*.

[1] In a Markov chain of order m, the present *state* is represented by the values of the process during the last m discrete time steps.

If $\mathbf{f}$ contains an explicit time dependence, e.g., through some external driving term, the mathematical literature does not consider this a dynamical system any more since time translation invariance is broken. The state vector alone (i.e., without the information about the actual time t) does not define the evolution uniquely. In many cases such as periodic driving forces, the system can be made autonomous by the introduction of additional degrees of freedom (e.g., a sinusoidal driving can be generated by an additional autonomous harmonic oscillator with a unidirectional coupling). Then, one can typically define an extended phase space in which the time evolution is again a unique function of the state vectors, even without introducing auxiliary degrees of freedom; just by introducing, e.g., a phase angle of the driving force. When admitting arbitrary time dependences of $\mathbf{f}$, however, this also includes the case of a noise driven stochastic system which is not a dynamical system in the sense of unique dependence of the future on some actual state vector.

In the autonomous case, the solution of the initial value problem of Eq. (3.2) is known to exist and to be unique if the vector field $\mathbf{f}$ is Lipshitz continuous. A sequence of points $\mathbf{x}_n$ or $\mathbf{x}(t)$, solving the above equations is called a *trajectory* of the dynamical system, with $\mathbf{x}_0$ or $\mathbf{x}(0)$, respectively, the *initial condition*. Typical trajectories will either run away to infinity as time proceeds or stay in a bounded area forever, which is the case we are interested in here.[2] The observed behaviour depends both on the form of $\mathbf{F}$ (or, respectively, $\mathbf{f}$) and on the initial condition; many systems allow for both types of solution. The set of initial conditions leading to the same asymptotic behaviour of the trajectory is called the *basin of attraction* for this particular motion.

On many occasions we will find the discrete time formulation more convenient. The formal solution of the differential equation Eq. (3.2) relating an initial condition to the end of the trajectory one unit of time later is sometimes called the *time one map* of $\mathbf{f}$. We will sometimes refer to it, since, after all, the time series we will have to deal with are only given at discrete time steps. Also the numerical integration of the differential equations, Eq. (3.2), with a finite time step Δt yields a map. For example, the Euler integration scheme yields

$$\mathbf{x}(t + \Delta t) \approx \mathbf{x}(t) + \Delta t \mathbf{f}(\mathbf{x}(t)). \tag{3.3}$$

When the time step Δt is small, the difference between consecutive values $\mathbf{x}(t)$, $\mathbf{x}(t + \Delta t)$ is small as well, which is characteristic for the particular kind of map which arises from differential equations. We will refer to such time series as

[2] In this book we do not discuss scattering problems, where trajectories approach an interaction region coming from infinity, and after being scattered they disappear towards infinity. However, there exists a class of very interesting phenomena called *chaotic scattering*. Deterministic but chaotic dynamics lead to irregular dependence of the scattering angle on the impact parameter.

"flow-like". The fundamental difference between flow-like and map-like data will be discussed below and in Section 3.5.

Example 3.1 The two linear differential equations

$$dx/dt = -\omega y, \qquad dy/dt = \omega x, \tag{3.4}$$

form a dynamical system with the periodic solution $x(t) = a\cos\omega(t - t_0)$, $y(t) = a\sin\omega(t - t_0)$. The solution will obviously stay finite forever. □

Example 3.2 (Hénon map). The map given by Hénon (1976):

$$x_{n+1} = a - x_n^2 + by_n, \qquad y_{n+1} = x_n, \tag{3.5}$$

yields irregular solutions for many choices of a and b. For $|b| \leq 1$ there exist initial conditions for which trajectories stay in a bounded region but, for example, when $a = 1.4$ and $b = 0.3$, a typical sequence of x_n will not be periodic but *chaotic*. We urge the reader to verify this claim by doing Exercise 3.1. □

The dynamical systems used as examples in this book do not only have bounded solutions but they are usually also *dissipative*, which means that on average a phase space volume containing initial conditions is contracted under the dynamics. Then we have on average $|\det \mathbf{J}_\mathbf{F}| < 1$ or $\operatorname{div}\mathbf{f} < 0$ respectively. ($\mathbf{J}_\mathbf{F}$ is the Jacobian matrix of derivatives of $\mathbf{F}$: $(\mathbf{J}_\mathbf{F})_{ij} = \partial/\partial x^{(j)} F_i(\mathbf{x})$.) For such systems, a set of initial conditions of positive measure will, after some transient time, be attracted to some sub-set of phase space. This set itself is invariant under the dynamical evolution and is called the *attractor* of the system. Simple examples of (non-chaotic) attractors are fixed points (after the transient time the system settles to a stationary state) and limit cycles (the system approaches a periodic motion). See Fig. 3.1. For an autonomous system with two degrees of freedom and continuous time, these (together with homoclinic and heteroclinic orbits connecting fixed points which are not discussed here) are the only possibilities (Poincaré–Bendixon theorem). More interesting attractors can occur for flows in at least three dimensions (see Fig. 3.2), where the mechanism of stretching and folding can produce *chaos*.

Example 3.3 The divergence of the flow introduced in Example 3.1 is zero, such that the dynamics are area preserving. The Hénon map, Example 3.2, is dissipative if the time independent determinant of the Jacobian, $|\det \mathbf{J}| = |b|$, is taken to be smaller than unity. The Lorenz system (see Exercise 3.2) is strongly dissipative, since its divergence, $-b - \sigma - 1$, for the usual choice of parameters is about -10. This means that on average a volume will shrink to e^{-10} of its original volume for each time unit. □

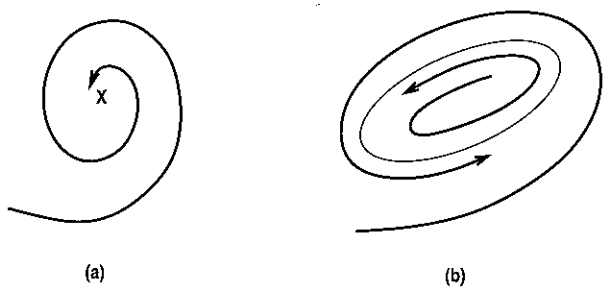


Figure 3.1 Fixed point (a) and limit cycle (b) attractors in the plane. Depending on the initial condition, a trajectory approaches the limit cycle (b) either from within or from without.

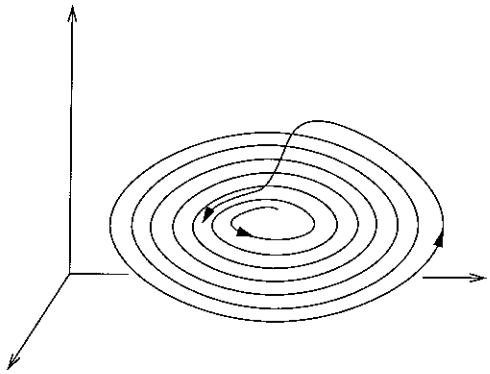


Figure 3.2 In three dimensional phase space a trajectory can be re-injected without violating determinism. Thus more complicated attractors are possible.

Characteristic for chaotic systems is that the corresponding attractors are complicated geometrical objects, typically exhibiting *fractal* structure. They are also called *strange attractors*. We will explain in Chapter 6, and in particular in Section 6.1, what we mean by a strange or fractal attractor.

So far we have illustrated that it is natural to describe a deterministic dynamical system as an object in phase space since this is the optimal way of studying its dynamical and geometrical properties. Since the dynamical equations (or the equations of motion) are defined in phase space, it is also most natural to use a phase space description for approximations to these equations. Such approximate dynamics will be important for predictions (Chapters 4 and 12), the determination of Lyapunov exponents (Chapters 5 and 11), noise filtering (Chapter 10) and most other applications (Chapter 15). For a deeper understanding of the nature of the underlying system, knowledge of the attractor geometry is desirable. How closely dynamics and geometry are related is expressed by theoretical results which relate Lyapunov exponents (dynamics) and dimensions (geometry). But practical

n the plane. Depending on
cle (b) either from within

an be re-injected without
s are possible.

ding attractors are compli-
ucture. They are also called
particular in Section 6.1,

e a deterministic dynam-
optimal way of studying
namical equations (or the
lso most natural to use a
ations. Such approximate
4 and 12), the determina-
filtering (Chapter 10) and
derstanding of the nature
ometry is desirable. How
y theoretical results which
(geometry). But practical

algorithms can also sometimes be formulated equivalently in terms of geometry or dynamics. (See the discussion about noise reduction methods, Section 10.3.)

3.2 Delay reconstruction

Having stressed the importance of phase space for the study of systems with deterministic properties, we have to face the first problem: what we observe in an experiment is not a phase space object but a time series, most likely only a sequence of scalar measurements. We therefore have to convert the observations into state vectors. This is the important problem of *phase space reconstruction* which is technically solved by the method of delays (or related constructions).

Most commonly, the time series is a sequence of scalar measurements of some quantity which depends on the current state of the system, taken at multiples of a fixed sampling time:

$$s_n = s(\mathbf{x}(n\Delta t)) + \eta_n. \tag{3.6}$$

That is, we look at the system through some measurement function s and make observations only up to some random fluctuations η_n, the *measurement noise*. Let us neglect the effect of noise at this level of presentation. (We will discuss its effect later in Section 10.2.)

A *delay reconstruction* in m dimensions is then formed by the vectors $\mathbf{s}_n$, given as

$$\mathbf{s}_n = (s_{n-(m-1)\tau}, s_{n-(m-2)\tau}, \ldots, s_{n-\tau}, s_n). \tag{3.7}$$

The time difference in number of samples τ (or in time units, $\tau\,\Delta t$) between adjacent components of the delay vectors is referred to as the *lag* or *delay time*. Note that for $\tau > 1$, only the time window covered by each vector is increased, while the number of vectors constructed from the scalar time series remains roughly the same. This is because we create a vector for every scalar observation, s_n, with $n > (m-1)\tau$. A number of embedding theorems are concerned with the question under which circumstances and to what extent the geometrical object formed by the vectors $\mathbf{s}_n$ is equivalent[3] to the original trajectory $\mathbf{x}_n$. In fact, under quite general circumstances the attractor formed by $\mathbf{s}_n$ is equivalent to the attractor in the unknown space in which the original system is living if the dimension m of the delay coordinate space is sufficiently large. To be precise, this is guaranteed if m is larger than twice the *box counting dimension* D_F of the attractor, i.e. roughly speaking, larger than twice the number of *active* degrees of freedom, regardless of how high the dimensionality of the true state space is. Depending on the application, even smaller

[3] In the sense that they can be mapped onto each other by a uniquely invertible smooth map. See Section 9.1.

m values satisfying $m > D_F$ can be sufficient. For theoretical details see Ding *et al.* (1993). One can hope to reconstruct the motion on attractors in systems such as hydrodynamic flows or in lasers, where the number of microscopic particles is huge, if only a few dominant degrees of freedom eventually remain as a result of some collective behaviour. In Chapter 9, we shall discuss the background of embedding and the theorems in more detail.

3.3 Finding a good embedding

When we start to analyse a scalar time series, we neither know the *box counting dimension*,[4] which is formally necessary to compute m, nor do we have any idea of how to choose the time lag τ. How we proceed depends on the underlying dynamics in the data, and on the kind of analysis intended. Most importantly, the embedding theorems *guarantee* that for ideal noise-free data, there exists a dimension m such that the vectors $\mathbf{s}_n$ are equivalent to phase space vectors. We will use this knowledge for the design of methods for the determination of the *dimension* of the attractor (Chapter 6), the *maximal Lyapunov exponent* (Chapter 5), and the *entropy* (Chapter 11). Generally, there are two different approaches for optimising the embedding parameters m and τ: either, one exploits specific statistical tools for their determination and uses the optimised values for further analysis, or one starts with the intended analysis right away and optimises the results with respect to m and τ. For example, dimension and Lyapunov estimates will be carried out by increasing the values of m until the typical behaviour of deterministic data appears.

For many practical purposes, the most important embedding parameter is the product $m\tau$ of the delay time and the embedding dimension, rather than the embedding dimension m or the delay time τ alone. The reason is that $m\tau$ is the time span represented by an embedding vector [see Kugiumtzis (1996) for a discussion]. For clarity, let us discuss the choice of m and τ separately. A precise knowledge of m is desirable since we want to exploit determinism with minimal computational effort. Of course, if an m dimensional embedding yields a faithful representation of the state space, every m' dimensional reconstruction with $m' > m$ does so as well. Choosing too large a value of m for chaotic data will add redundancy and thus degrade the performance of many algorithms, such as predictions and Lyapunov exponents. Due to the instability of chaotic motion, the first and last elements of a delay vector are the less related the larger their time difference. Therefore, taking a large value for m would not help much and it would risk "confusing" the algorithm.

[4] The box counting dimension is roughly the number of coordinates one needs to span the invariant subset on which the dynamics lives. We will be more specific in Chapter 6.

3.3.1 False neighbours

If we assume that the dynamics in phase space is represented by a smooth vector field, then neighbouring states should be subject to almost the same time evolution. Hence, after a short time interval into the future, the two trajectories emerging from them should be still close neighbours, even if chaos can introduce an exponential divergence of the two. This reasoning will be used very extensively in the next chapter on prediction. Here we want to refer to this property in order to discuss statistics which sometimes helps to identify whether a certain embedding dimension is sufficient for a reconstruction of a phase space.

The concept, called *false nearest neighbours*, was introduced by Kennel, Brown & Abarbanel (1992). We present it here with some minor modifications which avoid certain spurious results for noise [Hegger & Kantz (1999)].

The basic idea is to search for points in the data set which are neighbours in embedding space, but which should not be neighbours since their future temporal evolution is too different. Imagine that the correct embedding dimension for some data set is m_0. Now study the same data in a lower dimensional embedding $m < m_0$. The transition from m_0 to m is a projection, eliminating certain axes from the coordinate system. Hence, points whose coordinates which are eliminated by the projection differ strongly, can become "false neighbours" in the m dimensional space. The statistics to study is now obvious: for each point of the time series, take its closest neighbour in m dimensions, and compute the ratio of the distances between these two points in $m+1$ dimensions and in m dimensions. If this ratio is larger than a threshold r, the neighbour was false. This threshold has to be large enough to allow for exponential divergence due to deterministic chaos.

Now, when we denote the standard deviation of the data by σ and use the maximum norm, the statistics to compute is

$$X_{\mathrm{fnn}}(r) = \frac{\sum_{n=1}^{N-m-1} \Theta\left(\frac{|\mathbf{s}_n^{(m+1)} - \mathbf{s}_{k(n)}^{(m+1)}|}{|\mathbf{s}_n^{(m)} - \mathbf{s}_{k(n)}^{(m)}|} - r\right) \Theta\left(\frac{\sigma}{r} - |\mathbf{s}_n^{(m)} - \mathbf{s}_{k(n)}^{(m)}|\right)}{\sum_{n=1}^{N-m-1} \Theta\left(\frac{\sigma}{r} - |\mathbf{s}_n^{(m)} - \mathbf{s}_{k(n)}^{(m)}|\right)} \tag{3.8}$$

where $\mathbf{s}_{k(n)}^{(m)}$ is the closest neighbour to $\mathbf{s}_n$ in m dimensions, i.e., $k(n)$ is the index of the time series element k different from n for which $|\mathbf{s}_n - \mathbf{s}_k| = \min$. The first step function in the numerator is unity, if the closest neighbour is false, i.e., if the distance increases by a factor of more than r when increasing the embedding dimension by unity, whereas the second step function suppresses all those pairs, whose initial distance was already larger than σ/r. Pairs whose initial distance is larger than σ/r by definition cannot be false neighbours, since, on average, there is not enough space to depart farther than σ. Hence, these are invalid candidates for

the method and should not be counted, which is also reflected in the normalisation. There can be some false nearest neighbours even when working in the correct embedding dimension. Paradoxically, due to measurement noise there can be more false neighbours if more data is given. With more data, the closest neighbour is typically closer (there are more opportunities for a good recurrence on the attractor), whereas the chance for the distance of the $m+1$st components to be at least of the order of the noise level remains the same. For data with a very coarse discretisation (say, 8 bit), there can even be *identical* delay vectors in m dimensions, which are not identical in $m+1$ dimensions, so that the ratio of distances diverges. These appear to be false neighbours for any choice of r. Nonetheless, it is reasonable to study the false nearest neighbour ratio X_{fnn} as a function of r. The results of the false nearest neighbours analysis may depend on the time lag τ. Hence, if one wants to use this statistics for the distinction of chaos and noise, it is indispensable to verify the results with a surrogate data test (see Chapter 7).

Example 3.4 (False nearest neighbours of resonance circuit data). The electric resonance circuit described in Appendix B.11 creates rather noise free data on an attractor with a dimension D_F slightly above two. The false nearest neighbour statistics shown in Fig. 3.3 suggests that a 5 dimensional embedding is clearly enough (and this is also granted by the inequality $m > 2D_F$ from the embedding theorem). It is also evident that the few false neighbours found in 4 dimensions (for $N = 10000$ these are about 3 in absolute value) are hard to interprete. In Hegger *et al.* (1998) it was possible to do rather precise predictions and modelling in a four dimensional embedding, so that it seems that the false neighbours are noise artefacts. □

3.3.2 The time lag

A good estimate of the lag time $\tau = \tau \Delta t$ is even more difficult to obtain. The lag is not the subject of the embedding theorems, since they consider data with infinite precision. Embeddings with the same m but different τ are equivalent in the mathematical sense for noise-free data, but in reality a good choice of τ facilitates the analysis. If τ is small compared to the internal time scales of the system, successive elements of the delay vectors are strongly correlated. All vectors $\mathbf{s}_n$ are then clustered around the diagonal in the $\mathbb{R}^m$, unless m is very large. If τ is very large, successive elements are already almost independent, and the points fill a large cloud in the $\mathbb{R}^m$, where the deterministic structures are confined to the very small scales. The first zero of the autocorrelation function Eq. (2.5) of the signal often yields a good trade-off between these extrema. A more refined concept is called

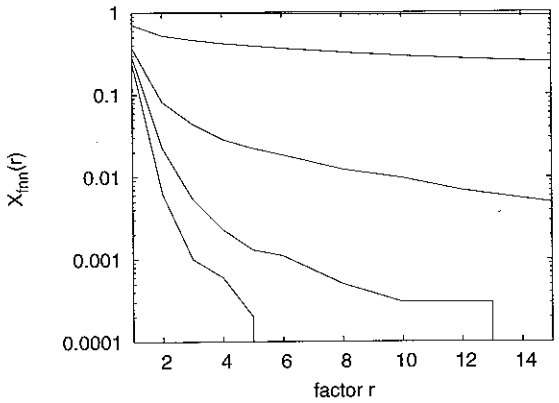


Figure 3.3 The relative number of false nearest neighbours as defined by Eq.(3.8) for electric resonance circuit data. The embedding dimension is $m = 1$ to 5 from top to bottom, with time lag 4.

mutual information and will be presented in Section 9.2. At this point we just give two recipes. The first is to study Example 3.5 in order to get a feeling for how a reasonable choice of τ can be verified visually. The second is that for a signal with a strong (almost-)periodic component, a time lag identical to one quarter of the period is a good first guess. All the nonlinear statistics in the next three chapters which rely on scaling behaviour suffer from reduced scaling ranges when τ has been chosen inappropriately, i.e., if τ is either unreasonably large or small.

Example 3.5 (Human ECG). Although the human electrocardiogram (ECG, see Appendix B.7) is probably not a deterministic signal, it can be interesting to view such signals in delay coordinates. Let us compare such representations for different delay times $\tau = \tau \Delta t$, where Δt is the sampling time, illustrated in Fig. 3.4. □

3.4 Visual inspection of data

Although this book contains a lot of refined methods for the characterisation and analysis of data, the first thing that we should do with a new data set is to look at it in several different ways. A plot of the signal as a function of time already gives the first hints of possible stationarity problems such as drifts, systematically varying amplitudes or time scales, and rare events. It allows us to select parts of the series which look more stationary. Quite often experimental data contain some faults which can be detected by visual inspection.

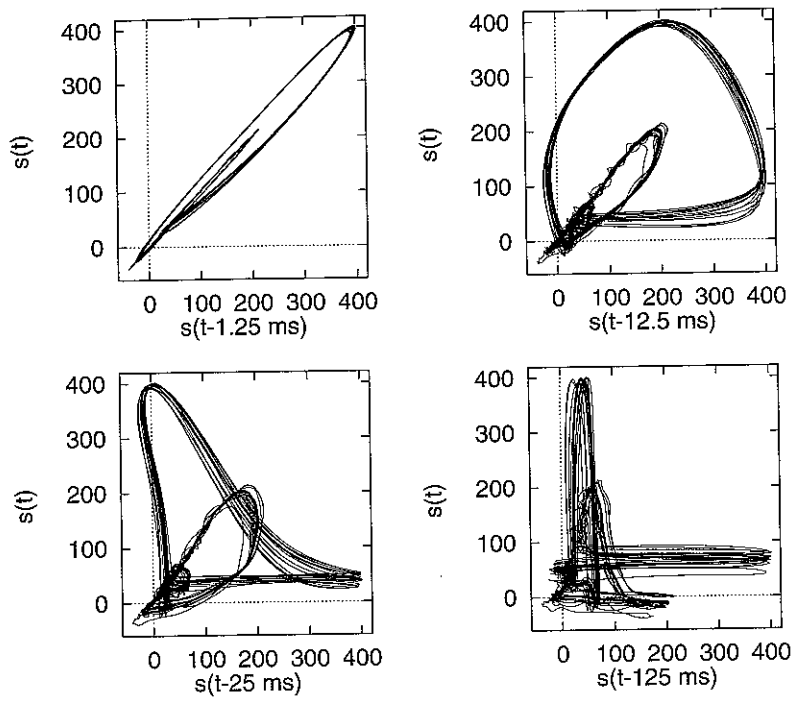
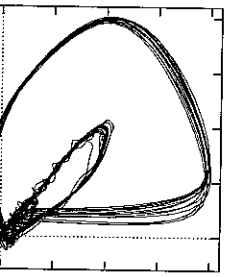


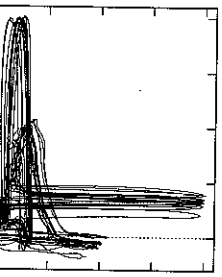
Figure 3.4 Delay representations of a human ECG signal, taken by Petr Saparin (1995). Upper left: $\tau = 1.25$ ms. All the data are close to the diagonal since consecutive values are very similar. Right: $\tau = 12.5$ ms. At this delay the large loop corresponding to the QRS complex (see Fig. B.4 for explanation) is now well unfolded. Lower left: $\tau = 25$ ms. The slower features, P- and T-waves, are better represented, although somewhat hidden by the QRS complex. Right: $\tau = 125$ ms. Larger delay times lead to unnecessarily complicated graphs.

The next step would be a two dimensional representation such as that shown in Fig. 1.5. Even if no clear structures are visible (as in the right hand panel), this gives a feeling about which time lag may be reasonable for an embedding of the data set.

Visual inspection can also reveal symmetries in the data or can guide us to a more useful representation of the data. An example of the latter is a measurement of the output power of a laser. It contains a nonlinear distortion of the physically more relevant variables, the electric and magnetic field strength inside the laser cavity: the power is proportional to the *square* of these quantities. Therefore, the data exhibit sharp maxima and smooth minima, and the signal changes much faster around the maxima than around the minima. Taking the square root of the data renders them more convenient for data analysis.

Sometimes one finds exact symmetries in the data, e.g., under change of sign of the observable. In this case, one can enlarge the data base for a purely geometric analysis by just replicating every data point by the symmetry operation. For an





al, taken by Petr Saparin
se to the diagonal since
. At this delay the large
explanation) is now well
- and T-waves, are better
plex. Right: $\tau = 125$ ms.
phs.

tion such as that shown in
ght hand panel), this gives
embedding of the data set.
data or can guide us to a
ne latter is a measurement
istortion of the physically
l strength inside the laser
quantities. Therefore, the
ignal changes much faster
e square root of the data

., under change of sign of
se for a purely geometric
nmetry operation. For an

analysis of the dynamics one can apply the symmetry operation to the time series as a whole. Since almost all nonlinear methods exploit local neighbourhood relationships, both tricks double the data base. When modelling the dynamics of the system the class of functions chosen should respect the symmetry.

3.5 Poincaré surface of section

If we consider the phase space of a system of m autonomous differential equations, we find that, locally, the direction tangential to the flow does not carry much interesting information. The position of the phase space point along this direction can be changed by re-parametrising time. It has no relationship to the *geometry* of the attractor and does not provide any further information about the dynamics. We can use this observation to reduce the phase space dimensionality by one, at the same time turning the continuous time flow into a discrete time map.

The method, called the *Poincaré section*, is the following. First form a suitable oriented surface in phase space (hatched area in Fig. 3.5). We can construct an invertible map on this surface by following a trajectory of the flow. The iterates of the map are given by the points where the trajectory intersects the surface in a specified direction (from above in Fig. 3.5). Note that the discrete "time" n of this map is the intersection count and is usually not simply proportional to the original time t of the flow. The time a trajectory spends between two successive intersection points will vary, depending both on the actual path in the reconstructed state space and on the surface of section chosen. Sometimes, after applying this technique, a disappointingly small number of points remain. Experimentalists tend to adjust the sampling rate such that about 10–100 observations are made per typical cycle of the signal. Each cycle yields at most one point in the Poincaré map (if there are

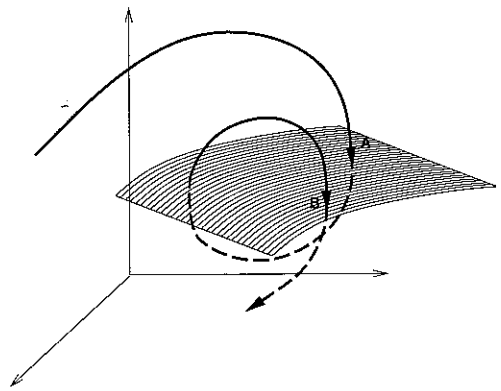


Figure 3.5 Poincaré section of a flow in three dimensions. The successive intersections A, B, ... of the trajectory with the surface of sections define iterates of an invertible two dimensional map.

considerably less intersections, the surface is inappropriate). However, if the data base in the Poincaré map appears to be poor, keep in mind the fact that the flow data themselves are not much better since the additional information they contain is largely redundant with respect to the chaotic properties of the signal.

Apart from the construction of intersections we can also collect all minima or all maxima of a scalar time series. As we shall argue in Section 9.3, the (numerical) time derivative of the signal is a "legal" coordinate in a reconstructed state space. Hence, in the hypothetical reconstructed space spanned by vectors $(s(t), \dot{s}(t), \ddot{s}(t), \ldots)$, intersections of the trajectory with the surface given by $\dot{s}(t) = 0$ are precisely given by the minima (or maxima) of the time series. The minima (or maxima) are interpreted as the special measurement function which projects onto the first component of a vector applied to the state vectors inside this surface. Hence, they have to be embedded with a time delay embedding with lag unity themselves in order to form the invariant set of this particular Poincaré map. The TISEAN package contains a utility to form Poincaré sections (called `poincare`) and one that finds extremal points, called `extrema`. Finding good sections often requires some experimentation with parameter settings.

The simplest class of non-autonomous systems that we can handle properly using methods from nonlinear dynamical systems consists of those which are driven periodically. Their phase space is the extended space containing the phase of the driving force as an additional variable. The most natural surfaces of section are those of constant phase. The resulting Poincaré map is also called a *stroboscopic map*. In this case the system always has the same time span between intersections, which is a very useful feature for quantitative analysis.

Let us note in passing that whenever we study periodic solutions of a time continuous system (limit cycles), the discrete period of a periodic orbit of the corresponding Poincaré map depends on the details of the surface of section chosen, since by moving the surface one can reduce the number of intersection points of the limit cycle with the surface. The choice of the proper surface of section will then be a crucial step in the analysis of the data.

Example 3.6 (Phase portrait of NMR laser data). These data comprise the chaotic output of a periodically driven system (Appendix B.2). The sampling interval is $\frac{1}{15}$th of the period of the driving force, $\Delta t = T_{\text{force}}/15$. Let us consider three out of the many different types of Poincaré maps. The stroboscopic view is the most natural in this case and has been used to produce the data shown in the introduction, Fig. 1.3. Apart from a complication to be explained in Example 9.4, $s_n = s(t = nT_{\text{force}} + t_0)$. Figure. 3.6 shows a section through a three dimensional delay embedding space using linear interpolation (left) across a diagonal plane,

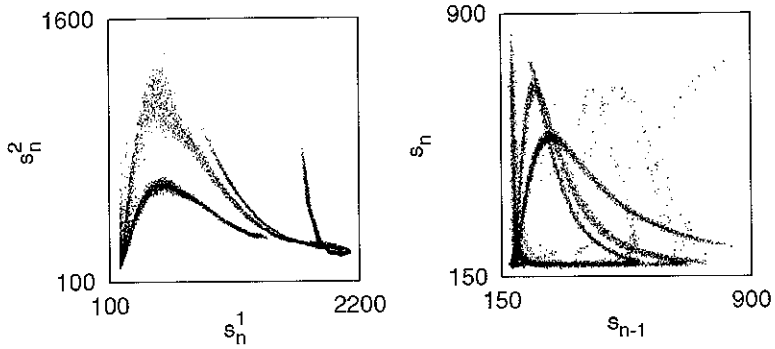


Figure 3.6 Phase portraits of the NMR laser data (Example 3.6). The left hand panel was obtained by a section through embedding space, see text. Right hand panel: successive minima of the parabolically interpolated series.

and the collection of all minima of the continuous time series after local parabolic interpolation (right). For the former, the section was taken such that $s_n^{(1)} = s(t_=)$, where $t_=$ is defined by $s(t_=) = s(t_= + \tau)$. The delay was taken to be $\tau = 4$. The second coordinate is $s_n^{(2)} = s(t_= - 2\tau)$. Thus the new time series is a series of two dimensional vectors. Alternatively, one could have used $s_n^{(1)}$ as a scalar time series and plot it as a two dimensional delay representation. The surface we selected is not optimal since in the upper left hand part of the attractor the noise is blown up considerably. In this part the surface of section cuts the attractor almost tangentially, such that the precise positions of the intersection points are strongly affected by noise and by the interpolation. The average time distance between intersection points is about 14. The deviation from the expected 15 is due to false crossings caused by the noise. The right hand panel uses a section at $\dot{s} = 0$, $\ddot{s} \geq 0$, i.e. every minimum of the parabolically interpolated series. The plot shows a time delay embedding of the time series given by the minima of these local parabolae. The average time distance between successive elements of the new series is about 15 and thus this section is as suitable as the stroboscopic view. Noise on the data causes some spurious extrema (the dots scattered outside the main structures).

The attractor looks different in all three cases since its geometrical details depend on the precise properties of the Poincaré section. Nevertheless, all three attractors are equivalent in the sense that they are characterised by identical values of the dimensions, Lyapunov exponents, and entropies. Numerical estimates of these quantities from finite and noisy data may differ however. □

3.6 Recurrence plots

The attempt of time series analysis to extract meaningful information from data to a good deal relies on redundancies inside the data. At least if data are aperiodic and no simple rule for their time dependence can be discerned, then approximate repetitions of a certain events can help us to construct more complicated rules. Assuming determinism and believing that a chosen delay embedding space forms a state space of the system, an approximate repetition is called a *recurrence*, i.e., the return of the trajectory in state space to a neighbourhood of a point where it has been before. Such recurrences exist in the original space for all types of motion which are not transient. A system on a fixed point is trivially recurrent for all times. In a system on a limit cycle, each point returns exactly to itself after one revolution. A system on a chaotic attractor returns to an arbitrarily small neighbourhood of any of its points. This is guaranteed by the invariance of the set which forms the support of the attractor. If, however, the system never returns to all points which we find in the initial part of the time series, then this indicates that this was a transient – the initial condition was outside the invariant set and the trajectory relaxes towards this set. Also non-stationarities through time dependent system parameters can cause a lack of recurrence.

A very simple method for visualising recurrences is called a *recurrence plot* and has been introduced by Eckmann *et al.* (1987): Compute the matrix

$$M_{ij} = \Theta(\epsilon - |\mathbf{s}_i - \mathbf{s}_j|) \,, \tag{3.9}$$

where $\Theta(.)$ is the Heaviside step function, ϵ is a tolerance parameter to be chosen, and $\mathbf{s}_i$ are delay vectors of some embedding dimension. This matrix is symmetric by construction. If the trajectory in the reconstructed space returns at time i into the ϵ-neighbourhood of where it was at time j then $M_{ij} = 1$, otherwise $M_{ij} = 0$. One can plot M_{ij} by black and white dots in the plane of indices for visual inspection. This is the recurrence plot. Numerical schemes for its quantitative characterisation have also been proposed. These are similar to many other statistical tools based on neighbourhoods in embedding space and hence will not be discussed here. See Casdagli (1997) for details.

So what can we learn from a visual inspection of the matrix M_{ij} by a recurrence plot? It gives hints on the time series and on the embedding space in which we are working. In a deterministic system, two points which are close should have images under the dynamics which are also close (even if they are not as close due to dynamical instability). Hence, one expects that black dots typically appear as short line segments parallel to the diagonal. If in addition there are many isolated dots, these rather indicate coincidental closeness and hence a strong noise component on the data, or an insufficient embedding dimension. If there are mostly scattered

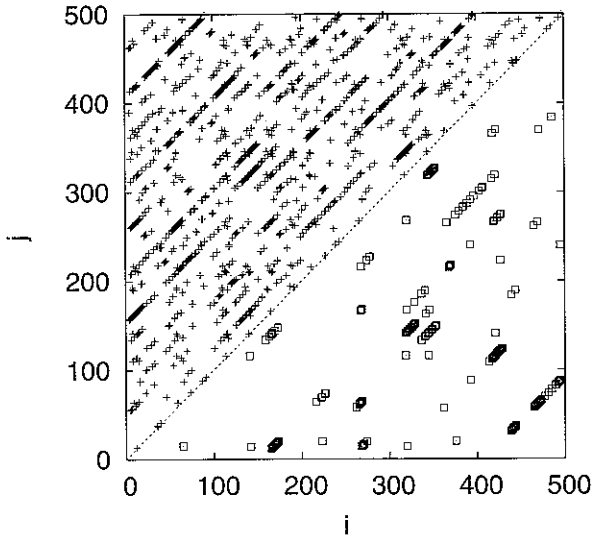


Figure 3.7 A recurrence plot of a short segment of the data from the electric resonance circuit (Appendix B.11). The upper left triangle represents embedding dimension two, the lower right is for dimension three. In two dimensions, this slightly more than two dimensional attractor intersects itself, hence there are lots of false neighbours. Many of them create isolated points. Due to the periodic forcing of this system, recurrences in the full phase space can only occur at temporal distances $|i - j|$ which are integer multiples of the driving period of about 25 samples.

dots, the deterministic component is absent or at least weak. Black dots in high dimensions are also black in lower dimensional embeddings. Hence, the relative number of lines increases when increasing the embedding dimension, since isolated dots disappear.

Example 3.7 (Recurrence plot of data from an electric resonance circuit). In Fig. 3.7 we show a recurrence plot of the NMR laser data in embedding dimensions 2 and 4. □

The irregularity of the arrangement of the line segments indicates chaos. If a signal was periodic, theses should form a periodic pattern. Stationarity of the whole time series requires that the density of line segments is uniform in the $i - j$-plane. See Example 13.8 for a non-stationary example. The temporal spacing between parallel lines can also give hints to the existence of unstable periodic orbits inside the chaotic attractor.

The essential drawback of this nice visual tool lies in the fact that for a time series of length N the matrix M_{ij} has N^2 elements, so that a huge number of pixels must be drawn. A compressed version of the recurrence plot was baptised

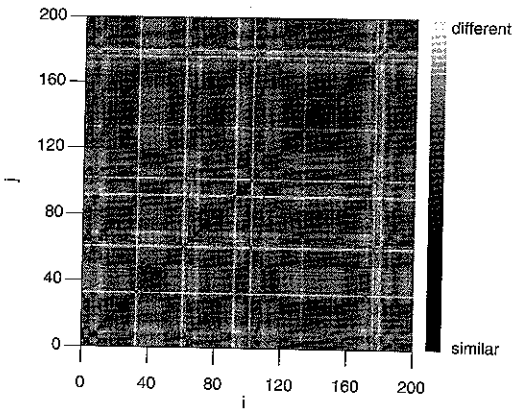


Figure 3.8 A meta-recurrence plot of human ECG data.

meta-recurrence plot in Manuca & Savit (1996). They subdivide the time series in segments and define the matrix elements M_{ij} to represent distances between the time series segments i and j. One possible distance measure is the *cross prediction error*. See the discussion in Schreiber (1999).

Example 3.8 (Meta-recurrence plot of ECG data). For ECG data, one can align segments along the main heart beat and then take a Euclidean distance. We define a segment by taking the last 100 ms before and the 500 ms after the upward zero crossing in the QRS-complex (see Appendix B.7). The resulting distance matrix is plotted in grey scale in Fig. 3.8. The non-stationarity is clearly visible. More importantly, almost stationary episodes can be recognised as light blocks on the diagonal. □

Further reading

Original sources about phase space embeddings are Takens (1981), Casdagli *et al.* (1991), and Sauer *et al.* (1991). Discussions of the proper choice of embedding parameters are contained for example in Fraser & Swinney (1986), Liebert & Schuster (1989), Liebert *et al.* (1991), Buzug & Pfister (1992), Kennel *et al.* (1992), and in Kugiumtzis (1996). See also Grassberger *et al.* (1991). *Recurrence plots* have been introduced by Eckmann *et al.* (1987). There is a nice more recent account by Casdagli (1997).

Exercises

3.1 Create different two dimensional phase portraits of a time series of the Hénon map (Example 3.2) using delay times $\tau = 1, 2, \ldots$ (here the sampling time is $\Delta t = 1$).

a.

subdivide the time series in
esent distances between the
easure is the *cross prediction*

For ECG data, one can align
clidean distance. We define
00 ms after the upward zero
ne resulting distance matrix
ity is clearly visible. More
ised as light blocks on the

kens (1981), Casdagli *et al.*
roper choice of embedding
winney (1986), Liebert &
1992), Kennel *et al.* (1992),
991). *Recurrence plots* have
ice more recent account by

me series of the Hénon map
e sampling time is $\Delta t = 1$).

Which picture gives the clearest information? Rewrite the map in delay coordinates with unit delay.

3.2 Create a scalar series of flow data numerically by integration of the Lorenz system [Lorenz (1963)]:

$$\dot{x} = \sigma(y - x) ,$$
$$\dot{y} = rx - y - xz ,$$
$$\dot{z} = -bz + xy ,$$

with the parameters $\sigma = 10$, $r = 28$, and $b = 8/3$. This model was designed to describe convective motion of the Rayleigh–Bénard type, where x is the velocity of the fluid, y is the temperature difference between ascending and descending fluid, and z is the deviation of the temperature profile form linearity. For a laser system governed by the same equations and experimental data, see Appendix B.1. As an integrator, use, for example, a Runge–Kutta routine with a small step size. Sample the data at such a rate that you record on average about 25 scalar measurements of the x-coordinate during a single turn of the trajectory on one leaf of the attractor. Plot the attractor in two dimensional delay coordinates with different time lags. Convince yourself by visual inspection that the reconstruction is best when the lag is about one-quarter of the mean cycle time. Compute the autocorrelation function and the time delayed mutual information (to be defined in Section 9.2) to confirm this impression.

3.3 Numerically integrate the Lorenz system and perform a Poincaré section. Record (y, z) every time the x-coordinate equals zero and its derivative is negative.

3.4 Use the time series of the x-variable of Exercise 3.2 and collect all the maxima. Plot the series of the maxima with delay one and compare it to the attractor of Exercise 3.3.