

***Embedding and Time Series Analysis***

Huke, Jeremy P. and Muldoon, Mark R.

2015

MIMS EPrint: **2016.50**

Manchester Institute for Mathematical Sciences  
School of Mathematics

The University of Manchester

Reports available from: <http://eprints.maths.manchester.ac.uk/>

And by contacting: The MIMS Secretary  
School of Mathematics  
The University of Manchester  
Manchester, M13 9PL, UK

ISSN 1749-9097

# EMBEDDING AND TIME SERIES ANALYSIS

J.P. HUKÉ, M.R. MULDOON

## INTRODUCTION

The 1970's and 80's saw a tremendous wave of interest—across the sciences and beyond—in the subject of nonlinear dynamics. Under the heading of ‘chaos theory’ the subject even gripped the public imagination, leading to popular books and television programmes and even a mention in the film *Jurassic Park*. One of the central ideas driving this interest was the realization that the complex, unpredictable behaviour known as chaos might be widespread in physical systems, and possibly even in biological, economic and social systems as well. This kind of behaviour, with its characteristic *sensitive dependence on initial conditions*, had been shown to occur in a range of simple mathematical systems, many of which were conceived as models of physical or biological phenomena. As well as triggering much work on the mathematical theory of nonlinear dynamical systems, this also raised the intriguing question of whether chaotic behaviour could actually be observed in the broad range of experimental situations that the simple models hinted at. But while physicists and engineers were very familiar with experiments designed to investigate the various periodicities within a system, how should they treat the experimental data (or devise the experiments themselves) so as to reveal the characteristic features of chaos, such as the aforementioned sensitive dependence on initial conditions, or the *strange attractors*, with their fractal structures, that live in the state spaces of some chaotic systems? A new sort of data analysis seemed to be called for.

For some simple laboratory systems (such as the forced, damped pendulum) computer simulations based on model equations might provide fairly compelling evidence for chaotic behaviour in the physical system itself; but even here one would want to compare measured data with the simulations: is the chaotic behaviour actually manifested at the parameter values suggested by the simulations? When one varies a parameter, does the transition from regular, periodic behaviour to chaos happen in the same way in the experiment as in the simulation? In other cases, simulations might be much less persuasive; for example, one of the simple models that sparked the widespread interest in chaos was the *logistic map*, a simple iterative scheme producing a sequence of numbers in which each number is calculated from the previous one with the formula:  $x_{n+1} = ax_n(1 - x_n)$ , where  $a$  is a parameter of the system. For values of  $a$  slightly less than 4 the sequence of  $x_n$ 's varies chaotically. The logistic map had long been used as a model for the time evolution of the population of certain sorts of insect—but clearly such a simple model was not likely to reproduce experimental observations with quantitative accuracy. So while chaotic behaviour undeniably occurs in the logistic map, the same could hardly be inferred for any given insect population. What was needed was some way of interrogating population data directly, to look directly for evidence of chaos.

**Dynamical systems.** The mathematical theory of dynamical systems—which Steve Smale has described as “The Mathematics of Time”—recognizes two kinds of system; the first kind, so-called *discrete time* systems, consist of a set of allowed *states*, and a mapping from the set to itself that determines how the states change over time. Thus, if  $F$  is the mapping and  $x_0$  is the starting state, the next state is  $x_1 = F(x_0)$ , the state after that is  $x_2 = F(x_1) = F(F(x_0))$  and so on: the states are produced sequentially from  $x_n = F(x_{n-1})$ . The logistic map just mentioned is an example: the state space there is the interval  $[0, 1]$ . The other kind of system is *continuous time*: here the evolution of the state is governed by differential equations—so the state space must have sufficient structure for differential equations to be meaningful: often it is  $\mathbb{R}^n$  or some subset of it. Here the paradigmatic example is classical mechanics, where the state consists of a list of particle positions and velocities, and the differential equations are, of course, Newton’s equations of motion. Of course, continuous time systems give rise to discrete time systems in a natural way: instead of recording the state  $x(t)$  at all real-valued times  $t$  we record it at integer multiples of some fixed time interval  $\tau$ , producing a sequence  $x_n = (n\tau)$ . By integrating the differential equations for a time  $\tau$  we find a function  $F$  that maps  $x(n\tau)$  to  $x((n+1)\tau)$ ;  $F$  is the *time- $\tau$  map* of the system. Systems such as fluid flows, governed by partial differential equations, can also be regarded as continuous time systems. Do these mathematical systems provide any hints as to useful ways of treating experimental data?

**States, measurement and injectivity.** To begin to answer this let us first consider how the process of experimental measurement fits into the mathematical picture. Performing an experiment commonly involves fixing the value of some properties of the system: a voltage at some point in an electrical circuit, for example, or the population level of some species in an ecosystem. In the classical mechanics example, where the state consists of particle positions and velocities, a measurement might record one of the positions, or it might produce some more global property such as the total angular momentum or, for a system undergoing small amplitude vibrations, the phase of one of the normal modes. In any case the result of the measurement is the value of some function of the current state. So more generally we say that, to each kind of measurement we can make on a dynamical system, there corresponds a function on the state space: the result of the measurement is the value of the function at the current state. To be precise, let us say the state space is  $S$  and the function corresponding to our measurement—the *measurement function*—is  $f$ . We take it that the result of each measurement is a real number so  $f: S \rightarrow \mathbb{R}$ .

It’s possible that measuring simultaneously several different properties—evaluating several different measurement functions—may provide enough information to determine the state uniquely. That is, there may be only one state consistent with the measured values. Gathering the values of  $m$  measurement functions into a vector produces a function from the state space to  $\mathbb{R}^m$ ; if we want the states to be uniquely identified by the measurements we need this function to be injective (that is, one-to-one). If the  $m$  measurement functions are  $f_1, f_2, \dots, f_m$ , assembling them into a vector  $(f_1, f_2, \dots, f_m)$  produces a mapping  $\Phi: S \rightarrow \mathbb{R}^m$  and it is this mapping whose injectivity we are seeking.

**State space reconstruction.** We will return shortly to the question of whether  $\Phi$  is injective or not, but for the moment let us see what we can do when it is. If  $x$  is the current state then the  $m$ -vector  $\Phi(x)$  is the list of measurements that we record—we

can see  $\Phi(x)$ . But if we have injectivity then we can use  $\Phi(x)$  as a proxy for  $x$ . And the correspondence extends beyond individual states: as the state in  $S$  moves along its trajectory, the image  $\Phi(x)$  will move along a corresponding trajectory in  $\Phi S \subset \mathbb{R}^m$ . If  $x$  moves round a periodic orbit then  $\Phi(x)$  does likewise (and has the same period). And if there is a set  $A$  in  $S$  that attracts all the orbits, then  $\Phi A$  will attract all orbits in  $\Phi S$ ; the difference is that the periodic orbits and attracting sets in  $\Phi S$  are now things that we can see. Thus by the simple device of plotting our measurements as  $m$ -vectors we can create a picture of what is happening in state space, and so reveal the dynamics of the system.

This procedure is sometimes referred to as *reconstructing the state space*. We can even use the picture to make predictions: for example, if the original system is discrete time with dynamics given by  $x_{n+1} = F(x_n)$ , then in the reconstruction the point  $\Phi(x)$  is followed by  $G(\Phi(x))$  where  $G$  is given by  $\Phi \circ F \circ \Phi^{-1}$ . And as  $G$  acts in the reconstructed state space, mapping  $m$ -vectors to  $m$ -vectors, we can hope to estimate it from data. While  $G$  does not tell us the next state in  $S$  (directly) it does predict the next set of measurements, and this has various uses. Of course, the picture we make by reconstruction is in a sense distorted because  $S$  is passed through the mapping  $\Phi$ ; thus states which are far apart in  $S$  may be brought close together in the reconstruction, and *vice versa*. On the other hand, if  $\Phi$  has further properties in addition to just being injective, more properties may be carried over from the original system to the reconstruction. For example, if  $\Phi$  is differentiable—and the system is continuous time—then corresponding to the differential equation that governs the behaviour on  $S$  is a different, differential equation that governs the behaviour on the reconstructed state space. And if  $\Phi$  is continuously differentiable, the fractal dimensions of compact sets such as strange attractors in  $S$  are preserved in the reconstruction.

A general mapping  $f: A \rightarrow B$  for which the image  $fA$  retains some important properties of the set  $A$  is often referred to as an *embedding*—we are thus here interested in the circumstances in which  $\Phi$  is an embedding of the dynamical system.

Note that to do the sort of reconstruction we have been describing we did not need to say much about the function  $\Phi$  or its components  $f_1, f_2$  etc. All we needed was the assumption of injectivity, and—for the preservation of properties such as dimension—some kind of regularity (e.g. differentiability) for  $\Phi$ . For example, we might not actually know the topology of the state space  $S$ , but so long as the measurement functions are continuous we may expect the topology to be preserved in the reconstruction (technically this requires that  $S$  be compact, or some equivalent assumption). To take a simple case, the state space for a system of  $k$  nonlinearly coupled phase oscillators is a  $k$ -torus, so reconstruction produces data that lies on a  $k$ -torus in  $\mathbb{R}^m$ . We could in principle use data to confirm the presence of this torus and so determine the number of oscillators.

We see then that this approach to analyzing experimental data, suggested by the dynamical systems picture, has a highly geometrical flavour, in which the data are used to produce a point cloud in  $\mathbb{R}^m$ , and properties of the measured system are investigated by examining the geometry, subsets, and dynamics of the point cloud.

**The method of delays.** We must now, however, return to some of the underlying assumptions on which we built the story above. One of these is that we have so far imagined that whenever we conduct an experiment we measure simultaneously

several different properties of the system. In practice, this may not be at all easy: the system may be difficult to prepare and manipulate and the measurements delicate or costly to perform; it may be as much as we can manage to measure just a single quantity. Have we then any hope of making the kind of analysis described above?

Packard, Crutchfield, Farmer and Shaw [5] suggested that, instead of  $m$  simultaneous measurements of different quantities we use measurements of the *same* quantity taken at  $m$  different times (and hence, usually, in  $m$  different states). There is then just a single measurement function  $f: S \rightarrow \mathbb{R}$ , and to each trajectory  $x_0, x_1, \dots, x_n, \dots$  there corresponds a sequence (or *times series*) of measured values

$$f(x_0), f(x_1), \dots, f(x_n), \dots$$

The  $m$ -vectors are then assembled from short subsequences, the  $n$ -th such being

$$(f(x_n), f(x_{n+1}), \dots, f(x_{n+m-1})).$$

For a discrete time system we can write this as

$$(f(x_n), f \circ F(x_n), \dots, f \circ F^{m-1}(x_n))$$

which we can again think of as a mapping  $\Phi_d: S \rightarrow \mathbb{R}^m$ , with the  $k$ -th component of  $\Phi_d$  being  $f \circ F^{k-1}$ . An entirely analogous setup holds for continuous time systems as well. The  $m$ -vectors are called *delay vectors* (the subscript  $d$  on  $\Phi_d$  connotes ‘delay’) and the approach is sometimes called the *method of delays*. Assuming that delays work as well as simultaneous measurements, the method of delays turns reconstruction into a much more practical proposition, and the vast majority of applications have employed it.

#### THE EMBEDDING THEOREMS OF WHITNEY AND TAKENS

The other main matter that we have not yet fully addressed is the question of the injectivity of  $\Phi$  and  $\Phi_d$ . We can see immediately that there is no possibility that these mappings will be injective for entirely arbitrary measurement functions: if the  $f_i$ 's are constant functions, or if they all happen to be the same function then clearly  $\Phi$  could only be injective on very restricted state spaces! Less trivially, if the state space is  $\mathbb{R}^n$  then we must surely make  $m \geq n$  to stand much chance of injectivity. We will therefore have to be content with some statement to the effect that  $\Phi$  and  $\Phi_d$  are injective for a ‘large’ (in some sense) set of measurement functions belonging to some suitable regularity class—and that in addition we must use ‘enough’ functions (or delays).

A suitable setting for such statements is *differential topology* which concerns itself with differentiable mappings between manifolds. Manifolds are higher dimensional generalizations of smooth surfaces such as the sphere, torus or Klein bottle; they are sets in which every point has a neighbourhood that looks like a subset of  $\mathbb{R}^n$  in the sense that one can find good local coordinates. The state spaces of dynamical systems are very often manifolds—especially for continuous time systems because the manifold structure is needed for differential equations on the state space to be properly defined. A map between manifolds can be regarded (locally) as a map from  $\mathbb{R}^n$  to  $\mathbb{R}^{n'}$ , so all the usual tools of calculus can be used.

We can also compare two functions on the same manifold: we regard the functions as ‘close together’ if, at every point of the manifold, the values of the functions, and their derivatives (and perhaps higher derivatives as well) are close—where we use

the coordinate systems to quantify how close the values (and derivatives) are. It turns out that this idea of ‘close’ functions can be used to give a precise sense to the somewhat vague statement about ‘large’ sets of functions mentioned above. We do this as follows: consider all the functions from some manifold  $M$  to  $\mathbb{R}$  (or  $\mathbb{R}^m$ ) which belong to some particular regularity class (for example, they have continuous first derivatives at all points). A subset of this collection of functions is said to be ‘open and dense’ if two conditions are satisfied: (i) if  $f$  is a function in the subset then all functions sufficiently close to  $f$  are also in the subset, and (ii) if  $g$  is any function (which may or may not be in the subset) then arbitrarily close to  $g$  we can find functions which are in the subset. It is this notion of an open and dense subset that we use when we come to consider the injectivity of the functions  $\Phi$  and  $\Phi_d$ .

A fundamental result from differential topology, due to Whitney, addresses our concerns quite directly: it states that if  $M$  is a compact manifold of dimension  $n$ ,  $m > 2n$ , and  $C^1(M, \mathbb{R}^m)$  is the set of continuously differentiable functions from  $M$  to  $\mathbb{R}^m$ , then the subset of functions that are embeddings of  $M$  is open and dense in  $C^1(M, \mathbb{R}^m)$ . (‘Embedding’ here means not only that the function is injective but that it is a *diffeomorphism*: both the function and its inverse are differentiable.) The mappings in  $C^1(M, \mathbb{R}^m)$  are just the sorts of functions we have called  $\Phi$  above; Whitney’s theorem thus provides some justification for hoping that the  $m$  simultaneous measurement functions approach will produce an embedding.

Is there a similar result underpinning the method of delays? In 1981 Floris Takens showed that there is [7]. Recall that  $\Phi_d$  is formed from a measurement function  $f$  and the mapping  $F: S \rightarrow S$  that specifies the dynamics. Takens’ theorem states that if  $M$  is a compact manifold of dimension  $n$ ,  $m > 2n$ ,  $C^2(M, \mathbb{R})$  is the set of twice continuously differentiable real-valued functions, and  $\chi^2(M)$  is the set of twice continuously differentiable diffeomorphisms from  $M$  to itself, then the set of pairs  $(f, F)$  in  $C^2(M, \mathbb{R}) \times \chi^2(M)$  for which the corresponding  $\Phi_d$  is an embedding, is open and dense.

### THE METHOD OF DELAYS IN PRACTICE

Though the statement of Takens’ Theorem sounds rather technical, it was widely considered to provide a sound theoretical foundation for the method of delays and its appearance triggered a flood of applications. But while the theorem provides some justification for delay embedding, it is silent about a number of practical issues. For example, in constructing the delay vector for state  $x$  we took measurements at  $x, F(x), F^2(x)$ , etc., but we could equally well have used  $x, F^2(x), F^4(x) \dots$ , or any other  $k$ -th iterate of  $F$ . This issue is particularly relevant if  $F$  is the time- $\tau$  map of a continuous time system; in principle we could use any non-zero value for  $\tau$ , but some values are likely to work better than others. For example, if  $\tau$  is too small then the successive states—and the corresponding measurements in the delay vector—will be very similar; thus all the points in the reconstruction will lie very close to the diagonal of  $\mathbb{R}^m$  and most of the important information about the system will be contained in the small deviations of the points away from this line: Figure 1 illustrates a mild version of this effect for delay embeddings of simulated data from a damped, driven pendulum. On the other hand, if  $\tau$  is too large then sensitive dependence on initial conditions may mean that  $\Phi_d(x)$  depends very sensitively on  $x$ , causing the reconstructed state space  $\Phi_d S$  to be highly stretched or convoluted. The

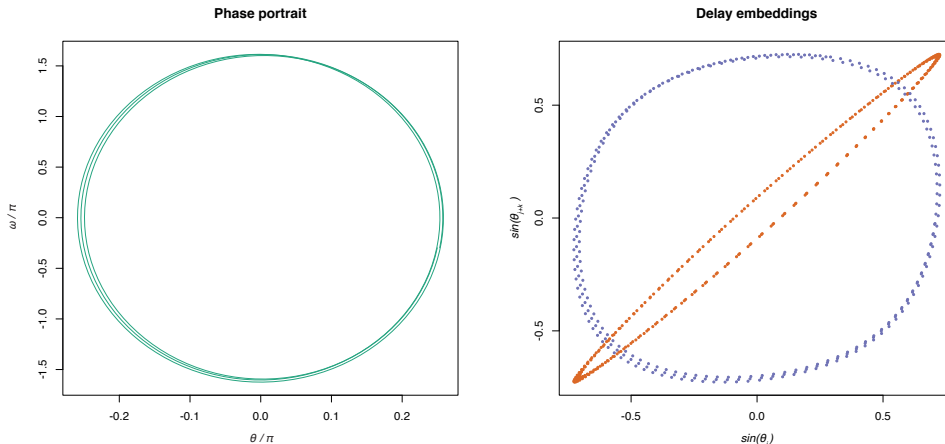


FIGURE 1. The method of delays applied to a mechanical system, the damped, driven pendulum governed by  $\ddot{\theta} = -g \sin(\theta) - \gamma \dot{\theta} + 25 \cos(2\pi t)$ . The panel at left is the phase portrait of a trajectory relaxing onto the system's attracting periodic solution while the panel at right shows two delay embeddings of the same trajectory. The narrow ellipse (orange dots) is the image of this periodic orbit under them delay embedding map  $\Phi_d(\theta) = (f(\theta), f(F(\theta)))$  with measurement function  $f(\theta) = \sin(\theta)$  while the rounded square (purple dots) was produced by  $\Phi_d(\theta) = (f(\theta), f(F^{25}(\theta)))$ . In both cases  $F$  is the time- $\tau$  map with  $\tau = 0.009$

success or otherwise of the data analysis may therefore be strongly influenced by the choice of the time between elements of the delay vectors.

Another problem is the choice of the number of delays,  $m$ . Takens' theorem tells us that  $m$  must be larger than twice the intrinsic dimension of the dynamical system, but we may not know in advance what this is (indeed it may be one of the things we hope to find). Choosing  $m$  too small may thus fail to produce an embedding, but choosing it too large leads to data scattered sparsely in a space of many dimensions. Again, a good choice may make the difference between a useful and an unsuccessful investigation.

As the widespread use of delay embedding got under way a considerable literature developed regarding these more practical issues. David Broomhead made many important contributions, including work with Greg King that turned out to be one of Dave's best-cited papers, [4]. Eventually the experience accumulated by all these workers culminated in the publication of several textbooks on the subject, including those by Abarbanel [1] and Kantz & Schreiber [6].

Although, as described above, the initial motivation for the method of delays was to look for chaotic behaviour in experimental systems—and more specifically to detect attractors with non-integer dimension in the state spaces of such systems (see, for example, Takens' original paper [7]), the potential use of the method in a broader range of signal processing activities was soon recognized. Suppose, for example, that the measurements  $f(F^k(x))$  are contaminated with a small amount of additive noise. The effect of this on the delay vectors will be to make small random shifts so that they are no longer confined to the image set  $\Phi_d S$ . So long as the shifts are small,

however, we can attempt to estimate the original, uncontaminated delay vectors. Two ways of doing this suggest themselves: firstly, we estimate the location of the set  $\Phi_d S$  from the noisy data. In practice we would do this locally: choosing one of the delay vectors we identify the other close by data points and find the  $n$  dimensional plane which most nearly contains them; then we project the data points to lie on this plane. We repeat the procedure at points all over the reconstructed state space to produce a cleaned data set. The second approach recognizes the above-mentioned fact that the reconstruction can be used to make a (short term) predictor. Noise will cause the measurements to deviate from predictions made using uncontaminated data. Of course, uncontaminated data is not likely to be available, but we can estimate the predictor using noisy data, adjust the data to be consistent with the resulting predictions, and then re-estimate the predictor with the cleaned data, and repeat the process, hoping it will converge towards the noise-free measurements. Both of these approaches have been extensively investigated, and various elaborations described.

Standard approaches to core signal processing tasks (such as filtering, signal detection and separation, compensation of distortion, etc.) rely heavily on linear methods. Dave Broomhead was particularly interested in the use of embedding techniques to find new approaches to such tasks, able to cope with (and perhaps exploit) the non-linear nature of signal generating mechanisms or of the media through which the signals passed. His work in this area included novel techniques for nonlinear signal separation [3] and for the identification of digital signals subjected to distortion and intersymbol interference by propagation through strongly nonlinear channels [2].

#### REFERENCES

- [1] H.D.I. Abarbanel (1996), *Analysis of Observed Chaotic Data*, Institute for Nonlinear Science, Springer-Verlag New York.
- [2] D.S. Broomhead, J.P. Huke, M.R. Muldoon, J. Stark (2004), Iterated Function System Models of Digital Channels, *Proc. Roy. Soc. Lond. A* **460**:3123–3142.
- [3] D.S. Broomhead, J.P. Huke, and M.A.S. Potts (1996), Cancelling deterministic noise by constructing nonlinear inverses to linear filters, *Physica D* **89**:439–458.
- [4] D.S. Broomhead and G.P. King (1986), Extracting qualitative dynamics from experimental data, *Physica D* **20**:217–236.
- [5] N.H. Packard, J.P. Crutchfield, J.D. Farmer, and R.S. Shaw (1980), Geometry from a time series, *Physical Review Letters*, **45**:712–716.
- [6] H. Kantz and T. Schreiber (2003), *Nonlinear Time Series Analysis*, Cambridge University Press, 2nd edition.
- [7] F. Takens (1981), Detecting strange attractors in turbulence, in *Dynamical Systems and Turbulence*, (Springer Lecture Notes in Mathematics, vol. 898) edited by D. A. Rand and L.-S. Young, Springer-Verlag, Berlin.

THE SCHOOL OF MATHEMATICS, ALAN TURING BUILDING, THE UNIVERSITY OF MANCHESTER,  
MANCHESTER M13 9PL, UK