Nonlinear thoughts about linear signal processing

Broomhead, D. S. and Huke, J. P. and Muldoon, M. R. and Brown, A. G.

2002

MIMS EPrint: **2006.381**

ISSN 1749-9097

Nonlinear Thoughts about Linear Signal Processing

D.S. Broomhead, J.P. Huke, M.R. Muldoon
University of Manchester Institute of Science and Technology

and

A.G. Brown
Defence Evaluation Research Agency

Abstract

Recent work on modelling digital channels using iterated function systems suggests a general approach to the theory of signal processing in digital communications which uses so-called *delay methods* developed for deterministic nonlinear timeseries analysis. Here we make the connection between this work and the more conventional approach to digital communications by casting linear channel models as iterated function systems and showing how the use of delay methods gives a nice connection with the theory of observability in the control of linear systems.

1 Introduction

The culture of signal processing is steeped in the mathematics of linear systems. Linear models are used across the range of signal processing applications and can be very effective. In contrast, the mathematics of nonlinear processes has had comparatively little impact. Nonlinearity—it is often suggested—leads to a confusion of special cases with no unifying theoretical picture. This is an unfortunate state of affairs since it is easy to anticipate circumstances in which nonlinearities unavoidably—by accident or design—play a significant role in the transmission of information. Indeed, we might imagine that a better developed theory of nonlinear signal processing might encourage the design of novel systems capable of exploiting nonlinear phenomena.

The assumption of linearity permits the invocation of a well-developed mathematical framework and with it, the possibility of making generalisations about signals: that they can be added together, for example, and that filtering sums of signals is equivalent to summing the separately filtered signals. The disadvantage is, of course, that these generalities may not be relevant. We would like, therefore, to develop an approach to signal processing which has a general utility but which is based on less specialised assumptions 3; 4. Of course, *something* should be assumed, and in our approach—which focuses on digital signal processing—it is the discreteness of the alphabet of input symbols which is exploited. Digital technology imposes discrete structures on continuous natural processes; by understanding better the implications of this we hope to make progress.

The starting point of this work is the inclusion of a model of the signal source within the model of the digital channel. As a result, a class of mathematical objects known as *iterated function systems* (IFS) 7; 2 arises in a natural way. Moreover it is possible, for these systems, to adapt methods for the analysis of time series data which were first developed for deterministic, nonlinear dynamical systems 1; 13; 10; 9; 11. This marriage of the theory of IFS with so-called *delay methods* for time series analysis is at the heart of our approach to digital signal processing.

In this paper we focus on establishing a connection with the usual theory of linear digital channels. It would make a pleasing picture if our results were to take a sensible form when restricted to the linear case. We shall show that this is indeed the case and that linear digital channels fall naturally within the ambit of the IFS-based theory. Actually, it is a peculiarity of the mathematical arguments employed here that we shall need a special theorem to deal with the particular case of linear channels. This will be the main result of the paper. During the discussion it will emerge that an IFS model of a digital linear channel is not actually linear. We shall show, however, that there is an elegant theoretical foundation on which we can build quite simple—nonlinear—equalisation algorithms.

2 State Space Models of Linear Filters

In the following we make a strong appeal to geometry to provide a picture of the basic concepts. In this spirit it is helpful to cast linear channel models in the language of state space 6.

A linear, mth order, all-pole filter subject to a sequence of inputs $\{b_n\}$ can represented by the following system of non-homogeneous linear difference equations

$$\mathbf{x}_{n+1} = \mathbf{A}\mathbf{x}_n + \mathbf{b}_{n+1} \tag{2.1}$$

where the states, $\mathbf{x}_n \in \mathbb{R}^m$, of the filter can be thought of as sequences of m numbers specifying the contents of a tapped delay line, and the input vectors, $\mathbf{b}_n \in \mathbb{R}^m$, have the form $\mathbf{b}_n = (b_n, 0, \ldots, 0)^T$. The matrix $\mathbf{A}$—which specifies the filter—has the form of a companion matrix

$$\mathbf{A} = \begin{pmatrix} a_0 & a_1 & \ldots & a_{m-2} & a_{m-1} \\ 1 & 0 & \ldots & 0 & 0 \\ 0 & 1 & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & 1 & 0 \end{pmatrix}$$

where the usual filter coefficients are given by the top row, $\mathbf{a}^T$. The structure of $\mathbf{A}$ has the effect of shifting all the components of $\mathbf{x}_n$ down one place (the mth component is thereby lost) and replacing the first component with the linear combination $\mathbf{a}^T \cdot \mathbf{x}_n$.

The output of the filter is generated by making measurements corresponding to a linear function $v : \mathbb{R}^m \to \mathbb{R}$ of the state of the filter. This is equivalent to forming the scalar product of the state with a fixed vector $\mathbf{v}^T$, $v(\mathbf{x}) = \mathbf{v}^T \cdot \mathbf{x}$. The sequence of observations $\{v(\mathbf{x}_n)\}$ is then the output of a pole-zero filter given the input sequence $\{b_n\}$.

Let us now shift the viewpoint slightly. Our particular interest here is in digital channels and so we can assert that the possible values taken by the inputs, $\{b_n\}$, are drawn from a finite alphabet (containing, say, p symbols). A different interpretation of equation (2.1) is that the channel state evolves in one sampling interval under the action of one of p different maps $\mathbf{w}_b : \mathbb{R}^m \to \mathbb{R}^m$ defined as follows

$$\mathbf{w}_b(\mathbf{x}) = \mathbf{A}\mathbf{x} + \mathbf{b} \tag{2.2}$$

Assume, for simplicity, that the sequence of symbols input to the channel is an independent, identically distributed random process. (This is a reasonable initial assumption—efficiently coded data will appear random—but is not crucial to what follows.) The channel state $\mathbf{x}_n$ then evolves under a random iteration procedure according to which one of the p maps is selected at random at each time step and applied to the current state. Thus the nth state is obtained from the initial state by composition of a random sequence of maps:

$$\mathbf{x}_n = \mathbf{w}_{b_n} \circ \mathbf{w}_{b_{n-1}} \circ \cdots \circ \mathbf{w}_{b_1}(\mathbf{x}) \tag{2.3}$$

and the corresponding output is given by

$$v_n = \mathbf{v}^T \cdot \mathbf{w}_{b_n} \circ \mathbf{w}_{b_{n-1}} \circ \cdots \circ \mathbf{w}_{b_1}(\mathbf{x}) \tag{2.4}$$

From this point of view, the digital channel is seen as an IFS. There is now a considerable amount of interest among pure and applied mathematicians in this kind of dynamical system (see, for example, the recent review by Diaconis and Freedman 5). A few basic results will suffice here. We shall assume that in some suitable norm the maps of the IFS are contractions, so that for each map $\mathbf{w}_b$

$$\|\mathbf{w}_b(\mathbf{x}) - \mathbf{w}_b(\mathbf{y})\| < \|\mathbf{x} - \mathbf{y}\|$$

for all $\mathbf{x}$ and $\mathbf{y}$ in some closed bounded subset of $\mathbb{R}^m$. This assumption is essentially one about the stability of the channel. It implies, for instance, that if the channel is repeatedly subjected to the same input symbol, the output will converge to a constant value which is independent of the initial state of the channel. (There has been a lot of recent interest in obtaining results for IFSs under weaker conditions than strict contractivity, and it is possible that these results could be of relevance to the modelling of digital channels. This, however, is work for the future.)

For the present purposes we note that with the above assumptions the IFS has a unique attractor, $\mathcal{A}$, which is a compact invariant subset of the region of $\mathbb{R}^m$. Supported on this set is a unique ergodic probability measure. We note also that $\mathcal{A}$ satisfies the following equation

$$\mathcal{A} = \bigcup_b \mathbf{w}_b(\mathcal{A}) \tag{2.5}$$

That is to say, the attractor is the union of p sets, each of which is the image of the attractor itself under one of the mappings in the IFS. This result has (at least) two interesting consequences: the first is that $\mathcal{A}$ is often a fractal set since it is the union of contracted copies of itself, each of which is a union of contracted copies, and so on; the second—related—result is that every point in $\mathcal{A}$ has an "address". A way to see this is to think of the *backward iteration* of the IFS

$$\bar{\mathbf{x}}_n = \mathbf{w}_{b_1} \circ \mathbf{w}_{b_2} \circ \cdots \circ \mathbf{w}_{b_n}(\mathbf{x}_0) \tag{2.6}$$

(note the reverse ordering of the subscripts compared with (2.3)). Since the maps are contracting on a closed bounded subset of $\mathbb{R}^m$, this process generates—for each choice of symbol sequence $\{b_k \ : \ k = 1, 2, \ldots\}$—a convergent sequence $\{\bar{\mathbf{x}}_k \ : \ k = 1, 2, \ldots\}$ whose limit is a point in $\mathcal{A}$ which is independent of the choice of $\mathbf{x}_0$. For any given point $\mathbf{x} \in \mathcal{A}$, any symbol sequence giving a convergent sequence under backward iteration which has $\mathbf{x}$ as its limit can be regarded as an address of $\mathbf{x}$. Each point in $\mathcal{A}$ has at least one address 2. If the images $\mathbf{w}_b(\mathcal{A})$ are all disjoint then the address is unique and we say that $\mathcal{A}$ is *totally disconnected*. The implication of this for digital channels is that if we can at any time identify where we are on the attractor of the channel, then implicitly this gives information about the history of inputs to the channel. In the case that the channel has a totally disconnected attractor then there is a unique sequence of input symbols which produces a given channel state. Of course, it would be necessary to measure the channel state with infinite precision to get a complete history, but as we shall see less complete measurements nonetheless provide useful information.

2.1 an example

A simple example will be useful to illustrate the various stages of the development. The constraints imposed by the need to represent the results graphically limit this to a 2nd-order IFS model of a binary channel. Specifically, we use equation (2.2) with

$$\mathbf{A} = \begin{pmatrix} 0.8 & -0.5 \\ 1 & 0 \end{pmatrix} \tag{2.7}$$

and $b = \pm 1$. We take $v^T = (1, 0)$. Random iteration of this model gives the attractor shown in figure 1 where two scales of grey have been used to label the sets $\mathbf{w}_{+1}(\mathcal{A})$ and $\mathbf{w}_{-1}(\mathcal{A})$, that is, the parts of the attractor which correspond respectively to the last input being $+1$ and -1. We note that the attractor does

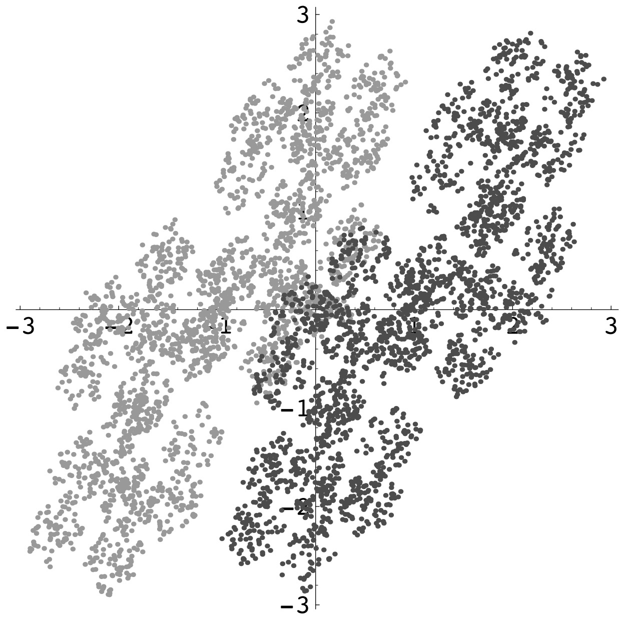

FIG. 1. Attractors of a 2nd order linear recursive channel—see equation (2.2)—with $\mathbf{A}$, defined in equation (2.7) and $b = \pm 1$. The darker points are in the set $\mathbf{w}_{+1}(\mathcal{A})$ and the lighter points are in $\mathbf{w}_{-1}(\mathcal{A})$.

not appear to be totally disconnected—there appears to be a region of overlap of the two differently shaded regions—and so we expect that points in the attractor will not be uniquely addressable. Despite this, the channel can be equalised by virtue of being an all-pole system. To see this we note simply that according to equation (2.1)

$$\mathbf{b}_{n+1} = \mathbf{x}_{n+1} - \mathbf{A}\mathbf{x}_n \tag{2.8}$$

that is, a suitable linear combination of two successive state vectors recovers the input to the channel.

We have assumed so far that the output of the channel is the first component of $\mathbf{x}$, or, in terms of the observation function introduced earlier, that $\mathbf{v}^T = (1, 0)$. A more general choice of $\mathbf{v}^T = (\cos\theta, \sin\theta)$ with $\theta \in (0, \pi)$ but not equal to $\pi/2$, will produce an output time series which is harder to invert since the linear inverse of an FIR filter is an IIR filter and hence requires an infinite history of the FIR output. Figure 2 shows a typical output time series obtained from the pole-zero filter defined by equation (2.7) with $\mathbf{v}^T = (\frac{1}{2}, \frac{\sqrt{3}}{2})$, driven by an equiprobable independent sequence of inputs with $b = \pm 1$.

FIG. 2. Typical time series of output from the pole-zero filter defined by equation (2.7) with $\mathbf{v}^T = (\frac{1}{2}, \frac{\sqrt{3}}{2})$ driven by an equiprobable independent sequence of inputs with $b = \pm 1$.

3 Delay embedding of linear channels

The second part of our development is the introduction of the method of delays (for a nice description see the book by Ott, Sauer and Yorke 8). The essentials of this are that there is a smooth dynamical system defined on a state space and a smooth measurement which is a real-valued function of the state. The basic results of this theory establish a link between the dynamical system and a construction based on time series data obtained by making successive measurements on the system. As an example, consider a dynamical system consisting of a linear map $\mathbf{A}$ which acts on a vector space $\mathbb{R}^m$. Starting with an arbitrary initial state $\mathbf{x}_0 \in \mathbb{R}^m$, repeated application of the map gives a sequence of new states, $\{\mathbf{x}_n = \mathbf{A}^n \mathbf{x}_0\}$, of the system. If at each time step we record only the projection of the state onto some fixed vector $\mathbf{v}^T$ we obtain a scalar time series $\{v_n = \mathbf{v}^T \cdot \mathbf{x}_n\}$. The construction of interest is based on the tapped delay line, that is, we consider vectors of the form: $(v_n, v_{n+1}, \ldots, v_{n+d-1})^T$. This vector can be thought of as being a function of the point $\mathbf{x}_n \in \mathbb{R}^m$ since

$$(v_n, v_{n+1}, \ldots, v_{n+d-1})^T = (\mathbf{v}^T \cdot \mathbf{x}_n, \mathbf{v}^T \cdot \mathbf{A}\mathbf{x}_n, \ldots, \mathbf{v}^T \cdot \mathbf{A}^{d-1}\mathbf{x}_n)^T$$

To be more formal, we use this tapped delay line approach to define a map $\Phi : \mathbb{R}^m \to \mathbb{R}^d$ taking points from the state space of the dynamical system to the space of states of the d-tap delay line.

$$\Phi(\mathbf{x}) = (\mathbf{v}^T \cdot \mathbf{x}, \mathbf{v}^T \cdot \mathbf{A}\mathbf{x}, \ldots, \mathbf{v}^T \cdot \mathbf{A}^{d-1}\mathbf{x})^T \tag{3.1}$$

This map—which is clearly linear—arises in linear control theory in connection with the *observability* of a system. A well-known result from control theory asserts that Φ is full rank if all the eigenvalues of $\mathbf{A}$ are distinct and none of its eigenvectors is orthogonal to $\mathbf{v}$. Thus, if $d \geq m$, the image $\Phi(\mathbb{R}^m)$ is an

m-dimensional linear subspace of $\mathbb{R}^d$. This result is a statement of the 'usual' or 'generic' situation in the sense that special conditions must hold for it not to be the case. It provides a strong link between the original dynamical system and the tapped delay line data by showing that the information preserved in the tapped delay line representation is that which is preserved when making a change of coordinates. For example, consider the relationship $\mathbf{x}_{n+1} = \mathbf{A}\mathbf{x}_n$ and write $\mathbf{y}_n = \Phi(\mathbf{x}_n)$. Then it follows that $\mathbf{y}_{n+1} = \Phi(\mathbf{A}\Phi^{-1}(\mathbf{y}_n))$; the states of the tapped delay line evolve according to a linear map $\Phi\mathbf{A}\Phi^{-1}$ which, since it is similar to $\mathbf{A}$, has the same spectrum as $\mathbf{A}$. (Note that the inverse $\Phi^{-1}\mathbf{y}$ is meaningful whenever $\mathbf{y} \in \Phi(\mathbb{R}^m)$.)

The theorems of Aeyels 1, Takens 13 and Sauer, Yorke and Casdagli 11 extend this analysis to nonlinear dynamical systems and nonlinear measurement functions. In this case the map corresponding to Φ is nonlinear and an *embedding* of the state space for generic choices of the measurement function. This means that the derivative of Φ is well-defined and full rank at every point in the state space of the dynamical system and, in addition, that the map is invertible (in the linear case these properties are equivalent). Delay embedding, even in this nonlinear case, preserves the information preserved by a (nonlinear) smooth change of coordinates.

The development of these ideas to make them applicable to digital signal processing requires that we enlarge their scope to include iterated function systems. We have reported results in this direction elsewhere 12; 4 and, in 4, described how to extend IFS models to include oversampling of channels and how to exploit the resulting structure through a further development of the method of delays.

Here we shall focus on linear channels sampled at the baud rate and ask, what happens if we apply the tapped delay line idea to the output of a digital linear channel? Equation (2.4) expresses the output of the channel in terms of the sequence of input symbols and the initial state of the channel. Using this, we can define, *for each input symbol sequence*, a delay map $\Phi_\Omega : \mathbb{R}^m \to \mathbb{R}^d$ by

$$\Phi_\Omega(\mathbf{x}) = (\mathbf{v}^T \cdot \mathbf{x}, \mathbf{v}^T \cdot \mathbf{w}_{b_1}(\mathbf{x}), \ldots, \mathbf{v}^T \cdot \mathbf{w}_{b_{d-1}} \circ \mathbf{w}_{b_{d-2}} \circ \cdots \circ \mathbf{w}_{b_1}(\mathbf{x}))^T \tag{3.2}$$

where the subscript Ω labels the input sequence: $\Omega = (b_1, b_2, \ldots, b_{d-1})$. Introducing the explicit form of the $\{\mathbf{w}_b\}$ given in equation (2.2), reveals that the delay map is affine

$$\Phi_\Omega(\mathbf{x}) = \Phi(\mathbf{x}) + \Phi_\Omega(\mathbf{0}) \tag{3.3}$$

where Φ—which is independent of Ω—is the linear delay map defined in equation (3.1). The remaining term is a fixed offset which depends on Ω, but is independent of the channel state $\mathbf{x}$

$$\Phi_\Omega(\mathbf{0}) = (0, \mathbf{v}^T \cdot \mathbf{b}_1, \mathbf{v}^T \cdot (\mathbf{b}_2 + \mathbf{A}\mathbf{b}_1), \ldots, \mathbf{v}^T \cdot (\mathbf{b}_{d-1} + \mathbf{A}\mathbf{b}_{d-2} + \ldots + \mathbf{A}^{d-2}\mathbf{b}_1))^T \tag{3.4}$$

The fact that the observability matrix, Φ, arises naturally here makes the point that delay methods reduce to well-established theory in the special case

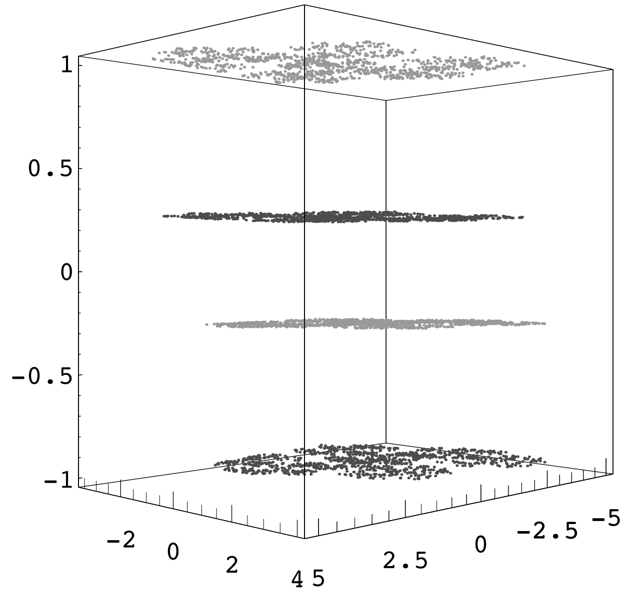

FIG. 3. The attractor shown in figure 1 mapped using the delay maps Φ_Ω : $\mathbb{R}^2 \to \mathbb{R}^3$ based on time series data shown in figure 2. The shading corresponds to that used in figure 1, for the darker points the latest symbol was $+1$ and for the lighter points the latest symbol was -1 .

of linear channels. The linear theory shows that for generic choices of $\mathbf{v}^T$, the rank of Φ is m when $d \geq m$ and, therefore, that the map Φ_Ω is an embedding for each Ω. Thus, $\Phi_\Omega(\mathcal{A})$ is $\mathcal{A}$, apart from a smooth change of coordinates. This has important implications for channel equalisation since there will be a correspondence between the addresses of points in $\mathcal{A}$ and addresses of points in $\Phi_\Omega(\mathcal{A})$ which is constructed using the channel output.

The additional complication that IFSs bring to the use of delay embedding is that each sequence Ω generates a different embedding Φ_Ω. For a p symbol alphabet and using d delays there are p^{d-1} of these. The question is, therefore, are the images of the attractor under the different delay maps all disjoint? The answer is given by the following theorem

Theorem 3.1 *If* $\mathbf{A}$ *has all distinct eigenvalues, then for generic choices of* $\mathbf{v}$ *each of the delay maps* $\Phi_\Omega : \mathbb{R}^m \to \mathbb{R}^d$*, with* $d \geq m$*, is an embedding. Moreover, if* $d > m$*, the generic case is that the images* $\Phi_\Omega(\mathbb{R}^m)$ *and* $\Phi_{\Omega'}(\mathbb{R}^m)$ *are disjoint when* $\Omega \neq \Omega'$.

The theorem has two parts. The first is just the observability condition already described. The second can be shown using a dimension-counting argument which often arises in this sort of proof. In this particular case the argument is

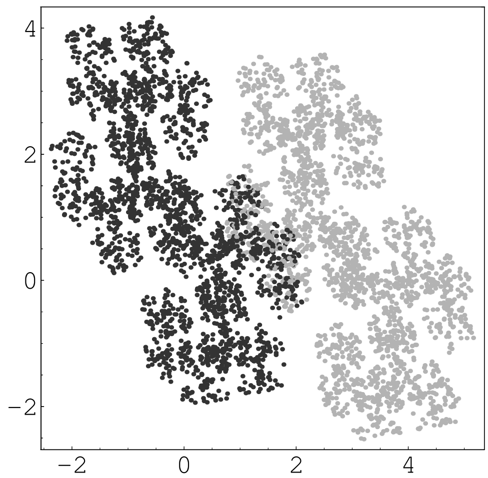

FIG. 4. A plot of the uppermost of the 4 sheets evident in figure 3. In this figure, the darker points are in the set $\Phi_{(-1,-1)}(\mathbf{w}_{+1}(\mathcal{A}))$ and the lighter points are in $\Phi_{(-1,-1)}(\mathbf{w}_{-1}(\mathcal{A}))$

based on the observation that a sufficient condition for the result to hold is that there is no $\mathbf{x} \in \mathbb{R}^m$ such that $\Phi(\mathbf{x}) = \Phi_{\Omega'}(\mathbf{0}) - \Phi_{\Omega}(\mathbf{0})$ for any pair of symbol sequences $\Omega \neq \Omega'$. Figure 3 shows what happens typically in the case of the example described in section 2.1. The delay maps are constructed using time series data as shown in figure 2. Choosing $d = 3 > m = 2$ and recalling that the channel input is binary ($p = 2$), we anticipate 2^2 images of the attractor as, indeed, are seen in figure 3. The fact that $m = 2$ implies that each image should be a subset of a plane—that is, a displaced copy of $\Phi(\mathbb{R}^2)$. These should all be parallel because they are simply translations of one another. Again, this is evident from the figure. In figure 4, the image of $\mathcal{A}$ under the action of one of the Φ_Ω is shown. Since this plot is essentially of $\Phi\mathcal{A}$, it is interesting to make a comparison with the untransformed form of $\mathcal{A}$ shown in figure 1.

Our example also illustrates the meaning of the statement that "generically" the images $\Phi_\Omega(\mathbb{R}^m)$ and $\Phi_{\Omega'}(\mathbb{R}^m)$ are disjoint when $\Omega \neq \Omega'$. In figure 5 we show how the choice of $\mathbf{v}$ changes the positions of the different image planes corresponding to the different values of Ω. This is done by calculating the points of intersection of the planes with their common normal. These are shown as a function of θ which parameterises $\mathbf{v}$ through $\mathbf{v} = (\cos\theta, \sin\theta)$. The interpretation of "generic" here is that at any value of θ except an exceptional set, $\mathcal{E}$, of isolated values, the four planes intersect their common normal at four different points.

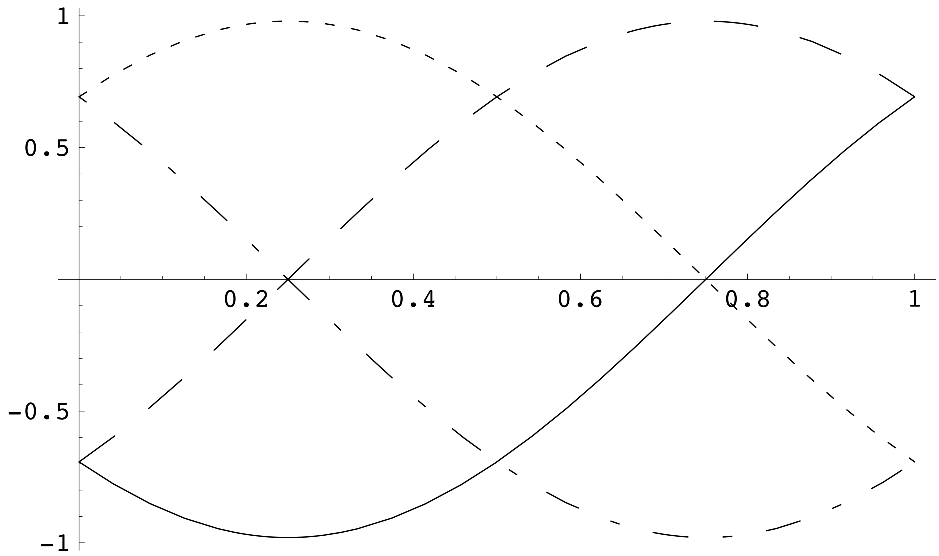

FIG. 5. The projections onto their common normal of the four parallel planes $\Phi_\Omega(\mathbb{R}^2)$ (where $\Omega = (+1,+1)$ (solid), $(-1,+1)$ (dashed), $(+1,-1)$ (dash-dotted), and $(-1,-1)$ (dotted)) plotted as a function of θ where the measurement function is $\mathbf{v} = (\cos\theta, \sin\theta)$ with $\theta \in [0,\pi]$.

The figure shows that $\mathcal{E} = \{0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}, \pi\}$.

There is a subset of $\mathcal{E}$, $\{0, \frac{\pi}{2}, \pi\}$, which is benign in the sense that for θ in this set the delay vectors are essentially state space vectors of the all-pole channel model specified by equation (2.7). Recall that the state space dynamical system described in section 2 is based on the structure of the tapped delay line just as is the method of delays. Therefore, a measurement function $\mathbf{v}$ that picks out a single component of the natural basis of the state space will give rise to delay vectors which are state vectors or—in the case of using $d > m$—vectors such that every m consecutive components are state vectors. It follows in these cases that all of the information about $\mathcal{A}$ is contained unambiguously in the delay plot. This is not true of the remaining points in $\mathcal{E}$ which correspond to measurements which confuse the time ordering of symbols.

4 Some remarks on the equalisation of IFS channels

Given that we observe the output sequence of a digital channel and are able to construct—at least in principle—an object such as is shown in figure 3, how is this of any help in reconstructing the corresponding input sequence? Leaving aside any attempt to do blind equalisation let us assume that we know the input sequence which generated some part of the output. In the first instance, we can use this information to label each of the planes shown in figure 3. By focussing on each separately, estimates of their common normal can be obtained by, for example, forming a matrix of the differences between delay vectors in a chosen plane and computing its singular value decomposition. Alternatively, if the channel model—specified by $\mathbf{A}$, $\mathbf{b}$ and $\mathbf{v}$—is known, then there is a direct way to obtain this information using equation (3.4).

Assuming without loss of generality, that we choose $d = m + 1$, the common normal of the planes is unique. Let us call this $\mathbf{n}$. A direct way to equalise the channel is to compute the projections of the delay vectors onto $\mathbf{n}$ and then to identify which of the projections of the image planes they correspond to. Formally, this amounts to the construction of a map from the delay space to the set of symbol sequences, $g : \mathbb{R}^d \to \{\Omega\}$. If we denote the projection of $\Phi_\Omega(\mathbb{R}^m)$ onto $\mathbf{n}$ as $\mathbf{n}_\Omega$ then $g(\mathbf{x}) = \tilde{\Omega}$ where $\tilde{\Omega}$ minimises $|\mathbf{n} \cdot \mathbf{x} - \mathbf{n}_\Omega|$. Of course, there are more sophisticated ways of doing this which can, for example, take into account noise on the output. The point is, however, that this (nonlinear) function equalises our linear pole-zero channel using $\mathbf{x}$ which is a finite history of the channel output. If, on the other hand, we were to try to invert the channel without assuming the discreteness of the input, we would require an infinite impulse response filter.

There is another, different, way of achieving the same end. This requires that we find the inverse of the Φ_Ω. The most direct way to do this is to assume that the channel models, equation (2.2), are known or have been estimated using the input/output data. This allows us to compute explicitly the components of the affine form given in equation (3.3). The inverse of the delay maps are then simply $\Phi^{-1}(\mathbf{y} - \Phi_\Omega(0))$ (where Φ^{-1} can be computed as the pseudo-inverse of Φ). In practice, something like the g defined above will be needed to decide which of the offsets to subtract from the delay vector. The result of applying the inverses of the delay maps in our example is a set which is indistinguishable from the original attractor shown in figure 1. In particular the addressing of the points of the attractor is preserved by this process. In order to recover the input sequence, however, we need only apply the inverse filter given in equation (2.8). Moreover, it is not clear that the extra complication involved in finding the inverse of the delay maps could ever lead to a method which is superior to simply finding g.

Finally, we come to nonlinear channels. It is clear that a function like g, which maps delay vectors to symbols, can be constructed in this more general case, provided that the images $\Phi_\Omega(\mathbb{R}^m)$ are disjoint. We do not need to assume planarity of the embedded images of the channel attractor, we could, for instance, use a radial basis function expansion to fit the characteristic functions of the different sets of image points. The issue here is whether or not there is a result for nonlinear channels which is analogous to Theorem 3.1. The answer is that there is a result for nonlinear systems which is like the first part of Theorem 3.1 12. Indeed, it is possible to show that—generically—distinct points in the attractor do not become identified by different delay maps. However, there is a counter-example which limits our scope in the nonlinear case. It is easy to write down a hyperbolic IFS—which must be a model of some nonlinear channel—for which the images of the attractor under different Φ_Ω intersect. This property holds for any continuous measurement function and is stable in the sense that it holds also for any small perturbation of the IFS. Physically, such a "difficult" channel is required to have a state, say $\mathbf{x}_*$, which evolves to the same new state, $\mathbf{x}'_*$, following the input of either of two different symbols. It is apparent that this is an extremely undesirable property for a communications channel to have, but unfortunately it is a possibility with nonlinear channels. For this reason we view this example is a limitation on our ability to make the simple generalisation of Theorem 3.1 rather than a practical limitation on the use of delay methods for nonlinear signal processing. In this context, it is worth recalling that the approach described in 4—which uses an oversampling technique—does not suffer from this mathematical difficulty. In this case a more detailed model of the way the channel is driven must be used. For channels which can be thought of as being driven by short pulses the issue of multiple embeddings of the attractor does not arise. Naturally, a channel which is as ambiguous as in our counter-example is likely to cause practical difficulties in this case too.

5 Conclusions

This paper has been about drawing connections between different approaches to digital signal processing. We have related the use of iterated function systems as models of digital channels to the more familiar state space models of linear channels, and we have contrasted the use of delay methods applied to IFS models with the more familiar linear methodology. Our main point has been that where the two approaches talk about the same thing there is a fundamental connection which is essentially the issue of *observability* of the channel. We have not discussed at any length the exciting prospect of a systematic and general theory of nonlinear signal processing of digital channels that the IFS work represents. This has been mentioned elsewhere 3; 4, and will be discussed in more detail in future publications.

Bibliography

Aeyels D. (1981). Generic observability of differentiable systems, *SIAM J. Control and Optimization*, **19**, 595–603.

Barnsley M. (1988). *Fractals Everywhere*, Academic Press Inc., San Diego.

Broomhead D.S., Huke J.P., Muldoon M.R. (1998). Fractals, linear channels and delay methods in *Mathematics in Signal Processing IV* edited by J.G. McWhirter, I.K. Proudler, Clarendon Press, Oxford, 55–65.

Broomhead D.S., Huke J.P. and Muldoon M.R. (2000). Digital channels, *Proceedings of IEEE 2000 Adaptive Systems for Signal Processing, Communications and Control Symposium*, 123–128.

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

Kailath T., Sayed A.H., Hassabi B. (2000). *Linear Estimation*, Prentice-Hall

Inc., New Jersey.

Hutchinson J.E. (1981). Fractals and self-similarity, *Indiana University Journal of Mathematics*, **30**, 713–747.

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

Packard N., Crutchfield J., Farmer D., Shaw R. (1980). Geometry from time series, *Phys.Rev. Lett.*, **45**, 712–715.

Ruelle D. (1989). *Chaotic Evolution and Strange Attractors*, Cambridge University Press, Cambridge

Sauer T., Yorke J.A., Casdagli M. (1991). Embedology, *J. Stat. Phys.*, **65**, 579–616.

Stark J., Broomhead D.S., Davies M.E., Huke J.P. (1997) Takens' embedding theorems for forced and stochastic systems, *Nonlinear Analysis*, **30**, 5303–5314, and (2001) Delay embedding for forced systems: II. Stochastic Forcing, in preparation

Takens F. (1981). Detecting strange attractors in turbulence, *Lecture Notes in Mathematics* edited by Rand D.A., Young L.-S. Springer, Berlin, 366