Iterated Function System Models of Digital Channels

Broomhead, D. S. and Huke, J. P. and Muldoon, M. R. and Stark, J.

2004

MIMS EPrint: **2005.1**

ISSN 1749-9097

Iterated Function System Models of Digital Channels

By D.S. Broomhead[1], J.P. Huke[1], M.R. Muldoon[1], J. Stark[2]

[1]*Department of Mathematics,*
University of Manchester Institute of Science and Technology,
P.O. Box 88,
Manchester M60 1QD, UK

[2]*Department of Mathematics,*
Imperial College London,
180 Queen's Gate,
London SW7 2BZ, UK

This paper introduces a new class of models of digital communications channels. Physically, these models take account of the digital nature of the input. Mathematically, they are iterated function systems. As a consequence of making explicit assumptions about the role of discreteness in the models, it is possible to make general statements about the behaviour of these channels without needing to assume that they are linear. We provide the mathematical background necessary to understand the behaviour of these models and prove a number of results about their observability. We also provide a number of examples intended to demonstrate their connection with linear state space models, and to suggest how the nonlinear theory might be developed towards applications.

Keywords: iterated function systems, digital channel models

1. Introduction

This paper concerns the modelling of digital communications systems; in particular, it introduces a new approach to the modelling of digital channels that is sufficiently general to incorporate nonlinear channel models. We aim to persuade the reader of two things: that there is a theory that allows nonlinearity to be dealt with in a general way without descending into a morass of special cases; and that this theory can be viewed as a natural development of the already familiar state space modelling of channels. The paper is a synthesis of two areas of recent mathematical development: work on a class of stochastic dynamical systems known as *iterated function systems* (IFS) (for a good self-contained introduction, see the book by Barnsley (1988), or for more mathematical detail Falconer 1990, Diaconis & Freedman 1999 or Kigami 2001); and *delay embedding* ideas, which have been developed in the dynamical systems community over the past two decades (see the books by Ott *et al.* 1994 and Kantz & Schreiber 1997). IFSs will be used to formulate a new channel model which is much richer than those used hitherto. The new channel model should really be seen as modelling both the channel and the transmitted signal.

Delay embedding will provide the tools that establish the relationship between the output of the channel, its internal state and ultimately the input that generated them.

In the next section the idea of IFSs will be introduced in the context of digital channel models. In addition, some examples will be developed which will serve as illustrations both here and in subsequent sections. In §3 some of the basic mathematical ideas that underpin the theory of IFSs will be described. In the following section, §4, linear systems will be discussed from this point of view. In §5 it will be shown how delay methods—which later will be developed in the general nonlinear context—reduce to the standard state space approach in the linear case. In §6 the full nonlinear theory will be developed and results proved about the information that can be obtained by processing the output time series of a general IFS channel. Finally, in §7, the implications of these results for digital signal processing will be discussed.

2. IFSs as Models of Digital Channels

A communications channel is a physical system which can—at least in principle—be modelled by differential equations derived from the laws of physics. For example, one may motivate ordinary differential equation (ODE) models of linear channels by considering the response of passive electrical circuits to externally applied driving signals. Similarly, the telegraph equation—a linear partial differential equation (PDE)—is a good model of signal transmission along an insulated wire. In optics, propagation of the envelope of an intense light pulse can be modelled by another PDE, the nonlinear Schrödinger equation. Here the Kerr effect—the dependence of the refractive index of the fibre on the electric field intensity—necessitates the use of a nonlinear model.

A channel can be very complicated and include the transmitter, receiver and amplification/repeater stages as well as the actual medium through which the signal propagates. We do not expect, therefore, to be able to write down and solve exact physical models (and, indeed, such explicit descriptions of channels are not commonly employed in the communications literature). However, the fact that we assume the existence of such a model allows us to make a number of further, basic assumptions which concern the representation of the state of the channel and the way that the state evolves. In particular, we shall generally assume that the state of the channel can be thought of as a point in a suitable state space, and that the state evolves according to a flow generated by the underlying differential equation model. For example, if the model consists of an ODE (or a system of ODEs) the state space will be $\mathbb{R}^n$ (for some positive integer n), or possibly some more general finite dimensional manifold; if the model is a PDE the state space will be a function space. Note that the differential equation will need to be non-autonomous to take account of the input to the channel. We are considering *digital* communications and so the input consists of a sequence of symbols drawn from some finite alphabet. The nature of these symbols is arbitrary, though in practice they would almost always be (possibly complex) numbers.

We take the state space of the channel be a compact, m-dimensional (finite m) differentiable manifold, $\mathcal{M}$. We assume that time is divided up into consecutive periods of length τ and that one symbol is input during each period—the number of symbols in the alphabet is K. For each symbol there is a system of ODEs on $\mathcal{M}$ that describe the evolution of the channel's state while that symbol is being fed in. That is, we have a collection of vector fields $X_k : \mathcal{M} \times [0, \tau) \to T\mathcal{M}$ where $k \in \{1, 2, \ldots, K\}$ labels the symbol and the dynamics of the channel are governed by

$$\frac{dx}{dt} = X_{k_n}(x, t - n\tau) \qquad \text{for} \qquad n\tau \leq t < (n+1)\tau.$$

Here $x \in \mathcal{M}$ is the state and k_n is the symbol input during the nth period $[n\tau, (n+1)\tau)$.

We assume the X_k are sufficiently well-behaved that the system of ODEs $\dot{x} = X_k(x, t)$ has unique solutions; then, integrating from $t = 0$ to $t = \tau$ gives a diffeomorphism $\mathbf{w}_k : \mathcal{M} \to \mathcal{M}$. (In this paper we consider sampling the channel at the input symbol rate; this corresponds to integrating the ODEs for the whole symbol period τ. Oversampling the channel would necessitate subdividing this interval; this possibility will be discussed elsewhere.) So in this picture an alphabet of input symbols corresponds via the channel model to a set of mappings of the state space. Sometimes it will be possible to assume that the input symbols are represented by forcing terms, $\chi_k : [0, \tau] \to T\mathcal{M}$, which are added to a symbol independent vector field

$$X_k(x, t) = X(x) + \chi_k(t) \tag{2.1}$$

as in the following examples.

Example 2.1 (A 2nd order linear recursive channel). *The model is a damped harmonic oscillator (damping constant γ and undamped natural frequency ω_0) which is forced by a sequence of non-overlapping pulses, $s_k(t)$, each occupying an interval of constant length τ.*

$$\frac{d^2}{dt^2}u + \gamma\frac{d}{dt}u + \omega_0^2 u = \sum_{l=0}^{\infty} s_{k_l}(t - l\tau)$$

We shall generally think of this as a system of 1st order ODEs:

$$\frac{d}{dt}\begin{pmatrix} u \\ p \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -\omega_0^2 & -\gamma \end{pmatrix}\begin{pmatrix} u \\ p \end{pmatrix} + \sum_{l=0}^{\infty} \chi_{k_l}(t - l\tau) \tag{2.2}$$

where the $\chi_k(t) = (0, s_k(t))^T$ are the compactly supported input pulses. In this simple example, the state space of the channel is $\mathbb{R}^2$.

For concreteness we will take the input to be binary so $K = 2$ and (calling the pulse shapes s_- and s_+ rather than s_1 and s_2) take $s_\pm(t)$ to be ± 1 on the interval $[0, \tau)$ and zero elsewhere. In this case the corresponding diffeomorphisms are

$$\mathbf{w}_\pm\begin{pmatrix} u \\ p \end{pmatrix} \;=\; A\begin{pmatrix} u \\ p \end{pmatrix} \pm B \tag{2.3}$$

where the 2×2 matrix A and the vector B may be obtained via simple integrations. Provided that $\gamma > 0$, the maps $\mathbf{w}_\pm$ are affine contractions of the state space with a common linear part.

Example 2.2 (A 2nd order nonlinear recursive channel). *This model is derived from example 2.1 by the inclusion of a cubic contribution to the restoring force so that the oscillator is now of Duffing type (Thompson & Stewart 2002).*

$$\frac{d^2}{dt^2}u + \gamma\frac{d}{dt}u + \omega_0^2 u + u^3 = \sum_{l=0}^{\infty} s_{k_l}(t - l\tau)$$

In this example we force the system with a sequence of delta functions uniformly spaced in time and with amplitudes $\pm\chi$. Our two pulse shapes are: $s_\pm(t) = \pm\chi\delta(t)$. As before we can take the state space of the channel to be $\mathbb{R}^2$.

We cannot find closed forms for the maps $\mathbf{w}_\pm$ in this case. This will usually be the case even for simple nonlinear equations. For the illustrations given below we find $\mathbf{w}_\pm$ by numerical integration.

This picture, in which an alphabet of input symbols corresponds via the channel model to a set of mappings of the state space, becomes more complicated if the physical model is a PDE. In this case the state space is an infinite-dimensional function space and the dynamics are generally given by a semi-flow. However, for certain dissipative PDEs, it can be shown that there exists an attracting, finite-dimensional, invariant submanifold—called the *inertial manifold*—of the state space. The PDE restricted to this manifold is a system of ODEs (see for example the books Temam 1988 and Constantin *et al.* 1989).

What happens when a sequence of symbols $k_1, k_2, k_3, \ldots$ is input to the channel? If the initial state of the channel is x_0 then we see from the above discussion that the state after the first symbol period will be $\mathbf{w}_{k_1}(x_0)$. Since this is the state at the beginning of the second symbol period, the state after two symbols have been input will be $\mathbf{w}_{k_2} \circ \mathbf{w}_{k_1}(x_0)$, and so on. In other words the effect of inputting a sequence of symbols is to apply the corresponding sequence of maps to the channel state. Since a digital message is a random sequence of symbols, the dynamics of the channel is given by random composition of the maps $\{\mathbf{w}_k\}_{k=1}^K$.

The maps $\{\mathbf{w}_k\}_{k=1}^K$ are determined by the physical properties of the channel as embodied, for example, by the differential equation that describes the channel's time evolution, so the properties of the channel—for instance, whether it is saturating or whether it is stable—will be reflected in the maps. This is important because it may be possible to use such properties of the maps to prove things about the system. One of the most useful properties is that of contraction (we saw that the maps in the above examples have this property). We can regard contractivity of the maps as relating to the stability of the channel in the following sense: if $\mathbf{w}_k$ is a contraction then the effect of repeatedly inputting the k-th symbol many times is that the channel converges to a fixed state that (in the limit) does not depend on its initial state. If all the maps are contractions then the state always remains in some bounded region of the state space, whatever the input sequence.

We noted above that the assumption that the channel is described by differential equations requires the maps $\{\mathbf{w}_k\}_{k=1}^K$ to be diffeomorphisms—in particular they are invertible. This invertibility reflects the fact that memory of the initial condition is never completely lost. However in some physical channels it may be that, to a good approximation, some (or all) information about the initial condition of the channel is lost in finite time. This would happen, for example, if the contraction of one of the maps were so great that the images of two different states were indistinguishable. (The decay rate of excitations, or at any rate certain kinds of excitations, would then be very short compared with the symbol period.) Thus under certain circumstances it may be advantageous to consider maps which are not invertible. In fact, communications engineers commonly model channels as finite impulse response (FIR) filters, and as we shall see below (§7) this corresponds to using non-invertible $\mathbf{w}_k$ maps.

We have—so far—referred only to the states of the channel. However, $\mathcal{M}$ will generally be multidimensional (as in the case of example 2.1 where two variables, u and p, are required to specify the state) and so we must consider what is meant by the output of the channel in this kind of model. We say that the output depends on the current state of the channel, that is, that there exists a function $v : \mathcal{M} \to \mathbb{R}$ such that if the state of the channel at a given time is $x \in \mathcal{M}$ then the corresponding output is $v(x)$. Later we shall need to assume that v is a smooth function.

This completes our (abstract) picture of the digital communication channel. The output of the channel is a sequence of real-valued measurements made on a random dynamical system which consists of a state space $\mathcal{M}$, and a finite collection of maps $\mathbf{w}_k : \mathcal{M} \to \mathcal{M}$, one for each symbol. At each time step (symbol period) one of these maps is chosen at random and applied to the current channel state to generate the next state. The appropriate mathematical structure for describing this situation is the *iterated function system* (IFS), which is described in the next section.

The abstract picture sketched above was motivated by consideration of physical channels governed by differential equations and driven by a discrete symbol set. In some situations the channel might be more appropriately described as a fully discrete system: for example, the channel may be represented as a hidden Markov model, the state space being a finite set with probabilistic dynamics. Though the motivation is then not applicable, with suitable interpretations of $\mathcal{M}$ and the $\mathbf{w}_k$ the IFS picture still applies, though such an extension is beyond the scope of this paper.

3. Some Theory of IFSs

Formally, an iterated function system consists of a state space, and a collection of maps of this space. In each time step the state evolves under the action of a map chosen at random from the collection. In our case we shall assume that the state space $\mathcal{M}$ is equipped with a metric and that with respect to this it is complete. We assume further that the maps $\{\mathbf{w}_k\}_{k=1}^{K}$ are contractions on $\mathcal{M}$. This gives us a special kind of IFS—a *hyperbolic* IFS—about which much is known (Barnsley 1988). In particular, it can be shown that there exists a unique compact subset of $\mathcal{M}$, let us say $\mathcal{A}$, which satisfies the following relationship:

$$\mathcal{A} = \bigcup_{k=1}^{K} \mathbf{w}_k(\mathcal{A}) \tag{3.1}$$

This set is attracting in the sense that any starting state in $\mathcal{M}$ approaches $\mathcal{A}$ as time goes on, so after a transient period the dynamics of the system become confined to $\mathcal{A}$.

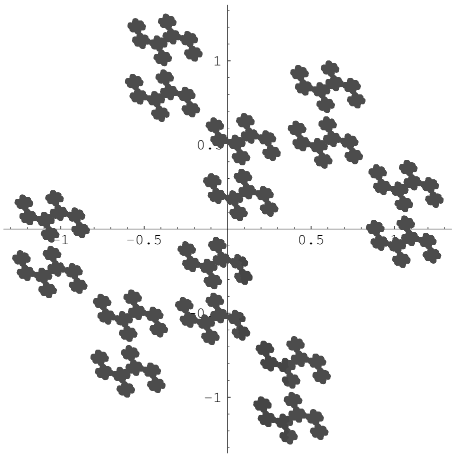

Figure 1. The attractor $\mathcal{A}$ for the system of example 2.1 with parameters $\tau = 1.0$, $\gamma = 1.0$ and $\omega_0 = \pi/3$.

An interpretation of equation (3.1) is that $\mathcal{A}$ is made up of K contracted images of itself, one corresponding to each input symbol. For any state in $\mathcal{A}$, if the symbol k is input the resulting next state is in $\mathbf{w}_k(\mathcal{A})$. Thus the set $\mathbf{w}_k(\mathcal{A})$ consists of all the states that the system can be in given that the last input was symbol k. This is not the same as saying that $\mathbf{w}_k(\mathcal{A})$ is the set of states for which the last input *was* symbol k since some states may be in more than one of the $\mathbf{w}_k(\mathcal{A})$ sets. However, if $\mathbf{w}_k(\mathcal{A}) \cap \mathbf{w}_{k'}(\mathcal{A}) = \emptyset$ for all $k \neq k'$, then no state can be in more than one such set and so the states uniquely identify the last input symbol. Such a situation is obviously convenient if the object is channel equalization (that is, the reconstruction of the input sequence from the sequence of outputs). IFSs which have this property are called *non-overlapping*; the attractor of a non-overlapping IFS is totally disconnected.

Figure 1 shows the attractor $\mathcal{A}$ for the system of example 2.1. Equation (3.1) indicates that the attractor is self-similar, and it is often a complicated fractal set like the one in the figure. It is easy to view the set in the figure as the union of two contracted copies of itself, as in equation (3.1): the part of the set above the line $p = -u$ is the image of $\mathcal{A}$ under $\mathbf{w}_+$, the part below is the image under $\mathbf{w}_-$. Clearly, the parameter values chosen here are such that the IFS is non-overlapping. Thus it is possible to determine the last symbol input from the state of the system, by noting whether this state is in $\mathbf{w}_+(\mathcal{A})$ or $\mathbf{w}_-(\mathcal{A})$. The line $p = -u$ can be used as a decision boundary: points above this line correspond to the last input being s_+, those below to the last symbol being s_-.

Whether or not an IFS is non-overlapping depends on the contractivity of the maps $\{\mathbf{w}_k\}$. This in turn depends on the physical properties of the channel—in particular the rate at which transients decay—and the time between inputs. For a given channel, the faster that symbols are input the weaker the contraction, and so there will be a symbol input rate above which the IFS will be overlapping. (In example 2.1, assuming that $2\omega_0 > \gamma$ (*i.e.* the underdamped case), the contraction factor is $e^{-\frac{\gamma\tau}{2}}$. For sufficiently short symbol periods τ the contraction will be small (the exponential factor will be close to 1) and the IFS will fail to be non-overlapping.

However, the larger the damping factor γ the smaller τ can be before overlapping occurs. In fact it is sufficient (though not always necessary) that $e^{-\frac{\gamma\tau}{2}}$ be less than $\frac{1}{2}$ for this IFS to be non-overlapping.)

The recursive structure of equation (3.1) allows more refined decompositions such as:

$$\mathcal{A} = \bigcup_{k,k'=1}^{K} \mathbf{w}_k \circ \mathbf{w}_{k'}(\mathcal{A}) \tag{3.2}$$

that is, a decomposition of $\mathcal{A}$ obtained by applying all possible pairs of $\mathbf{w}_k$ to $\mathcal{A}$. These sets can be labelled with the pairs k, k' corresponding to the last 2 symbols input to the channel. For a hyperbolic IFS this process can be refined to the limit where each state in $\mathcal{A}$ can be labelled by an infinite string of k's. (This process is known, for obvious reasons, as *backward iteration* (Diaconis & Freedman 1999).) Such a string is called an *address* of the state. If the IFS is non-overlapping each state in $\mathcal{A}$ has a unique address.

The discussion above suggests that each state retains information about the infinite history of inputs that gave rise to it. However, even if it were possible to observe the states directly, we can only do so with a limited accuracy and can therefore only obtain a limited amount of information about the history of inputs.

4. Linear Input-Output Systems as IFSs

The majority of previous work on the modelling of communications channels has concentrated on linear models. In this section and the next we shall describe how such linear models—for the case of *digital* signalling—can be considered as iterated function systems of the kind described above. The linear models are special cases of the IFS models because the $\mathbf{w}_k$ maps they give rise to are always affine, but we shall see that some of the important questions that arise when we seek to use IFS models for signal processing applications are already present in the linear case. In particular we will need to investigate how to exploit the models when only the outputs (not the states) are known: this is done for the linear case in the next section.

The channel models currently in most common use fall into the class of linear input-output models (Kailath *et al.* 2000). These models have the form

$$\sum_{j=0}^{p} \alpha_j y_{n+1-j} = \sum_{j=0}^{q-1} \beta_j u_{n-j} \tag{4.1}$$

where we can take α_0 to be unity, and α_p and β_{q-1} are non-zero. To find the relationship between these and the IFS models discussed above we convert to a state space form in the usual way (Kailath *et al.* 2000): we let r be the larger of p and q and set the upper limits of the sums in equation (4.1) to r on the left hand side, and $r-1$ on the right, adding terms with zero coefficients to the left or right as necessary. This model can be written in the form of a system of linear equations

$$x_{n+1} = Ax_n + Bu_n \tag{4.2}$$

$$y_{n+1} = Cx_{n+1} \tag{4.3}$$

where

$$A = \begin{bmatrix} -\alpha_1 & -\alpha_2 & -\alpha_3 & \dots & -\alpha_{r-1} & -\alpha_r \\ 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & 0 \end{bmatrix} \tag{4.4}$$

$B = (1, 0, 0, \dots, 0)^T$ and $C = (\beta_0, \beta_1, \beta_2, \dots, \beta_{r-1})$. ($A$ is $r \times r$, B is $r \times 1$ and C is $1 \times r$)†. For the case of digital inputs we think of the u_n's as all being drawn from a finite set, corresponding to the finite set of possible input symbols; say s_1, s_2, …, s_K are the possible input values. Thus equation (4.2) takes the form

$$x_{n+1} = Ax_n + Bs_{k_n} \tag{4.5}$$

In this case the model (4.2,4.3) can be regarded as an IFS model, with x_n as the state at time n, and $\mathbb{R}^r$ (with the Euclidean metric) as the state space. The maps of the IFS are the functions $\mathbf{w}_k(x) = Ax + Bs_k$, one for each input symbol as usual, and equation (4.5) becomes $x_{n+1} = \mathbf{w}_{k_n}(x_n)$, so that at every time step we find the new state by applying one of the maps $\mathbf{w}_k$. Recall from §2 that the output of the IFS channel is given by a function on the state space; equation (4.3) specifies this function for the linear model.

As noted above the $\mathbf{w}_k$ maps for the linear model have the form $\mathbf{w}_k(x) = Ax + Bs_k$. The attractor of the IFS consisting of these maps will be a (often fractal) set in the state space $\mathbb{R}^r$. In the case where $p < q$, $\alpha_r = 0$ and A is rank deficient: in fact it is clear from (4.4) that A has rank $r - 1$ ($= q - 1$). The range of A is thus an $r - 1$ dimensional subspace (say V_1), and the range of $\mathbf{w}_k$ is the $r - 1$ dimensional hyperplane produced by translating V_1 by $\mathbf{b}_k = Bs_k$. Hence for any $x \in \mathbb{R}^r$, $\mathbf{w}_k(x)$ lies in one of K parallel hyperplanes. (These hyperplanes are distinct since B does not lie in the range of A.) From (3.1) we see that the attractor in this case consists of K pieces, each a translation of the others, each piece lying in one of the hyperplanes. The hyperplane in which a given state lies uniquely identifies the latest input symbol.

If α_{r-1} is also zero we can say more about the attractor. Note that $\mathbf{w}_k \circ \mathbf{w}_{k'}(x) = A^2 x + ABs_k + Bs_{k'} = A^2 x + A\mathbf{b}_k + \mathbf{b}_{k'}$. Note also that A^2 has rank $r - 2$, so that the range of A^2 is an $r - 2$ dimensional subspace, V_2 say, which is a subspace of V_1: the K translates of V_2 produced by adding the vectors $A\mathbf{b}_k$ also all lie in V_1. The $r - 1$ dimensional hyperplanes parallel to V_1 (produced by adding the vectors $\mathbf{b}_{k'}$) thus each contain a copy of these K translates of V_2. In the last paragraph we found that the attractor consisted of K parallel pieces; equation (3.2) indicates that each of these pieces itself consists of K pieces, each of which is a translate of the others. Thus the attractor consists of K^2 pieces, all identical apart from translation.

By repeating this argument for $\alpha_{r-2} = 0$, …, $\alpha_{p+1} = 0$ we see that the attractor consists of K^{r-p} pieces, identical apart from translation, each lying in an p dimensional hyperplane of $\mathbb{R}^r$. The hyperplanes have a hierarchical organization: there are K hyperplanes, of dimension $r - 1$, each of which contains K hyperplanes of dimension $r - 2$, and so on down to the p dimensional hyperplanes containing

† We assume that (4.2,4.3) is observable; this question is discussed further in the next section.

the pieces of the attractor. Within the p dimensional hyperplanes the attractor may be fractal, and connected or disconnected. Note that identifying which $r - i$ dimensional hyperplane contains the current state uniquely identifies the i latest input symbols.

In applications, p is often in fact taken to be zero (in which case all the elements in the top row of A are zero); this corresponds to modelling the channel as an FIR filter (Clark 1985). In this case the attractor consists of only finitely many (in fact K^r) points, namely all the points of the form $(s_{k_1}, s_{k_2}, \ldots, s_{k_r})^T$. Whatever the initial state of the channel this attractor is reached in finite time (in r symbol periods), and the current state on the attractor obviously identifies the r latest input symbols.

5. Delay Embedding of Linear Systems

We have seen that the linear input-output system (4.1) falls into the class of IFS models. The main differences between (4.1) and the more general iterated function systems we want to consider are of course that the state space of the IFS can be a more general space than $\mathbb{R}^n$ (it need not even be a vector space), and the maps $\mathbf{w}_k$ need not be of the simple affine form $\mathbf{w}_k(x) = Ax + Bs_k$. Before we move on to these nonlinear systems, however, it will be helpful to consider certain systems which, while more general than (4.1), are still linear. The main question to be addressed here is how to derive information about the system when the only knowledge available is the output sequence: that is, the sequence of states is not known. This problem is a standard one in the theory of linear systems; here we shall set it in the context of iterated function systems to make clear its connections to the corresponding problem for the nonlinear case, which is treated in the next section.

Suppose that the digital channel has the form of a general discrete time, linear, time invariant system:

$$x_{n+1} = Ax_n + Bu_n \tag{5.1}$$

$$y_{n+1} = Cx_{n+1} \tag{5.2}$$

These equations have the same form as (4.2) and (4.3), but now the input and output can be in principle be vectors: $u_n \in \mathbb{R}^m$ and $y_n \in \mathbb{R}^p$; $x_n \in \mathbb{R}^r$ is the state vector; and A, B and C are appropriately dimensioned but otherwise arbitrary matrices. For the case of digital inputs we again think of the u_n's as all being drawn from a finite set, corresponding to the finite set of possible input symbols; say s_1, s_2, …, s_K are the possible input vectors. Thus equation (5.1) again takes the form

$$x_{n+1} = Ax_n + Bs_{k_n} \tag{5.3}$$

and this model can be regarded as an IFS in just the same way as (4.2,4.3): $\mathbb{R}^r$ is the state space: the maps of the IFS are $\mathbf{w}_k(x) = Ax + Bs_k$ as before; and (5.2) defines the measurement function.

The usual assumption is that the only information available to us is the sequence of outputs $\{y_n\}$. What can we learn about the sequence of states and in particular the sequence of inputs from this information? To answer this question we shall assume that the output is a sequence of scalar values *i.e.* $p = 1$. (This is inessential:

the same general principles apply for multivariate observations.) We can approach the problem by trying to use sequences of consecutive output values to *represent* the state of the system. Thus we define a d-dimensional *delay vector* of observations

$$\mathbf{y}_n = (y_n, y_{n+1}, \ldots, y_{n+(d-1)})^T$$

and ask how this is related to the state of the system, x_n, and the inputs.

In terms of the maps $\{\mathbf{w}_k\}$ and the state x_n, the delay vector $\mathbf{y}_n$ can be written as follows

$$\mathbf{y}_n = (Cx_n, C\mathbf{w}_{k_n}(x_n), C\mathbf{w}_{k_{n+1}} \circ \mathbf{w}_{k_n}(x_n), \ldots, C\mathbf{w}_{k_{n+d-2}} \circ \ldots \circ \mathbf{w}_{k_n}(x_n))^T \quad (5.4)$$

This is a mapping from $\mathbb{R}^r$ to $\mathbb{R}^d$, parameterized by the $(d-1)$-tuple of input symbols $(k_n, k_{n+1}, \ldots, k_{n+d-2})$; there are K^{d-1} such maps. Note that the map relating the n-th state x_n to the n-th delay vector $\mathbf{y}_n$ depends not only on the n-th input symbol, but on the subsequent $d-2$ input symbols as well. We write this as $\mathbf{y}_n = \Phi_{k_n,\ldots,k_{n+d-2}} x_n$, with the function $\Phi_{k_n,\ldots,k_{n+d-2}} : \mathbb{R}^r \to \mathbb{R}^d$ defined by equation (5.4). To simplify the notation further we shall denote a general $(d-1)$-tuple of input symbols by Ω—as we have noted there are K^{d-1} possible $(d-1)$-tuples; we can call this set $\mathcal{K}$: thus Ω is an element of $\mathcal{K}$. We say $\Omega_n = (k_n, k_{n+1}, \ldots, k_{n+d-2})$, and then write $\mathbf{y}_n = \Phi_{\Omega_n} x_n$.

Using the expressions for the $\mathbf{w}_k$ maps in terms of A, B and s_k we can write Φ_{Ω_n} in terms of these quantities. Since the $\mathbf{w}_k$ maps are all affine, with common linear part, it turns out that Φ_{Ω_n} is affine, and its linear part is independent of the input symbols. In particular $\Phi_{\Omega_n} x = \Phi x + \Psi_{\Omega_n}$ where

$$\Phi = \begin{bmatrix} C \\ CA \\ CA^2 \\ \vdots \\ CA^{d-1} \end{bmatrix} \quad (5.5)$$

and

$$\Psi_{\Omega_n} = \begin{bmatrix} 0 \\ CBs_{k_n} \\ C(Bs_{k_{n+1}} + ABs_{k_n}) \\ \vdots \\ C(Bs_{k_{n+d-2}} + ABs_{k_{n+d-3}} + A^2Bs_{k_{n+d-4}} + \ldots + A^{d-2}Bs_{k_n}) \end{bmatrix} \quad (5.6)$$

The delay vectors therefore all lie on a finite collection of parallel affine subspaces of $\mathbb{R}^d$; there will be K^{d-1} of these, one for each offset vector Ψ_{Ω_n}. Ideally we would like these subspaces to be distinct: for this to be so we must have at least that $d \geq \text{rank } \Phi + 1$, so from now on we will assume that $d \geq r + 1 \geq \text{rank } \Phi + 1$. Even with this condition it is possible for two of the subspaces to be coincident: this happens if the difference between two of the Ψ_{Ω_n} vectors lies in the image of Φ; however such a situation would be non-generic and could be removed by small changes in A, B, C or s_k (see Appendix Appendix A). If the K^{d-1} subspaces are indeed all distinct it is possible to identify the n-th input signal just by noting in which affine subspace $\mathbf{y}_n$ lies. As each input symbol arrives the vector of delays moves to a new subspace, but although there are K^{d-1} subspaces, at any given time the delay vector can move to one of only K different subspaces.

We noted before that after transients have decayed the system state becomes confined to an attractor, $\mathcal{A}$—a compact subset of the state space. Each of the affine subspaces contains an image of the attractor, namely $\Phi_{\Omega_n}\mathcal{A}$, and the delay vectors correspondingly become confined to these copies of $\mathcal{A}$, (in fact these images are all identical apart from translation). Assuming that Φ is full rank, each of the images is equivalent to $\mathcal{A}$ under an invertible affine transformation (in particular there is a one-to-one correspondence between the states in $\mathcal{A}$ and the delay vectors in each image). Hence each image shares the topological properties and many of the geometrical properties of $\mathcal{A}$: it is connected (or disconnected) if $\mathcal{A}$ is; it is a fractal set if $\mathcal{A}$ is, and has the same dimension as $\mathcal{A}$; it can be partitioned and addressed in just the same way as $\mathcal{A}$.

If Φ is full rank, and the offsets Ψ_{Ω_n} are such that the affine subspaces are distinct, the delay vectors give a rather complete representation of the state space, and of the evolution of the states as symbols are input. How likely is it that these conditions will hold?

Conditions under which Φ will be full rank are well known: this matrix is commonly encountered in linear systems theory and is known as the *observability matrix* (Kaczorek 1992). Simple sufficient conditions are that A has distinct eigenvalues and that C is not a left eigenvector of A. These conditions are generically satisfied, and so we may generally assume that Φ will be full rank. Further it is shown in Appendix A that for generic choices of B, $\Omega \neq \Omega'$ implies that $\Psi_\Omega - \Psi_{\Omega'}$ is not in the range of Φ, and so the affine subspaces are distinct.

(a) *Example 2.1 revisited*

From equation (2.3) we see that the system of example 2.1 is an example of the type of linear system specified in equation (5.1). We can generate an output from this system by specifying a measurement C as in equation (5.2). In the following we use the parameter values as for figure 1 *i.e.* $\tau = 1.0$, $\gamma = 1.0$ and $\omega_0 = \pi/3$. In addition we take $C = (1, 0)$.

Figure 2 shows the result of using the method of delays with $d = 3$ on the output from this system. Since $K = 2$ we expect there to be 2^2 parallel affine subspaces each containing a copy of the attractor from figure 1. Figure 2 is plotted using a coordinate system which takes the common normal of the subspaces—that is the normal to the range of Φ—as the vertical axis; the four attractor images can be clearly distinguished. The four images in figure 2 are the images of $\mathcal{A}$ under the four possible delay maps $\Phi_{(+,+)}, \Phi_{(-,+)}, \Phi_{(+,-)}$ and $\Phi_{(-,-)}$, in this order working from the top. The ordering of the images as well as their actual positions in delay space are dependent upon the choice of C. In figure 3 we plot a single sheet from figure 2: the one shown contains the image $\Phi_{(+,+)}\mathcal{A}$. This is just a linear transformation of the attractor in figure 1. By observing which sheet a delay vector lies in, we can identify the last two input symbols and since the images have addresses in the same way as $\mathcal{A}$ itself—since the delay map is one-to-one—knowing where a delay vector

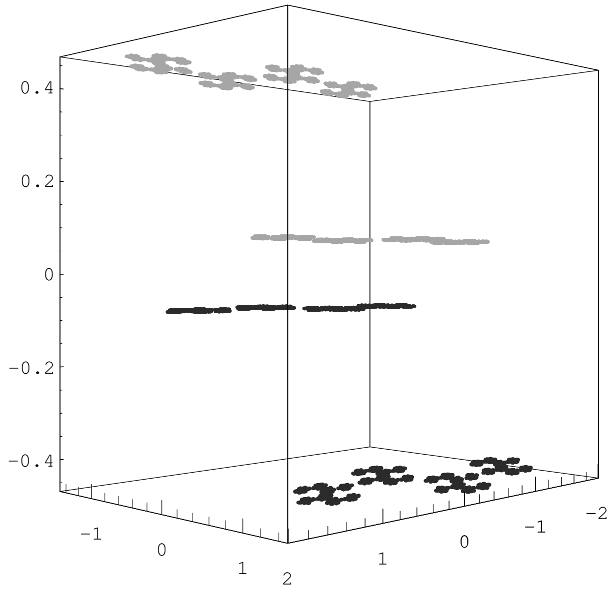

Figure 2. A delay plot ($d = 3$) of the attractor for the system of example 2.1 with parameters $\tau = 1.0$, $\gamma = 1.0$ and $\omega_0 = \pi/3$. The measurement function used was $C = (1, 0)$.

lies within a sheet in principle gives information about the previous history of input symbols.

In contrast, if we use $d = 2$ delays to study the output of this second order system, the result is ambiguous because the (two-dimensional) affine subspaces produced by the two delay maps Φ_+ and Φ_-, are necessarily coincident. This is illustrated in figure 4 which shows how the images $\Phi_+\mathcal{A}$ and $\Phi_-\mathcal{A}$ overlap. Clearly delay vectors found in the centre of the plot cannot be associated unambiguously with either input.

6. Delay Embedding of Nonlinear Systems

Recall from §2 our general picture of a digital channel: the state of the channel is an element of a state space $\mathcal{M}$ (usually a manifold), and at each time step (input symbol period) a map $\mathbf{w}_k$ is chosen from a finite collection of K maps, corresponding to the K possible input symbols, and applied to the current state to generate a new state: thus we have $x_{n+1} = \mathbf{w}_{k_n}(x_n)$, where k_n labels the symbol input at time n. A new output value is also generated from x_{n+1} using the measurement function v: thus we have $y_n = v(x_n)$. Just as in the linear case discussed above we can define d-dimensional delay vectors by

$$\mathbf{y}_n = \left(y_n, y_{n+1}, \ldots, y_{n+(d-1)}\right)^T$$

The delay vector $\mathbf{y}_n$ can be written in terms of the state x_n as follows

$$\mathbf{y}_n = (v(x_n), v\circ\mathbf{w}_{k_n}(x_n), v\circ\mathbf{w}_{k_{n+1}}\circ\mathbf{w}_{k_n}(x_n), \ldots, v\circ\mathbf{w}_{k_{n+d-2}}\circ\ldots\circ\mathbf{w}_{k_n}(x_n))^T \quad (6.1)$$

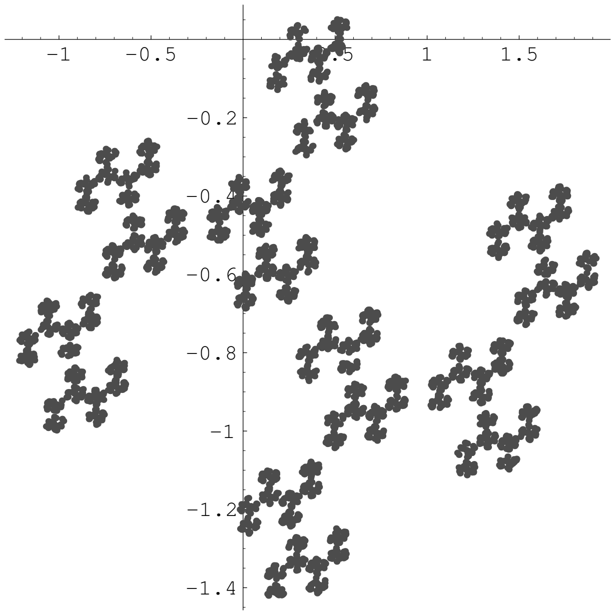

Figure 3. The image $\Phi_{(+,+)}\mathcal{A}$ of the attractor for the system of example 2.1 with parameters $\tau = 1.0$, $\gamma = 1.0$ and $\omega_0 = \pi/3$. This corresponds to the topmost sheet shown in figure 2.

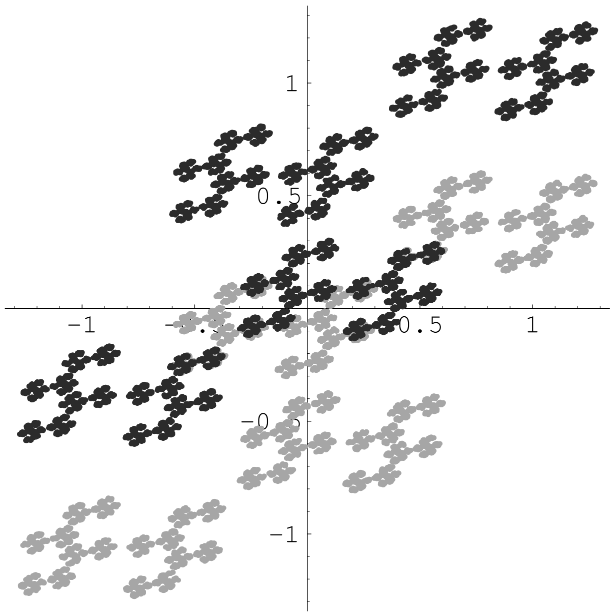

Figure 4. A delay plot ($d = 2$) of the attractor for the system of example 2.1 with parameters $\tau = 1.0$, $\gamma = 1.0$ and $\omega_0 = \pi/3$. The measurement function used was $C = (1, 0)$.

(cf. equation (5.4)). Again we can write this as $y_n = \Phi_{\Omega_n}(x_n)$ where $\Phi_{\Omega_n} : \mathcal{M} \to \mathbb{R}^d$ is the *delay map* defined in (6.1), and $\Omega_n = (k_n, k_{n+1}, \ldots, k_{n+d-2})$ is a vector of input symbols as before. We found in the linear case that, so long as d is large enough, we could expect the delay maps to reproduce the state space faithfully in the sense that Φ_{Ω_n} was injective, (and in fact affine). Thus each delay map produced a copy of the attractor, and the delay vector $\mathbf{y}_n$ moved around on these copies according the sequence of input symbols. It can be shown that analogous results hold in the nonlinear case as well. The delay maps will not now be affine, but if $\mathcal{M}$ is a compact differentiable manifold, on which the $\mathbf{w}_k$ maps are diffeomorphisms, and v is smooth then the delay maps will (generically) be *embeddings*, that is, smooth maps from $\mathcal{M}$ to $\mathbb{R}^d$ that are diffeomorphisms onto their images. (Again, d will need to be large enough for this to be true: in the nonlinear case 'large enough' means $d \geq 2m + 1$, where m is the dimension of $\mathcal{M}$.) Thus for each Ω, $\Phi_\Omega\mathcal{M}$ will be an m-dimensional submanifold of $\mathbb{R}^d$, and the corresponding image of the attractor $\Phi_\Omega\mathcal{A}$ shares the topological and many of the geometric properties of $\mathcal{A}$: it is a fractal set if $\mathcal{A}$ is, with the same dimension; it can be partitioned and addressed in the same way as $\mathcal{A}$, and so on. Taking $\mathcal{M}$ to be a manifold is consistent with our assumption that the channel is to be modelled by a set of differential equations. The $\mathbf{w}_k$'s, which are derived from the flow produced by the differential equations, will be diffeomorphisms, as we saw in §2.

The proof of these assertions for the nonlinear case is somewhat technical: details can be found in Stark *et al.* 2003, where the following theorem is proved.

Theorem 6.1 (Takens' theorem for IFSs). *Let $\mathcal{M}$ be a compact manifold of dimension $m \geq 1$ and say $d \geq 2m + 1$, $r \geq 1$; let $C^r(\mathcal{M}, \mathbb{R})$ be the space of C^r real-valued functions on $\mathcal{M}$ (the 'measurement functions'), and $D^r(\mathcal{M})$ the space of C^r diffeomorphisms of $\mathcal{M}$. Let $\mathcal{S} = \{1, 2, \ldots, K\}$ (the 'alphabet of symbols') and $\mathcal{K} = \mathcal{S}^{d-1}$. Also let $\mathcal{F} = [D^r(\mathcal{M})]^K$. For every (c, v) in an open and dense set of $\mathcal{F} \times C^r(\mathcal{M}, \mathbb{R})$ the 'delay map' $\Phi_{(c,v,\Omega)}$ is an embedding for every $\Omega \in \mathcal{K}$, where $\Phi_{(c,v,\Omega)}$ is defined by*

$$\Phi_{(c,v,\Omega)}(x) = (v(x), v \circ \mathbf{w}_{k_1}(x), v \circ \mathbf{w}_{k_2} \circ \mathbf{w}_{k_1}(x), \ldots, v \circ \mathbf{w}_{k_{d-1}} \circ \ldots \circ \mathbf{w}_{k_1}(x))^T$$

where $\Omega = (k_1, k_2, \ldots, k_{d-1})$ and $c = (\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_K)$.

In the discussion of the linear system in the previous section we concluded not only that the individual delay maps Φ_Ω each give a faithful copy (*i.e.* an embedding) of the state space, but also that the copies arising from *different* delay maps usually do not intersect (that is, $\Phi_\Omega\mathcal{M}$ and $\Phi_{\Omega'}\mathcal{M}$ are generically disjoint if $\Omega \neq \Omega'$): this is because the images $\Phi_\Omega\mathcal{M}$ and $\Phi_{\Omega'}\mathcal{M}$ form parallel affine subspaces of $\mathbb{R}^d$. The theorem just quoted provides the analogue for the nonlinear case of the delay maps giving faithful copies of the state space $\mathcal{M}$, but nothing has been said so far about the intersection of $\Phi_\Omega\mathcal{M}$ and $\Phi_{\Omega'}\mathcal{M}$ for different delay maps. Of course the delay maps are now nonlinear so we cannot expect their images to be anything like parallel. Nor can we use the fact that two m-dimensional submanifolds of $\mathbb{R}^{2m+1}$ will generically have empty intersection, because the images are not *arbitrary* submanifolds of $\mathbb{R}^d$: they must be of the form $\Phi_\Omega\mathcal{M}$ for an allowed delay map. It turns out that there are cases where the images of two different delay maps intersect and that the intersection persists under small changes in both the diffeomorphisms $\{\mathbf{w}_k\}_{k=1}^K$ and the measurement function v. In particular, suppose there is $x \in \mathcal{M}$, and two diffeomorphisms $\mathbf{w}_1$ and $\mathbf{w}_2$ such that $\mathbf{w}_1(x) = \mathbf{w}_2(x)$; thus $\mathbf{w}_1\mathcal{M}$ intersects $\mathbf{w}_2\mathcal{M}$ at $\mathbf{w}_1(x)$, and if this intersection is transversal it cannot be eliminated by small changes in $\mathbf{w}_1$ and $\mathbf{w}_2$. Now let $\Omega = (1, k_2, k_3, \ldots, k_{d-1})$ and $\Omega' = (2, k_2, k_3, \ldots, k_{d-1})$, then it is clear that $\Phi_\Omega(x) = \Phi_{\Omega'}(x)$ whatever the measurement function. Thus the images $\Phi_\Omega\mathcal{M}$ and $\Phi_{\Omega'}\mathcal{M}$ have a point of intersection, and this point cannot be eliminated by small changes to the $\mathbf{w}_k$ maps.

If two different $\mathbf{w}_k$'s map a state x to the same image this means that the channel can find itself in a state in which its response to two different input symbols is the same. This situation is clearly undesirable in a communications system, so one expects this possibility to have been designed out of any practical system. In fact there is a situation in which we can be sure that this problem will not arise. We take the state space to be $\mathbb{R}^n$, and assume, as in §2, that the evolution of the state while the k-th symbol is input is governed by the differential equations $\dot{x} = X_k(x,t)$. As noted before (see equation (2.1)), the right hand side can take the form of a time independent vector field (describing the state evolution when there is no input) plus a forcing term depending on the input symbol:

$$X_k(x,t) = X(x) + \chi_k(t)$$

The forcing corresponding to each symbol may last only a short fraction of the symbol period. If the pulses are sufficiently sharp and strong that they can be treated impulsively, so that the forcing is effectively a delta function occurring at the start of the symbol period, we have $X_k(x,t) = X(x) + \alpha_k\delta(t)$, where α_k characterises the k-th symbol. (Example 2.2 is of this kind.) Integrating this we find $w_k(x) = \phi_\tau(x + \alpha_k)$, where ϕ_τ is the time τ map of the unforced vector field X. Since α_k is assumed different to $\alpha_{k'}$ for $k \neq k'$, and ϕ_τ is a diffeomorphism, we see that, for any x, $\mathbf{w}_k(x) \neq \mathbf{w}_{k'}(x)$ for $k \neq k'$. So in this case the problem of persistent intersection of delay map images will not occur. We note that the practice of modelling data transmission as the driving of the channel by a sequence of delta functions is a common one in conventional signal processing: see for example Bissell & Chapman 1992.

(a) *Delay Embedding for Example 2.2*

Example 2.2 provides a simple model of a nonlinear channel. The extent to which the nonlinearity affects the behaviour of the channel depends upon the amplitude of the input χ. For small enough χ (at a given τ) the amplitude u will remain small and the equation will be effectively linear. For illustrative purposes we take χ to be moderately large, by which we mean that if the system is close to zero when a pulse arrives the resulting displaced state experiences a restoring force with similar linear and nonlinear contributions. The attractor for this case is shown in figure 5. As with example 2.1 we find that this attractor appears to be totally disconnected; in particular the line $p = u$ may be used as a boundary which separates the two components $\mathbf{w}_+\mathcal{A}$ and $\mathbf{w}_-\mathcal{A}$.

To apply the method of delays we need to specify a measurement function. In the figures below we use a linear measurement $v\colon \mathbb{R}^2 \to \mathbb{R}$ where $v(u,p) = \cos(\frac{\pi}{32})u - \sin(\frac{\pi}{32})p$. Figure 6 shows the delay plot of this output using $d = 3$. It is

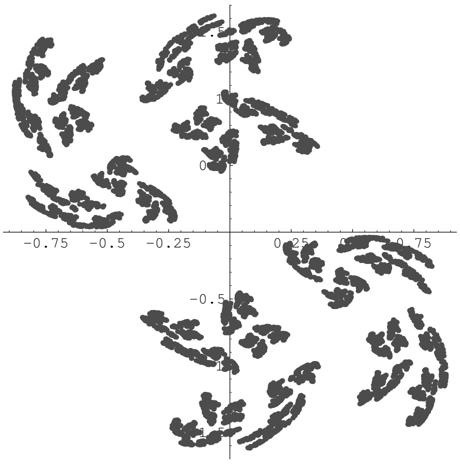

Figure 5. The attractor for the system of example 2.2 with parameters $\tau = 1.0$, $\gamma = 1.0$, $\omega_0 = \pi/3$ and $\chi = 1$.

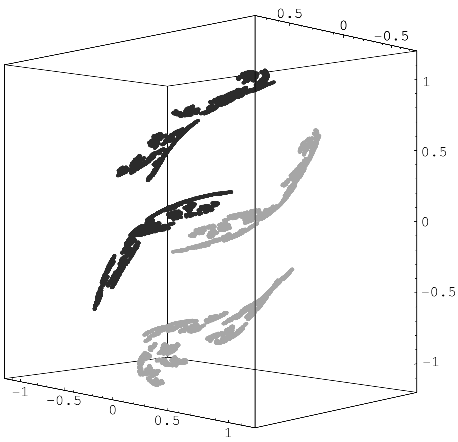

Figure 6. A delay plot ($d = 3$) of the attractor for the system of example 2.2 with parameters $\tau = 1.0$, $\gamma = 1.0$, $\omega_0 = \pi/3$ and $\chi = 1$. The measurement function used was $v(u, p) = \cos(\frac{\pi}{32})u - \sin(\frac{\pi}{32})p$.

plotted using a similar coordinate system to the one used in figure 2, here based on the linear system that approximates the nonlinear system when the state remains close to the origin. As expected there are four copies of the attractor, though they clearly no longer lie in affine subspaces. It is clear from the figure that the four images of the attractor are disjoint and so, as with the linear case, we can decide the latest symbol input to the channel by noting the image in which the delay vector currently lies. We could in principle use this for channel equalization.

In fact three delays are not enough to ensure that the images will be embeddings of the attractor; the theorem above would require us to use $d = 5$ to guarantee this, although using fewer may result in the delay maps being embeddings if, as here, the nonlinearity is not too great. Figure 7 shows a similar plot using a larger value

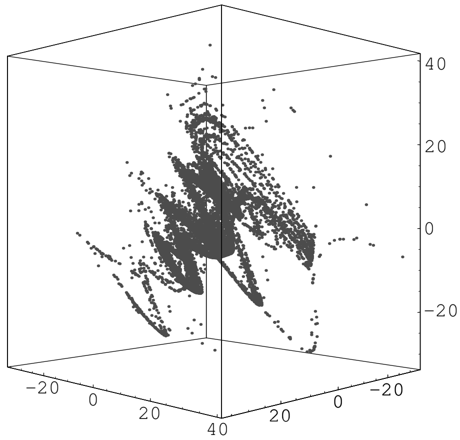

Figure 7. A delay plot ($d = 3$) of the attractor for the system of example 2.2 with parameters $\tau = 1.0$, $\gamma = 1.0$, $\omega_0 = \pi/3$ and $\chi = 10$. The measurement function used was $v(u,p) = \cos(\frac{\pi}{32})u - \sin(\frac{\pi}{32})p$.

of χ so that the nonlinear terms are much more significant. This figure shows the image of only one of the delay maps rather than the four shown in the previous figure. In this case it is not at all clear that the data lies on a submanifold of the delay space. It may be that more delays are needed to achieve this.

7. Implications for Signal Processing

The essential features of the IFS model of a digital channel are: the state space $\mathcal{M}$, (which need not be a vector space); the maps $\mathbf{w}_k$ of the state space to itself, (which again need not be linear), one for each symbol in the alphabet; and the attractor $\mathcal{A}$ that these maps generate (and which is specified in terms of the maps by equation (3.1)). When it comes to using the model in signal processing applications a further layer is added: measurements are made on the channel states and are described by the function v, and the relationships between these measurements and the underlying state space need to be examined. We have seen in the preceding sections that by constructing delay vectors of outputs various aspects of the state space are reproduced in delay space $\mathbb{R}^d$; in particular there are several (K^{d-1}) copies of $\mathcal{M}$ embedded in $\mathbb{R}^d$, each copy labelled by a $d-1$-tuple of input symbols. Each copy is in one-to-one correspondence with $\mathcal{M}$ and each copy obviously contains a copy of the attractor $\mathcal{A}$. The delay space is also equipped with analogues of the maps $\mathbf{w}_k$, though as with the state space itself the version in $\mathbb{R}^d$ is somewhat more complicated than the original collection $\{\mathbf{w}_k\}_{k=1}^K$: each copy of $\mathcal{M}$ has K maps defined on it, the image of each map being one of the copies of $\mathcal{M}$ (note that these images are all different).

The picture provided by the model is thus a geometrical one. Of course, defining the model is only part of the story: we also need to know how to use the model in the processing of digital signals passed through our channel. We anticipate that this will need considerable development: here we restrict ourselves to some broad comments, concentrating on equalization (which for our purposes means the identification of the input sequence from the channel outputs). We have remarked several times above that the location of the current delay vector gives information about the latest input symbols; indeed, the more information we have about this location the more we can infer (in principle) about the input sequence: identifying which copy of $\mathcal{M}$ the delay vector lies in specifies the latest $d-1$ symbols, and more precise information about where in the attractor the delay vector lies (*i.e.* its address) potentially gives information both about these symbols and earlier ones. The key to using the delay vectors actually to detect the input signals of course lies in knowing where in delay space the copies of $\mathcal{M}$ are situated. This information is not known *a priori*: it must be derived in some way from observations made on the channel. The information we seek to obtain from these observations may be more or less precise: at its most basic we could simply try to divide up the delay space into regions, each of which contains one of the copies; then identifying which region a given delay vector lies in would identify the copy (all delay vectors being assumed to lie in one copy or another). Indeed it may be enough for a region to contain several copies: say, all those K^{d-2} copies sharing a given symbol as the latest one. In fact, the use of a feedforward transversal filter to equalize the linear FIR channel can be viewed as a division of delay space into regions in just this way (see Gibson *et al.* 1991): in this case the space is partitioned by one or more parallel hyperplanes. Even for the linear FIR channel, however, the regions may not be separable by hyperplanes, and Gibson *et al.* 1991 suggest the use of nonlinear region boundaries implemented using (for example) multilayer perceptrons (Gibson *et al.* 1991) or radial basis functions (Chen *et al.* 1991).

There are various ways we could go about trying to locate the copies in delay space, depending on what information there is available. In the particularly simple case of an FIR channel, the attractor (which plays an even more dominant role than usual in this case since transients disappear in finite time) consists of a finite collection of points, and can be found directly from the channel output. If the channel is assumed to be linear (described by (5.1) and (5.2)) then we know that each of the copies is an r-dimensional affine subspace, and that all the copies are identical apart from translations. An efficient approach to locating the copies in that case would be to estimate the parameters of the channel (the matrices A, B and C of (5.1) and (5.2), or the coefficients in (4.1)) which then determine what happens in delay space through the maps Φ_Ω of §5. How we go about estimating the parameters will depend on whether or not a training sequence is to be used.

For a nonlinear channel the copies of state space are no longer r-dimensional affine subspaces: they are now r-dimensional submanifolds of delay space. The geometry of these submanifolds may be quite complicated, and they may intersect. Determining their positions is now a rather more challenging problem. Without going into too much detail we note that one way to view this is as a pattern classification problem, with the delay vectors as patterns and the copies as classes. In this spirit, collecting delay vectors from the channel output supplies us with a sample of data points from the copies of $\mathcal{M}$. If we know the corresponding inputs (if, say, we are using a training sequence) then we can deduce to which copy each data point belongs. We can attempt to delineate the copies by, for example, the use of clustering techniques, combined perhaps with the use of level set representations of the submanifolds, or local parameterizations (Kirby 2001). In the absence of knowledge of the inputs we shall not only need to determine the region of space in which each copy lies, but also its 'label': the $d-1$-tuple of symbols to which it corresponds. One way to approach this latter problem is to observe the sequence in which the copies are visited: if a delay vector in copy A is succeeded by one in copy B the label of B is related to that of A by shifting the symbols in A's label one place to the left (dropping the leftmost) and adding a new symbol at the right. (Note this means that if B is the same as A, all the symbols in A's label must be the same.) By observing the delay vectors for long enough we can attempt to devise a consistent set of labels. In fact the sequence of delay vectors contains much more subtle information: the structure of the attractor implied by equation (3.1) means that *where* in a copy of the attractor a delay vector lies carries information about the labels of the copies it has previously visited: thus the more information we can deduce about the structure of the attractor the more easily we can assign the labels.

A particular difference between linear and nonlinear channels concerns the possible intersection of the copies of the state space, in $\mathbb{R}^d$. It was noted in §4 that in the linear case such intersections are non-generic—they can only happen for special choices of the system parameters, and even then most perturbations (however small) of the parameters will produce systems without intersections. The simplicity of this situation results, of course, from the fact that the copies are necessarily parallel affine subspaces. In the nonlinear case the copies are not constrained in this way, and, as well as intersecting each other, can in principle have self-intersections. The 'Takens' Theorem for IFS's' quoted in §6 shows that, in fact, self-intersections are non-generic, but it says nothing about intersections between two different copies (that is, two images $\Phi_\Omega\mathcal{M}$ and $\Phi_{\Omega'}\mathcal{M}$ where $\Omega \neq \Omega'$). As we have seen such intersections can in fact be persistent under perturbations. These intersections clearly pose difficulties for equalization: there will now be delay vectors whose latest input symbols cannot be identified simply by determining which copy of $\mathcal{M}$ they lie in (since they lie in more than one); and (probably more significantly) there are likely to be problems in locating and distinguishing the copies in delay space using data. Whether or not a particular channel actually suffers from these problems depends on the maps $\mathbf{w}_k$ of the IFS. As described at the end of §6 there is a case in which it is clear that intersections will not occur: this is when the input symbols are introduced as sharp pulses.

The above is not intended to do more than hint at some of the problems and approaches that arise when we consider using the IFS model in applications (particularly channel equalization). Other questions also arise: how best should we use the output to estimate the dimensions r and d? How should we assess the extent to which the channel is in fact nonlinear, or non-recursive? How can we use the self-similar nature of the attractor to inform the model in delay space? We intend to develop algorithms based on the IFS model in subsequent work.

Appendix A. Non-intersection of Hyperplanes

Here we show that if Ω and Ω' are distinct elements of $\mathcal{K}$ then (generically) the vector $\Psi_\Omega - \Psi_{\Omega'}$ does not lie in the range of Φ. We can do this for the case $p=1$ and $m=1$ as follows (the argument for larger p and m is similar). We assume that Φ has $r+1$ rows (that is, we ignore any rows of Φ below the $r+1$-th: clearly if the proposition is true for $d=r+1$ it will be true for all larger values of d).

If Φ is full rank then the row vectors C, CA, …, CA^{r-1} are linearly independent (Kaczorek 1992). Hence there is a unique $r+1$-vector $\Lambda = (\lambda_1, \lambda_2, \ldots, \lambda_r, 1)$ such that $\Lambda\Phi = 0$. The condition that $\Psi_\Omega - \Psi_{\Omega'}$ lies outside the range of Φ is equivalent to the condition that the matrix $[\Phi : \Psi_\Omega - \Psi_{\Omega'}]$ is full rank (*i.e.* rank $r+1$). From equations (5.5) and (5.6) we see that this matrix has the form

$$\begin{bmatrix} C & 0 \\ CA & CB\delta_{k_n} \\ CA^2 & C(B\delta_{k_{n+1}} + AB\delta_{k_n}) \\ \vdots & \vdots \\ CA^r & C(B\delta_{k_{n+r-1}} + AB\delta_{k_{n+r-2}} + \ldots + A^{r-1}B\delta_{k_n}) \end{bmatrix}$$

where $\delta_{k_n} = s_{k_n} - s_{k'_n}$. It is clear that this matrix fails to have full rank if and only if $\Lambda(\Psi_\Omega - \Psi_{\Omega'}) = 0$. Using equation (5.6) this becomes

$$[\lambda_2, \lambda_3, \ldots, \lambda_r, 1] \begin{bmatrix} \delta_{k_n} & 0 & 0 & \cdots & 0 \\ \delta_{k_{n+1}} & \delta_{k_n} & 0 & \cdots & 0 \\ \delta_{k_{n+2}} & \delta_{k_{n+1}} & \delta_{k_n} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \delta_{k_{n+r-1}} & \delta_{k_{n+r-2}} & \delta_{k_{n+r-3}} & \cdots & \delta_{k_n} \end{bmatrix} \begin{bmatrix} C \\ CA \\ CA^2 \\ \vdots \\ CA^{r-1} \end{bmatrix} B = 0 \tag{A 1}$$

Writing the product of the first three matrices of the above equation as the $(1 \times r)$ matrix V this becomes $VB = 0$; if $V \neq 0$ this equation is clearly not satisfied for almost all choices of B. Hence it is sufficient for us to show that $V \neq 0$.

Assume to begin with that $\delta_{k_n} \neq 0$. Since the rows of the third matrix are linearly independent and the second matrix is full rank, there is no choice of $\lambda_2, \lambda_3, \ldots, \lambda_r$ for which V is zero. If, on the contrary, $\delta_{k_n} = 0$ we work with the reduced system

$$[\lambda_3, \ldots, \lambda_r, 1] \begin{bmatrix} \delta_{k_{n+1}} & 0 & \cdots & 0 \\ \delta_{k_{n+2}} & \delta_{k_{n+1}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \delta_{k_{n+r-1}} & \delta_{k_{n+r-2}} & \cdots & \delta_{k_{n+1}} \end{bmatrix} \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{r-2} \end{bmatrix} B = 0 \tag{A 2}$$

and note that if $\delta_{k_{n+1}} \neq 0$ the same argument applies. If $\delta_{k_{n+1}} = 0$, we continue the same process until we find the first $\delta_k \neq 0$ (since $\Omega \neq \Omega'$ there must always be at least one such δ_k).

References

Barnsley, M. 1988 *Fractals Everywhere*, San Diego: Academic Press.

Bissell, C. C. & Chapman, D. A. 1992 *Digital Signal Transmission*, Cambridge: Cambridge University Press.

Chen, S., Gibson, G. J., Cowan, C. F. N. & Grant, P. M. 1991 Reconstruction of binary signals using an adaptive radial-basis-function equalizer. *Signal Processing* **22**, 77–93.

Clark, A. P. 1985 *Equalizers for Digital Modems*, London: Pentech Press.

Constantin, P., Foias, C., Nicolenko, B. & Temam, R. 1989 *Integral Manifolds and Inertial Manifolds for Dissipative Partial Differential Equations.* Applied Mathematical Sciences **70**. Springer-Verlag.

Diaconis, P. & Freedman, D. 1999 Iterated random functions. *SIAM Review* **41**, 45–76.

Falconer, K. 1990 *Fractal Geometry: Mathematical foundations and applications*, New York: John Wiley and Sons.

Gibson, G. J., Siu, S. & Cowan, C. F. G. 1991 The application of nonlinear structures to the reconstruction of binary signals. *IEEE Trans. Sig. Proc.* **39**, 1877–1884.

Horn, R. A. & Johnson, C. R. 1991 *Topics in Matrix Analysis*, Cambridge: Cambridge University Press.

Kaczorek, T. 1992 *Linear Control Systems, Vol. 1: Analysis of Multivariable Systems*, Taunton: Research Studies Press.

Kailath, T., Sayed, A. H. & Hassabi, B. 2000 *Linear Estimation*, New Jersey: Prentice-Hall.

Kantz, H. & Schreiber, T. 1997 *Nonlinear Time Series Analysis*, Cambridge: Cambridge University Press.

Kigami, J. 2001 *Analysis on Fractals*, Cambridge Tracts in Mathematics **143**, Cambridge University Press.

Kirby, M. 2001 *Geometric Data Analysis*, New York: John Wiley and Sons.

Ott, E., Sauer, T. & Yorke, J. A. (eds) 1994 *Coping with Chaos*, New York: John Wiley and Sons.

Stark, J., Broomhead, D. S., Davies, M. E. & Huke, J. P. 2003 Delay embeddings for forced systems: II stochastic forcing. *J. Nonlinear Sci.* **13**, 519–577.

Temam, R. 1988 *Infinite Dimensional Dynamical Systems in Mechanics and Physics*, Applied Mathematical Sciences **68**. Springer-Verlag.

Thompson, J. M. T. & Stewart, H. B. 2002 *Nonlinear Dynamics and Chaos*, Chichester: John Wiley and Sons.