The Dynamical Systems Approach to Nonlinear Signal Processing

J.P. Huke

Department of Mathematics
UMIST
PO Box 88
Manchester M60 1QD

Abstract

Dynamical systems theory is the mathematical theory of systems that change with time. Here some of the basic ideas from this theory are introduced and their relationship to the description of physical systems is described. We shall concentrate on nonlinear systems; these can show novel kinds of behaviour not found in linear systems, and require a different set of concepts to analyse them. The description of such systems in terms of (nonlinear) state spaces is discussed, and how this leads to a geometrical theory in which the dynamics are represented by trajectories and the sets that they trace out; (these sets often have a very complex, fractal, structure).

Central to the analysis of signals from such systems is the idea of 'embedding'. An embedding is a kind of model of the system made using a time series of measurements. We can capture many of the important features of the system using such models, even though the measurements made at any particular time are not enough to specify the state completely; (for example, an embedding can be used to make a predictor for the system). Embeddings can also be used to process the signals themselves, for example to perform noise reduction, signal separation, or control.

Introduction: signals from nonlinear systems

Both dynamical systems theory and the theory of signal processing by now have substantial histories, and there has long been a connection, in that signal processing has drawn extensively on the theory of linear systems. However, in the past twenty years or so the two fields have come together in a new way, which centres around nonlinear systems. This connection has arisen because of recent advances in the understanding of the mathematics of nonlinear systems, and the desire to extend this understanding to physical and experimental systems. In particular, the realisation that the irregular behaviour known as 'chaos' is common in (mathematical) nonlinear systems immediately raises the question of whether such behaviour can be observed in experimental systems. To answer this necessarily involves some kind of data analysis: our observations provide us with a quantity of data which we must work with in trying to identify chaotic behaviour. We can compare this with the situation where we are instead looking for periodic behaviour; in that case there is a widely studied and used body of techniques (from periodogram analysis onwards) on which we can call. But those techniques will not help us very much when it comes to chaotic behaviour; and chaotic behaviour is only one aspect of nonlinear dynamics. Hence we are led to the following questions: given a nonlinear experimental system, what can we hope to learn about the system from observations made on it? And how should we process the experimental data to extract the information?

Once physicists and engineers began to try to answer these questions they began to put together a set of techniques for processing signals from nonlinear systems. The fact that they were working with experimental data led them to address issues such as how to deal with noise, what the effects of filtering are, and so on. These issues are common to all signal processing, but it turns out that their implications for nonlinear signal processing can be quite different to the linear case. The techniques for dealing with such signals are much less well studied than linear processing techniques, and they are still evolving rapidly, with some initially popular approaches falling into disuse, and new approaches emerging all the time. The lack of a textbook or monograph literature is indicative of this; despite a huge growth in the number of texts dealing with nonlinear dynamics, the number dealing with the signal processing aspects remains very small [1].

The work on analysing signals from nonlinear experimental systems also leads to other questions. What about signals of unknown origin? Can the techniques be applied to tell us anything about these? Could we use the techniques not only to learn about systems, but to control them? Do the signals have special properties that might make them useful in applications such as communications? All of these have spawned a considerable amount of research activity, and there are many others as well. At the moment, things are changing very fast, and it is too early to say how the new fields will take their place beside the established ones. Below we shall outline the basic ideas of dynamical systems theory, and draw the connections between them and the corresponding signal processing problems.

Dynamical systems: basic ideas

Dynamical systems theory is the mathematical theory of systems that change with time. Historically, the moving particles and bodies of classical mechanics led to the abstract idea of a dynamical system, but the modern theory is not just a mathematical language for Newtonian mechanics: it can be used to describe many kinds of time evolving systems, including biological, ecological and economic ones, as well as those from the more conventional areas of physics and engineering. A dynamical system has two important ingredients: there is a set of points, or *states*, and a specification of how states change over time. The states will often be vectors in $\mathbb{R}^n$; for a system of Newtonian particles, the state consists of the positions and momenta of the particles, so is a $6N$ dimensional vector if there are N particles; for an electrical network, the state consists of currents and/or voltages at various points in the network. The description of how the states change with time takes one of two forms: it can be a set of differential equations if time is continuous; or it can be a set of difference equations if time moves in unit steps. Our system of Newtonian particles is described by differential equations (namely Newton's laws); an example of a system in which time is discrete is a digital electronic system controlled by a clock. In fact, continuous time systems can be converted to discrete time systems by sampling at a uniform rate: here the system is observed not at all times but only at integer times. Dynamical systems theory concerns both continuous and discrete time systems; mostly, the real interest lies with continuous time systems because these are usually the kind that describe physical phenomena, but discrete time systems can be conceptually easier to handle, and so it is often useful first to convert a continuous time system to a discrete time one (say by sampling), and then to analyse this.

Mathematically then, a (discrete time) dynamical system consists of two things: a set M, and a transformation $T: M \to M$ mapping M to itself. This is all there is to the basic structure. M is called the *state space* and its points are the *states.* The important point is that it makes sense to iterate the mapping T, that is, the composition

$$T^n \equiv \underbrace{T \circ T \circ \ldots \circ T}_{n \text{ times}}$$

is defined. T describes how states change in one time step: $x \in M$ changes to Tx. By repeated applications of T we find that after n time steps the state has changed to $T^n x$; so for each state x we can construct a set of points

$$\{T^n x : n = 0, 1, \ldots\}$$

which is called the *trajectory* (or *orbit*) of x. Note that if we say x_n is the n-th point on the trajectory (starting at $n = 0$, so $x_n = T^n x$), we have $x_{n+1} = Tx_n$ for all n, illustrating how T maps each point one time step forward.

Not surprisingly, we shall not be able to say much of interest if this is all the structure we have; but usually of course we will know more about M than that it is just a set. The most common case is where M is $\mathbb{R}^n$ or a subset of it called a *differential manifold.* We need not go into the details of defining manifolds here: we can think of them as the analogues in $\mathbb{R}^n$ of the smooth curves and surfaces we meet in advanced calculus. (The simplest kind of manifold is $\mathbb{R}^n$ itself). The important fact about a manifold is that around each point is a region which can be given a coordinate system, so locally the manifold looks like $\mathbb{R}^n$; these coordinate systems allow us to generalize the usual ideas of differentiation on $\mathbb{R}^n$ to differentiation on manifolds. Thus we can have differentiable functions between manifolds. (Differentiable functions are sometimes called *smooth*). An important class of such functions is the *diffeomorphisms*, which are invertible functions where both the function and its inverse are differentiable. We often think of two manifolds that are related by a diffeomorphism as being equivalent, because the diffeomorphism is basically a smooth nonlinear change of coordinates from one to the other. For example, all smooth simple closed curves in the plane are diffeomorphic to the circle.

So now we can draw a picture of the sort of dynamical system we have in mind, see figure 1; the box represents $\mathbb{R}^n$, of which the manifold M is a subset—a surface in this picture. The trajectory of the point x is the set of points visited sequentially as we iterate the mapping T.

One obvious question we can ask about a dynamical system is what happens to the images of x, $T^n x$, as n increases? Some points do not move at all when T is applied; such points satisfy $Tx = x$ and are called *fixed points*: systems in these states do not change with time. Other points do change, but return to the same state after some number of time steps; they satisfy $T^k x = x$ for some k and are called *periodic points.* (Clearly, if a trajectory returns to a point, it must then repeat the sequence of states it visited after the point was last encountered, and so return again to the same point, and so on indefinitely. So the trajectory consists of a finite sequence of points repeated infinitely often.) Although T usually has some fixed and periodic points, most points are not of this kind: their trajectories never return exactly to the same place. What happens to these as n gets large? Typically we find that there are subsets of the state space which many (if not all) points approach and become confined to as we keep applying the map T. Such sets are known as *attractors.* Attractors can be quite varied: a single fixed point can be an attractor (and is

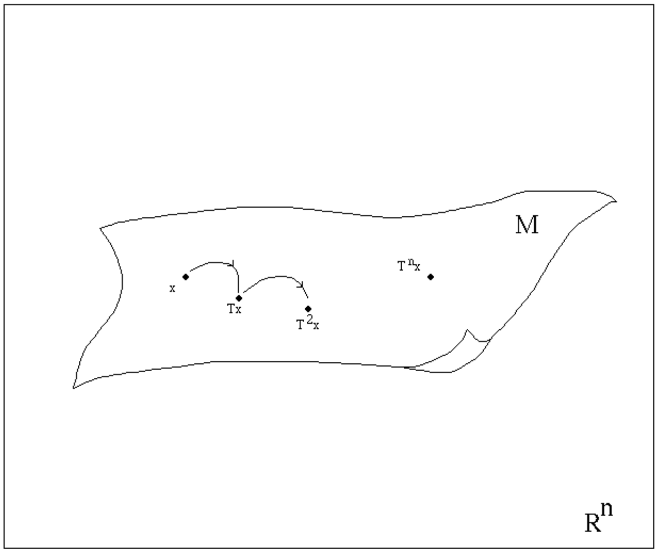

Figure 1: A differentiable dynamical system. M is the state space (a manifold). The first few points on the trajectory of x are shown.

then called an *attracting fixed point*), or a periodic orbit can be attracting. On the other hand, the attractor can be a very complicated set with a fractal structure.

To illustrate these ideas consider the map

$$T\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} ax(1-x-y) \\ axy \end{pmatrix} \tag{1}$$

which we can write in iterative form as

$$\begin{aligned} x_{n+1} &= ax_n(1-x_n-y_n) \\ y_{n+1} &= ax_ny_n \end{aligned}$$

where a is a constant.

T maps $\mathbb{R}^2$ to itself, so here $\mathbb{R}^2$ is the state space. Actually we have a family of maps: one for each value of the parameter a; we will choose different values to show different behaviours. To begin with, choose $a = 2.5$. Figure 2 shows the trajectory of a typical point $x_0 = (0.25, 0.04)$. This trajectory moves around the plane but eventually approaches the point $(0.4, 0.2)$, which is an attracting fixed point. All trajectories starting sufficiently close to $(0.4, 0.2)$ approach it in the long term, but points starting far away may do something else: they may for instance move away to infinity. (By solving $Tx = x$ we find that $(0, 0)$ is also a fixed point, but it is not attracting because there are points arbitrarily close to it which move away from it as n increases.)

Now consider the map where $a = 3.2$. This map has an attracting periodic orbit, shown in figure 3; the periodic orbit consists of six periodic points. Figure 4 shows the trajectory from $(0.21, 0.41)$, which approaches the periodic orbit. In figure 5 the same trajectory is shown, but now each point is joined to the one following it with a line: it is clear that the trajectory spirals out towards the periodic orbit.

Finally let us consider the map obtained by setting $a = 3.43$. Though this map still has fixed points and periodic orbits, they are not attracting: instead a typical point is attracted to the much more complicated set shown in figure 6. This attractor has a fractal

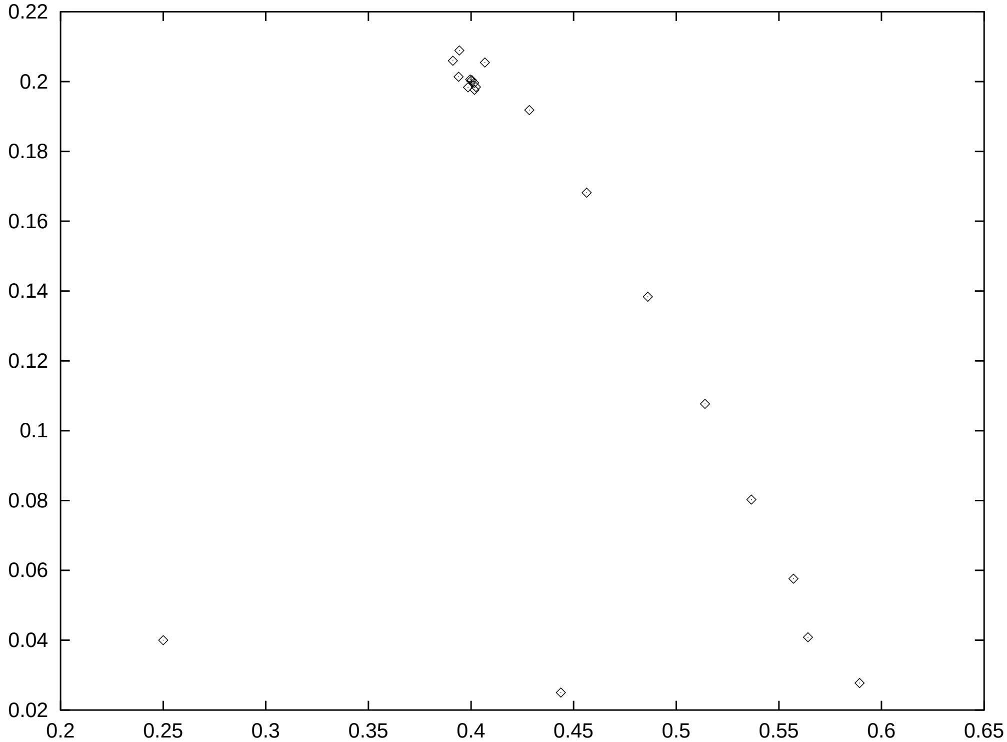

Figure 2: The trajectory of $(0.25, 0.04)$ under the map T given by (1) (with $a = 2.5$). The trajectory moves left to right across the lower part of the figure, before moving towards the top centre, and eventually approaching the fixed point $(0.4, 0.2)$.

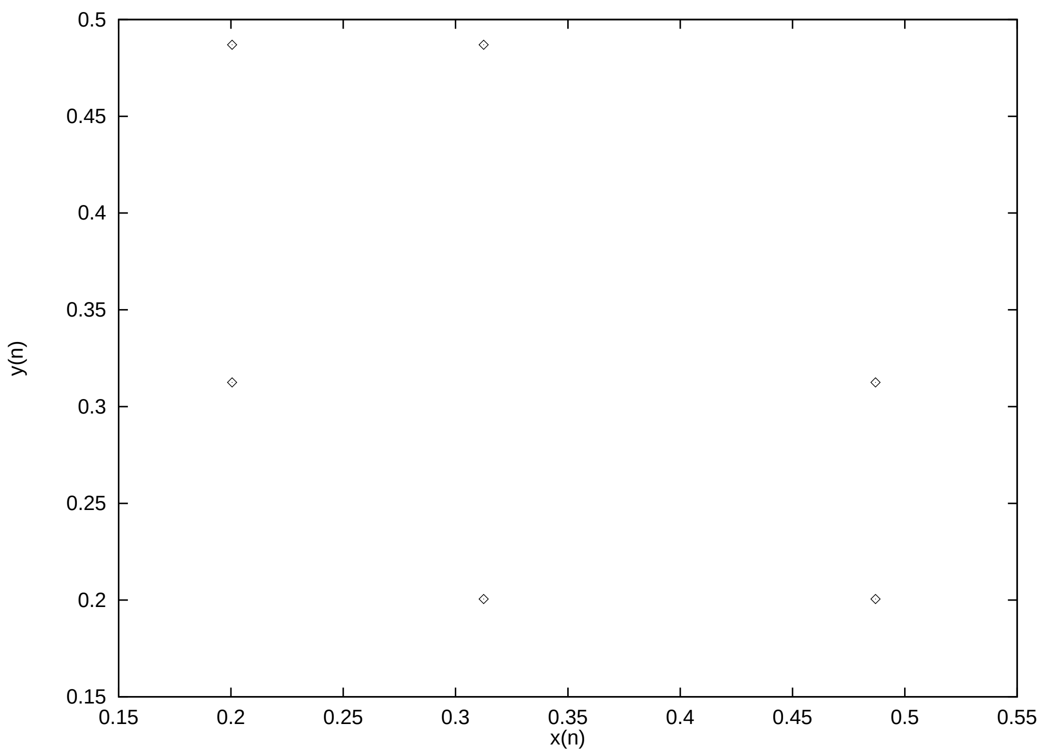

Figure 3: A periodic orbit of the map T given by (1) (with $a = 3.2$). The points are visited in anticlockwise order.

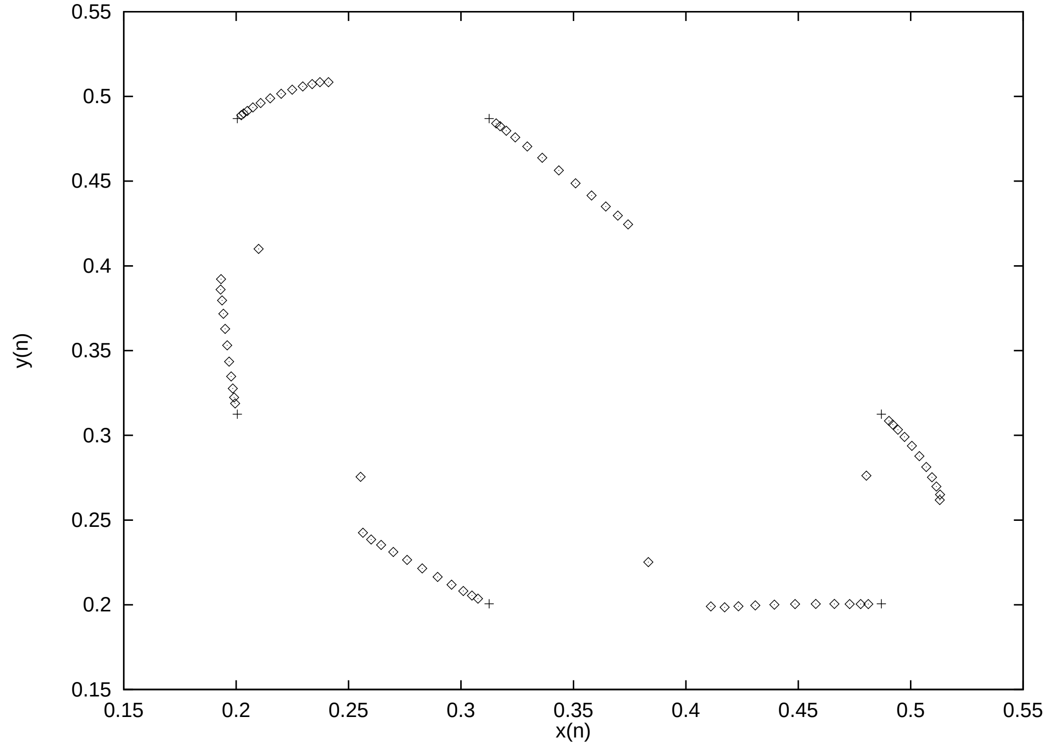

Figure 4: The trajectory of $(0.21, 0.41)$ under the same map as in figure 3. The trajectory approaches the periodic orbit shown in the previous figure, which is marked here by $+$'s.

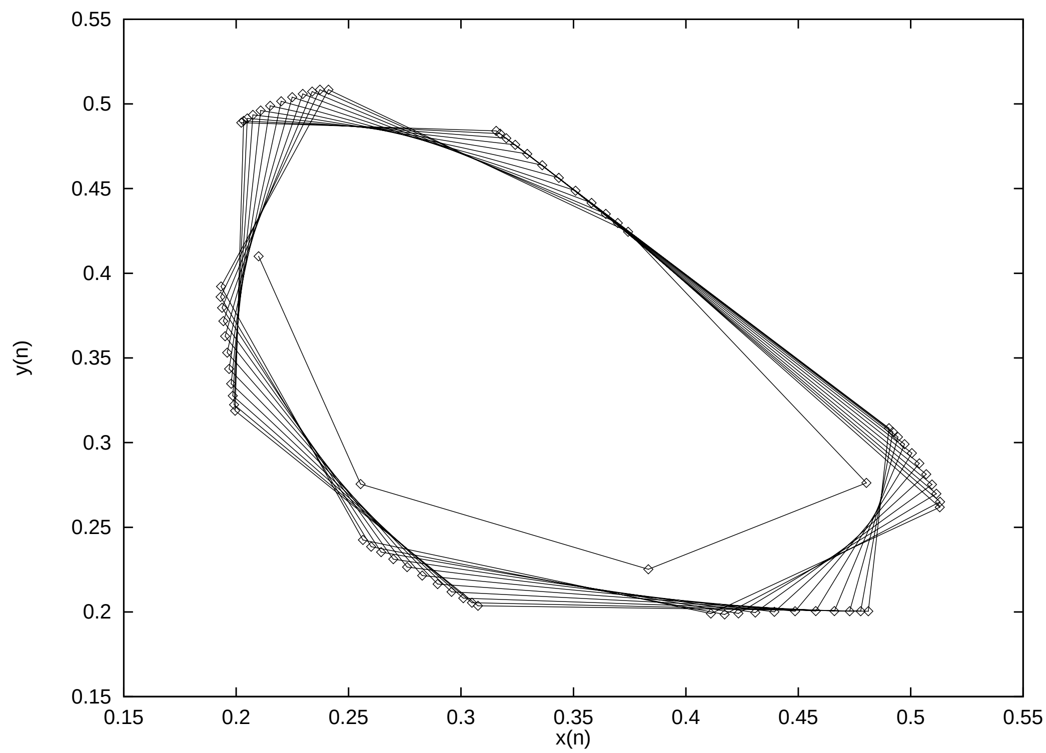

Figure 5: The same trajectory as in figure 4. The points on the trajectory have been joined up to make the path they follow clearer.

structure: if we magnify parts of the set we find more and more of the stranded structure is revealed at smaller scales. The trajectory represented in figure 6 does not repeat itself: as n increases the points fill out the set more and more densely. Furthermore, the order in which different regions of the attractor are visited is quite irregular. This kind of trajectory is called *chaotic*.

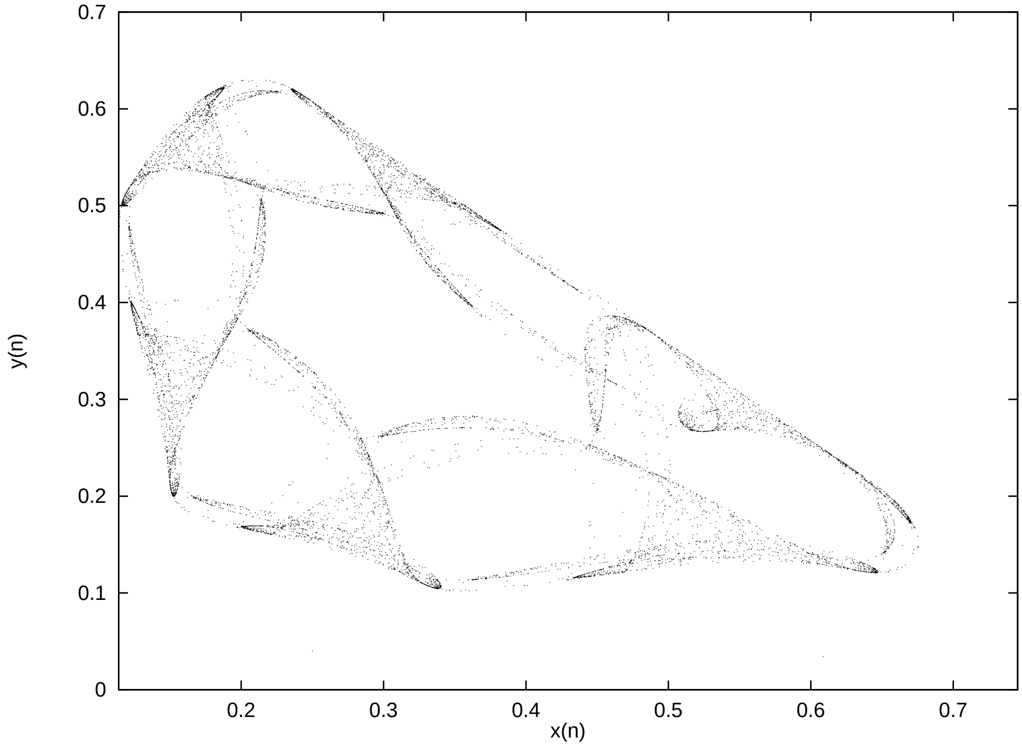

Figure 6: A chaotic attractor for the map T of equation (1) (with $a = 3.43$). 10000 points of the trajectory are plotted, but it is clear that some regions of the attractor are only sparesly populated: a longer trajectory would fill in the attractor more fully.

Of course, the dynamical systems of interest to us can be much more complicated than that of equation (1). Usually, the map T will embody the physical (or other) laws that govern the time evolution of the system. Sometimes, as in the case of classical mechanics, we know these laws in some detail; in other cases, such as ecological systems, our knowledge may be much more limited. Dynamical systems theory, however, treats all such systems on a similar footing, and one of its great appeals is that it allows us to make general statements relevant to a broad class of scientific phenomena. This broad view makes it attractive as an approach to signal processing, where we need to deal with many different kinds of signals from a variety of sources. (The statistical approach to time series has been very valuable to signal processing for similar reasons.) Corresponding to this broad view of nonlinear systems, the theory tends not to focus attention on particular trajectories, but to consider whole sets of them. It concerns itself with questions such as: what sort of attractors can exist in systems? What are the properties (for example, topological or statistical properties) of specific sets of trajectories such as those that form a given attractor, or which approach it? If we consider a collection of systems which are in some sense similar (such as those in (1), which differ only in the value of a parameter), how can the properties of trajectories vary among this collection?

Dynamical systems theory thus ranges widely from the abstract to the concrete. Our main interest here, though, is how we can use it as the basis for an approach to the processing of signals from nonlinear systems.

Signals from nonlinear dynamical systems

One of the purposes of signal processing is to extract information about a system from measurements made on it. How do these signals relate to the picture of a dynamical system described in the last section? Recall that points in M specify the state of the system; in a physical context the state will be some set of variables such as positions and momenta, or currents and voltages. Usually when we perform experiments on the system we do not acquire complete information about the state—we might only measure one of the variables, or perhaps some single quantity which depends on many of the variables. This suggests we introduce a real valued function $h: M \to \mathbb{R}$ into the theory to represent the experimental procedure. The number $h(x)$ is the result of the experiment if the system is in state x when the experiment is performed. h is sometimes called the *observation function*; different experiments correspond to different h's. (Note that we are assuming that the outcome of the experiment depends *only* on the state of the system: there are no external influences which affect the result. This is of course an idealization: experiments will always be disturbed by noise and other external effects, and we shall say more about this later.)

Given an observation function h (i.e. the specification of an experimental procedure), each trajectory of the dynamical system gives rise to a sequence of real numbers:

trajectory:	x	Tx	T^2x	$\ldots$	T^nx	$\ldots$
observations:	$h(x)$	$h(Tx)$	$h(T^2x)$	$\ldots$	$h(T^nx)$	$\ldots$

We usually call this a *time series.* (Note that we are still working with discrete time here, so that the measurements form a sequence of numbers, but a similar situation occurs in continuous time systems: there the state varies as a function of continuous time t, which we can write as $x(t)$; the measurements then form the function $h(x(t))$, a real-valued function of time.)

Our signal analysis problem is then: how can we get information about the system if we are presented only with the time series? The most penetrating information would enable us to analyse the system: can we tell if there is an attractor, or several? Can we say what kind they are, and what their topological properties are, for example? The fact that we have access only to the time series, not the trajectory directly, prompts us to ask: is there information we can get out of the time series which is invariant in the sense that it would turn out the same whatever observation function we used?

Perhaps less ambitiously, we might be interested in recognizing the system, rather than actually analysing it. Given two time series, might we be able to tell if they have come from the same system, perhaps using different observation functions?

And finally, if we can successfully gain information from the time series, could we use this to process the signal itself, to try to remove noise for example, or to separate signals from different systems? Does the time series contain the information we might need to control the system? In what ways could we control it?

As we shall see below, dynamical systems theory offers us some ways of approaching these questions.

The method of delays and Takens' theorem

As mentioned above, one of the most important classes of dynamical systems consists of those in which M is a differential manifold, and the map T is a diffeomorphism; (this is the case that arises, for instance, when the underlying dynamics are described by differential equations, and we convert to a discrete time system by sampling). For this case, a strong connection was established in the early 1980's between dynamical systems and time series measured from them. Before we spell out this connection we must say something about one way that we can treat time series information.

Recall that the observation function h converts a trajectory into a time series of numbers. Let us write this sequence as $h_0, h_1, h_2, \ldots$, where $h_n = h(T^n x) = h(x_n)$ is the n-th number in the sequence. We can collect consecutive values together to form *delay vectors*:

$$\begin{aligned} \boldsymbol{h}_0 &= (h_0, h_1, \ldots, h_{d-1}) \\ \boldsymbol{h}_1 &= (h_1, h_2, \ldots, h_d) \\ \vdots \quad & \quad \vdots \end{aligned}$$

Obviously, these are vectors in $\mathbb{R}^d$. Note that this simple procedure converts the time series of numbers into a time series of vectors. Note also that not only is h_n a function of x_n, but the whole vector $\boldsymbol{h}_n$ is determined by x_n, because

$$\begin{aligned} \boldsymbol{h}_n &= (h_n, h_{n+1}, \ldots, h_{n+d-1}) \\ &= (h(x_n), h(x_{n+1}), \ldots, h(x_{n+d-1})) \\ &= (h(x_n), h(Tx_n), \ldots, h(T^{d-1}x_n)) \end{aligned}$$

What we have done is to create a vector valued observation function, which we can write

$$\Phi_{(T,h)}: M \to \mathbb{R}^d$$

defined by

$$\Phi_{(T,h)}(x) = (h(x), h(Tx), \ldots, h(T^{d-1}x))$$

The method of delays constructs for us the values of $\Phi_{(T,h)}$ from the original time series. ($\Phi_{(T,h)}$ is sometimes called the *delay map*; for convenience we often shorten $\Phi_{(T,h)}$ to just Φ, but remember that it depends on T and h.)

We are now in a position to describe the connection between systems and time series mentioned above. The most common expression of this connection is known as *Takens' theorem*; roughly speaking this theorem says that, for most mappings T and observation functions h, the map $\Phi_{(T,h)}$ is an *embedding*, if d is large enough. What the word 'embedding' means is this: by assumption M is a differential manifold; we say that a function $\Phi: M \to \mathbb{R}^d$ is an embedding if the image $\Phi M \subset \mathbb{R}^d$ is also a manifold, and if the corresponding map between manifolds, $\Phi: M \to \Phi M$ is a diffeomorphism (in particular, it is invertible). The basic point is that ΦM can be considered to be a copy of M, with Φ acting essentially as a (nonlinear) change of coordinates.

This means that there is a one-to-one correspondence between states x of the system and the delay vectors $\Phi(x)$ to which they give rise. So the time series of delay vectors is essentially a copy of the original trajectory; we can use the time series of observations to create ('reconstruct' is the commonly used word) a trajectory in $\mathbb{R}^d$ which reflects the original trajectory on which the observations were made. We can draw a picture of this: see figure 7.

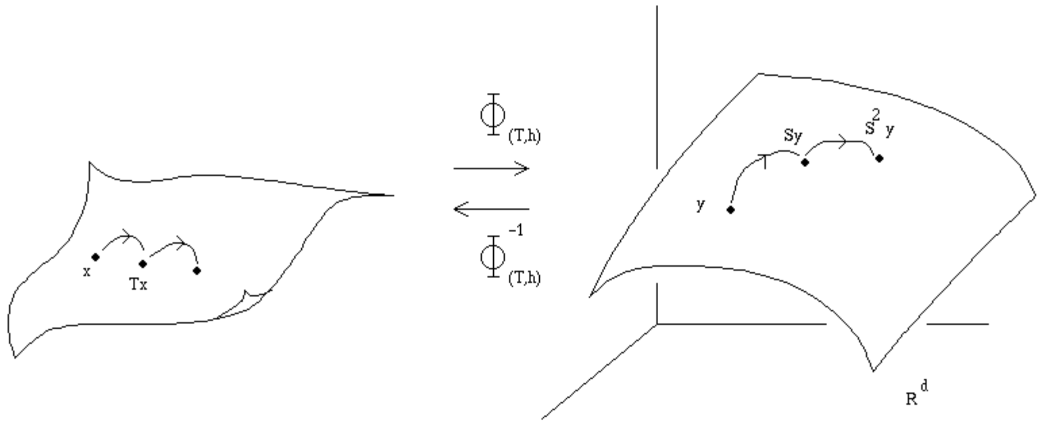

Figure 7: The original dynamical system and its reconstruction in delay space. $\Phi_{(T,h)}$ is the delay map connecting the two systems.

Not only is ΦM a manifold equivalent to the original M, but there is a corresponding dynamical system on ΦM, which is just the original transformation T in the new coordinates:

$$S = \Phi \circ T \circ \Phi^{-1}$$

(The fact that Φ is invertible, recall, is part of its being a diffeomorphism.) Thus the system $S: \Phi M \to \Phi M$ in delay space is a copy of the original system $T: M \to M$. The difference, of course, is that we know points in ΦM from our measurements, and so we have direct information about S.

The above description of Takens' theorem is rather rough. For the mathematically inclined, a precise statement of the theorem is as follows [2]:

Theorem 1 *Let M be a compact manifold of dimension m. For pairs (T, h), with T a diffeomorphism of M and h a smooth, real-valued function on M, it is a generic property that $\Phi_{(T,h)}: M \to \mathbb{R}^d$ is an embedding, if $d > 2m$.*

Most of the technicalities here do not need to concern us. Perhaps one thing worth mentioning is what is meant by 'most' in the statement that most T and h functions make $\Phi_{(T,h)}$ an embedding. It is clear that the delay map cannot be invertible for every T and h: if, for example, h were a constant, then every value of the time series would be the same, and the images of all $x \in M$ under $\Phi_{(T,h)}$ would be the same point. So we have to rule out some observation functions (and, it turns out, some T's) for the method of delays to work. In non-technical language what Takens' theorem states is that if, for some given T and h, $\Phi_{(T,h)}$ is an embedding, then the delay map will also be an embedding for all T' and h' sufficiently close to T and h. Furthermore, if for given T and h, $\Phi_{(T,h)}$ is *not* an embedding then we can find T' and h', arbitrarily close to T and h, such that $\Phi_{(T',h')}$ *is* an embedding.

(Incidentally, the last remark gives a hint about how we can prove this kind of theorem: the basic approach is to show how, given T and h that do not give rise to an embedding, we can create close by functions that do lead to an embedding. This is done by adding small functions to T and h. The details, however, are rather technical. Happily, these details are unnecessary for actually using the method of delays.)

We may be able to use the method of delays to reconstruct a dynamical system even if it does not satisfy the conditions of Takens' theorem (if, for example, the function T is not invertible). The important thing is whether the delay map $\Phi_{(T,h)}$ is a diffeomorphism; if it is, the reconstructed system is still a copy of the original in the way described above, and can be used for extracting information about the original system. (However, without knowing T and h we cannot tell in advance if $\Phi_{(T,h)}$ is a diffeomorphism—we must try to test this using the observed data.)

Since Takens' original paper there have been a number of other 'embedding theorems' which extend the method to cases not covered by the original theorem [3, 4, 5, 6, 7]. This work addresses questions such as: can the method of delays still be used if we filter the time series first? What is the effect of having a random or noisy system instead of the completely deterministic one we have been considering? However, there are still plenty of unanswered questions concerning the best way to go about processing time series data to learn about dynamical systems.

We can give a simple illustration of the method of delays using the dynamical system given by equation (1) above. For an observation function h we choose the y component of the points: $h(x, y) = y$. The trajectory pictured in figure 6 (for example) then gives rise to a time series as described above—in this case the time series is just the sequence of y components. Part of this time series is shown in figure 8; we see there is a fairly regular oscillation (this reflects the period 6 orbit of the system, which is still present, but not attracting), but the sequence has irregular features typical of chaotic motion.

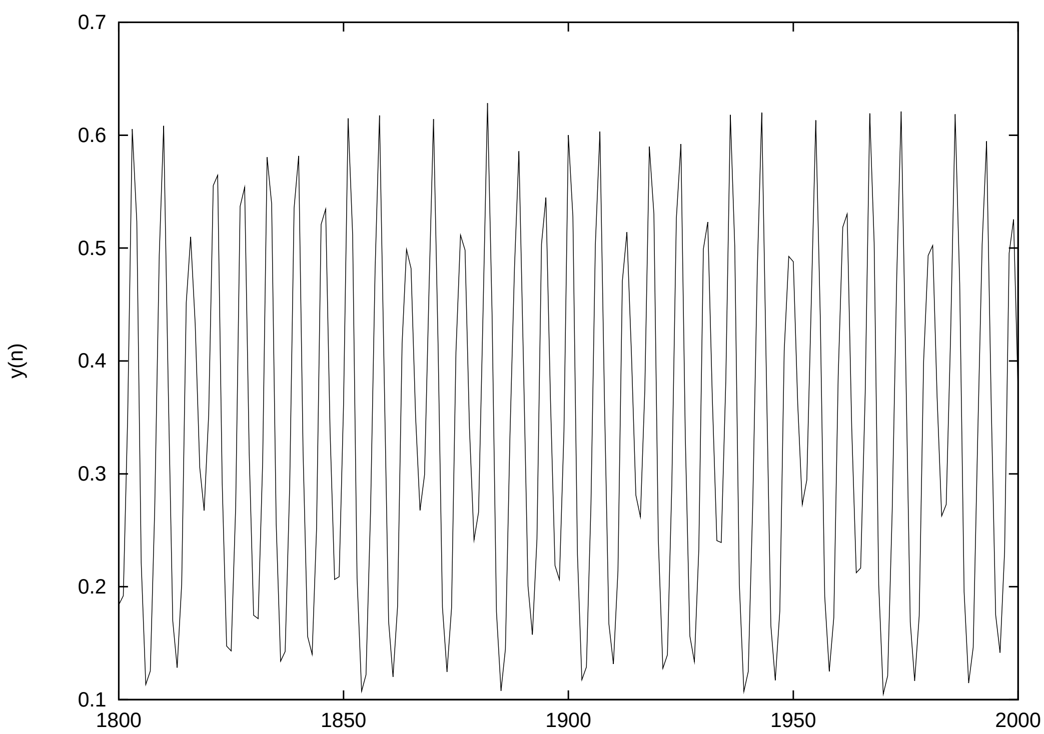

Figure 8: A segment of the time series of y components of the trajectory of figure 6.

The time series values can be gathered into delay vectors. How long should these vectors be? Takens' theorem tells us that—for maps satisfying the conditions of the theorem—it is enough to take $2m + 1$ delays, where m is the dimension of the state space M. However, it may not be necessary to use this many, and often we will not know the dimension of M in advance anyway. We can proceed by experimenting with different numbers of delays. In figure 9 we plot the sequence of delay vectors with three delays (that is, we plot (y_n, y_{n+1}, y_{n+2}) for $n = 0, 1, \ldots$). This figure should be compared with figure 6; it is reasonably clear that the points in the delay plot are confined to a two dimensional surface (which is ΦM in the above notation), and that the trajectory is a nonlinearly (but smoothly) distorted copy of that in figure 6. The relationship between the trajectories is even clearer if we just take two delays: see figure 10.

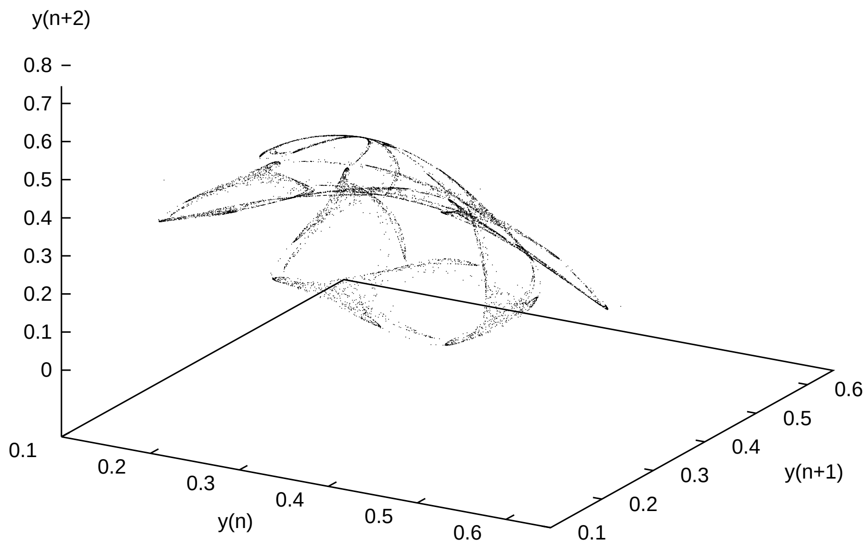

Figure 9: Plot of delay vectors constructed from the time series in figure 8. The attractor is a copy of that in figure 6.

Finding properties of the system from measured data

If the method of delays produces an embedding there is, as we have seen, a strong relationship between the set of delay vectors in $\mathbb{R}^d$ and the original dynamical system (namely, they are diffeomorphic). This means that many properties of the original system are preserved in the reconstructed one. The preserved properties are essentially those which are invariant under coordinate changes, for example, the topology of M and of its attractors (but not, say, the sizes of the attractors). The pattern of trajectories is preserved: a periodic orbit in M, for example, gives rise to a periodic orbit in the delay reconstruction, of the same period. Certain statistical properties such as the various fractal dimensions of attractors are preserved. And the characteristic numbers of the fixed points and periodic orbits (i.e. the eigenvalues of the Jacobian of T about these points, which determines their stability to small perturbations), are also preserved, as are the Lyapunov numbers (which are basically characteristic numbers for the whole attractor).

Since all these quantities are the same for the system in delay space as for the original system, we can hope to learn about the original values by estimating them from the time

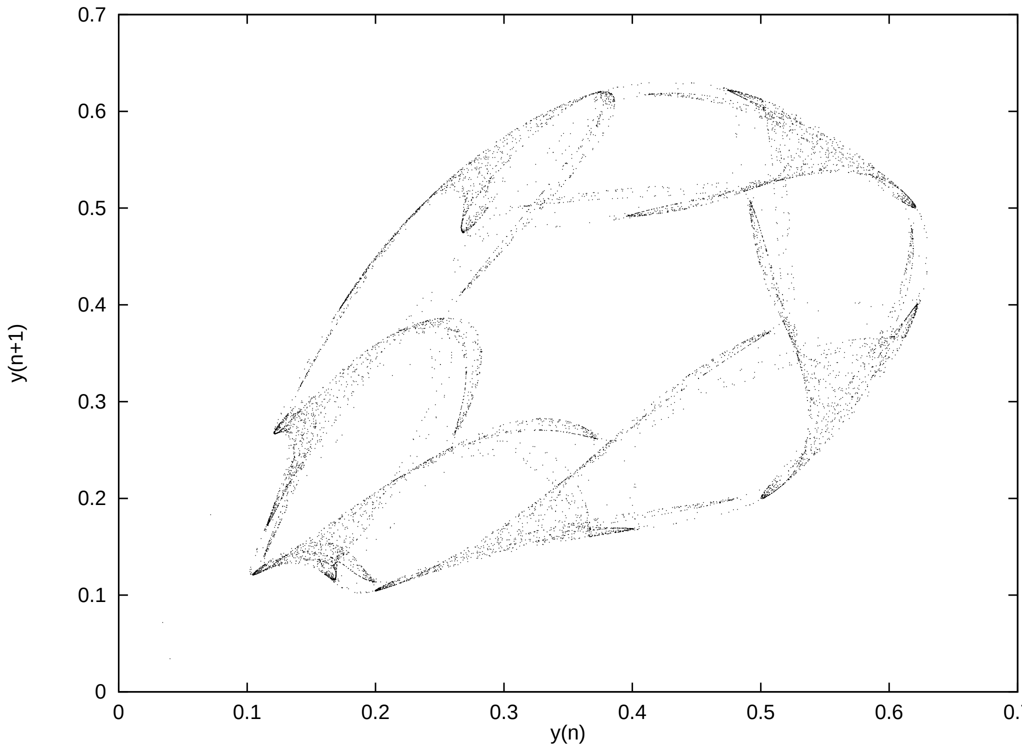

Figure 10: Delay vectors from the time series of figure 8, now using just 2 delays.

series data. We can easily construct delay vectors from such data and plot them in $\mathbb{R}^d$; then we can estimate the manifold in which such data lie and examine its dimension and topology. We can estimate the map S by fitting the data, because we know where each of our delay vectors maps to, and we can attempt to find the eigenvalues of its derivative. And there are various other things we can do.

This possibility has spawned a great industry. Delay reconstructions of all kinds of time series have been produced and attempts made to calculate invariants from them—including from some time series for which there is very little reason to believe Takens' theorem has any relevance. Many of these studies had to do with the attempt to identify chaotic behaviour in physical or other systems. Even if Takens' theorem is applicable however, (meaning, even if the time series is observed from a system satisfying the conditions of the theorem), the calculation of invariants may be difficult, requiring large amounts of clean data. Finding the fractal dimension of an attractor, for example—not long ago one of the most common ways of trying to show a system is chaotic—often requires unreasonably large numbers of data points. Other invariants, however, may be easier to find.

Processing signals from nonlinear systems

Apart from the analysis of experimental systems, there is another side to delay reconstructions: since they give us a handle on the systems that generate the time series, we can use them to process the time series and manipulate the systems themselves. So they can be useful for various signal processing activities, such as the separation of signals from different sources which have become mixed together; the extraction of signals from noise or interference; the detection of known signals, and so on. Standard techniques of signal processing are based on quite different assumptions about signals than the ones discussed here, so the possibility exists that the special properties of differentiable dynamical systems—in particular the embedding property—might allow us to achieve our signal processing aims better than the standard methods, in those cases where the new assumptions apply.

Noise reduction for chaotic time series

As an example of how state space reconstruction can be used to process as well as analyse signals, we will consider a simple noise reduction problem. Whenever we take measurements on a system we must expect that the observations will be contaminated with noise. So, for example, instead of observing the noiseless time series h_n from the dynamical system $T: M \to M$, we might have recorded the sequence $z_n = h_n + \eta_n$, where η_n is some random noise process. To illustrate the effect of this noise, consider the map of the plane $T: \mathbb{R}^2 \to \mathbb{R}^2$ defined by

$$T\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 - ax^2 + y \\ bx \end{pmatrix} \tag{2}$$

or in iterative form:

$$\begin{aligned} x_{n+1} &= 1 - ax_n^2 + y_n \\ y_{n+1} &= bx_n \end{aligned}$$

where a and b are constants. (This is the 'Hénon map', and is one of the most commonly used examples in applied dynamical systems theory.) For $a = 1.4$ and $b = 0.3$ the map has a chaotic attractor. We create a time series using the observation function $h(x, y) = x$; a typical segment of such a time series is shown in figure 11: it shows the irregular behaviour of chaotic motion. (In fact, the time series plotted here has been normalized for convenience to have zero mean and unit variance.) Applying the method of delays to this time series produces figure 12; two delays are sufficient to show the reconstructed attractor quite clearly.

We model the noise by an independent, identically distributed (i.i.d.) gaussian sequence η_n, whose standard deviation is 0.1 times that of the x_n signal. A typical realization of the noisy signal $z_n = x_n + \eta_n$ is shown in figure 13, together with the clean signal from figure 11.

Although the time series does not appear to be greatly affected by the noise, we can see that the noise does in fact have a major effect by plotting the delay vectors of the z_n sequence: figure 14. Clearly much of the structure of the attractor has been smeared out, and it is easy to believe that the noisy time series is much less useful for estimating properties such as fractal dimension or characteristic numbers. Furthermore, conventional filtering is not very effective at recovering the situation, because removing frequency components of the signal also tends to distort the attractor.

In delay space, the effect of the noise is to add a random vector to each of the delay vectors in the reconstructed attractor. So long as the amplitude of the noise is not too high, the delay vectors will not be moved very far away from their true (noiseless) positions. We know from Takens' theorem that the true positions lie on a subset of the delay space which is the image under the delay map Φ of the original state space M. In the example here, M is $\mathbb{R}^2$ and so the noiseless delay vectors will lie on a 2 dimensional surface in delay space. The random perturbations due to noise, however, point in all directions, and so the delay vectors are moved off the 2-d surface. So long as they are not moved too far, however, it should still be possible to estimate from the data where this surface lies. And we can then attempt to remove some of the noise by moving the delay vectors back on to the surface.

In general, the shape of the surface can be quite complicated, so in practice we proceed as follows. A point is chosen from the delay vectors, and a small ball is constructed, with

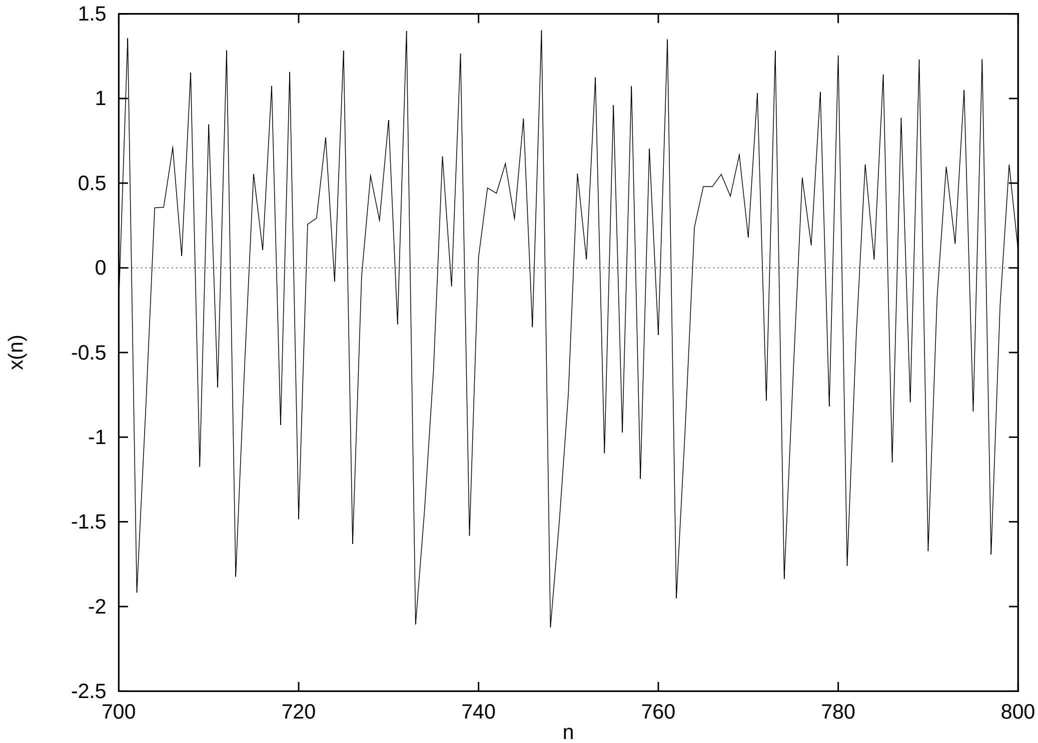

Figure 11: Typical segment of a times series of x_n values from the Hénon map, equation (2).

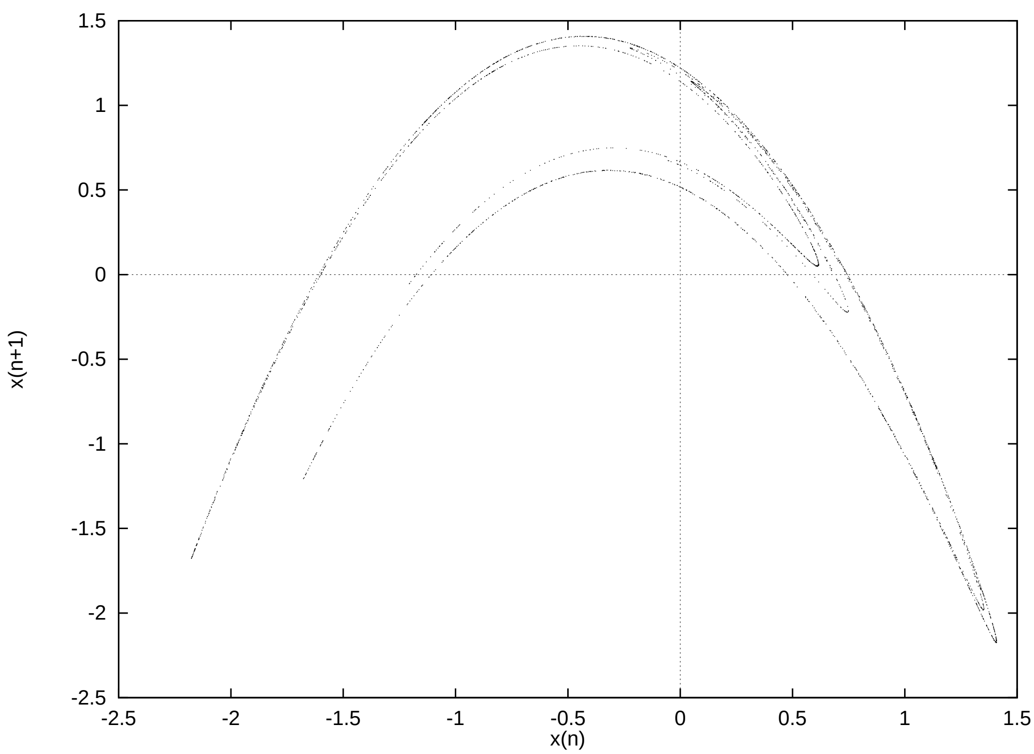

Figure 12: Delay vectors from the Hénon map. 3000 points are shown.

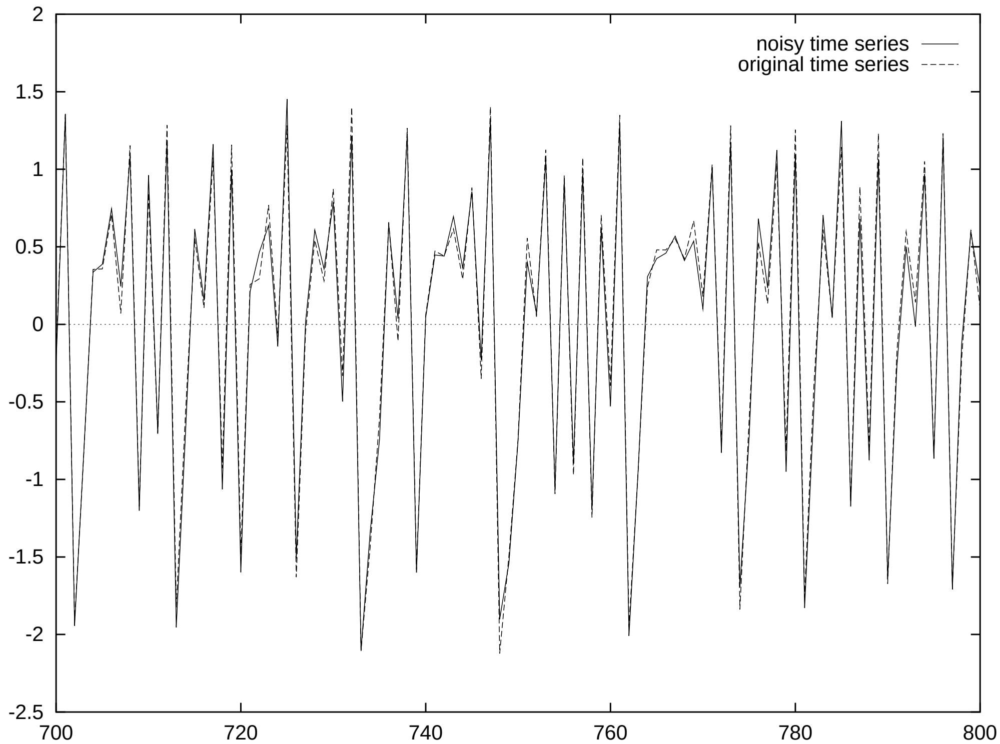

Figure 13: A segment of the noisy times series created by adding an i.i.d. gaussian sequence to the time series of figure 11.

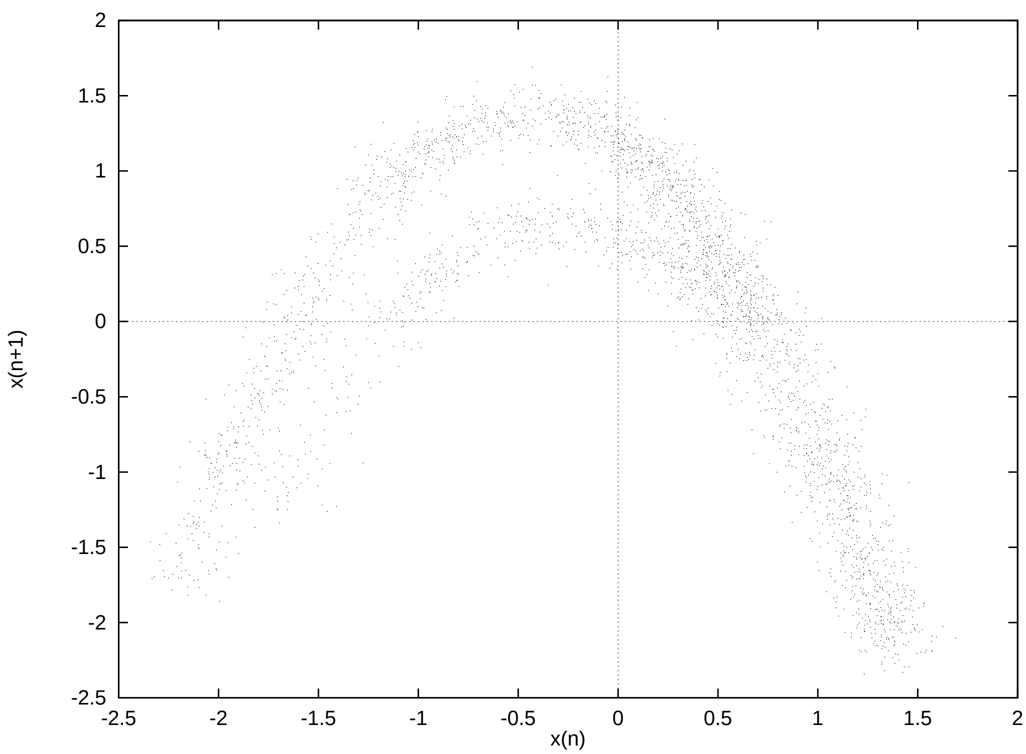

Figure 14: Delay vectors from the noisy z_n sequence.

centre at the chosen point. Since the surface we are looking for is smooth, in a small region around any of its points it will appear essentially planar; so if we look at all the delay vectors contained in the ball we should find that they lie approximately in a plane, passing through or near the centre of the ball. (For this to be true, however, it will be necessary for the ball to be larger than the perturbations due to noise. On the other hand, the ball cannot be made too large otherwise the surface will not look planar inside it; hence this approach cannot be expected to work with high amplitude noise.) The location of the plane can be estimated from the data in the ball by a technique such as principal components analysis. Once it is found, we can adjust the position of the chosen point by projecting it on to the plane (or at least moving it towards the plane.)

The whole procedure can be repeated for each delay vector in turn. The result should be a new set of delay vectors with less noise than the originals. More noise can possibly be removed by repeating the process, and in practice we iterate around several times until no further noise seems to be being removed. This procedure was first proposed by Cawley and Hsu [9], and by Sauer [8].

To apply the process to our example, we need to choose a ball size, and the dimension of the delay space in which to work. The radius of the balls is taken to be twice the standard deviation of the noise. (Often, of course, one will not know this standard deviation, and it may be necessary to experiment with various ball sizes.) The number of delays is taken to be 5; thus, within each ball we are looking for the 2 dimensional linear subspace of $\mathbb{R}^5$ that lies closest to the delay vectors. (Higher dimensional delay spaces can be used, and may lead to more noise reduction, but more data is generally needed to estimate the positions of the local planes.) The process was iterated 5 times. The resulting delay vectors are plotted in figure 15. While some noise remains, it is clear that the adjusted points much better represent the original attractor, and even some of the fractal structure has been recovered.

The local projective method described here is not the only approach to the noise reduction of chaotic time series: see [1] and [10] for more information. What the approaches have in common however, is that they work in the reconstructed state space rather than the time or frequency domains; this enables them to exploit the special properties of deterministic, but nonlinear, dynamical systems.

Other aspects of nonlinear signal processing

Noise reduction of nonlinear time series is just one of a range of techniques and applications emerging from the field of nonlinear dynamical systems. One area that has proved particularly surprising is that of the linear filtering of chaotic signals. It turns out that recursive (IIR) filtering can significantly alter the properties of such signals, in particular by changing the characteristics (such as dimension) of their attractors. Although such properties can in principle be found from data using the method of delays, filtering can make this fail. (This happens essentially because the filter itself becomes part of the observed system, its output reflecting not only the system from which the original signal came, but a mixture of this system and the filter.) Hence when dealing with such signals the choice of filter needs to take account of considerations beyond the usual ones of desired spectral characteristics. One application of the interaction between linear filters and nonlinear deterministic system has been to signal separation [11].

One area where there has been very rapid growth is in the modelling and control of nonlinear systems using delay space methods. Modelling chaotic systems from data has exploited a variety of techniques, from global nonlinear ones such as radial basis functions or neural networks, to local ones including low-order polynomial fitting. Such models have been

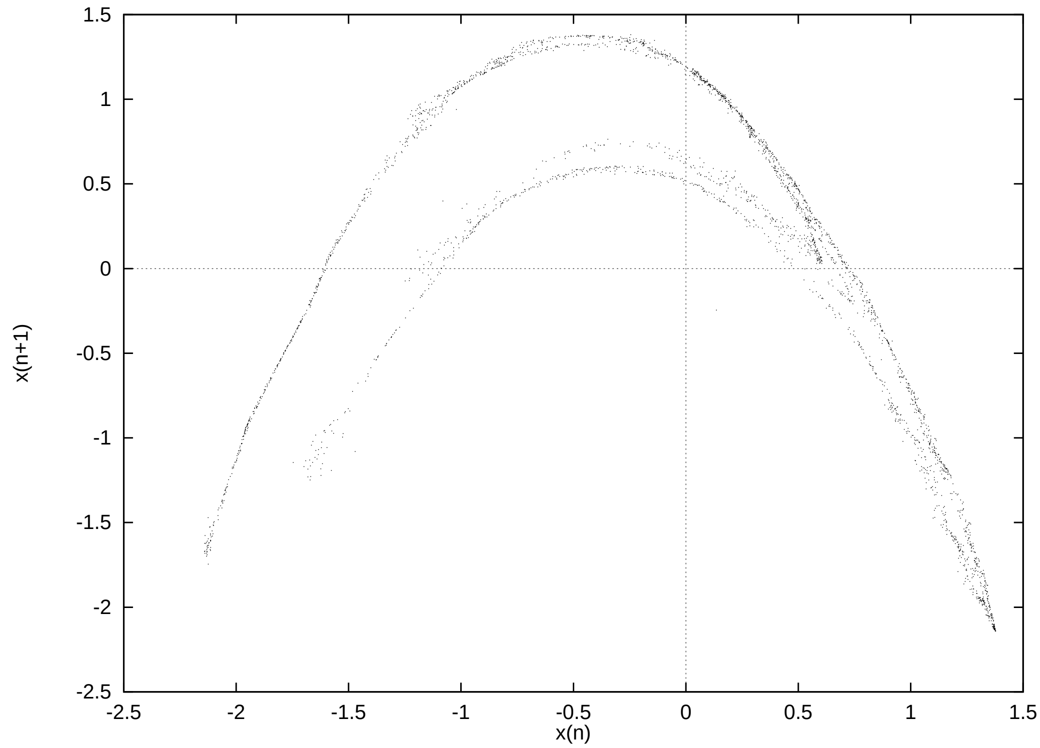

Figure 15: Delay vectors from the Hénon map sequence after noise reduction has been applied.

used to make predictors for nonlinear systems, which can be applied to noise reduction and data compression as well as control. The properties of chaotic attractors have themselves suggested new types of control [12].

Another growth area has been the synchronization (or more generally, coupling) of nonlinear systems; much of the interest in this field has stemmed from potential applications in communications. There is now a huge literature on this field: see [13] for a review.

A more fundamental area, in which further progress needs to be made, is that of stochastic dynamical systems. At the moment the treatment of noise in the processing of signals from nonlinear systems is somewhat *ad hoc*, and an integrated theory of random dynamical systems would probably put the techniques on a firmer basis. Stochastic dynamical systems do not obey Takens' embedding theorem in the same way as deterministic ones (but see [5, 6]) and so the methods developed for the latter are not immediately applicable to the former. In particular, systems with large amplitude noise are not covered by the current approaches.

However, one kind of random system for which the deterministic theory does not need much modification is the *iterated function system*; here the single map T in the deterministic system discussed above is replaced by a finite collection of maps $T_1, T_2, \ldots, T_n$, and at each time one of these is chosen at random to move the state on one time step. So long as the number of maps is finite, a version of Takens' theorem applies. These systems promise to have a number of applications. One immediate one is to channel equalization in digital signalling systems [14]. Here, the symbols in the signalling alphabet correspond to the different maps, and the channel itself is the dynamical system, whose behaviour is determined by the symbols input to it. Delay methods can be applied to the output of the channel to determine the sequence of maps (and hence the input sequence) that was applied to it.

References

[1] H. Kantz and T. Schreiber, 1997, *Nonlinear Time Series Analysis*, Cambridge University Press, Cambridge.

[2] F. Takens, 1981, Detecting strange attractors in turbulence, in *Dynamical Systems and Turbulence*, (Springer Lecture Notes in Mathematics, vol. 898) edited by D. A. Rand and L.-S. Young, Springer-Verlag, Berlin.

[3] T. Sauer, J. A. Yorke and M. Casdagli, 1991, *Embedology*, Journal of Statistical Physics, **65**, 579-616

[4] D.S. Broomhead, J.P. Huke and M.R. Muldoon, 1992, *Linear Filters and Nonlinear Systems*, Journal of the Royal Statistical Society B, **54**, 373–382

[5] J. Stark, D.S. Broomhead, M.E. Davies and J. Huke, 1997, *Takens embedding theorems for forced and stochastic systems*, Nonlinear Analysis, Theory, Methods and Applications, **30**, 5303–5314

[6] M.R. Muldoon, D.S. Broomhead, J.P. Huke and R. Hegger, 1998, *Delay embedding in the presence of dynamical noise*, Dynamics and Stability of Systems, **13**, 175–186

[7] J. Stark, 1999, *Delay embeddings for forced systems I*, Journal of Nonlinear Science, **9**, 255–332

[8] T. Sauer, 1992, *A noise reduction method for signals from nonlinear systems*, Physica D, **58**, 193-201.

[9] R. Cawley and G.-H. Hsu,1992, *Local-geometric-projection method for noise reduction in chaotic maps and flows*, Phys. Rev. A, **46**, 3057-3082.

[10] E. J. Kostelich and T. Schreiber, 1993, *Noise reduction in chaotic time-series data: A survey of common methods*, Phys. Rev. E, **48**, 1752-1763.

[11] D.S. Broomhead, J.P. Huke and M.A.S. Potts, 1996, *Cancelling Deterministic Noise by Constructing Nonlinear Inverses to Linear Filters*, Physica D **89** 439–458.

[12] E. Ott, T. Sauer and J.A. Yorke (eds.), 1994, *Coping with Chaos*, Wiley, New York.

[13] L.M. Pecora, T.L. Carroll, G.A. Johnson, D.J. Mar and J.F. Heagy, 1997, *Fundamentals of synchronization in chaotic systems, concepts and applications*, Chaos, **7**, 520–543.

[14] D.S. Broomhead, J.P. Huke and M.R. Muldoon, 1998, *Fractals, Linear Channels and Delay Methods* in 'Mathematics in Signal Processing: Proceedings of the Fourth Institute of Mathematics Conference', eds. J.G. McWhirter and I.K. Proudler, Oxford University Press.