Embedding Nonlinear Dynamical Systems: A Guide to Takens' Theorem

Huke, J. P.

2006

MIMS EPrint: **2006.26**

Manchester Institute for Mathematical Sciences
School of Mathematics

The University of Manchester

ISSN 1749-9097

Embedding Nonlinear Dynamical Systems: A Guide to Takens' Theorem

J. P. Huke

Abstract

The embedding theorem of Takens forms a bridge between the theory of nonlinear dynamical systems and the analysis of experimental time series. This memorandum describes the theorem, and gives a detailed account of its proof. The necessary differential topology is briefly reviewed, and then a proof of the theorem is presented; this proof follows broadly the argument of Takens, although it differs in some details. Some extensions to the theorem, which facilitate its use in applications, are described. The memo concludes with a brief discussion of what the theorem implies about time series, viewed as the raw material for signal processing algorithms.

Defence Research Agency
Farnborough, Hants., GU14 6TD, UK

Contents

1 Introduction

One of the most surprising lessons of dynamical systems theory is that the phase spaces of simple nonlinear systems may contain strange attractors [1, 2]. For certain simple systems, such as pendula, direct observations of the phase space are possible, and the existence of strange attractors can be experimentally investigated. But physical systems in which all the relevant dynamical variables can be simultaneously monitored are relatively few, and such systems are usually limited to the laboratory bench. At first sight, the search for these attractors in systems where the dynamical variables cannot be measured, or are unknown or infinite in number, seems problematic. Yet many such searches have been made, including attempts to show chaos in situations, such as epidemiological [3] or economic systems, where even the applicability of dynamical systems theory is unproven.

The possibility that these attempts might actually be successful is suggested by a branch of applied dynamical systems theory referred to by Sauer, Yorke and Casdagli [4] as 'embedology'; this is concerned with the extraction of information about features of phase space from time series of general measurements made on an evolving system. The central plank of this theory is a result—suggested by several people [5] and eventually proved by Takens [6]—which shows how a time series of measurements of a single observable can often be used to reconstruct qualitative features of the phase space of the system. The technique described by Takens, the *method of delays*, is so simple that it can be applied to essentially any time series whatever, and has made possible the wide-ranging search for chaos mentioned above.

Of course, embedology is not a finished subject: how the amount of derivable information varies with the quantity and quality of the data, and how the method of delays may best be used, are matters of continuing debate [7, 8], as are the range of applicability of the method and its recent extensions. But these debates may be rather hampered by the inaccessibility of the theorem which justifies the method of delays: the method may be simple but the proof seems discouragingly mathematical. A disinclination to wade through what appears to be a particularly technical and specialized argument is understandable, but there remains the lurking worry that if we just quote the theorem we may be missing the point. And if we wish to use or adapt the theorem for our own purposes, for example to address the problem of filtered time series [9], we have little choice but to familiarize ourselves with the mechanics of its proof.

The problem with Takens's proof is that it is couched in the language of differential topology, a subject not well known to the wider physics and engineering communities. Indeed, the important point of the theorem depends on the topological notion of *genericity* ('open and denseness'), itself an unfamiliar notion to many. 'Generic' may be interpreted loosely as 'mostly' or 'apart from special exceptions', but such interpretations fail to catch the topological nature of the property. (Furthermore, a measure-theoretic property called *prevalence* may have a better claim on these interpretations, and recent work has extended the theorem essentially by replacing 'generic' with 'prevalent' [4].) Even those with enough familiarity with classical mechanics to have some notion of how a differentiable manifold is defined might be forgiven for finding Takens's paper something of a challenge. It is with a view to helping them meet this challenge that the following exposition is offered.

We will go through the proof of Takens' theorem in some detail. Necessarily this will involve us in quite a lot of differential topology, but actually the level needed is not very high: everything is covered in elementary books on the subject, and we shall refer to these freely [10, 11, 12]. Presenting the proof in this way should help us to appreciate exactly what the significance of the theorem is, and will make clear some aspects—such as why functions are required to be C^2, how this can be relaxed to C^1, and why the number of delays needed is $2m+1$—which have caused some bafflement in the past. We will also be in a position to use the same style of proof for our own purposes, such as the filtering problem.

2 Some Differential Topology

For the sake of completeness we start with some ideas and results from elementary differential topology, but those for whom the idea of a manifold is completely new might be advised to consult first an introductory account of the subject [10, 11, 12]. (That by Chillingworth is very good.) Even introductory texts will probably assume some basic topology [13, 14] (though Chillingworth is a notable exception), and so will we.

A *manifold* M is a (separable, Hausdorff) topological space which is locally like $\mathbb{R}^m$, that is, every point has an open neighbourhood which is homeomorphic to an open subset of $\mathbb{R}^m$. (The manifold is said to be of *dimension* m.) A pair (U, h), where $U \subset M$ is open and $h : U \to \mathbb{R}^n$ is a homeomorphism onto its range is called a *chart*, with U as the *chart domain* (or *coordinate neighbourhood*), and h as the *coordinate function*. A collection of charts whose domains cover M is an *atlas*.

If two charts (U, h) and (V, g) have overlapping domains, the *coordinate transformations*

$$hg^{-1} : g(U \cap V) \to \mathbb{R}^m \text{ and } gh^{-1} : h(U \cap V) \to \mathbb{R}^m$$

are functions from open subsets of $\mathbb{R}^m$ to $\mathbb{R}^m$. The charts are C^r*-related* if both these functions are continuously differentiable r times. An atlas is C^r*-differentiable* if all its charts are C^r-related to each other (where there is overlap). A *differential structure* is the set of all charts which are C^r-related to those in a particular atlas—it is itself an atlas and is defined so as to avoid giving the impression that some coordinate systems are to be preferred over others. (A manifold with a differential structure is *differentiable*, and is called a C^r manifold.) In fact, when showing that a manifold has particular properties, we may choose any convenient atlas to work with: the results carry over automatically to all C^r-related charts, and hence to the differential structure. A certain amount of elementary differential topology is concerned with finding appropriate atlases with which to prove desired results, and this is also true of Takens's proof.

A function $f : M \to N$ between C^r manifolds is C^s*-differentiable* ($s \leq r$) if, for every point $p \in M$ there are charts (U, h) and (V, g) of M and N respectively, with $p \in U$ and $f(p) \in V$, such that $gfh^{-1} : h(U \cap f^{-1}V) \to \mathbb{R}^n$ is s times continuously differentiable at $h(p)$. Although this definition is couched in terms of particular charts containing p and $f(p)$ it is quickly seen that it does not matter which ones we use. The *Jacobi matrix* of f at p with respect to h and g is the matrix of partial derivatives $Dgfh^{-1}(h(p))$; clearly this does depend on the charts we use, although its rank does not. If the derivative at $h(p)$ is injective, then f is said to be *immersive at* p; and a function which is immersive everywhere is an *immersion*. An immersion which carries M homeomorphically onto its image is an *embedding*. If M is compact, it can be shown that any injective immersion is an embedding. If the derivative is surjective then f is said to be *submersive at p*.

If N is an n-dimensional manifold, and $M \subset N$ is an m-dimensional manifold, $m \leq n$, then M is a *submanifold* of N if at every point of M there is a chart which can be obtained from a chart (V, g) of N, by restricting g to $V \cap M$, and dropping the last $n - m$ coordinates. If A and B are manifolds, and $f : A \to B$ is an embedding, then $f(A)$ is a submanifold of B, and $f : A \to f(A)$ is a *diffeomorphism* (i.e. a differentiable function with a differentiable inverse). Two manifolds which are diffeomorphic can be considered the same apart from a smooth (and invertible) change of coordinates. For our purposes, the importance of embedding is that it allows us to identify a subset of $\mathbb{R}^n$ which is diffeomorphic to the phase space of the system we are studying. By investigating this subset, we can clearly learn a great deal about the system itself.

We shall need some basic results from differential topology. We will just quote these results here without proof—any introductory text will supply the details. The first result is the following. If M is a differentiable manifold, and $\{U_\mu : \mu \in \Lambda\}$ is an open cover, then there is an atlas $\{(V_\nu, g_\nu) : \nu \in \mathbb{N},\ g_\nu : V_\nu \to V'_\nu\}$ with the following properties:

1. For every $\nu \in \mathbb{N}$ there is a $\mu \in \Lambda$ such that $V_\nu \subset U_\mu$; and every point in M has a neighbourhood that intersects only finitely many V_ν. ($\{V_\nu : \nu \in \mathbb{N}\}$ is a *locally finite refinement* of $\{U_\mu : \mu \in \Lambda\}$.)
2. $V'_\nu = \{x \in \mathbb{R}^n : \|x\| < 3\} = B(3)$
3. The sets $W_\nu = g_\nu^{-1}\{x \in \mathbb{R}^m : \|x\| < 1\} = g_\nu^{-1} B(1)$ still cover M.

An atlas such as this is called a *good atlas, subordinate to* $\{U_\mu : \mu \in \Lambda\}$. Since we shall only be concerned with compact manifolds, we can assume that any atlas has a finite number of charts.

Good atlases are used in conjunction with *differentiable bump functions*. For any positive radius r, and $\epsilon > 0$, we can construct a function $\lambda : \mathbb{R}^m \to \mathbb{R}$ with the properties:

$0 \leq \lambda(x) \leq 1$ for all $x \in \mathbb{R}^m$,

$\lambda(x) = 1 \Leftrightarrow x \in \overline{B(r)}$, (An overbar represents the closure of a set.)

$\lambda(x) = 0 \Leftrightarrow \|x\| \geq r + \epsilon$,

and λ is infinitely differentiable.

Another basic tool is the *partition of unity*. If A is a closed subset of M, (which may be M itself) and $\{U_i : i \in \Lambda\}$ is an open cover of A, then there is a set of functions $\lambda_i : M \to [0, 1], i \in \Lambda$, with the following properties:

1. λ_i is C^∞,
2. the support of λ_i is contained in U_i, $i \in \Lambda$,
3. $\{\text{support}\lambda_i\}_{i \in \Lambda}$ is locally finite,
4. $\sum_{i \in \Lambda} \lambda_i(x) = 1$, for every $x \in A$.

This set of functions is known as a C^∞ *partition of unity, subordinate to* $\{U_i : i \in \Lambda\}$. Partitions of unity allow us to create global properties by making local adjustments, a device used several times in the proof of Takens' theorem.

We will need to use the following two results at various points, so we present them as lemmas:

Lemma 1 *If M and N are manifolds with dimensions m and n respectively, $m < n$, and $f : M \to N$ is a C^1 function, then $N - f(M)$ is dense in N.*

Lemma 2 *Let M and N be manifolds with dimensions m and n, and $m > n$, and $f : M \to N$ be a C^1 function. Let $q \in N$. If f is submersive at at every p such that $f(p) = q$, then the set $f^{-1}(q)$ is a submanifold of M, with dimension $m - n$.*

If we are to discuss the genericity of embeddings, then we must first endow some function spaces with topologies [15]. Let $C^r(M,N)$ be the set of C^r maps from M to N. The C^s *topology*, $(s \leq r)$ on $C^r(M,N)$ is generated by a sub-base consisting of sets defined as follows. Let $f \in C^r(M,N)$, and (U,h) and (V,g) be charts on M and N; let $K \subset U$ be a compact set such that $f(K) \subset V$, and let $0 < \epsilon \leq \infty$. The set $\mathcal{N}^s(f;(U,h),(V,g),K,\epsilon)$ consists of those functions $\hat{f} \in C^r(M,N)$ for which $\hat{f}(K) \subset V$ and

$$\|D^k g\hat{f}h^{-1}(x) - D^k gfh^{-1}(x)\| < \epsilon \tag{1}$$

for all $x \in h(K)$, $k = 0,\ldots,s$. We always use the Euclidean norm.

(This looks more complicated than it really is. Roughly speaking, to decide whether we should regard two functions as close together, we break up the domain M into pieces which can be transformed to subsets of $\mathbb{R}^m$, and then assess how close the functions are on each piece, in the usual way.) Note that in this topology, any apparent dependence on the charts is removed by simply considering them all. We shall want to use this topology in two ways. To show that a set of functions, (for example immersions), is open in $C^r(M,N)$, we need to find, for each member f of the set, charts $\{(U_i,h_i)\}$ and $\{(V_i,g_i)\}$, compact sets $\{K_i\}$ and numbers $\{\epsilon_i\}$, so that all members of $\bigcap_i \mathcal{N}^s(f;(U_i,h_i),(V_i,g_i),K_i,\epsilon_i)$ are members of the set. To show that a set is dense, we have to show that for every $f \in C^r(M,N)$, every neighbourhood of f contains a member of the set. To do this it is enough to find a good atlas for M, as described above, any convenient atlas for N, and to show that for any $\epsilon > 0$, $\bigcap_\nu \mathcal{N}^s(f;(U_\nu,h_\nu),(V_\nu,g_\nu),\overline{W}_\nu,\epsilon)$ contains a member of the set. In our case N will often be $\mathbb{R}^n$ (for some n), for which we can always choose the chart $(\mathbb{R}^n, \mathit{identity})$. The set of C^r functions from M to itself which are also diffeomorphisms (have C^r inverses) is called $\mathit{Diff}^r(M)$; it is clearly a subset of $C^r(M,M)$, and can be given the subspace topology.

Finally we note that any manifold may be regarded as a complete metric space: this fact can be used to simplify some of the topological arguments we need in the proof of Takens' theorem.

3 Proof of Takens' Embedding Theorem

Let us first state the thing we are setting out to prove: this is Theorem 1 of Takens' paper [6].

Theorem 1 (Takens) *Let M be a compact manifold of dimension m. For pairs (ϕ, y), with $\phi \in \mathit{Diff}^2(M)$, $y \in C^2(M,\mathbb{R})$, it is a generic property that the map $\Phi_{(\phi,y)} : M \to \mathbb{R}^{2m+1}$, defined by*

$$\Phi_{(\phi,y)}(x) = (y(x), y(\phi(x)), \ldots, y(\phi^{2m}(x)))$$

is an embedding.

Here 'generic' means open and dense, and we use the C^1 topology. We refer to the functions $y \in C^2(M,\mathbb{R})$ as *measurement* functions.

The discussion given by Takens is largely directed towards establishing a slightly different version of the theorem, which runs as follows:

Theorem 2 (Takens, unstated) *Let M be as above. Let $\phi : M \to M$ be a diffeomorphism, with the properties: firstly, that the periodic points of ϕ with periods less than or equal to $2m$ are finite in number, and secondly that if x is any periodic point with period $k \leq 2m$ then the eigenvalues of the derivative of ϕ^k at x are all distinct. Then for generic $y \in C^2(M,\mathbb{R})$, the map $\Phi_{(\phi,y)} : M \to \mathbb{R}^{2m+1}$, defined as in Theorem 1 is an embedding.*

Note that Theorem 1 is concerned with open and dense sets of diffeomorphisms, while in Theorem 2 attention is focused on one particular ϕ. This ϕ is not, however, arbitrary: it has to satisfy the conditions concerning the periodic points with period less than or equal to $2m$. We shall discuss later on how to get to Theorem 1 having established Theorem 2, but for the moment the latter is our target. This version of the theorem also helps to make clear the connections between Takens' theorem and the later work of Sauer, Yorke and Casdagli, and we will say more about this later as well.

Naturally enough, the proofs of Theorems 1 and 2 come in two parts: one part establishes the openness of the embeddings and the other their denseness. However the proof of denseness draws several times upon the fact that certain sets of functions, such as immersions and embeddings, are open, and so we need to start with that. (Takens himself discusses openness very briefly at the end of his proof.)

3.1 Openness of the set of embeddings

Theorem 2 promises genericity in the set $C^2(M, \mathbb{R})$ of measurement functions. Every function y in this space gives rise to a delay map $(y, \ldots, y\phi^{2m})$, so we can define a mapping $\mathcal{F}^{(2)} : C^2(M, \mathbb{R}) \to C^2(M, \mathbb{R}^{2m+1})$ by $y \mapsto \Phi_{(\phi,y)}$. We shall need to show that this mapping is continuous. This is essentially straightforward, but since it involves using the perhaps unfamiliar topology of $C^2(M, \mathbb{R})$ we shall proceed in stages.

Lemma 3 *The function $F_1 : C^2(M, \mathbb{R}) \to C^2(M, \mathbb{R})$ defined by $y \mapsto y \circ \phi$ is continuous.*

Proof. Let $\{(U_i, h_i); i \in \Lambda\}$ be a finite good atlas for M, and say $W_i = h_i^{-1}B(1)$. Given any neighbourhood in $C^2(M, \mathbb{R})$ of $y \circ \phi$ there is a neighbourhood $\mathcal{N}$ of the form $\cap_i \mathcal{N}^1(y \circ \phi; (U_i, h_i), (\mathbb{R}, id), \overline{W_i}, \epsilon')$ contained within it—we need only choose ϵ' small enough. We will show that there is a neighbourhood $\mathcal{N}(\epsilon) = \cap_i \mathcal{N}^1(y; (U_i, h_i), (\mathbb{R}, id), \overline{W}_i, \epsilon)$ of y such that if $\hat{y} \in \mathcal{N}(\epsilon)$ then $F_1(\hat{y}) \in \mathcal{N}$; that is, $F_1\mathcal{N}(\epsilon) \subset \mathcal{N}$, so that F_1 is continuous. All we need is to show that this is true for sufficiently small ϵ.

The sets W_i, $i \in \Lambda$, cover M, and since ϕ is a diffeomorphism, so do the sets $\phi^{-1}W_i$, $i \in \Lambda$, and also the sets $\phi^{-1}W_i \cap W_j$, $i, j \in \Lambda$. If $\phi^{-1}\overline{W}_i \cap \overline{W}_j$ is not empty for some particular i, j, the derivative $Dh_i\phi h_j^{-1} : h_j(\phi^{-1}\overline{W}_i \cap \overline{W}_j) \to \mathbb{R}^{m\times m}$ is continuous and has a compact domain. (To see this it must be recalled $\overline{W}_i \subset U_i$.) Hence the norms of these derivatives are bounded: we can find a constant $A_{i,j}$ such that $\|Dh_i\phi h_j^{-1}(u)\| < A_{i,j}$ for all $u \in h_j(\phi^{-1}\overline{W}_i \cap \overline{W}_j)$. And since there are only finitely many of these intersections, we can find a single A which is an upper bound for $\{A_{i,j} : i, j \in \Lambda\}$.

Now choose $\epsilon < \min\{\epsilon', \epsilon'/A\}$. Let $\hat{y} \in \mathcal{N}(\epsilon)$ and let $x \in \overline{W}_j$. Then $x \in \phi^{-1}\overline{W}_i \cap \overline{W}_j$ for some $i \in \Lambda$. Let $x' = \phi(x)$, so that $x' \in \overline{W}_i$. Then

$$\begin{aligned}
\|\hat{y} \circ \phi h_j^{-1}(h_jx) - y \circ \phi h_j^{-1}(h_jx)\| &= \|\hat{y}(x') - y(x')\| \\
&= \|\hat{y}h_i^{-1}(h_ix') - yh_i^{-1}(h_ix')\| \\
&< \epsilon \\
&< \epsilon'.
\end{aligned}$$

Further

$$\begin{aligned}
D\hat{y} \circ \phi h_j^{-1}(h_jx) - Dy \circ \phi h_j^{-1}(h_jx) &= D\hat{y}h_i^{-1}h_i\phi h_j^{-1}(h_jx) - Dyh_i^{-1}h_i\phi h_j^{-1}(h_jx) \\
&= D\hat{y}h_i^{-1}(h_ix)Dh_i\phi h_j^{-1}(h_jx) - Dyh_i^{-1}(h_ix)Dh_i\phi h_j^{-1}(h_jx) \\
&= (D\hat{y}h_i^{-1}(h_ix) - Dyh_i^{-1}(h_ix))Dh_i\phi h_j^{-1}(h_jx)
\end{aligned}$$

where the second line uses the Chain Rule. This means

$$\begin{aligned}
\|D\hat{y} \circ \phi h_j^{-1}(h_jx) - Dy \circ \phi h_j^{-1}(h_jx)\| &\leq \|D\hat{y}h_i^{-1}(h_ix) - Dyh_i^{-1}(h_ix)\|\|Dh_i\phi h_j^{-1}(h_jx)\| \\
&< \epsilon A \\
&< \epsilon'.
\end{aligned}$$

Hence the conditions 1 are satisfied, and $\hat{y} \circ \phi \in \mathcal{N}$. It follows that F_1 is continuous.

Lemma 4 *The function $F_n : C^2(M, \mathbb{R}) \to C^2(M, \mathbb{R})$ defined by $(\phi, y) \mapsto y \circ \phi^n$ is continuous.*

Proof. This is done by induction. We already know that F_1 is continuous. Assume that F_{n-1} is continuous, and note that $F_ny = y \circ \phi^n = (y \circ \phi^{n-1}) \circ \phi = F_1(y \circ \phi^{n-1}) = F_1(F_{n-1}y)$, so F_n is the composition of continuous functions.

Proposition 1 *$\mathcal{F}^{(2)}$ is continuous.*

Proof. Let us say $F_0 : C^2(M,\mathbb{R}) \to C^2(M,\mathbb{R})$ is the identity on $C^2(M,\mathbb{R})$, and also that $F : C^2(M,\mathbb{R}) \to [C^2(M,\mathbb{R})]^{2m+1}$ is specified by its components: $F = (F_0, F_1, \ldots, F_{2m})$. Then $\mathcal{F}^{(2)} = T \circ F$, where $T : [C^2(M,\mathbb{R})]^{2m+1} \to C^2(M,\mathbb{R}^{2m+1})$ maps each $2m+1$-tuple of real-valued functions $(f_0, f_1, \ldots, f_{2m})$ to the vector-valued function having these functions as its components. The previous lemma shows that F is continuous, so it only remains to show that T is. This is quite similar to lemma 3: for each $2m+1$-tuple $(f_0, f_1, \ldots, f_{2m})$ in $[C^2(M,\mathbb{R})]^{2m+1}$ we have to show that for every neighbourhood $\mathcal{N}$ of $T(f_0, f_1, \ldots, f_{2m}) = f$ there is a neighbourhood $\mathcal{N}(\epsilon)$ of the $2m+1$-tuple such that $T\mathcal{N}(\epsilon) \subset \mathcal{N}$.

Let $\{(U_i, h_i), i \in \Lambda\}$ be a finite good atlas for M, with $W_i = h_i^{-1}B(1)$. Given any neighbourhood $\mathcal{N}$ in $C^2(M,\mathbb{R}^{2m+1})$ of f, there is a neighbourhood of the form $\cap_i \mathcal{N}^1(f; (U_i, h_i), (\mathbb{R}^{2m+1}, id), \overline{W}_i, \epsilon')$ contained within it. Now choose $\epsilon < \epsilon'/(2m+1)$, and consider the neighbourhood $\mathcal{N}(\epsilon)$ of $(f_0, \ldots, f_{2m})$ defined by

$$\mathcal{N}(\epsilon) = \bigotimes_{j=0}^{2m} \cap_i (f_j; (U_i, h_i), (\mathbb{R}, id), \overline{W}_i, \epsilon).$$

Let $(\hat{f}_0, \ldots, \hat{f}_{2m}) \in \mathcal{N}(\epsilon)$, $\hat{f} = T(\hat{f}_0, \ldots, \hat{f}_{2m})$, and $x \in \overline{W}_i$. Then

$$\begin{aligned} \|\hat{f}h_i^{-1}(h_i x) - f h_i^{-1}(h_i x)\| &\leq \sum_{j=0}^{2m} |\hat{f}_j h_i^{-1}(h_i x) - f_j h_i^{-1}(h_i x)| \\ &< \sum_{j=0}^{2m} \epsilon \\ &< \epsilon'. \end{aligned}$$

Also

$$\begin{aligned} \|D\hat{f}h_i^{-1}(h_i x) - Df h_i^{-1}(h_i x)\| &\leq \sum_{j=0}^{2m} \|D\hat{f}_j h_i^{-1}(h_i x) - Df_j h_i^{-1}(h_i x)\| \\ &< \sum_{j=0}^{2m} \epsilon \\ &< \epsilon'. \end{aligned}$$

Hence $\hat{f} \in \mathcal{N}$, so T is continuous.

This brings us to the main result of this section.

Proposition 2 *Let M be a compact manifold, $\phi : M \to M$ a diffeomorphism, and K a compact subset of M. Then the set of functions y such that the delay map $\Phi_{(\phi,y)} : M \to \mathbb{R}^{2m+1}$ is immersive on K is open in $C^2(M,\mathbb{R})$. The same is true of injective immersions of K.*

Proof. We first note that the set, S, of maps $f : M \to \mathbb{R}^{2m+1}$ which are immersive on K is open in $C^2(M,\mathbb{R}^{2m+1})$. This is a standard result of differential topology so we need not prove it here: Hirsch [15] gives a full treatment of these matters. It is clear that the measurement functions giving delay maps which are immersive on K are just those which lie in the inverse image of S under $\mathcal{F}^{(2)}$. Since $\mathcal{F}^{(2)}$ is continuous, by the previous proposition, this inverse image is open. Just the same argument applies to injective immersions of K.

We shall have many occasions to use this result below; we can use it immediately to prove the openness part of Theorem 2: all we need to do is take K to be M.

Our statement of Propostion 2 is much more specific than it need be. A glance at the argument above shows that no use is made of the fact that the number of delays is $2m+1$: the proposition is true for any number of delays. There is no need for ϕ to be a diffeomorphism: any differentiable function from M to M will do; and there is no need for ϕ and y to be C^2: it is sufficient that they be C^1, (remember that we always work in the C^1 topology).

3.2 Denseness of the set of embeddings

Having shown that the measurement functions giving rise to embeddings of M are open in $C^2(M, \mathbb{R})$ we now wish to show that they are also dense. The general strategy for doing this is again a standard method of differential topology. Given any $y \in C^2(M, \mathbb{R})$ and any neighbourhood $\mathcal{N}$ of y, we have to show that there exists $y' \in \mathcal{N}$ such that $\Phi_{(\phi,y')}$ is an embedding of M. This is done by showing how to construct a suitable y' explicitly.

We construct new measurement functions by adding functions to y:

$$y' = y + \sum_{i=1}^{N} a_i \psi_i \tag{2}$$

where N is finite, the a_i's are real, and $\psi_i : M \to \mathbb{R}$ is differentiable. To ensure that y' is C^2 we need to make ψ_i C^2; in fact we will always be able to assume that they are C^∞.

We adjust the measurement function in the manner of equation (2) several times during the course of the construction. Each such adjustment endows y' with some desirable property, for instance that of giving an immersive delay map, or embedding[1] some compact subset of M. These properties are used when we make subsequent adjustments. The question immediately arises: how do we ensure that the properties we have given to the measurement function are preserved under subsequent adjustments? For example, if $\Phi_{(\phi,y)}$ is an immersion of M, and y' is given by (2), how do we ensure that $\Phi_{(\phi,y')}$ is still an immersion? This example hints at the answer we need: according to Proposition 2, there is a neighbourhood of y such that so long as y' lies within this neighbourhood, y' will share the properties of y.

Given any $y \in C^2(M, \mathbb{R})$, and any neighbourhood $\mathcal{N}$ of y, we make a suitable choice for ψ_i, $i = 1, \ldots, N$, and for a_i, $i = 1, \ldots, N$, and form y' according to (2). We choose the ψ_i's and a_i's so that y' is a member of $\mathcal{N}$, and y' has the desired property (which in the first place will be immersivity at a periodic point). (Of course, we must demonstrate that such a choice is possible.) This property is shared by all the measurement functions in some open set, $\mathcal{O}$, containing y', and we can assume $\mathcal{O}$ is contained in $\mathcal{N}$. We choose another set of ψ_i's and a_i's and construct a third function y'', which lies in $\mathcal{O}$, and which has some further desired property (for example it is immersive at another periodic point). Since y'' lies in $\mathcal{O}$ it still has the first property. We carry on producing new measurement functions, each within an open neighbourhood of the previous one until, after a finite number of adjustments, we generate an embedding of M. The final function lies in $\mathcal{N}$.

It is apparent that for each adjustment to y, and every neighbourhood of y, we must show that it is possible to find sets $\{\psi_i\}$ and $\{a_i\}$ such that y' lies in the neighbourhood, and has the property we want. Finding $\{\psi_i\}$ and $\{a_i\}$ so that y' lies in the neighbourhood is discussed in the next paragraph; finding them so that y' has the desired property naturally depends on the property we are after: the rest of the proof is concerned with showing that we can find suitable ψ_i's and a_i's to endow the measurement function with the characteristics we want—ultimately that of giving rise to a delay map which is an embedding of M.

To make the argument outlined above, we shall need the following lemma.

Lemma 5 *Let $y : M \to \mathbb{R}$ be C^2, and let $\psi_i : M \to \mathbb{R}$, $i = 1, \ldots, N$ be C^2 for all i, where N is finite. Let $a = (a_1, \ldots, a_N)^T$ be a member of $\mathbb{R}^N$. For each neighbourhood $\mathcal{N}$ of y there is some $\delta > 0$ such that if $\|a\| < \delta$ the function y' defined by equation (2) lies in $\mathcal{N}$.*

Proof. As usual, let $\{(U_i, h_i) : i \in \Lambda\}$ be a finite good atlas for M. Then there exists $\epsilon > 0$ such that the open set $\cap_i \mathcal{N}^1(y; (U_i, h_i), (\mathbb{R}, id), \overline{W}_i, \epsilon)$ is a subset of $\mathcal{N}$. For each j, $1 \le j \le N$ and each $i \in \Lambda$, the function $\psi_j h_i^{-1} : h_i(\overline{W}_i) \to \mathbb{R}$ is well defined, and being a continuous function on a compact domain the magnitudes of its values are bounded, say by $B_{i,j}$. Since there are only finitely many such functions, there exists an upper bound B of the set $\{B_{i,j} : 1 \le j \le N,\ i \in \Lambda\}$. It is clear from (2) that if $x \in \overline{W}_i$ then

$$\|y' h_i^{-1}(h_i x) - y h_i^{-1}(h_i x)\| \quad = \quad \|\sum_{j=1}^{N} a_j \psi_j h_i^{-1}(h_i x)\|$$

[1] We shall sometimes say a compact set A is 'embedded' by a function f if f is an injective immersion on A, even though A may not be manifold.

$$
\begin{aligned}
&\leq \sum_{j=1}^{N} |a_j||\psi_j h_i^{-1}(h_i x)| \\
&\leq \sum_{j=1}^{N} |a_j| B_{i,j} \\
&\leq B \sum_{j=1}^{N} |a_j|.
\end{aligned}
$$

The derivatives $D\psi_j h_i^{-1}$ are similarly continuous functions, so we can make the same argument to give

$$
\|Dy'h_i^{-1}(h_i x) - Dyh_i^{-1}(h_i x)\| \leq B' \sum_{j=1}^{N} |a_j|.
$$

It is clear from the last two inequalities that there is some $\delta > 0$ such that if $\|a\| < \delta$ then $\|y'h_i^{-1}(h_i x) - yh_i^{-1}(h_i x)\|$ and $\|Dy'h_i^{-1}(h_i x) - Dyh_i^{-1}(h_i x)\|$ will both be less than ϵ, for all $x \in \overline{W}_i$ and $i \in \Lambda$. Hence y' will then be a member of $\mathcal{N}$.

How shall we adjust the measurement function to yield an embedding? The procedure described below has four stages; each stage produces a new measurement function whose delay map is an injective immersion on successively larger parts of M. If we say P_k is the set of periodic points of ϕ with period less than or equal to k, then we shall create new functions which give a delay map which is

1. immersive at every point in P_{2m} (and hence an injective immersion on some compact neighbourhood of P_{2m}),
2. an immersion on the whole of M,
3. an injective immersion on orbit segments,
4. an injective immersion on M.

Each of these stages requires one or more adjustments of the form (2); for each one it must be shown that $\|a\|$ in (2) can be made arbitrarily small and yet still give a y' having the desired property. As we shall see, the main tool for doing this is lemma 1.

3.3 Stage 1: Periodic orbits

It is clear that if we try to embed M using delay maps into $\mathbb{R}^{2m+1}$, any periodic points whose period is less than or equal to $2m$ will present us with special problems. For these points, not all the coordinates of their images under $\Phi_{(\phi,y)}$ can be different. In particular, for fixed points all the coordinates are equal, implying that all the images of fixed points lie on the diagonal of $\mathbb{R}^{2m+1}$. This degeneracy causes some delay maps to fail to be embeddings: for example, if ϕ is the identity then any corresponding delay map will not be an embedding whatever the measurement function. Nor will increasing the number of delays help to repair these failures.

These difficulties are the reason why in Theorem 2 the condition is imposed that the number of periodic points of ϕ with period $2m$ or less shall be finite. (That is, P_{2m} shall be a finite set.) Given this condition, there is clearly an open neighbourhood of each $x_i \in P_{2m}$ containing no other point in P_{2m}. By taking a smaller neighbourhood if necessary, we can assume that it lies within some chart domain U_i, and indeed that it is homeomorphic, under the chart map h_i, to an open ball B_i in $\mathbb{R}^m$ centred at $h_i(x_i)$. The Hausdorff property implies we can choose these neighbourhoods so that they do not intersect each other. Now as noted above, the images of the fixed points under $\Phi_{(\phi,y)}$ all lie on the diagonal of $\mathbb{R}^{2m+1}$. More explicitly, if x is a fixed point,

$$
\Phi_{(\phi,y)}(x) = (y(x), y(x), \ldots, y(x)).
$$

To make sure that none of the fixed points have the same image, we must adjust y so that it takes a different value at every fixed point. More generally, if y is such that it takes a different value for every $x_i \in P_{2m}$, then no two of these points map to the same image (at least the first components of the images will be different). Since P_{2m} is a finite set, the measurement functions with this property form an open subset of $C^2(M, \mathbb{R})$, by Proposition 2.

To see how to adjust y so that it becomes an injection on P_{2m}, let x_1 and x_2 be points in P_{2m} at which y takes the same value. Let (U_1, h_1) be the chart containing x_1 mentioned in the previous paragraph, so that $h_1 x_1$ is the centre of $B(3)$. Define a function $\psi : M \to \mathbb{R}$ by

$$\psi(x) = \begin{cases} \lambda(h_1 x) & \text{for } x \in h_1^{-1}B(3) \\ 0 & \text{otherwise} \end{cases}$$

where $\lambda : \mathbb{R}^m \to \mathbb{R}$ is a bump function having support in $B(3)$, and equal to 1 on $B(1)$. It is clear that ψ is a C^∞ function on M. Then say

$$y' = y + a\psi$$

where a is a real number. (This is the form that (2) takes in this case.) Clearly for every $a > 0$, $y'(x_1)$ and $y'(x_2)$ are different (they differ by a). Lemma 5 shows that we can find a suitable y' in any neighbourhood of y by taking a sufficiently small. A similar argument deals with the case when more than two x_i's have the same image, though then we have to make more than one perturbation. Taken with the previous paragraph, this shows that for generic y, $\Phi_{(\phi,y)}$ is injective when restricted to P_{2m}.

Our objective in Stage 1 is to make $\Phi_{(\phi,y)}$ an injective immersion on P_{2m} so we must now turn to the immersivity part. We require $D\Phi_{(\phi,y)} h_i^{-1}(h_i x_i)$ to be full rank at every $x_i \in P_{2m}$. Consider first the fixed points of ϕ. If x_1 is a fixed point the k-th row of $D\Phi_{(\phi,y)} h_1^{-1}(h_1 x_1)$ is $Dy\phi^{k-1}h_1^{-1}(h_1 x_1)$, which by the chain rule (and using the fact that x_1 is a fixed point) is $Dyh_1^{-1}(h_1 x_1) Dh_1 \phi^{k-1} h_1^{-1}(h_1 x_1)$. Letting the row vector $v = Dyh_1^{-1}(h_1 x_1)$, and $J = Dh_1 \phi h_1^{-1}(h_1 x_1)$, the k-th row is vJ^{k-1}. So the question of immersivity at x_1 boils down to whether the set $\{v, vJ, vJ^2 \dots vJ^{2m}\}$ contains m linearly independent vectors.

Propostion 2 shows that the measurement functions which make $\Phi_{(\phi,y)}$ an immersion at x_1 form an open subset of $C^2(M, \mathbb{R})$.

To create an immersion at x_1 we note first that by assumption J has distinct eigenvalues λ_j, and hence linearly independent eigenvectors. We expand v in terms of these vectors $v = \sum \alpha_j e_j$, so that $vJ^{k-1} = \sum \alpha_j \lambda_j^{k-1} e_j$. (If the eigenvalues are all real, then so are the α_j's, otherwise we may need to complexify to do this: see for example Hirsch and Smale [19].) In this basis the vectors v, vJ etc. take the form

$$\begin{bmatrix} \alpha_1 & \alpha_2 & \dots & \alpha_m \\ \alpha_1\lambda_1 & \alpha_2\lambda_2 & \dots & \alpha_m\lambda_m \\ \vdots & & & \vdots \\ \alpha_1\lambda_1^{2m} & \alpha_2\lambda_2^{2m} & \dots & \alpha_m\lambda_m^{2m} \end{bmatrix}$$

We consider the first m of these, discarding the later rows to leave a square matrix; the first m rows will be linearly independent if and only if the determinant of this matrix is zero. If the α_i's are real, then considered as a function of these coefficients this determinant represents a mapping from $\mathbb{R}^m$ to $\mathbb{R}$: by expanding it out we see that it is a polynomial function, and that vectors v for which the matrix is not full rank correspond to sets of α's which are zeroes of the polynomial. Now it is well known that if a polynomial does not vanish identically, its zeroes form a closed, nowhere dense set. To see that the determinant is not identically zero we need only consider $(\alpha_1, \alpha_2, \dots, \alpha_m) = (1, 1, \dots, 1)$ in which case the matrix takes the form

$$\begin{bmatrix} 1 & 1 & \dots & 1 \\ \lambda_1 & \lambda_2 & \dots & \lambda_m \\ \vdots & & & \vdots \\ \lambda_1^m & \lambda_2^m & \dots & \lambda_m^m \end{bmatrix}$$

which we know (from the theory of Vandermonde determinants) to be full rank. Thus the v for which $\{v, vJ, vJ^2, \dots, vJ^{2m}\}$ contains m linearly independent vectors form an open dense set. (This argument can easily be extended to the case where the λ's and α's are not real: it relies on the fact that since v and J are real the non-real λ_i's and α_i's come in complex conjugate pairs.) Thus given any v we can find v' such that $\{v', v'J, v'J^2, \ldots, v'J^{2m}\}$ span $\mathbb{R}^m$ and the difference $a \equiv v' - v$ has arbitrarily small norm.

Define C^∞ functions $\psi_j : M \to \mathbb{R}$, $1 \leq j \leq m$ by

$$\psi_j(x) = \begin{cases} \mu_j(x)\lambda(h_1 x) & \text{for } x \in h_1^{-1}B(3) \\ 0 & \text{otherwise} \end{cases}$$

where $\mu_j : U_1 \to \mathbb{R}$ is the j-th coordinate function: that is, $\mu_j(x)$ is the j-th coordinate of $h_1(x)$. (We shall use functions like this several times; the important to note is that for any $u = h_1 x \in \overline{B(1)}$ we have $\frac{\partial \psi_j h_1^{-1}}{\partial u_k}(u) = \delta_{kj}$. This follows because the bump function has zero derivative on $\overline{B(1)}$.) Also define y' by equation (2), ($N = m$ here of course). Then $Dy'h_1^{-1}(h_1x_1) = Dyh_1^{-1}(h_1x_1) + a$ where a is the row vector whose components are a_j. We have just seen that we can arrange for $\{Dy'h_1^{-1}(h_1x_1)J^k : k = 0 \ldots 2m\}$ to span $\mathbb{R}^m$, and hence for $\Phi_{(\phi,y')}$ to be immersive, while $\|a\|$ is less than any prescribed value, so that lemma 5 implies that a y' giving rise to an immersion at x_1 can be found in every neighbourhood of y.

Since the number of fixed points of ϕ is finite, a finite number of such adjustments to y will give a y' for which $\Phi_{(\phi,y')}$ is immersive at all of them, and by the usual argument, such a y' can be found in every neighbourhood of y.

In considering immersivity we have concentrated so far on fixed points of ϕ, but the arguments extend easily to the periodic points. For example, say x_1, x_2 are a pair of period 2 points, ($\phi(x_1) = x_2$, $\phi(x_2) = x_1$). The important observation is that since these points are distinct, we can find disjoint open sets containing x_1 and x_2, homeomorphic to open balls B_1 and B_2 centred at $h_1(x_1)$ and $h_2(x_2)$. This means that we can use bump functions, as above, to perturb y independently at these two points.

Evidently

$$\Phi_{(\phi,y)}(x_1) = (y(x_1), y(x_2), y(x_1), \ldots, y(x_1)).$$

Consider the question of immersivity at x_1, that is, the rank of $D\Phi_{(\phi,y)}h_1^{-1}(h_1x_1)$. The $2i+1$-th row of the matrix is equal to

$$Dy\phi^{2i}h_1^{-1}(h_1x_1) = Dyh_1^{-1}h_1\phi^{2i}h_1^{-1}(h_1x_1) = Dyh_1^{-1}(h_1x_1)Dh_1\phi^{2i}h_1^{-1}(h_1x_1)$$

which we write as vJ^i, where $v = Dyh_1^{-1}(h_1x_1)$ and $J = Dh_1\phi^2h_1^{-1}(h_1x_1)$. The $2i$-th row is

$$Dy\phi^{2i-1}h_1^{-1}(h_1x_1) = Dy\phi^{-1}\phi^{2i}h_1^{-1}(h_1x_1) = Dy\phi^{-1}h_1^{-1}(h_1x_1)Dh_1\phi^{2i}h_1^{-1}(h_1x_1)$$

or wJ^i where $w = Dy\phi^{-1}h_1^{-1}(h_1x_1)$. As before we transform to the eigenbasis of J, and say $v = \sum \alpha_j e_j$ and $w = \sum \beta_j e_j$. $D\Phi_{(\phi,y)}h_1^{-1}(h_1x_1)$ now takes the form

$$\begin{bmatrix} \alpha_1 & \alpha_2 & \cdots & \alpha_m \\ \beta_1 & \beta_2 & \cdots & \beta_m \\ \alpha_1\lambda_1 & \alpha_2\lambda_2 & \cdots & \alpha_m\lambda_m \\ \beta_1\lambda_1 & \beta_2\lambda_2 & \cdots & \beta_m\lambda_m \\ \vdots & & & \vdots \\ \beta_1\lambda_1^m & \beta_2\lambda_2^m & \cdots & \beta_m\lambda_m^m \\ \alpha_1\lambda_1^{m+1} & \alpha_2\lambda_2^{m+1} & \cdots & \alpha_m\lambda_m^{m+1} \end{bmatrix}.$$

We rearrange the rows of this matrix to give

$$\begin{bmatrix} \alpha_1 & \alpha_2 & \cdots & \alpha_m \\ \alpha_1\lambda_1 & \alpha_2\lambda_2 & \cdots & \alpha_m\lambda_m \\ \vdots & & & \vdots \\ \alpha_1\lambda_1^{m+1} & \alpha_2\lambda_2^{m+1} & \cdots & \alpha_m\lambda_m^{m+1} \\ \beta_1 & \beta_2 & \cdots & \beta_m \\ \beta_1\lambda_1 & \beta_2\lambda_2 & \cdots & \beta_m\lambda_m \\ \vdots & & & \vdots \\ \beta_1\lambda_1^m & \beta_2\lambda_2^m & \cdots & \beta_m\lambda_m^m \end{bmatrix}$$

and then eliminate the lower rows to make the matrix square. The corresponding determinant is a polynomial function, this time $\mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}$, which is not identically zero because we can choose

$$(\alpha_1, \alpha_2, \ldots, \alpha_m, \beta_1, \beta_2, \ldots, \beta_m) = (1, 1, \ldots, 1, \lambda_1^{m+2}, \lambda_2^{m+2}, \ldots, \lambda_m^{m+2})$$

to yield a Vandermonde determinant. It follows that the set of pairs (v, w) which make the derivative full rank is open and dense in $\mathbb{R}^m \times \mathbb{R}^m$, and hence given any v and w we can find v' and w' such that $\|(v, w) - (v', w')\|$ is arbitrarily small, and the set $\{v', v'J, \ldots, v'J^{m+1}, w', w'J, \ldots, w'J^m\}$ spans $\mathbb{R}^m$.

Now suppose that y is a measurement function for which $\Phi_{(\phi,y)}$ is not immersive at x_1. We can define new functions $\psi_i : M \to \mathbb{R}$, $\chi_i : M \to \mathbb{R}$, $1 \leq i \leq m$ by

$$\psi_i(x) = \begin{cases} \mu_{1,i}(x)\lambda(h_1 x) & \text{for } x \in h_1^{-1}B_1 \\ 0 & \text{otherwise} \end{cases}$$

and

$$\chi_i(x) = \begin{cases} \mu_{2,i}(x)\lambda(h_2 x) & \text{for } x \in h_2^{-1}B_2 \\ 0 & \text{otherwise} \end{cases}$$

where $\mu_{1,i}$ is the i-th coordinate function of h_1, and $\mu_{2,i}$ that of h_2. Then say

$$y'(x) = y(x) + \sum_{i=1}^{m} a_i \psi_i(x) + \sum_{i=1}^{m} b_i \chi_i(x).$$

This of course is just another version of equation (2): our task is to show that there are vectors (a, b) of arbitrarily small norm such that the corresponding $\Phi_{(\phi,y')}$ is immersive at x_1. The result of the previous paragraph is that we can make $\Phi_{(\phi,y')}$ immersive if we can choose (a, b) so that $Dy'h_1^{-1}(h_1 x) = v'$ and $Dy'\phi^{-1}h_1^{-1}(h_1 x_1) = w'$. Note that since B_1 and B_2 are disjoint, there is no x for which $\psi_i(x)$ and $\chi_j(x)$ are both non-zero, whatever i and j may be. Hence the derivative of y' at x_1 is $Dy'h_1^{-1}(h_1 x_1) = Dyh_1^{-1}(h_1 x_1) + a = v + a$, so we set $a = v' - v$. The derivative $Dy'\phi^{-1}h_1^{-1}(h_1 x_1)$ is given by

$$\begin{aligned} Dy'\phi^{-1}h_1^{-1}(h_1 x_1) = Dy'h_2^{-1}h_2\phi^{-1}h_1^{-1}(h_1 x_1) &= Dy'h_2^{-1}(h_2 x_2) Dh_2\phi^{-1}h_1^{-1}(h_1 x_1) \\ &= (Dyh_2^{-1}(h_2 x_2) + b)A \\ &= Dy\phi^{-1}h_1^{-1}(h_1 x_1) + bA \\ &= w + bA \end{aligned}$$

where A is the matrix $Dh_2\phi^{-1}h_1^{-1}(h_1 x_1)$. Since A is invertible (ϕ is a diffeomorphism) we can choose $b = (w' - w)A^{-1}$. These choices make $\Phi_{(\phi,y')}$ immersive; evidently we can make the norm of (a, b) arbitrarily small by choosing the norm of $(v - v', w - w')$ sufficiently small.

Observe that to find an immersion at x_1 we have to perturb y both at x_1 and x_2. This arises because of the 'delay' nature of Φ. The measurement functions giving rise to delay maps which are immersive at x_1 are open, and so by making another, similar adjustment we can construct a measurement function which is immersive at both points x_1 and x_2.

The reader will need no further prompting to realise that these arguments can be extended to cover all the points of period less than or equal to $2m$. (If the period is greater than or equal to m, m components of the delay map can be perturbed independently.) The general conclusion is that by making a sequence of adjustments to y, we can construct a y' such that $\Phi_{(\phi,y')}$ is immersive at all the points in P_{2m}, and also injective on P_{2m}. Since this set is compact Proposition 2 shows that these measurement functions form an open subset of $C^2(M, \mathbb{R})$.

If $f : M \to \mathbb{R}^k$ is a C^1 function, and its derivative at some point $p \in M$ is injective, then there is some neighbourhood U of p such that the restriction of f to U embeds U in $\mathbb{R}^k$ [10]. This is, essentially, a consequence of the Inverse Function Theorem. Thus we can find a neighbourhood of each point $x_i \in P_{2m}$ which is embedded in $\mathbb{R}^{2m+1}$ by the immersive maps $\Phi_{(\phi,y)}$ that we have just been discussing. (This neighbourhood will depend on y.) Recalling that the manifold M can be considered as a metric space, we may take the neighbourhood of x_i to be an open ball $b_i(r_i, x_i)$ of radius r_i centred at x_i. (We use a lower case letter to label these balls to distinguish them from the open balls of $\mathbb{R}^m$, which are the supports for the bump functions we used above.) $\Phi_{(\phi,y)}$ is an immersion on the union of these balls (since this is a local property), and does not map two points in the same ball to the same image, but the images of different balls could intersect. However, by taking smaller radii it must be possible to find balls whose images do not intersect. (This follows from the continuity of $\Phi_{(\phi,y)}$: if it were not true we could find a sequence $\{z_i\}$ in one ball, converging to its centre, and a corresponding sequence $\{\hat{z}_i\}$ in another ball, converging to its centre, such that $\Phi_{(\phi,y)}(z_i) = \Phi_{(\phi,y)}(\hat{z}_i)$, implying that two points in P_{2m} map to the same image.) $\Phi_{(\phi,y)}$ is thus an injective immersion on the union of these smaller balls. Now consider closed balls each having half the radius of an open ball; the union of these is closed, hence compact, and is a subset of the union of open balls, so that $\Phi_{(\phi,y)}$ is an injective immersion of this compact set. This set depends on y, so we call it V_y[2]. V_y is clearly a compact neighbourhood of P_{2m}. We call the closed ball containing x_i, b_i: then $V_y = \cup_i b_i$.

3.4 Stage 2: Making an immersion

Let us take stock of where we are. We have concluded that for an open dense subset of $C^2(M, \mathbb{R})$ the delay map $\Phi_{(\phi,y)}$ is an injective immersion of P_{2m}, and that for each y in this set there is a compact neighbourhood V_y of P_{2m} such that $\Phi_{(\phi,y)}$ is an embedding of V_y. The next stage in the proof is to show that in every neighbourhood of y we can find another measurement function which gives rise to an immersion of M. The strategy for doing this is a standard procedure in differential topology, though the details depend on the 'delay' nature of Φ; the description here follows that of Bröcker and Jänich [12]. We start by covering M with compact sets, and show that by making arbitrarily small perturbations of the measurement function, we can produce a delay map which is an immersion of one of these sets. By making another perturbation we immerse another of the compact sets, and, as usual, use the openness of immersions to show that the second adjustment does not spoil the immersion created by the first. We proceed in this way, immersing the compact sets one by one, until the whole manifold is immersed.

As noted, the set V_y embedded by $\Phi_{(\phi,y)}$ will in general be different for different measurement functions. But Proposition 2 shows that delay embeddings of a compact set such as V_y are open. It follows that there is a neighbourhood of y in $C^2(M, \mathbb{R})$ (which we call $\mathcal{U}_y$), such that for every $\hat{y} \in \mathcal{U}_y$, $\Phi_{(\phi,\hat{y})}$ is an embedding of V_y. We intend to show that every neighbourhood of y contained in $\mathcal{U}_y$ contains an immersion of M.

We begin, as so often, by constructing a suitable atlas. First we select an arbitrary atlas of M. Since every point in P_{2m} lies in a chart domain we can, by intersecting this domain with the interior of b_i, find a new chart which is a subset of b_i. By shifting and scaling the chart map, and taking a smaller domain if necessary we can find a chart (U_i, h_i), with $U_i \subset b_i$ and $U_i = h_i^{-1}B(3)$. We can find such a chart for each x_i, and the chart domains are clearly disjoint. Note that the sets $W_i = h_i^{-1}B(1)$ form an open cover for P_{2m}, and that $\overline{W}_i$ is a compact subset of b_i.

Now consider the complement of P_{2m}, P_{2m}^c, which is an open set. Evidently for each element x in this set the points $\{x, \phi x, \phi^2 x, \ldots, \phi^{2m} x\}$ are all distinct. So we can find an open set $U_x \subset P_{2m}^c$ containing x such that $U_x, \phi U_x, \ldots, \phi^{2m} U_x$ are disjoint. By the usual argument (taking smaller sets if necessary) we can take U_x to be contained in—and hence to be—a chart domain, and finally we can find a chart (U_x, h_x), with $U_x = h_x^{-1}B(3)$.

The collection of sets $\{W_x = h_x^{-1}B(1) : x \in P_{2m}^c\} \cup \{W_i : x_i \in P_{2m}\}$ is clearly an open cover of M. From it we extract a finite subcover. Note that the subcover contains every set of $\{W_i : x_i \in P_{2m}\}$. We relabel the sets so that W_i, $1 \leq i \leq k$ are the sets containing the periodic points, and W_i, $k < i \leq l$ are the sets contained in P_{2m}^c. The corresponding charts (U_i, h_i) form the required atlas.

By construction, for every $\hat{y} \in \mathcal{U}_y$, $\Phi_{(\phi,\hat{y})}$ is an embedding of, and hence an immersion of, the compact set $\cup_{i=1}^k \overline{W}_i$. It remains to adjust the measurement function so as to make an immersion of the remaining $\overline{W}_i$'s.

Suppose that i is the smallest index greater than k for which $\Phi_{(\phi,y)}$ fails to be an immersion of $\overline{W}_i$. Let x be a member of U_i, let $\mu_j : U_i \to \mathbb{R}$, $j = 1, \ldots, m$ be the coordinate functions of h_i, (that

[2] This is slightly different to Takens' notation: he calls this set V.

is, $h_i x = (\mu_1 x, \ldots, \mu_m x))$, and let $u = h_i x$ and $u_j = \mu_j x$. Then the Jacobi matrix of $\Phi_{(\phi,y)}$ at $h_i x$ is

$$\begin{bmatrix} \frac{\partial y h_i^{-1}}{\partial u_1}(u) & \frac{\partial y h_i^{-1}}{\partial u_2}(u) & \cdots & \frac{\partial y h_i^{-1}}{\partial u_m}(u) \\ \frac{\partial y \phi h_i^{-1}}{\partial u_1}(u) & \frac{\partial y \phi h_i^{-1}}{\partial u_2}(u) & \cdots & \frac{\partial y \phi h_i^{-1}}{\partial u_m}(u) \\ \vdots & & & \vdots \\ \frac{\partial y \phi^{2m} h_i^{-1}}{\partial u_1}(u) & \frac{\partial y \phi^{2m} h_i^{-1}}{\partial u_2}(u) & \cdots & \frac{\partial y \phi^{2m} h_i^{-1}}{\partial u_m}(u) \end{bmatrix} \tag{3}$$

For some $u \in h_i \overline{W}_i (= \overline{B}(1))$ this matrix does not have full rank: we need to make it full rank by an arbitrarily small perturbation of y. We make this perturbation in stages. At each stage we alter the measurement function in such a way as to change one of the columns of (3), in particular to make it linearly independent of the columns to its left. We make such a change for each column in turn to produce, after a maximum of m such changes, a full rank matrix. Clearly, if we can make each of these changes arbitrarily small, the overall perturbation can also be made small.

To illustrate, let us suppose that the first s columns of (3) are linearly independent for all $u \in \overline{B}(1)$. Define as usual $\lambda : \mathbb{R}^m \to \mathbb{R}$ to be a bump function equal to 1 on $\overline{B}(1)$ and having support in $B(2)$. Also define $\psi : M \to \mathbb{R}$ by

$$\psi(x) = \begin{cases} \mu_{s+1}(x)\lambda(h_i(x)) & \text{if } x \in U_i \\ 0 & \text{otherwise.} \end{cases}$$

then $\psi(x) = \mu_{s+1}(x)$ if $x \in \overline{W}_i$, ψ has support in U_i, and $\psi \circ \phi^{-j}$ has support in $\phi^j U_i$. Now define $\psi_j : M \to \mathbb{R}$ by $\psi_j = \psi \circ \phi^{-j}$ for $0 \le j \le 2m$; the ψ_j's have disjoint support. Construct a measurement function y' according to

$$y' = y + \sum_{j=0}^{2m} a_{j+1} \psi_j.$$

Note that the bump function factors ensure that y' is a C^2 function $M \to \mathbb{R}$, and that $y'(x) = y(x)$ if x is not in one of the sets $U_i, \phi U_i, \ldots, \phi^{2m} U_i$.

What is the effect on the Jacobi matrix (3) of introducing the perturbations? It is clear that for $u \in h_i \overline{W}_i$, $\phi^k h_i^{-1}(u) \in \phi^k U_i$ so

$$\begin{aligned} y' \phi^k h_i^{-1}(u) &= y \phi^k h_i^{-1}(u) + a_{k+1}.\psi(\phi^{-k}\phi^k h_i^{-1}(u)) \\ &= y \phi^k h_i^{-1}(u) + a_{k+1}.\psi(h_i^{-1}(u)) \\ &= y \phi^k h_i^{-1}(u) + a_{k+1}.\mu_{s+1}(h_i^{-1}(u)) \\ &= y \phi^k h_i^{-1}(u) + a_{k+1}.u_{s+1} \end{aligned}$$

and hence

$$\frac{\partial y' \phi^k h_i^{-1}}{\partial u_{s+1}}(u) = \frac{\partial y \phi^k h_i^{-1}}{\partial u_{s+1}}(u) + a_{k+1}.$$

Thus we see that the only column of (3) which is affected by the perturbation is the $s+1$-th; and the effect on this column is to add the vector $(a_1, a_2, \ldots, a_{2m+1})^T$ to it. This is true whatever u is, so long as $u \in h_i \overline{W}_i$.

Now that we have seen how to adjust the $s+1$-th column of the Jacobi matrix, can we make it linearly independent of the first s columns? And if so, can we do this with an arbitrarily small perturbation? We can answer both these questions at once. Let $J_s(x)$ be the matrix formed from the first s columns of the Jacobi matrix of $\Phi_{(\phi,y)}$ at $x \in U_i$. Then by assumption, for $x \in \overline{W}_i$, $J_s(x)$ is full rank. Since J_s is a continuous function from U_i to the space of $(2m+1) \times s$ matrices, and the full rank matrices form an open subset of this space, there is an open set $X \subset U_i$, with $\overline{W}_i \subset X$, such that for every point in X, the first s columns of the Jacobi matrix of $\Phi_{(\phi,y)}$ are linearly independent.

Now define the function $\mathcal{S} : \mathbb{R}^s \times X \to \mathbb{R}^{2m+1}$ by

$$(\lambda_1, \ldots, \lambda_s, x) \mapsto \sum_{j=1}^{s} \lambda_j \begin{bmatrix} \frac{\partial y h_i^{-1}}{\partial u_j}(u) \\ \frac{\partial y \phi h_i^{-1}}{\partial u_j}(u) \\ \vdots \\ \frac{\partial y \phi^{2m} h_i^{-1}}{\partial u_j}(u) \end{bmatrix} - \begin{bmatrix} \frac{\partial y h_i^{-1}}{\partial u_{s+1}}(u) \\ \frac{\partial y \phi h_i^{-1}}{\partial u_{s+1}}(u) \\ \vdots \\ \frac{\partial y \phi^{2m} h_i^{-1}}{\partial u_{s+1}}(u) \end{bmatrix} \tag{4}$$

(see Bröcker and Jänich [12]). The function $\mathcal{S}$ is C^1, *because y and ϕ are assumed to be C^2*. Of course $s \leq m-1$ so that the dimension of $\mathbb{R}^s \times X$ is smaller than that of $\mathbb{R}^{2m+1}$. It follows from lemma 1 that the complement of $\mathcal{S}(\mathbb{R}^s \times X)$ is dense in $\mathbb{R}^{2m+1}$. In particular, we can find a vector $(a_1, \ldots, a_{2m+1})^T \in \mathbb{R}^{2m+1}$, with arbitrarily small norm, such that $(a_1, \ldots, a_{2m+1})^T \notin \mathcal{S}(\mathbb{R}^s \times X)$. With this choice for the numbers a_{j+1} in (3.4) the first $s+1$ columns of the Jacobi matrix of $\Phi_{(\phi,y')}$ must be linearly independent for all $x \in \overline{W}_i$.

So we have succeeded in finding a measurement function y' arbitrarily close to y which gives rise to a delay map of at least rank $s+1$ on $\overline{W}_i$. Obviously we can repeat the argument for each column of the Jacobi matrix in turn to find a measurement function giving rise to an immersion of $\overline{W}_i$. It may be worth emphasising the points that to make this argument we needed to assume that y and ϕ are C^2, and that the number of delays we use must be greater than $s+m$, which means that the minimum dimension for which immersions are dense is $2m$. As we shall see later, considerations of injectivity lead to the stronger condition that we must use $2m+1$ delays.

Since we can find immersions of $\overline{W}_i$ using arbitrarily small perturbations, lemma 5 shows that we can find one in $\mathcal{U}_y$ (which will be an embedding of $\cup_{j=1}^{k} \overline{W}_j$), and which is an immersion of any $\overline{W}_j$ with $j < i$. Finally, we can repeat the whole argument for $\overline{W}_{i+1}$ and so find an immersion of that as well. If we do this enough times we eventually immerse the whole of M.

We note in passing that this establishes the density of immersions, and hence their genericity, though we will make no explicit appeal to this result in the remainder of the proof. More important is the observation that immersions are locally embeddings. If $\Phi_{(\phi,y')}$ is an immersion of M, then for each point $x \in M$ there is an open neighbourhood N_x of x such that $\Phi_{(\phi,y')}$ is an embedding of N_x (the Inverse Function Theorem again). Thinking of M as a metric space, we can find a closed ball $\overline{\beta}_x$ centred at x and contained in N_x. The interiors of these balls form an open cover of M (since there is one for each x) from which we can extract a finite subcover. The corresponding finite collection of closed balls, say $\{\overline{\beta}_i : 1 \leq i \leq n'\}$, forms a compact cover, and each of the closed balls $\overline{\beta}_i$ is embedded by $\Phi_{(\phi,y')}$. If we select one of the balls, the set of measurement functions giving rise to embeddings of it is open, so the measurement functions giving embeddings of all the balls is also open, since it is the intersection of a finite number of open sets. We call this set $\mathcal{U}'_y$; it is a neighbourhood of y' and since $y' \in \mathcal{U}_y$ we can take $\mathcal{U}'_y \subset \mathcal{U}_y$. (Note the functions in $\mathcal{U}'_y$ are not usually embeddings of $\cup_{i=1}^{n} \overline{\beta}_i$: they embed each ball individually.) A well-known result from topology (Lebesgue's lemma [13]) tells us there is some number $\epsilon > 0$ such that a closed ball of radius ϵ centred at any point of M is contained in the interior of $\overline{\beta}_i$ for at least one i. It follows that every such ϵ-ball is embedded by $\Phi_{(\phi,y')}$, and indeed by all the maps $\Phi_{(\phi,\hat{y})}$ where $\hat{y} \in \mathcal{U}'_y$. If, following Takens, we call the metric on M, ρ, we can re-express this by saying that if $\hat{y} \in \mathcal{U}'_y$ then $\Phi_{(\phi,\hat{y})}$ is an immersion of M, an embedding of V_y, and $\Phi_{(\phi,\hat{y})}(x) \neq \Phi_{(\phi,\hat{y})}(x')$ whenever $x \neq x'$ and $\rho(x, x') \leq \epsilon$.

3.5 Stage 3: Embedding orbit segments

Having found an immersion of M we must now turn to the question of injectivity. For each $x \in M$ let us call the collection of points $\{x, \phi x, \ldots, \phi^{2m} x\}$ the *orbit segment of x*. We noted before that periodic orbits could cause particular problems to the making of an embedding, and for similar reasons pairs (x, x') where x' belongs to the orbit segment of x might also be problematic, since we cannot change the coordinates of $\Phi_{(\phi,y)}(x)$ without also changing some of those of $\Phi_{(\phi,y)}(x')$. Especially troublesome are points that lie on a periodic orbit of period less than or equal to $4m$; the segments $\{x, \phi x, \ldots, \phi^{2m} x\}$ and $\{x', \phi x', \ldots, \phi^{2m} x'\}$ can then overlap 'at both ends' (that is, $x = \phi^j x'$ for some $0 \leq j \leq 2m$ and $x' = \phi^k x$ for some $0 \leq k \leq 2m$). In view of these observations we shall tackle the matter of injectivity in two stages. The first of these creates a delay map with the property that no x (except those in P_{2m}) shares an image under $\Phi_{(\phi,y)}$ with another point on its orbit segment. In the second stage this limited form of injectivity is extended to injectivity on the whole of M.

We therefore start with the following lemma.

Lemma 6 *Let y' be such that $\Phi_{(\phi,y')}$ is an injective immersion on V_y. In every neighbourhood of y' in $C^2(M,\mathbb{R})$ there is a function, say y'' such that for every $x \in M$, and j in the range $1 \le j \le 2m$, $\Phi_{(\phi,y'')}(x) \neq \Phi_{(\phi,y'')}(\phi^j x)$ unless $x = \phi^j x$.*

Proof. We will establish this for each j in turn, so let us take j to be the smallest value for which the lemma is not already true.

Define the set S by $S = \cap_{i=0}^{2m} \phi^{-i} V_y$. Evidently for every $x \in S$ and $0 \le k \le 2m$, $\phi^k x \in V_y$, and because of the already established injectivity on V_y the lemma is already true for S. We need therefore only attend to points outside S.

Let T be the closure of the complement of S. Since S is a neighbourhood of P_{2m}, if $x \in T$, $x \notin P_{2m}$, and $\{x, \phi x, \ldots, \phi^{2m} x\}$ are all different, so we can find an open set U_x such that $U_x, \phi U_x, \ldots, \phi^{2m} U_x$ are all disjoint. We now have to consider two cases. If x is *not* a periodic point whose period lies between $2m+1$ and $4m$, then we can go further and find a U_x such that $U_x, \phi U_x, \ldots, \phi^{4m} U_x$ are all disjoint. As usual, we can assume that U_x is the domain of a good chart $(h_x U_x = B(3), W_x = h_x^{-1} B(1))$. Let us call this Case 1.

In the other case, (Case 2), x *is* a periodic point of period k, where $2m+1 \le k \le 4m$. We now find U_x such that $U_x, \phi U_x, \ldots, \phi^{k-1} U_x$ are all disjoint, and take U_x to be the domain of a good chart. We further define $X_x = W_x \cap \phi^{-k} W_x$: this set is open, and not empty since $x \in X_x$. (To simplify the notation, if x belongs to Case 1, write $X_x = W_x$.) We have arranged things so that, in Case 1, and Case 2 if $2m + j < k$, none of the sets $\phi^{2m+1}\overline{X}_x, \ldots, \phi^{2m+j}\overline{X}_x$ intersect $\cup_{l=0}^{2m} \phi^l U_x$; while in Case 2, if $2m + j \ge k$, none of $\phi^{2m+1}\overline{X}_x, \ldots, \phi^{k-1}\overline{X}_x$ intersect $\cup_{l=0}^{2m} \phi^l U_x$, and $\phi^k \overline{X}_x \subset \overline{W}_x, \phi^{k+1}\overline{X}_x \subset \phi \overline{W}_x, \ldots, \phi^{2m+j}\overline{X}_x \subset \phi^{2m+j-k}\overline{W}_x$.

From the collection $\{X_x : x \in T\}$ extract a finite cover of T; label the sets of this cover X_i, $i = 1, \ldots, N$, with (U_i, h_i) the corresponding charts. The procedure now is similar to the one followed to create an immersion: adjustments are made to the measurement function y' so that the property described in the lemma holds for all points in one of the $\overline{X}_i$. Further adjustments establish the property on more of the sets, and we show that these adjustments can all be made arbitrarily small, so preserving the property on those sets already dealt with. After a finite number of adjustments, the lemma is true on all of T.

Suppose that X_i is the next set with which we need to deal (that is, for $1 \le i' < i$, and every $x \in X_{i'}$, $\Phi_{(\phi,y')}(x) \neq \Phi_{(\phi,y')}(\phi^j x)$). Define $\psi : M \to \mathbb{R}$ by

$$\psi(x) = \begin{cases} \lambda(h_i x) & \text{for } x \in U_i \\ 0 & \text{otherwise} \end{cases}$$

and $\psi_l : M \to \mathbb{R}$, $l = 0, \ldots, 2m$ by $\psi_l = \psi \circ \phi^{-l}$. The support of ψ_l thus lies in $\phi^l U_i$. Finally define

$$y'' = y' + \sum_{l=0}^{2m} a_l \psi_l \tag{5}$$

(another version of equation (2)).

Now for all $x \in \overline{X}_i$, $x \in \overline{W}_i$ and $\phi^l x \in \phi^l \overline{W}_i$, $l = 0, \ldots, 2m$ so

$$\begin{array}{rcrcl} y''(x) & = & y'(x) & + & a_0 \\ y''(\phi x) & = & y'(\phi x) & + & a_1 \\ y''(\phi^2 x) & = & y'(\phi^2 x) & + & a_2 \\ & \vdots & & & \\ y''(\phi^{2m} x) & = & y'(\phi^{2m} x) & + & a_{2m}. \end{array}$$

The corresponding values for $\phi^j x$ depend on whether X_i is Case 1 or Case 2; we consider these separately:

Case 1, or Case2 with $2m + j < k$:

the points $\phi^j x, \phi^{j+1}x, \ldots, \phi^{2m}x$ lie in $\phi^j\overline{W}_i, \phi^{j+1}\overline{W}_i, \ldots, \phi^{2m}\overline{W}_i$ respectively, and the points $\phi^{2m+1}x, \ldots, \phi^{j+2m}x$ lie outside $\cup_{l=0}^{2m}\phi^l U_i$. Thus

$$\begin{array}{rcll} y''(\phi^j x) & = & y'(\phi^j x) & + \quad a_j \\ & \vdots & & \\ y''(\phi^{2m} x) & = & y'(\phi^{2m} x) & + \quad a_{2m} \\ y''(\phi^{2m+1} x) & = & y'(\phi^{2m+1} x) & \\ & \vdots & & \\ y''(\phi^{j+2m} x) & = & y'(\phi^{j+2m} x). & \end{array}$$

Case 2 with $2m + j \geq k$:

the points $\phi^j x, \phi^{j+1}x, \ldots, \phi^{2m}x$ lie in $\phi^j\overline{W}_i, \phi^{j+1}\overline{W}_i, \ldots, \phi^{2m}\overline{W}_i$ respectively, the points $\phi^{2m+1}x, \ldots, \phi^{k-1}x$ lie outside $\cup_{l=0}^{2m}\phi^l U_i$, and the points $\phi^k x, \ldots, \phi^{j+2m}x$ lie in $\overline{W}_i, \phi\overline{W}_i, \ldots, \phi^{j+2m-k}\overline{W}_i$ respectively. Thus

$$\begin{array}{rcll} y''(\phi^j x) & = & y'(\phi^j x) & + \quad a_j \\ & \vdots & & \\ y''(\phi^{2m} x) & = & y'(\phi^{2m} x) & + \quad a_{2m} \\ y''(\phi^{2m+1} x) & = & y'(\phi^{2m+1} x) & \\ & \vdots & & \\ y''(\phi^{k-1} x) & = & y'(\phi^{k-1} x) & \\ y''(\phi^{k} x) & = & y'(\phi^{k} x) & + \quad a_0 \\ & \vdots & & \\ y''(\phi^{j+2m} x) & = & y'(\phi^{j+2m} x) & + \quad a_{j+2m-k}. \end{array}$$

Let us concentrate on Case 2: the other is similar. We claim that we can find $a = (a_0, \ldots, a_{2m})^T$ with arbitrarily small norm, such that $\Phi_{(\phi,y'')}(x) \neq \Phi_{(\phi,y'')}(\phi^j x)$ for all $x \in \overline{X}_i$. Note that for any $x \in \overline{X}_i$

$$\begin{aligned} \Phi_{(\phi,y'')}(x) - \Phi_{(\phi,y'')}(\phi^j x) &= \Phi_{(\phi,y')}(x) - \Phi_{(\phi,y')}(\phi^j x) + \begin{bmatrix} a_0 - a_j \\ \vdots \\ a_{2m-j} - a_{2m} \\ a_{2m-j+1} \\ \vdots \\ a_{k-j} \\ a_{k-j+1} - a_0 \\ \vdots \\ a_{2m} - a_{2m+j-k} \end{bmatrix} \\ &= \Phi_{(\phi,y')}(x) - \Phi_{(\phi,y')}(\phi^j x) + Aa. \end{aligned}$$

The $(2m + 1) \times (2m + 1)$ matrix A has the following properties: every diagonal element is equal to 1; every row has at most one non-zero element apart from the one on the diagonal; and the same is true of every column. It is not hard to show that the rank of such a matrix is at least $m + 1$; let this rank be r, and let L be the r-dimensional subspace of $\mathbb{R}^{2m+1}$ which is the image of $\mathbb{R}^{2m+1}$ under A. Let $\pi : \mathbb{R}^{2m+1} \to L$ be the orthogonal projector onto L. Define $F : U_i \to L$ by $x \mapsto \pi(\Phi_{(\phi,y')}(x) - \Phi_{(\phi,y')}(\phi^j x))$. F is a C^2 function from the m-dimensional manifold U_i to the r-dimensional one L. Lemma 1 shows that there are vectors of arbitrarily small norm in L which are not contained in the image of U_i; let b be such a vector. If V is the orthogonal complement of the null space of A, there is a unique $b' \in V$ such that $b = Ab'$, and we can arrange for the norm of b' to be arbitrarily small by choosing $||b||$ to be sufficiently small. If we set $a = -b'$ in equation (5), then it follows that $\Phi_{(\phi,y'')}(x) - \Phi_{(\phi,y'')}(\phi^j x)$ cannot equal 0 for any $x \in X_i$ (for if it did we should have

$$0 = \Phi_{(\phi,y')}(x) - \Phi_{(\phi,y')}(\phi^j x) - Ab'$$

hence

$$0 = \pi(\Phi_{(\phi,y')}(x) - \Phi_{(\phi,y')}(\phi^j x)) - b$$

so that $b \in FU_i$). So $\Phi_{(\phi,y'')}(x) \neq \Phi_{(\phi,y'')}(\phi^j x)$ for all $x \in \overline{X}_i$.

Since $\overline{X}_i$ is compact, the functions in $C^2(M, \mathbb{R})$ which have the property mentioned in the last sentence form an open set. (Actually, this case is not quite covered by our usual Proposition 2, but it follows from very similar arguments, which the reader can easily supply.) Hence we can make a series of adjustments of the form (5) each of which establishes the property on one of the $\overline{X}_i$. A finite number of these is sufficient to generate a function y'' such that $\Phi_{(\phi,y'')}(x) \neq \Phi_{(\phi,y'')}(\phi^j x)$ for all $x \in T$. Finally, since T is compact, we can make further adjustments while retaining this property, and in particular can adjust y'' so that for all $x \in T$, $\Phi_{(\phi,y'')}(x) \neq \Phi_{(\phi,y'')}(\phi^j x)$ for every $1 \leq j \leq 2m$. We can, of course, make all these adjustments sufficiently small that $\Phi_{(\phi,y'')}$ is still an injective immersion on V_y.

So we have seen that for all $x \in T$, and $1 \leq j \leq 2m$, $\Phi_{(\phi,y'')}(x) \neq \Phi_{(\phi,y'')}(\phi^j x)$; also, for all $x \in S$, and $1 \leq j \leq 2m$, $\Phi_{(\phi,y'')}(x) \neq \Phi_{(\phi,y'')}(\phi^j x)$ unless $x = \phi^j x$. Together these observations establish the lemma.

As usual, the differentiability of y'' and ϕ allow us to extend the conclusion of the previous lemma: as well as concluding that x does not have the same image under $\Phi_{(\phi,y'')}$ as other points on its orbit segment, we can conclude that it does not have the same image as points on *nearby* orbit segments. This is formalized in the next lemma.

Lemma 7 *Let y'' be a function as promised in the previous lemma. There is a number $\delta > 0$, such that if $x, x' \in M$, $x \neq x'$, and $\rho(\phi^i x, \phi^j x') < \delta$ for some $0 \leq i, j \leq k$, then $\Phi_{(\phi,y'')}(x) \neq \Phi_{(\phi,y'')}(x')$.*

Proof. We aim to prove the lemma by contradiction. Let $\delta_n \to 0$ be a sequence of positive numbers tending to zero. If the lemma were not true we should be able to find, for each n, a pair of points x_n, x'_n, $x_n \neq x'_n$, and integers i_n, j_n, $0 \leq i_n, j_n \leq k$, such that $\rho(\phi^{i_n} x_n, \phi^{j_n} x'_n) < \delta_n$ and $\Phi_{(\phi,y'')}(x_n) = \Phi_{(\phi,y'')}(x'_n)$.

By the compactness of M we can find subsequences $\{x_n\}$ and $\{x'_n\}$ which have limits, say x and x' respectively. Also, since the number of values that i_n can take is finite, we can find an infinite subsequence in which all the i_n's have the same value, say i; we can similarly take $j_n = j$. By continuity $\phi^i x_n \to \phi^i x$ and $\phi^j x'_n \to \phi^j x'$, and since $\rho(\phi^i x_n, \phi^j x'_n) \to 0$ we see that $\phi^i x = \phi^j x'$; this means one of x, x' lies on the orbit segment of the other. But also by continuity we see that $\Phi_{(\phi,y'')}(x) = \Phi_{(\phi,y'')}(x')$, so by lemma 6 it follows that $x = x'$.

Since x_n and x'_n tend to the same limit, we can find n large enough so that $\rho(x_n, x'_n) < \epsilon$. ($\epsilon$ is the quantity described at the end of the last subsection.) But then $x_n \neq x'_n$, $\rho(x_n, x'_n) < \epsilon$ and $\Phi_{(\phi,y'')}(x_n) = \Phi_{(\phi,y'')}(x'_n)$, which is a contradiction.

So we have been partly successful at creating an injection on M: distinct points are not mapped to the same image by $\Phi_{(\phi,y'')}$ if their orbit segments are sufficiently close together. Naturally, we construct y'' so that it lies in $\mathcal{U}'_y$; and as usual, the properties possessed by y'', described in the preceding lemma, are shared by all the measurement functions in an open neighbourhood of y''. Pairs of points separated by larger distances, however, require a rather different approach.

3.6 Stage 4: Making an injection on M

There are several ways in which we might approach the problem of injectivity; we might consider the product space $M \times M$ and remove from it the diagonal $\Delta' = \{(x, x) : x \in M\}$: then $\Phi_{(\phi,y'')}$ is injective if and only if the image of $(M \times M) \backslash \Delta'$ under the function $(x, x') \mapsto (\Phi_{(\phi,y'')}(x), \Phi_{(\phi,y'')}(x'))$ does not intersect the diagonal, Δ, of $\mathbb{R}^{2m+1} \times \mathbb{R}^{2m+1}$. Since Δ is a submanifold of $\mathbb{R}^{2m+1} \times \mathbb{R}^{2m+1}$, we might hope to proceed using considerations of transversality. This is the approach followed by Takens.

Another thing we might try is to look at the map $(x, x') \mapsto (\Phi_{(\phi,y'')}(x) - \Phi_{(\phi,y'')}(x'))$. $\Phi_{(\phi,y'')}$ is injective if and only if the image of $(M \times M)\backslash\Delta'$ does not contain $\mathbf{0}$. This turns out to be very similar to the approach of Takens, and is the one we take here.

It has to be admitted that this part of the argument is rather involved; a version of it takes up the bulk of Takens' published form of his proof. In outline, we remove the set $\mathrm{int}V_y$ from M, and map the resulting set with $\phi, \phi^2, \ldots, \phi^{2m}$. Taking the union of the images, we find a covering of it, which has some special properties, and make a partition of unity. The functions in the partition form the ψ's of equation (2), and we find we can adjust the measurement function so that $\Phi_{(\phi,y'')}$ becomes injective on M.

Recall that we have found a y'' such that $\Phi_{(\phi,y'')}$ is an injective immersion of V_y, and $\Phi_{(\phi,y'')}(x) \neq \Phi_{(\phi,y'')}(x')$ if $x, x' \in M$, $x \neq x'$ and $\rho(\phi^i x, \phi^j x') < \delta$ for some $0 \leq i, j \leq k$.

The set $M\backslash\mathrm{int}\ V_y$ is compact, and so is the set Z defined by

$$Z = \bigcup_{j=0}^{2m} \phi^j(M\backslash\mathrm{int}\ V_y).$$

For reasons which will become clear shortly, we shall need a finite open covering of Z, $\{U_l, l = 1, \ldots, N\}$ with the following two properties

1. for each $l = 1, \ldots, N$ and $0 \leq i, j \leq 2m$, $\phi^{-i}U_l \cap \phi^{-j}U_l = \emptyset$ unless $i = j$,

2. for each $l = 1, \ldots, N$ the diameter of U_l is less than δ.

To see how to construct such a cover, note first that for every $x \in Z$, $x \notin P_{2m}$ so that the points $x, \phi^{-1}x, \ldots, \phi^{-2m}x$ are all distinct, and so we can find an open set U_x containing x such that $U_x, \phi^{-1}U_x, \ldots, \phi^{-2m}U_x$ are all disjoint. So the collection $\{U_x : x \in Z\}$ has property 1. And we can easily ensure that $U_x \subset b(x, \delta/2)$ so that it also has property 2. From $\{U_x : x \in Z\}$, which covers Z, we extract a finite subcover $\{U_l : l = 1, \ldots, N\}$. We build a partition of unity on Z, subordinate to this cover. Say the functions in the partition of unity are $\psi_l : M \to \mathbb{R}$, $l = 1, \ldots, N$.

This is a slightly different arrangement to those we have seen so far, but the adjustment we are going to make to y'' will still be of the form of equation (2), and in particular the conclusion of lemma 5 will still hold. As a matter of notation, we define

$$y_\epsilon = y'' + \sum_{l=1}^{N} \epsilon_l \psi_l \tag{6}$$

for $\epsilon \in \mathbb{R}^N$.

We saw above that $\Phi_{(\phi,y'')}$ is injective on M if and only if the image of $M \times M\backslash\Delta$ under the map $(x, x') \mapsto \Phi_{(\phi,y'')}(x) - \Phi_{(\phi,y'')}(x')$ does not contain $\mathbf{0}$. This suggests we should attempt to adjust y'' so that the image is shifted away from $\mathbf{0}$ if it contains it. It will turn out that be sure of being able to make such an adjustment the dimensionality of the codomain of $\Phi_{(\phi,y'')}$ will have to be sufficiently large: in particular, greater than or equal to $2m + 1$. Let us now go through this in detail.

Consider the subset of $M \times M$, W, given by

$$W = \{(x, x') : \rho(\phi^i x, \phi^j x') \geq \delta \text{ for all } 0 \leq i, j \leq k, \text{ and either } x \text{ or } x' \notin \text{ the interior of } V_y\}.$$

Note that W is closed. The reason for introducing this set rather than working with $M \times M$ directly is that certain points are eliminated (such as (x, x)) where the construction below fails to work—this is why these points were dealt with separately beforehand. The idea is to show that appropriate choices of ϵ in (6) will yield delay maps which carry W into $\mathbb{R}^{2m+1}\backslash\{\mathbf{0}\}$. To this end we consider the map $\Psi : M \times M \times \mathbb{R}^N \to \mathbb{R}^{2m+1}$ defined by

$$\Psi(x, x', \epsilon_1, \epsilon_2, \ldots, \epsilon_N) = \Phi_{(\phi,y_\epsilon)}(x) - \Phi_{(\phi,y_\epsilon)}(x') \tag{7}$$

and investigate the inverse image of $\mathbf{0}$ under Ψ, by using lemma 2 and submersivity of Ψ. We shall then argue that the y_ϵ so found give injective delay maps on M.

We need to establish the submersivity of Ψ at all the points in the set $W \times \{0\}$[3]—quite why this is appropriate will become clear later on. To specify the derivative of Ψ we need to assign charts to $M \times M \times \mathbb{R}^N$. Let $\{(h_p, V_p) : p \in \Lambda\}$ be an atlas for M, then $\{(g_{p,q}, V_p \times V_q \times \mathbb{R}^N) : p, q \in \Lambda\}$, where $g_{p,q}(x, x', \epsilon) = (h_p(x), h_q(x'), \epsilon)$, is an atlas for $M \times M \times \mathbb{R}^N$. Using these charts, (and writing $h_p(x) = u$, $h_q(x') = u'$) we can see from the definition (7) that the derivative $D\Psi g_{p,q}^{-1}(u, u', 0)$ takes the form

$$\begin{array}{|c|c|c|} \hline D\Phi_{(\phi,y')}h_p^{-1}(u) & -D\Phi_{(\phi,y')}h_q^{-1}(u') & A(x) - A(x') \\ \hline \end{array} \tag{8}$$

where $A(x)$ is a $(2m+1) \times N$ matrix whose elements are given by

$$A_{il}(x) = \frac{\partial y_\epsilon \phi^{i-1} h_p^{-1}}{\partial \epsilon_l}(u) = \psi_l \phi^{i-1} h_p^{-1}(u) = \psi_l \phi^{k-1}(x). \tag{9}$$

(As it turns out, the form of $A(x)$ does not depend on the chart map.) Our task is to show that if $(x, x') \in W$ the columns of (8) form a basis for $\mathbb{R}^{2m+1}$. The immersivity property of y'' implies that the first $2m$ columns of (8) span at least an m-dimensional subspace of $\mathbb{R}^{2m+1}$; but since we aim to establish injectivity by adjusting ϵ, we discard these columns and concentrate on the submatrix $A(x) - A(x')$. We will show that this matrix has $2m+1$ independent columns. From (9)

$$A_{il}(x) - A_{il}(x') = \psi_l \phi^{i-1}(x) - \psi_l \phi^{i-1}(x'). \tag{10}$$

We need the following observations. To begin with, we can show that each column of the matrix $A(x) - A(x')$ has at most one non-zero element. For if, for some l, there were distinct i,j such that $A_{il}(x) - A_{il}(x')$ and $A_{jl}(x) - A_{jl}(x')$ were both non-zero, this would mean at least one of $A_{il}(x), A_{il}(x')$ was non-zero, and at least one of $A_{jl}(x), A_{jl}(x')$ was non-zero. But $A_{il}(x)$ and $A_{jl}(x)$ cannot both be non-zero, for that would imply $\psi_l \phi^{i-1}(x) \neq 0$ and $\psi_l \phi^{j-1}(x) \neq 0$ so that both $\phi^{i-1}(x)$ and $\phi^{j-1}(x)$ are in the support of ψ_l and hence in the same element U_l of the cover. But then $\phi^{-(i-1)}U_l$ and $\phi^{-(j-1)}U_l$ are not disjoint, contradicting property 1. We can also see that $A_{il}(x)$ and $A_{jl}(x')$ cannot both be non-zero, for then $\phi^{i-1}(x)$ and $\phi^{j-1}(x')$ are both in U_l, implying $\rho(\phi^{i-1}(x), \phi^{j-1}(x')) < \delta$ and so $(x, x') \notin W$. Similar arguments dispense with the other two possibilities ($A_{il}(x')$ and $A_{jl}(x)$ both non-zero, or $A_{il}(x')$ and $A_{jl}(x')$ both non-zero.)

We now show that every row of $A(x) - A(x')$ has at least one non-zero element. Assume $(x, x') \in W$, then at least one of x, x' is in $M \backslash \mathrm{int}\ V_y$. For definiteness say that $x \in M \backslash \mathrm{int}\ V_y$; then $\phi^{i-1}(x) \in Z$ for $1 \leq i \leq 2m+1$. Hence $\sum_{l=1}^{N} \psi_l \phi^{i-1}(x) = 1$ (because the $\{\psi_l\}$ form a partition of unity), so that for every $1 \leq i \leq 2m+1$ there must be some l, $1 \leq l \leq N$ such that $\psi_l \phi^{i-1}(x) \neq 0$; that is, for every i there is an l such that $A_{il}(x)$ is not zero. But note also that if $A_{il}(x) \neq 0$, then $A_{il}(x')$ must be zero, for otherwise $\phi^{i-1}x$ and $\phi^{i-1}x'$ would both lie in the support of ψ_l and hence in a ball of radius $\delta/2$, implying $(x, x') \notin W$. So if $A_{il}(x)$ is not zero then $A_{il}(x) - A_{il}(x')$ is not zero either.

Since $A(x) - A(x')$ has at least one non-zero element in every row, but no more than one non-zero element in every column, two things follow: the matrix must have at least as many columns as rows; and it must be full rank. Hence the rank of $D\Psi g_{p,q}^{-1}(u, u', 0)$ must be $2m+1$, and Ψ is submersive at $(x, x', 0)$.

Actually we can go further than this. Since the derivative of Ψ is full rank at $(x, x', 0)$, by continuity there is an open subset of $M \times M \times \mathbb{R}^N$ containing $(x, x', 0)$ throughout which the derivative is full rank. These open sets form a cover of $W \times \{0\}$ and their union is an open set X such that the restriction of Ψ to X is a submersion. By Lebesgue's lemma there is an $\eta > 0$ such that every closed ball of radius η or less, centred at a point in $W \times \{0\}$ is contained in X. Note in particular that this means that if $\epsilon \in \mathbb{R}^N$ and $\|\epsilon\| \leq \eta$, then $W \times \{\epsilon\} \subset X$.

Since $\Psi|_X : X \to \mathbb{R}^{2m+1}$ is a submersion, lemma 2 shows that $\Psi|_X^{-1}(\mathbf{0})$ is a submanifold of X, of dimension $2m + N - (2m+1) = N - 1$. Consider the projection $\pi : X \to \mathbb{R}^N$, $(x, x', \epsilon) \mapsto \epsilon$, and its restriction, $\hat{\pi}$, to $\Psi|_X^{-1}(\mathbf{0})$. Suppose that there is some ϵ, with $\|\epsilon\| < \eta$, which is *not* in the range of $\hat{\pi}$: then there is no pair $(x, x') \in W$ such that $\Phi_{(\phi, y_\epsilon)}(x) = \Phi_{(\phi, y_\epsilon)}(x')$ (for if there were, since $(x, x', \epsilon) \in X$ then $(x, x', \epsilon) \in \Psi|_X^{-1}(\mathbf{0})$ implying ϵ is in the range of $\hat{\pi}$). Moreover, since $\hat{\pi}$ is a C^1 map from a manifold of dimension $N-1$ to $\mathbb{R}^N$, lemma 1 shows that ϵ's not in the range of $\hat{\pi}$ are dense in $\mathbb{R}^N$; in particular, we can find such ϵ's with arbitrarily small norm.

[3]Note that this $0 \in \mathbb{R}^N$, whereas in the previous paragraph $\mathbf{0}$ stood for the origin of $\mathbb{R}^{2m+1}$.

So we have seen that the image of $W \times \{\epsilon\}$ under Ψ does not contain $\mathbf{0}$. The pairs (x, x') in $M \times M$ missing from W are those for which $\rho(\phi^i x, \phi^j x') < \delta$ for some $0 \leq i, j \leq k$, or for which both $x, x' \in \text{ int } V_y$. But we already know that if ϵ is chosen small enough, then for all these missing pairs, $\Psi(x, x', \epsilon) \neq \mathbf{0}$ unless $x = x'$. So $\Psi(x, x', \epsilon) \neq \mathbf{0}$ for all $(x, x') \in M \times M$, unless $x = x'$. This means $\Phi_{(\phi, y_\epsilon)}$ is injective on M.

So $\Phi_{(\phi, y_\epsilon)}$ is an injective immersion, hence an embedding of M. We need only note that we can make the norm of ϵ sufficiently small to ensure that $y_\epsilon \in \mathcal{N}$ to complete the proof of Theorem 2.

The reader will not have failed to notice that for the above argument to work, the number of delays in the delay map must be at least $2m + 1$: if it is less than this, the dimension of $\Psi|_X^{-1}(\mathbf{0})$ is greater than or equal to N, and so we cannot apply lemma 1 to the projection $\hat{\pi}$. Of course, this does not by itself prove that $2m + 1$ delays is the minimum we can get away with for Takens' theorem to hold in general; but it is not too difficult to devise examples which demonstrate that this number is indeed the minimum.

3.7 Proof of Theorem 1

For a diffeomorphism of M having special properties (to do with the points in P_{2m}) we have seen that the measurement functions y such that $\Phi_{(\phi, y)}$ embeds M are open and dense in $C^2(M, \mathbb{R})$ (in the C^1 topology). Can we say anything about diffeomorphisms in general? Theorem 1 states that pairs (ϕ, y) such that $\Phi_{(\phi, y)}$ is an embedding are open and dense in $\mathit{Diff}^2(M) \times C^2(M, \mathbb{R})$. Roughly speaking, this means that given any C^2 diffeomorphism, and a measurement function, there is a diffeomorphism arbitrarily close to the first (and a measurement function, similarly) which give rise to an embedding of M. So the situation described by Theorem 2 is not exceptional. This happens because diffeomorphisms which satisfy the conditions demanded in Theorem 2 are themselves unexceptional, that is, are generic.

This genericity is the essential instrument for proving the denseness part of Theorem 1 from Theorem 2. The thing we need for this is provided by a well-known theorem in nonlinear dynamical systems theory, the *Kupka-Smale Theorem* [16, 17], which we may summarise as follows:

Theorem 3 (Kupka-Smale) *Let M be a compact manifold, and n be a finite positive integer. For generic $\phi \in \mathit{Diff}^2(M)$, the number of periodic points, with period n or less, is finite.*

This is really only part of the theorem. Another part of the theorem tells us that we make take the periodic points to be *hyperbolic*[4], and by a relatively simple extension we can conclude that for generic ϕ the fixed points have distinct eigenvalues, and the periodic points have distinct eigenvalues when considered as fixed points of ϕ^k. We need not take the trouble to prove the Kupka-Smale theorem here: an extensive discussion is given by Palis and de Melo [18].

Let A be the set of ϕ's, with finite numbers of periodic points with period $\leq 2m$, and having distinct eigenvalues; the Kupka-Smale theorem tells us that $A \subset \mathit{Diff}^2(M)$ is open and dense. Now let X and Y by topological spaces, and say V is a subset of $X \times Y$ with the following property: there is a dense subset A of X such that, for every $x \in A$ there is an open dense subset of Y, O_x such that $\{(x, y) : y \in O_x\} \subset V$. Then it is not difficult to see that V is dense in $X \times Y$. If we interpret X to be $\mathit{Diff}^2(M)$ and Y to be $C^2(M, \mathbb{R})$ then we immediately deduce the denseness part of Theorem 1.

The openness part does not fall into our hands so easily. Even if A in the last paragraph is open this does not necessarily mean that V will be. It seems that we will have to establish openness from scratch. We can do this in the same sort of way that we proved Proposition 2, but the details are more complicated now because we need to consider neighbourhoods in $\mathit{Diff}^2(M) \times C^2(M, \mathbb{R})$ rather than just $C^2(M, \mathbb{R})$. The argument is given below; its construction parallels that of subsection 3.1.

[4] A fixed point x of ϕ is hyperbolic if $Dg\phi g^{-1}(gx)$ has no eigenvalues of unit modulus. A periodic point x of period k is hyperbolic if it is a hyperbolic fixed point of ϕ^k.

3.8 The openness part of Theorem 1

For each positive integer p every pair in the space $Diff^2(M) \times C^2(M, \mathbb{R})$ gives rise to a delay map $(\phi, y) \mapsto (y, \ldots, y\phi^p)$, so for fixed p we can define a mapping

$$\mathcal{F}^{(2)} : Diff^2(M) \times C^2(M, \mathbb{R}) \to C^2(M, \mathbb{R}^n)$$

by $(\phi, y) \mapsto \Phi_{(\phi,y)}$[5]. We need to show that this mapping is continuous. As before, we shall tackle this in stages.

Lemma 8 *The function* F_1 *:*$Diff^2(M) \times C^2(M, \mathbb{R}) \to C^2(M, \mathbb{R})$ *defined by* $(\phi, y) \mapsto y \circ \phi$ *is continuous.*

Proof. Let $\{(U_i, h_i); i \in \Lambda\}$ be a finite good atlas for M, and say $W_i = h_i^{-1}B(1)$. Then the collection $\{\phi^{-1}W_i; i \in \Lambda\}$ is a cover of M. Let $\{(V_j, g_j); j \in \Theta\}$ be another finite good atlas, subordinate to the cover $\{\phi^{-1}W_i; i \in \Lambda\}$, with $X_j = g_j^{-1}B(1)$. Then for each V_j there is some W_i, which we call $W_{i(j)}$, such that $\phi V_j \subset W_{i(j)}$.

For any (ϕ, y), observe that the functions $yh_i^{-1} : h_i\overline{W_i} \to \mathbb{R}$ are uniformly continuous, so given $\epsilon > 0$ there is a $\delta_i > 0$ such that $|yh_i(u') - yh_i(u)| < \epsilon$ if $\|u' - u\| < \delta_i$; and since there is a finite number of these functions, we can find $\delta > 0$ which works for all $i \in \Lambda$.

The derivatives $Dyh_i^{-1} : h_i\overline{W}_i \to \mathbb{R}^m$ and $Dh_i\phi g_j^{-1} : g_j\overline{X}_j \to \mathbb{R}^{m\times m}$ are continuous and have compact domains. This means their norms are bounded: we can find constants A and B, such that $\|Dyh_i^{-1}(u)\| < A$ for all $u \in h_i\overline{W}_i$ and $i \in \Lambda$, and $\|Dh_i\phi g_j^{-1}(u)\| < B$ for all $u \in g_j\overline{X}_j$, $j \in \Theta$ and $i \in \Lambda$. Also, since the continuity is uniform, given $\epsilon > 0$ we can find $\delta > 0$ such that $\|Dyh_i^{-1}(u') - Dyh_i^{-1}(u)\| < \epsilon$ for all $\|u' - u\| < \delta$.

Now given any neighbourhood in $C^2(M, \mathbb{R})$ of $y \circ \phi$, there is a neighbourhood of the form $\mathcal{N} = \cap_j \mathcal{N}^1(y \circ \phi; (V_j, g_j), (\mathbb{R}, id), \overline{X_j}, \epsilon')$ contained within it. Choose δ sufficiently small that the following are satisfied:

$|yh_i^{-1}(u') - yh_i^{-1}(u)| < \epsilon'/2$ for all $\|u' - u\| < \delta$, $u, u' \in \overline{W}_i$, and $i \in \Lambda$,

$\|Dyh_i^{-1}(u') - Dyh_i^{-1}(u)\| < \epsilon'/3B$ for all $\|u' - u\| < \delta$, $u, u' \in \overline{W}_i$, and $i \in \Lambda$,

$\delta < \epsilon'/3A$, and $\delta < B$.

Also choose $\epsilon < \min\{\epsilon'/2, \epsilon'/6B\}$. Now consider the open neighbourhood $\mathcal{N}^1(\delta, \epsilon)$ of (ϕ, y) in $Diff^2(M) \times C^2(M, \mathbb{R})$, where

$$\mathcal{N}^1(\delta, \epsilon) = \bigcap_j \mathcal{N}^1(\phi; (V_j, g_j), (W_{i(j)}, h_{i(j)}), \overline{X}_j, \delta) \times \bigcap_i \mathcal{N}^1(y; (U_i, h_i), (\mathbb{R}, id), \overline{W}_i, \epsilon).$$

To show that F_1 is continuous, we show that if $(\hat{\phi}, \hat{y}) \in \mathcal{N}^1(\delta, \epsilon)$ then $F_1(\hat{\phi}, \hat{y}) \in \mathcal{N}$; to do this we have to show that the conditions (1) are satisfied.

Let $(\hat{\phi}, \hat{y}) \in \mathcal{N}^1(\delta, \epsilon)$, $j \in \Theta$, $x \in \overline{X}_j$ and $u = g_jx$. Then

$$|\hat{y}\hat{\phi}g_j^{-1}(u) - y\phi g_j^{-1}(u)| \leq |\hat{y}\hat{\phi}g_j^{-1}(u) - y\hat{\phi}g_j^{-1}(u)| + |y\hat{\phi}g_j^{-1}(u) - y\phi g_j^{-1}(u)| \tag{11}$$

If we say $u' = h_{i(j)}\hat{\phi}g_j^{-1}(u)$ then

$$|\hat{y}\hat{\phi}g_j^{-1}(u) - y\hat{\phi}g_j^{-1}(u)| = |\hat{y}h_{i(j)}^{-1}(u') - yh_{i(j)}^{-1}(u')| < \epsilon < \epsilon'/2 \tag{12}$$

since $\hat{y} \in \cap_i \mathcal{N}^1(y; (U_i, h_i), (\mathbb{R}, id), \overline{W}_i, \epsilon)$. Also, if $u'' = h_{i(j)}\phi g_j^{-1}(u)$ then

$$|y\hat{\phi}g_j^{-1}(u) - y\phi g_j^{-1}(u)| = |yh_{i(j)}^{-1}(u') - yh_{i(j)}^{-1}(u'')| < \epsilon'/2 \tag{13}$$

[5]Note that this mapping is not necessarily injective: to take a rather trivial example, consider the diffeomorphisms, f,g of the circle, defined by $f(\theta) = \theta$ and $g(\theta) = \theta + \pi$, together with the measurement function $y(\theta) = \sin 2\theta$; then $\Phi_{(f,y)} = \Phi_{(g,y)}$.

since $\|u' - u''\| = \|h_{i(j)}\phi g_j^{-1}(u) - h_{i(j)}\phi g_j(u)\| < \delta$. Combining (11), (12) and (13) gives

$$|\hat{y}\hat{\phi}g_j^{-1}(u) - y\phi g_j^{-1}(u)| < \epsilon'.$$

The derivatives are dealt with as follows:

$$\|D\hat{y}\hat{\phi}g_j^{-1}(u) - Dy\phi g_j^{-1}(u)\| \leq \|D\hat{y}\hat{\phi}g_j^{-1}(u) - Dy\hat{\phi}g_j^{-1}(u)\| + \|Dy\hat{\phi}g_j^{-1}(u) - Dy\phi g_j^{-1}(u)\|$$

$$= \|D\hat{y}h_i^{-1}h_i\hat{\phi}g_j^{-1}(u) - Dyh_i^{-1}h_i\hat{\phi}g_j^{-1}(u)\| + \|Dyh_i^{-1}h_i\hat{\phi}g_j^{-1}(u) - Dyh_i^{-1}h_i\phi g_j^{-1}(u)\|$$

(the j dependence of i has been suppressed in the last line). Using the Chain Rule (and u' and u'' as above) this becomes

$$\begin{aligned}
\|D\hat{y}\hat{\phi}g_j^{-1}(u) - Dy\phi g_j^{-1}(u)\| &\leq \|D\hat{y}h_i^{-1}(u')Dh_i\hat{\phi}g_j^{-1}(u) - Dyh_i^{-1}(u')Dh_i\hat{\phi}g_j^{-1}(u)\| \\
&+ \|Dyh_i^{-1}(u')Dh_i\hat{\phi}g_j^{-1}(u) - Dyh_i^{-1}(u'')Dh_i\phi g_j^{-1}(u)\| \\
&\leq \|D\hat{y}h_i^{-1}(u')Dh_i\hat{\phi}g_j^{-1}(u) - Dyh_i^{-1}(u')Dh_i\hat{\phi}g_j^{-1}(u)\| \\
&+ \|Dyh_i^{-1}(u')Dh_i\hat{\phi}g_j^{-1}(u) - Dyh_i^{-1}(u')Dh_i\phi g_j^{-1}(u)\| \\
&+ \|Dyh_i^{-1}(u')Dh_i\phi g_j^{-1}(u) - Dyh_i^{-1}(u'')Dh_i\phi g_j^{-1}(u)\|
\end{aligned}$$

$$\begin{aligned}
\leq \|D\hat{y}h_i^{-1}(u') - Dyh_i^{-1}(u')\|\|Dh_i\hat{\phi}g_j^{-1}(u)\| &+ \|Dh_i\hat{\phi}g_j^{-1}(u) - Dh_i\phi g_j^{-1}(u)\|\|Dyh_i^{-1}(u')\| \\
&+ \|Dyh_i^{-1}(u') - Dyh_i^{-1}(u'')\|\|Dh_i\phi g_j^{-1}(u)\|.
\end{aligned}$$

Now $\hat{\phi} \in \cap_j \mathcal{N}^1(\phi; (V_j, g_j), (W_{i(j)}, h_{i(j)}), \overline{X}_j, \delta)$ so $\|Dh_i\hat{\phi}g_j^{-1}(u) - Dh_i\phi g_j^{-1}(u)\| < \delta$ and since $\delta < B$ this implies $\|Dh_i\hat{\phi}g_j^{-1}(u)\| < 2B$. Hence the inequality above becomes

$$\|D\hat{y}\hat{\phi}g_j^{-1}(u) - Dy\phi g_j^{-1}(u)\| \leq \epsilon.2B + \delta.A + \frac{\epsilon'}{3B}.B < \epsilon'.$$

What all this shows is that if $(\hat{\phi}, \hat{y}) \in \mathcal{N}(\delta, \epsilon)$ then $\hat{y} \circ \hat{\phi} \in \mathcal{N}$. That is, $F_1\mathcal{N}(\delta, \epsilon) \subset \mathcal{N}$, and so F_1 is continuous.

Lemma 9 *The function F_n :Diff$^2(M) \times C^2(M, \mathbb{R}) \to C^2(M, \mathbb{R})$ defined by $(\phi, y) \mapsto y \circ \phi^n$ is continuous.*

Proof. This is done by induction. We already know that F_1 is continuous. Assume that F_{n-1} is continuous, and note that $F_n = F_{n-1} \circ G$, where $G : \textit{Diff}^2(M) \times C^2(M, \mathbb{R}) \to \textit{Diff}^2(M) \times C^2(M, \mathbb{R})$ is defined by $G(\phi, y) = (\phi, F_1(\phi, y))$. Lemma 8 shows that G is continuous, so F_n is the composition of continuous functions.

Proposition 3 *$\mathcal{F}^{(2)}$ is continuous.*

Proof. Let us say $F_0 : \textit{Diff}^2(M) \times C^2(M, \mathbb{R}) \to C^2(M, \mathbb{R})$ is given by $(\phi, y) \mapsto y$, and also that $F : \textit{Diff}^2(M) \times C^2(M, \mathbb{R}) \to [C^2(M, \mathbb{R})]^{p+1}$ is specified by its components: $F = (F_0, F_1, \ldots, F_p)$. Then $\mathcal{F}^{(2)} = T \circ F$, where T is as defined in Proposition 1. The previous lemma shows that F is continuous, so we need only show that T is. This was done in the proof of Proposition 1.

So now we can prove the analogue of Proposition 2:

Proposition 4 *Let M be a compact manifold, and K a compact subset of M. Then the set of pairs (ϕ, y) such that the delay map $\Phi_{(\phi,y)} : M \to \mathbb{R}^{2m+1}$ is immersive on K is open in* Diff$^2(M) \times C^2(M, \mathbb{R})$. *The same is true of injective immersions (embeddings) of K.*

This is proved using the continuity of $\mathcal{F}^{(2)}$, in just the same way as Proposition 2. This completes the proof of the openness part of Theorem 1 and so of Theorem 1 itself.

4 Extensions to Takens' Theorem

General though Theorem 1 is, it does not, of course, answer every question we might think to ask about the reconstruction of phase spaces of dynamical systems. Attempts have therefore been made to extend the theorem in various directions. Most notable among these attempts so far has been the work of Sauer, Yorke and Casdagli, who shifted fundamentally the viewpoint of the theorem, translating it from the setting of differential topology to a more geometrical and measure-theoretic one. This enabled them to make a number of detailed statements about aspects of reconstruction not addressed in Takens' paper. We will summarize these results below, but first we look at some extensions which stem more directly from Theorems 1 and 2.

4.1 Relaxing the C^2 Condition

We have assumed so far that all the diffeomorphisms ϕ and measurements functions y are C^2; but on the other hand we have only used the C^1 topologies on $\mathit{Diff}^2(M)$ and $C^2(M,\mathbb{R})$. Can the condition that the ϕ's and y's be twice continuously differentiable be relaxed to just once continuously differentiable?

It is not very hard to see that it can. It is well known [15] that the set of twice continuously differentiable functions is dense in $C^1(M,\mathbb{R})$ where the latter has the C^1 topology, and that correspondingly the set of C^2 diffeomorphisms is dense in $\mathit{Diff}^1(M)$. It follows that $\mathit{Diff}^2(M)\times C^2(M,\mathbb{R})$ is dense in $\mathit{Diff}^1(M)\times C^1(M,\mathbb{R})$. The C^1 topology on $C^2(M,\mathbb{R})$ is simply the induced topology when $C^2(M,\mathbb{R})$ is considered as a subset of $C^1(M,\mathbb{R})$, and similarly for $\mathit{Diff}^2(M)$ and $\mathit{Diff}^1(M)$, so that $\mathit{Diff}^2(M)\times C^2(M,\mathbb{R})$ has the induced topology from $\mathit{Diff}^1(M)\times C^1(M,\mathbb{R})$. Now Theorem 1 tells us that the set of pairs (ϕ, y) giving rise to delay maps which are embeddings of M is dense in $\mathit{Diff}^2(M)\times C^2(M,\mathbb{R})$; and since the latter is dense in $\mathit{Diff}^1(M)\times C^1(M,\mathbb{R})$ (and has the induced topology) these pairs must be dense in $\mathit{Diff}^1(M)\times C^1(M,\mathbb{R})$ as well.

To show that the set of pairs is also open, we need to show that the function

$$\mathcal{F}^{(1)}:\mathit{Diff}^1(M)\times C^1(M,\mathbb{R})\to C^1(M,\mathbb{R}^n)$$

defined by $(\phi,y)\mapsto\Phi_{(\phi,y)}$ is continuous. In sections 3.1 and 3.8 we showed this for $\mathcal{F}^{(2)}$, but examination of the arguments given there shows that it does not make use of the twice continuous differentiability of ϕ and y: exactly the same argument works for $\mathcal{F}^{(1)}$. It follows that Theorem 1 remains true if we replace $\mathit{Diff}^2(M)$ by $\mathit{Diff}^1(M)$ and $C^2(M,\mathbb{R})$ by $C^1(M,\mathbb{R})$, that is, if we allow ϕ and y to be merely C^1.

4.2 Compact Subsets of Non-Compact Manifolds

One of the most popular uses of phase space reconstruction has been the calculation of various fractal dimensions of chaotic attractors. If we think about using this approach for, say, the Hénon or Lorenz systems we can immediately see a problem. The phase space of the Hénon system is $\mathbb{R}^2$, which is not a compact manifold, so we cannot apply Takens' theorem to it directly. The Hénon attractor, though compact, is presumably not a submanifold of $\mathbb{R}^2$, so we cannot apply the theorem to it either. Nor indeed are there any compact submanifolds of $\mathbb{R}^2$ which contain the attractor[6]. As it stands, therefore, we do not seem to be able to use Theorem 2 to justify the use of the method of delays in the calculation of dimensions, (or anything else), of the Hénon attractor; and it seems likely that the same sort of difficulty will occur with experimental systems.

Any hopes we might have that the requirement that M be compact can simply be dropped from Theorems 1 and 2 are instantly damaged by recalling that embeddings of M in $\mathbb{R}^n$ are not necessarily generic whatever the value of n [15]. Clearly we are not going to be able to say such useful things about non-compact manifolds as we are about compact ones.

The resolution of this difficulty comes from the recognition that we may not require such general statements as Theorems 1 and 2 for non-compact sets. If we attempt to reconstruct the phase space from experimental (or even computed) data, then since the amount of this data will always be finite, it will always lie in some compact subset of the phase space—our data can never explore the whole

[6]unless we allow the manifold to have a boundary.

of phase space if the phase space is not compact. So there are regions of phase space about which we are unconcerned because we do not observe them.

There are two ways we might consider adjusting the theorems in the light of this observation. We could replace the measurement function with another which, while equal to the first on a compact set containing the data, may differ from it elsewhere, and which has some special property allowing us to deduce something similar to Theorem 2, even though M is not compact. This idea is hinted at by Takens, in a remark after the proof of Theorem 1 in which it is stated that the theorem holds for non-compact manifolds if we restrict our attention to 'proper' functions[7].

Perhaps rather simpler than this would be to demand only that the delay map be an 'embedding' of the compact subset rather than the whole of M. Then the reconstructed set that we create using the method of delays would be a faithful (diffeomorphic) copy of the part of the phase space that we have observed (for example, the attractor). This set would have the same dimensions and so on as the original in phase space. The version of Takens' theorem suggested by this approach is given below.

Theorem 4 *Let M be a m-dimensional manifold. Let A be a compact subset of M, and let $\phi : M \to M$ be a diffeomorphism, with the properties: firstly, that the periodic points of ϕ with periods less than or equal to $2m$ contained in A are finite in number, and secondly that if $x \in A$ is any periodic point with period $k \leq 2m$ then the eigenvalues of the derivative of ϕ^k at x are all distinct. Then the set of $y \in C^2(M, \mathbb{R})$ such that the map $\Phi_{(\phi,y)} : M \to \mathbb{R}^{2m+1}$, defined as in Theorem 1, is an injective immersion of A, is open.*

Furthermore, if O is any open set such that $\cup_{i=0}^{2m} \phi A \subset O$ (with O^c its complement), y any member of $C^2(M, \mathbb{R})$, and $\mathcal{N}$ a neighbourhood of y, then there is $y' \in \mathcal{N}$ such that $\Phi_{(\phi,y')}$ is an embedding of a neighbourhood of A, and $y'(x) = y(x)$ for all $x \in O^c$.

Remarks. We have to mention a technical point which was ignored in the statement of the theorem. When dealing with non-compact manifolds M, N there are *two* topologies that are routinely given to the spaces $C^k(M, N)$; the first of these, which is the one described in section 2, is called the *weak* topology: the second, in which we insist that conditions such as 1 are satisfied on countable numbers of compact sets rather than just a finite number, is called the *strong* topology. For compact manifolds, the weak and strong topologies are the same, but this is no longer true if M is not compact. As an example of the difference between these topologies we can observe that the set of C^1 functions $\mathbb{R} \to \mathbb{R}$ that are embeddings of $\mathbb{R}$ is open in the strong topology, but not in the weak.

But since Theorem 4 is concerned with compact sets, the distinction between the weak and the strong topologies loses its significance: the theorem is true whichever topology we use.

We will not go through the proof of this theorem, since it is little different to Theorem 2. We need only mention that when constructing the various open covers which provide the supports for the ψ_i functions in equation (2) we must restrict the open sets to lie in O; obviously this is always possible. To make an injective immersion on A we may need to adjust the measurement function on a neighbourhood of $\cup_{i=0}^{2m} \phi A$, as hinted at in the statement above. Usually though, we will be interested in invariant sets of ϕ, such as attractors, in which case $\cup_{i=0}^{2m} \phi A = A$.

For each delay map which is an injective immersion of A there is some open set U containing A for which the delay map is an embedding. The image of U is a submanifold of $\mathbb{R}^{2m+1}$, to which U is diffeomorphic. It follows that the dimensions (box-counting and Hausdorff) of the image of A are the same as those of A, and if A has an invariant measure, the dimensions of this measure are the same as those of the corresponding measure, induced on the image of A by the delay map. Furthermore, the characteristic exponents of A are also shared by its image [2]. So theorem 4 provides some justification for many of the applications of phase space reconstruction using the method of delays.

4.3 The work of Sauer, Yorke and Casdagli

The theorems drawn from differential topology to which we appealed in the proof of Takens' theorem, (such as that embeddings of M in $\mathbb{R}^{2m+1}$ are open in $C^2(M, \mathbb{R}^{2m+1})$), apply to general mappings between manifolds: there is no need for the codomain to be $\mathbb{R}^n$. $\mathbb{R}^n$ however has properties not

[7]That is, functions whose inverse images map compact sets to compact sets.

shared by other n dimensional manifolds: in particular, it is a vector space. It follows that we can also give $C^2(M, \mathbb{R}^{2m+1})$ a vector space structure in the usual way.

Sauer, Yorke and Casdagli [4] noted that this makes possible a measure theoretic approach to assessing whether functions having particular properties are common among all functions. We can define measures, analogous to Lebesgue measure, on finite dimensional real vector spaces by using the isomorphism of such spaces to $\mathbb{R}^n$; hence the notion of 'almost everywhere' extends simply to these spaces. We cannot do this for $C^2(M, \mathbb{R}^{2m+1})$ since it is not finite dimensional, so Sauer *et al.* defined a subset A of an infinite dimensional space S to be *prevalent* if there is a finite dimensional subspace L of S such that for every $v \in S$, $v+e \in A$ for almost every $e \in L$. They then addressed the question of whether measurement functions giving rise to embeddings are prevalent in $C^2(M, \mathbb{R}^{2m+1})$.

Sauer, Yorke and Casdagli were able to prove a theorem similar to Theorem 2, but in which 'generic' is replaced by 'prevalent'. They assumed that M is a subset of a suitable $\mathbb{R}^k$ (which involves no loss of generality because of Whitney's theorem) and used for L the set of polynomials in k variables up to order $2m+1$. (There is no corresponding version of Theorem 1, because $Diff^2(M)$ is not a vector space.)

To prove this theorem, Sauer *et al.* replaced lemma 1 with a more specific lemma taking account of the box-counting dimension of the domain. This enabled them to extend their theorem in several ways. Firstly, they were able to replace the assumption of finiteness of the set of periodic points by an assumption about the box-counting dimension of this set. Secondly, they were able to derive a theorem similar to Theorem 4 (though not quite the same) in which the number of delays required to embed the compact set depends on its box-counting dimension, rather than the dimension of a manifold which contains it. Thirdly, they were able to make some statements regarding the set of points at which delay maps fail to be injective or immersive, if an insufficient number of delays is used.

5 Consequences of Takens' Theorem for Signal Processing

For all its importance, the unforgiving differential topology in the proof of Takens' theorem in some ways obscures, rather than illuminates, the point. Let us take a fresh look at the situation in which we hope to make use of the theorem.

We are faced with some physical system, which might be something quite general. In the signal processsing context it could be anything from the engine of a vehicle to the sea surface, to a piece of electronic signal processing equipment, or even a communications channel. The state of this system is changing with time according to some deterministic law, which will be Newton's laws for a mechanical system. (For the moment we are thinking of *autonomous* systems, or those, such as periodically driven systems, that can be made autonomous by a relatively simple extension.) In many situations the dynamics of the system will be smooth—the state will not change abruptly or discontinuously with time. The dynamics may naturally occur in a discrete fashion, in a clocked system perhaps, but more usually the state will change continuously through time. In the latter case we assume that our *observations* of the system are discrete, that is, that our measurements are taken at distinct separated instants, or that we sample a continuous time record. During the experiment, measurements are taken at times $t_i, i = 1, \ldots$. Again for simplicity we assume the t_i to be evenly spaced in time.

As the system is deterministic, the state at any t_i is uniquely determined by the state at t_{i-1}. We assume that the reverse is also true, that is, that the motion is invertible. These conditions are equivalent to supposing that there is some function $\phi : M \to M$, which maps the state at time t_i to that at time t_{i+1}, and that this function is invertible, and both ϕ and its inverse are differentiable: in short that the dynamics is described by a diffeomorphism. Often the dynamics will be specified by some differential equation, and conditions which guarantee that ϕ is a diffeomorphism may be drawn from the theory of differential equations [19].

The mapping ϕ forms part of the bridge between our experiment and Takens' theorem; the other part is the measurement. We make an observation (or sample a record) at each t_i, the result of which is a single real number. (Ignore for the moment the inevitable presence of noise and quantization errors.) These observations could be signals from vibration monitoring equipment, sound records, samples of voltages from electrical machinery and so on. Very little about the nature of the experiment has to be specified; all that we need to assume is that the measurement at time t_i is determined solely by the state of the system at t_i, and that the dependence is smooth. Then the measurement is described by a differentiable function $y : M \to \mathbb{R}$.

Takens' theorem now tells us that if we take a time series of measurements and construct n-vectors from it by selecting n consecutive elements, (where n is greater than twice the dimension of M) then, if ϕ and y are generic, these vectors lie on a subset of $\mathbb{R}^n$ which is an embedding of M.

At first sight this might seem to be little enough cause for celebration, and indeed, if the number of observations we make is small this knowledge will be of little practical value. But if we have sufficient data that all of the state space is well sampled, then the information in our reconstructed space will be correspondingly complete and we can learn from it things about the system itself. The most basic thing we might hope to learn is the *dimension* of the state space, which is to say the number of degrees of freedom of the system. Another thing, almost as basic, is the *topology* of the space [21]. Though this information may seem rather qualitative, even this has its uses: for a multifrequency periodic system for example the state space should have the geometry of an n-torus, where n is the number of independent oscillators—and this will still be true even if the oscillators are coupled together in a highly nonlinear manner, leading to a very complex, unintelligible Fourier spectrum. And the *homology groups* described in [21] are characteristics of the system which might be used for recognition purposes.

While this topological information is important, it does not actually exploit all the properties which are preserved under embedding. Furthermore, it is often the case that the dynamics of the system do not cause it to explore all of its state space: it may become confined to subsets of the space known as *attractors*. If the attractors are themselves manifolds then we can simply proceed as before, but in general they will not be. Nevertheless, the reconstruction still contains important information about the attractor.

We can usually assume the attractor is associated with an *invariant measure* of ϕ on M [1]. If we say $N = \Phi M$ is the image of the state space under the embedding (from now on we abbreviate $\Phi_{(\phi,y)}$ to Φ), there is a corresponding measure on N induced by Φ, and because Φ is a diffeomorphism the induced measure has the same (fractal) dimensions as the original. Many of the early studies based on Takens' theorem were concerned with the calculation of dimensions from reconstructions made using the method of delays.

From the point of view of signal processing, a much more important point to note is that the embedding Φ not only yields a copy of the space M, but also of the dynamics. The function $\psi : N \to N$ defined by $\psi = \Phi\phi\Phi^{-1}$ is a dynamical system on N which inherits the properties of invertibility and smoothness from ϕ. It is clear that periodic orbits of ϕ are mapped by Φ into corresponding periodic orbits of ψ, that the same is true of dense orbits, and so on. These things follow from the fact that Φ is a homeomorphism. The diffeomorphic nature of Φ means that even more is true: it can be shown that the characteristic numbers of the periodic orbits are preserved by embedding, and so are the Liapunov numbers [2, 1].

Since we know that the reconstructed points in N are related by the map ψ, we might hope to estimate ψ from the data. That is, we can try to build a model of the dynamical system (ψ, N). If we can do this accurately, then we can begin to do signal processing in earnest. We can, for example, use the model for prediction: given an n-vector from the time series, we can use our estimate of ψ to compute the next vector—essentially to predict the next element of the time series. This opens the way to predictive coding and predictive noise cancellation. Of course, for signals from the highly nonlinear systems we are considering, these things are not so straightforward as in the linear case: they call for new ideas and algorithms; but models such as ψ appear to be an essential starting point. We can also use the model in other ways: for example for removing noise from signals by minimizing the inconsistency between the data and a suitable form of model. (Applications of this type have begun to appear in the literature [22, 23, 24].) More excitingly, we can turn the noise cancellation on its head, and explore the possibility that some signals currently treated as stochastic noise might in fact be better regarded as chaotic processes, and viewed in terms of a model such as ψ.

In general there is no way for us to use the model to derive ϕ, because the data depend not only on ϕ but on some unknown measurement function y. (The measurement may indeed vary each time we encounter the system, because we view it from different distances and orientations, or in different environments. Of course, quantities preserved by embedding will not vary in this way.) But in the signal processing context it will often be the output of the system which is of most relevance, rather than its state. It should even be possible to characterise systems 'up to diffeomorphism', that is, to decide whether two signals could be related by the smooth 'coordinate changes' that Takens' theorem tells us relate the method-of-delays reconstructions to the original system. If so, the reconstructions offer a way to perform system recognition and classification, which would be all the more powerful because of its independence of the way the system is measured.

All these things flow from the apparently abstract and abstruse theorem of Takens. It seems likely that most of the eventual implications that the theorem—and nonlinear dynamical systems theory generally—will have for signal processing remain to be thought of. But it is already clear that the possibilities are extensive, and that to exploit them we will have to come to terms with new, and perhaps unfamiliar, mathematical ideas and techniques.

6 Acknowledgements

The above discussion owes a great deal to the contributions of Mark Muldoon and David Broomhead.

References

[1] J.-P. Eckmann and D. Ruelle. Ergodic theory of chaos and strange attractors. *Rev. Mod. Phys.* **57** (1985) 617-656

[2] J. Guckenheimer and P. Holmes. *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields*, Springer-Verlag, New York, 1983.

[3] W. M. Schaffer and M. Kot. Differential systems in ecology and epidemiology, in *Chaos*, edited by A. V. Holden, Manchester University Press, Manchester, 1986.

[4] T. Sauer, J. A. Yorke and M. Casdagli. Embedology. *J. Stat. Phys.* **65** (1991) 579-616

[5] N. H. Packard, J. P. Crutchfield, J. D. Farmer and R. S. Shaw. Geometry from a time series. *Phys. Rev. Lett.* **45** (1980) 712-716

[6] F. Takens. 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, 1981.

[7] M. Casdagli, S. Eubank, J. D. Farmer and J. Gibson. State space reconstruction in the presence of noise. *Physica D* **51** (1991) 52-98

[8] J. F. Gibson, J. D. Farmer, M. Casdagli and S. Eubank. An analytic approach to state space reconstruction. *Physica D* **57** (1992) 1-30

[9] D. S. Broomhead, J. P. Huke and M. R. Muldoon. Linear Filters and Nonlinear Systems. *J. R. Statist. Soc.* **54** (1992) 373-382

[10] D. R. Chillingworth. *Differential topology with a view to applications*, Pitman Publishing, London, 1986.

[11] L. Auslander and R. E. MacKenzie. *Introduction to differentiable manifolds*, Dover, New York, 1977.

[12] Th. Bröcker and K. Jänich. *Introduction to differential topology*, Cambridge University Press, Cambridge, 1982.

[13] M. A. Armstrong. *Basic Topology*, Springer-Verlag, New York, 1983.

[14] G. F. Simmons. *Introduction to Topology and Modern Analysis*, McGraw-Hill, New York, 1963.

[15] M. W. Hirsch. *Differential Topology*, Springer-Verlag, New York, 1976.

[16] I. Kupka. Contribution à la thèorie des champs gènèriques. *Contrib. Diff. Eqs.* **2** (1963) 457-484

[17] S. Smale. Stable manifolds for differential equations and diffeomorphisms. *Ann. Scuola Norm. Sup. Pisa* **18** (1963) 97-116

[18] J. Palis, Jr. and W. de Melo. *Geometric Theory of Dynamical Systems*, Springer-Verlag, New York, 1982.

[19] M. W. Hirsch and S. Smale. *Differential Equations, Dynamical Systems, and Linear Algebra*, Academic Press, San Diego, 1974.

[20] M. Spivak. *Calculus on Manifolds* Benjamin, New York, 1965.

[21] M. R. Muldoon, R. S. MacKay, J. P. Huke and D. S. Broomhead. Topology from time series. *Physica D* in press.

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

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

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