Journal of Statistical Physics, Vol. 65, Nos. 3/4, 1991

Embedology

Tim Sauer,[1] James A. Yorke,[2] and Martin Casdagli[3]

Received May 10, 1991

Mathematical formulations of the embedding methods commonly used for the reconstruction of attractors from data series are discussed. Embedding theorems, based on previous work by H. Whitney and F. Takens, are established for compact subsets A of Euclidean space R^k. If n is an integer larger than twice the box-counting dimension of A, then almost every map from R^k to R^n, in the sense of prevalence, is one-to-one on A, and moreover is an embedding on smooth manifolds contained within A. If A is a chaotic attractor of a typical dynamical system, then the same is true for almost every *delay-coordinate map* from R^k to R^n. These results are extended in two other directions. Similar results are proved in the more general case of reconstructions which use moving averages of delay coordinates. Second, information is given on the self-intersection set that exists when n is less than or equal to twice the box-counting dimension of A.

KEY WORDS: Embedding; chaotic attractor; attractor reconstruction; probability one; prevalence; box-counting dimension; delay coordinates.

1. INTRODUCTION

In this work we give theoretical justification of data embedding techniques used by experimentalists to reconstruct dynamical information from time series. We focus on cases in which trajectories of the system under study are asymptotic to a compact attractor. We state conditions that ensure that the map from the attractor into reconstruction space is an embedding, meaning that it is one-to-one and preserves differential information. Our approach integrates and expands on previous results on embedding by Whitney[(29)] and Takens.[(27)]

[1] Department of Mathematical Sciences, George Mason University, Fairfax, Virginia 22030.

[2] Institute of Physical Science and Technology, University of Maryland, College Park, Maryland 20742.

[3] Santa Fe Institute, Santa Fe, New Mexico 87501. Current address: Tech Partners, 4 Stamford Forum, 8th Floor, Stamford, Connecticut 06901.

0022-4715/91/1100-0579$06.50/0

Whitney showed that a generic smooth map F from a d-dimensional smooth compact manifold M to R^{2d+1} is actually a diffeomorphism on M. That is, M and $F(M)$ are diffeomorphic. We generalize this in two ways: first, by replacing "generic" with "probability-one" (in a prescribed sense), and second, by replacing the manifold M by a compact invariant set A contained in R^k that may have noninteger box-counting dimension (boxdim). In that case, we show that almost every smooth map from a neighborhood of A to R^n is one-to-one as long as

$$n > 2 \cdot \mathrm{boxdim}(A)$$

We also show that almost every smooth map is an embedding on compact subsets of smooth manifolds within A. This suggests that embedding techniques can be used to compute positive Lyapunov exponents (but not necessarily negative Lyapunov exponents). The positive Lyapunov exponents are usually carried by smooth unstable manifolds on attractors. We give precise definitions of one-to-one, embedding, and almost every in the next section.

Takens dealt with a restricted class of maps called delay-coordinate maps. A delay-coordinate map is constructed from a time series of a single observed quantity from an experiment. Because of this, a typical delay-coordinate map is much more likely to be accessible to an experimentalist than a typical map. Takens[27] showed that if the dynamical system and the observed quantity are generic, then the delay-coordinate map from a d-dimensional smooth compact manifold M to R^{2d+1} is a diffeomorphism on M.

Our results generalize those of Takens[27] in the same two ways as for Whitney's theorem. Namely, we replace generic with probability-one and the manifold M by a possibly fractal set. Thus, for a compact invariant subset A of R^k, under mild conditions on the dynamical system, almost every delay-coordinate map F from R^k to R^n is one-to-one on A provided that $n > 2 \cdot \mathrm{boxdim}(A)$. Also, any manifold structure within A will be preserved in $F(A)$. These results hold for lower box-counting dimension (see Section 4) if boxdim does not exist. The ambient space R^k can be replaced by a k-dimensional smooth manifold in the general case. In addition, we have made explicit the hypotheses on the dynamical system (discrete or continuous) that are needed to ensure that the delay-coordinate map gives an embedding. In particular, only C^1 smoothness is needed. For flows, the delay must be chosen so that there are no periodic orbits whose period is exactly equal to the time delay used or twice the delay. (A finite number of periodic orbits of a flow whose periods are p times the delay are allowed for $p \geqslant 3$.) Further, we explain what happens in the case that $n \leqslant 2 \cdot \text{boxdim}(A)$. In that case we put bounds on the dimension of the self-intersection set, which is the set on which the one-to-one property fails. Finally, we give more general versions of the delay-coordinate theorem which deals with filtered delay coordinates, which allow more versatile and useful applications of embedding methods.

There are no analogues of these results where the box-counting dimension is replaced by Hausdorff dimension (see Theorem 4.7 and the discussion that follows). In an Appendix to this work written by I. Kan, examples are described of compact subsets of R^k, for any positive integer k, which have Hausdorff dimension $d=0$, and which are difficult to project in a one-to-one way. The requirement $n > 2d$ discussed above translates in this case to $n > 0$. However, *every* projection of such a set to R^n, $n < k$, fails to be one-to-one.

In Section 2 we explain the new version of the Whitney and Takens embedding theorems. In Section 3 we discuss filtered delay coordinates. Section 4 contains proofs of the results.

2. HOW TO EMBED MANIFOLDS AND FRACTAL SETS

2.1. Fractal Whitney Embedding Prevalence Theorem

Assume Φ is a flow on Euclidean space R^k, generated, for example, by an autonomous system of k differential equations. Assume further that all trajectories are asymptotic to an attractor A. The study of long-time behavior of the system will involve the study of the set A.

In a typical scientific experiment, the phase space R^k cannot be explicitly seen. The experimenter tries to infer properties of the system by taking measurements. Since each state of the dynamical system is uniquely specified by a point in phase space, a measured quantity is a function from phase space to the real number line. If n independent quantities $Q_1,..., Q_n$ can be measured simultaneously, then with each point in phase space is associated a point in Euclidean space R^n. We can then talk about the *measurement function*

$$F(\text{state}) = (Q_1,..., Q_n)$$

which maps R^k to R^n.

For example, suppose all trajectories in phase space R^k are attracted to a periodic cycle. Thus, A is topologically a circle lying in R^k. Imagine that two available measurement quantities Q_1 and Q_2 are plotted in the plane. Then there is a measurement map F from A to R^2 given by $F(\text{state}) = (Q_1, Q_2)$. To what extent are the properties of the hidden attractor A preserved in the observable "reconstruction space" R^2?

The answer depends on how the circle is mapped to R^2 under F. Consider the case where $R^k = R^3$ and Q_1 and Q_2 are simply the two coordinate functions x_1 and x_2. In Fig. 1a, the relative position of the points is preserved upon projection, and we may view $F(A)$ as a faithful reconstruction of the attractor A. If distinct points on the attractor A map under F to distinct points on $F(A)$, we say that F is *one-to-one* on A.

In the case of Fig. 1b, on the other hand, two different states of the dynamical system have been identified together in $F(A)$. In the reconstruction space, which is all the experimenter actually sees, the two distinct states cannot be distinguished, and information has been lost.

The one-to-one property is useful because the state of a deterministic dynamical system, and thus its future evolution, is completely specified by a point in phase space. Suppose that at a given state x one observes the value $F(x)$ in the reconstruction space, and that this is followed 1 sec later by a particular event. If F is one-to-one, each appearance of the measurements represented by $F(x)$ will be followed 1 sec later by the same event. This is because there is a one-to-one correspondence between the attractor points in phase space and their images in reconstruction space. There is predictive power in finding a one-to-one map.

Perhaps the measurements $F(x)$ would not be repeated precisely. Yet if the map F is reasonable, similar measurements will predict similar events. This approach to prediction and noise reduction of data is being pursued by a number of research groups.

Although most of the interest lies in the case that A is an attractor of a dynamical system, the main question can be posed in more generality. Let A be a compact subset of Euclidean space R^k, and let F map R^k to another Euclidean space R^n. Under what conditions can we be assured that A is "embedded" in R^n by typical maps F?

The precise definition of embedding involves differential structure. A one-to-one map is a map that does not collapse points, that is, no two points are mapped to the same image point. An embedding is a map that does not collapse points or tangent directions. Thus, to define embedding, we need to be working on a compact set A that has well-defined tangent spaces.

Let A be a compact smooth differentiable manifold. (Here, as in the remainder of the paper, the word *smooth* will be used to mean continuously differentiable, or C^1.) A smooth map F on A is an *immersion* if the derivative map $DF(x)$ (represented by the Jacobian matrix of F at x) is one-to-one at every point x of A. Since $DF(x)$ is a linear map, this is equivalent to $DF(x)$ having full rank on the tangent space. This can happen whether or not F is one-to-one. Under an immersion, no differential structure is lost in going from A to $F(A)$.

An *embedding* of A is a smooth diffeomorphism from A onto its image $F(A)$, that is, a smooth one-to-one map which has a smooth inverse. For a compact manifold A, the map F is an embedding if and only if F is a one-to-one immersion. Figure 1a shows an example of an embedding of a circle into the plane. Figure 1b shows an immersion that is not one-to-one, and Fig. 1c shows a one-to-one map that fails to be an immersion.

Whether or not a typical map from A to R^n is an embedding of A depends on the set A, and on what we mean by "typical." A celebrated result of this type is the embedding genericity theorem of Whitney,[29] which says that if A is a smooth manifold of dimension d, then the set of maps into R^{2d+1} that are embeddings of A is an open and dense set in the C^1-topology of maps.

The fact that the set of embeddings is *open* in the set of smooth maps means that given each embedding, arbitrarily small perturbations will still be embeddings. The fact that the set of embeddings is *dense* in the set of maps means that every smooth map, whether it is an embedding or not, is arbitrarily near an embedding. One would like to conclude from Whitney's

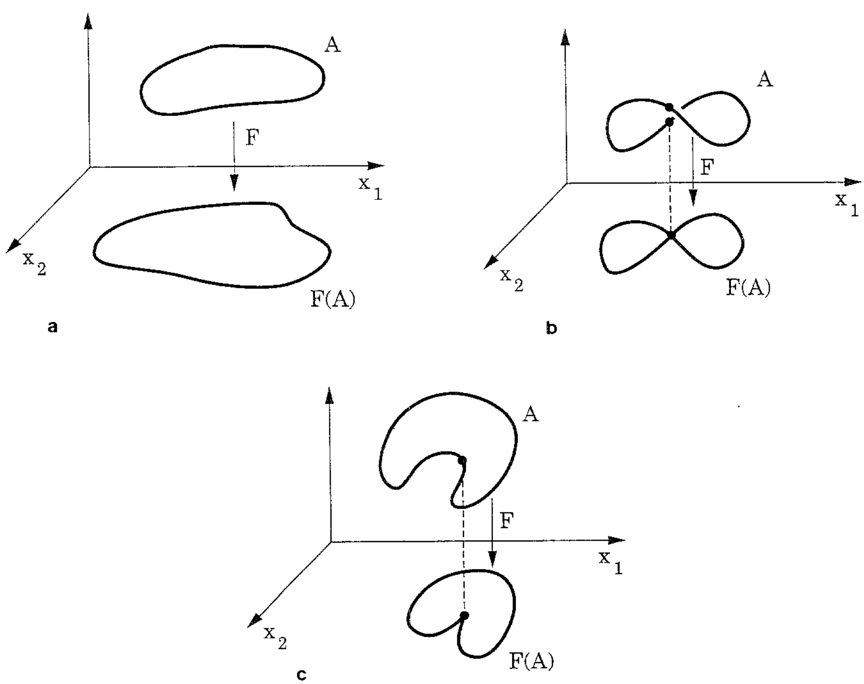

Fig. 1. (a) An embedding F of the smooth manifold A into R^2. (b) An immersion that fails to be one-to-one. (c) A one-to-one map that fails to be an immersion.

theorem that $n=2d+1$ simultaneous measurements are typically sufficient to reconstruct a d-dimensional state manifold A in the measurement space R^n.

However, open dense subsets, even of Euclidean space, can be thin in terms of probability. There are standard examples, many from recent studies in dynamics, of open dense sets that have arbitrarily small Lebesgue measure, and therefore arbitrarily small probability of being realized.

A well-known example is the phenomenon of Arnold tongues. Consider the family of circle diffeomorphisms

$$g_{\omega,k}(x)=x+\omega+k\sin x \mod 2\pi$$

where $0\leqslant\omega\leqslant 2\pi$ and $0\leqslant k<1$ are parameters. For each k we can define the set

$$\mathrm{Stab}(k)=\{0\leqslant\omega\leqslant 2\pi\text{: } g_{\omega,k} \text{ has a stable periodic orbit}\}$$

For $0<k<1$, the set Stab(k) is a countable union of disjoint open intervals of positive length, and is an open dense subset of $[0, 2\pi]$. However, the total length (Lebesgue measure) of the open dense set Stab(k) approaches zero as $k\to 0$. For small k, the probability of landing in this open dense set is very small. See ref. 3 for more details.

With such examples in mind, an experimentalist would like to make a stronger statement than that the set of embeddings is an open and dense set of smooth maps. Instead, one would like to know that the particular map that results from analyzing the experimental data is an embedding *with probability one.*

The problem with such a statement is that the space of all smooth maps is infinite-dimensional. The notion of probability one on infinite-dimensional spaces does not have an obvious generalization from finite-dimensional spaces. There is no measure on a Banach space that corresponds to Lebesgue measure on finite-dimensional subspaces. Nonetheless, we would lilke to make sense of "almost every" map having some property, such as being an embedding. Following ref. 24, we propose the following definition of prevalence.

Definition 2.1. A Borel subset S of a normed linear space V is *prevalent* if there is a finite-dimensional subspace E of V such that for each v in V, $v+e$ belongs to S for (Lebesgue) almost every e in E.

We give the distinguished subspace E the nickname of *probe space.* The fact that S is prevalent means that if we start at any point in the ambient space V and explore along the finite-dimensional space of directions specified by E, then almost every point encountered will lie in S.

Notice that any space containing a probe space for S is itself a probe space for S. In other words, if E' is any finite-dimensional space containing E, then perturbations of any element of V by elements of E' will be in S with probability one. This is a simple consequence of the Fubini theorem.[22]

From this fact it is easy to see that a prevalent subset of a finite-dimensional vector space is simply a set whose complement has zero measure. Also, the union or intersection of a finite number of prevalent sets is again prevalent. We will often use the notion of prevalence to describe subsets of functions. It follows from the definition that prevalent implies dense in the C^k-topology for any k. More generally, prevalent implies dense in any normed linear space.

When a condition holds for a prevalent set of functions, it is usually illuminating to determine the smallest, or most efficient, probe subspace E. This corresponds to the minimal amount of perturbation that must be available to the system in order for the condition to hold with virtual certainty.

As stated above, for subsets of finite-dimensional spaces the term prevalent is synonomous with "almost every," in the sense of outside a set of measure zero. When there is no possibility of confusion, we will say that "almost every" map satisfies a property when the set of such maps is prevalent, even in the infinite-dimensional case. For example, consider convergent Fourier series in one variable, which form an infinite-dimensional vector space with basis $\{e^{inx}\}_{n=-\infty}^{\infty}$. In the sense of prevalence, almost every Fourier series has nonzero integral on $[0, 2\pi]$. The probe space E in this case is the one-dimensional space of constant functions. If E' is a vector space of Fourier series which contains the constant functions, then for every Fourier series f, the integral of $f+e$ will be nonzero for almost every e in E'.

With this definition, we introduce a prevalence version of the Whitney embedding theorem.

Theorem 2.2 (Whitney Embedding Prevalence Theorem). Let A be a compact smooth manifold of dimension d contained in R^k. Almost every smooth map $R^k \to R^{2d+1}$ is an embedding of A.

In particular, given any smooth map F, not only are there maps arbitrarily near F that are embeddings, but in the sense of prevalence, *almost all* of the maps near F are embeddings. The probe space E of Definition 2.1 is the $k(2d+1)$-dimensional space of linear maps from R^k to R^{2d+1}. This theorem, which is proved in Section 4, gives a stengthening of the traditional statement of the Whitney embedding theorem.

It is quite interesting that Whitney later proved the different result that under the same circumstances, there *exists* an embedding into R^{2d}. (This could be called the Whitney embedding existence theorem.) However, an existence theorem is of little help to an experimentalist, who needs information about maps near the particular one that happens to be available. Knowledge that an embedding exists sheds little information on the particular F under study.

The example of Fig. 1b shows that the dimension $2d+1$ of Theorem 2.2 is the best possible. The map F is not one-to-one on the twisted circle A, thus does not embed A into R^2. Further, no nearby map (even in the C^0-topology) embeds A. On the other hand, if a given map of the circle A into R^3 was not one-to-one, there would necessarily be a prevalent set of nearby maps that are embeddings.

The first main goal of this section was to express Whitney's embedding theorem (and Takens' theorem; see below) in this probabilistic sense. The second is to extend Whitney's theorem to sets A that are not manifolds. Here we use the fractal dimension known as box-counting dimension.

The box-counting (or capacity) dimension of a compact set A in R^n is defined as follows. For a positive number ε, let A_ε be the set of all points within ε of A, i.e., $A_\varepsilon = \{x \in R^n\colon |x-a| \leqslant \varepsilon \text{ for some } a \in A\}$. Let $\mathrm{vol}(A_\varepsilon)$ denote the n-dimensional outer volume of A_ε. Then the *box-counting dimension* of A is

$$\operatorname{boxdim}(A) = n - \lim_{\varepsilon \to 0} \frac{\log \mathrm{vol}(A_\varepsilon)}{\log \varepsilon}$$

if the limit exists. If not, the upper (respectively, lower) box-counting dimension can be defined by replacing the limit by the lim inf (resp., lim sup). When the box-counting dimension exists, the approximate scaling law

$$\mathrm{vol}(A_\varepsilon) \approx \varepsilon^{n-d}$$

holds, where $d = \operatorname{boxdim}(A)$.

There are several equivalent definitions of box-counting dimension. For example, R^n can be divided into ε-cubes by a grid based, say, at points whose coordinates are ε-multiples of integers. Let $N(\varepsilon)$ be the number of boxes that intersect A. Then

$$\operatorname{boxdim}(A) = \lim_{\varepsilon \to 0} \frac{\log N(\varepsilon)}{-\log \varepsilon}$$

with similar provisions for upper and lower box-counting dimension. The scaling in this case is

$$N(\varepsilon) \approx \varepsilon^{-d}$$

Even if we know the box-counting dimension of an attractor A, Theorem 2.2 gives no estimate on the lowest dimension for which almost every map is an embedding. Suppose we know that A is the invariant set of a flow on R^{100}, and that the box-counting dimension of A is 1.4. In the absence of any knowledge about the containment of A in a smooth manifold of dimension less than 100, the use of Theorem 2.2 to get a one-to-one reconstruction requires the use of maps into R^{201}. In fact, the smallest smooth manifold that contains the 1.4-dimensional attractor may indeed have dimension 100. But as the next result shows, one can do much better: almost every reconstruction map into R^3 will be one-to-one on A.

Theorem 2.3 (Fractal Whitney Embedding Prevalence Theorem). Let A be a compact subset of R^k of box-counting dimension d, and let n be an integer greater than $2d$. For almost every smooth map F: $R^k \to R^n$,

1. F is one-to-one on A
2. F is an immersion on each compact subset C of a smooth manifold contained in A.

The proof of the one-to-one half of the fractal Whitney embedding prevalence theorem may be sketched as follows. Choose any bounded finite-dimensional space E of smooth maps F so that varying F by elements of E results in perturbing $F(x)-F(y)$ throughout R^n for each pair $x \neq y$ in A. For example, the probe space E can be taken to be the space of linear maps from R^k to R^n. Then the probability (measured in E) that the perturbed $F(x)$ and $F(y)$ lie within ε is on the order of ε^n. Similarly, if B_1 and B_2 are ε-boxes on A, the probability that $F(B_1)$ and $F(B_2)$ intersect is on the order of ε^n. Here we assume that there is a bound on the magnification of F, as when F is a smooth map near the compact set A. The set A can be covered by essentially ε^{-d} boxes of size ε, and the number of pairs of boxes is proportional to $(\varepsilon^{-d})^2$. The probability that no distinct pair of boxes collide in the image $F(A)$ is proportional to $(\varepsilon^{-d})^2\varepsilon^n = \varepsilon^{n-2d}$. If $n > 2d$, this probability of choosing a perturbation of F that fails to be one-to-one is negligible for small ε. More precise details of the proof, as well as the immersion part, are in Section 4.

2.2. Fractal Delay Embedding Prevalence Theorem

Despite the beauty of Whitney's embedding theorem, it is rare for a scientist to be able to measure a large number of independent quantities simultaneously. In fact, it is a rather subtle problem to decide whether two different simultaneous measurements are indeed independent. These problems can be sidestepped to some degree by introducing the use of *delay coordinates*. In this approach, only one measurable quantity is needed.

In a typical experiment, the single measurable quantity is sampled at intervals T time units apart. The resulting list of samples $\{Q_t\}$ is called a *time series*. Think of the measurable quantity as an observation function h on the state space R^k on which the dynamical system Φ is acting. Each reading $Q_t = h(x_t)$ is the result of evaluating the observation function h at the current state x_t.

Definition 2.4. If Φ is a flow on a manifold M, T is a positive number (called the *delay*), and h: $M \to R$ is a smooth function, define the *delay-coordinate map* $F(h, \Phi, T)$: $M \to R^n$ by

$$F(h, \Phi, T)(x) = (h(x), h(\Phi_{-T}(x)), h(\Phi_{-2T}(x)),..., h(\Phi_{-(n-1)T}(x)))$$

To start with a simple example, let A be a periodic orbit of the flow Φ. We found above that in the absence of dynamics, three independent coordinates are required to embed A in reconstruction space, or more precisely, that almost every smooth map $F = (f_1, f_2, f_3)$ from a neighborhood of A to R^3 is an embedding on A.

Now the situation is different. Instead of three functions f_1, f_2, f_3 that must be independent, there is a single function h, and the corresponding map $F(h, \Phi, T)$ pictured in Fig. 2. We want to know that for almost every function h from A to the real numbers R, the delay-coordinate map $F(h, \Phi, T)$ from A into R^n is an embedding. It should be stressed that this does not follow from Theorems 2.2 and 2.3. The maps $F(h, \Phi, T)$ form a restricted subset of all maps; whether they contain enough variation to perturb away self-crossings of A needs to be determined. In fact, the general

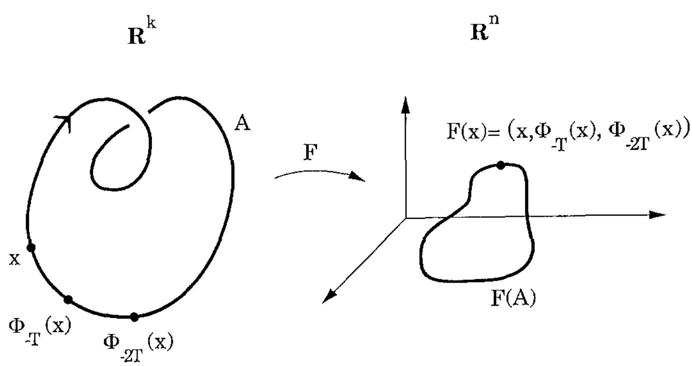

Fig. 2. The attractor on the left is mapped using delay coordinates into the reconstruction space on the right.

answer is that they do not contain enough variation. Extra hypotheses on the dynamical system Φ are required to ensure that almost every h leads to an embedding of A.

To see the need for extra hypotheses, consider the case that A is a periodic orbit of a continuous dynamical system whose period is equal to the sampling interval T. Topologically, A is a circle. In this case, $F(h, \Phi, T)$ cannot be one-to-one for *any* observation function h. Let x be a point on the topological circle A. Since the period of A is T, $h(x) = h(\Phi_{-T}(x)) = \cdots = h(\Phi_{-(n-1)T}(x))$, so that $F = F(h, \Phi, T)$ maps x to the diagonal line $\{(x_1,..., x_n): x_1 = \cdots = x_n\}$ in R^n. A circle cannot be mapped continuously to a line (in this case, the diagonal line in R^n) in a one-to-one fashion. See Fig. 3.

The one-to-one property also fails when A is a periodic orbit of period $2T$. Define the function $d(x) = h(x) - h(\Phi_{-T}(x))$ on A. The function d is either identically zero or it is nonzero for some x on A, in which case it has the opposite sign at the image point $\Phi_{-T}(x)$, and changes sign on A. In any case, $d(x)$ has a root x_0 on A. Since the period of A is $2T$, we have $h(x_0) = h(\Phi_{-T}(x_0)) = h(\Phi_{-2T}(x_0)) = \cdots$. Then $F(h, \Phi, T)$ maps x_0 and $\Phi_{-T}(x_0)$ to the same point in R^n. If x_0 and $\Phi_{-T}(x_0)$ are distinct, this says that F is not one-to-one. If $x_0 = \Phi_{-T}(x_0)$, then the orbit actually has period T, and F fails to be one-to-one as above. In the presence of periodic orbits of period $2T$, $F(h, \Phi, T)$ cannot be one-to-one for any observation function h.

On the other hand, when A is a periodic orbit of period $3T$, or any period not equal to T or $2T$, there is no such problem. In this case the delay-coordinate map of a periodic orbit A into R^n is an embedding for almost every observation function h as long as the reconstruction dimension is at least three. The statement for more general attractors A is as follows.

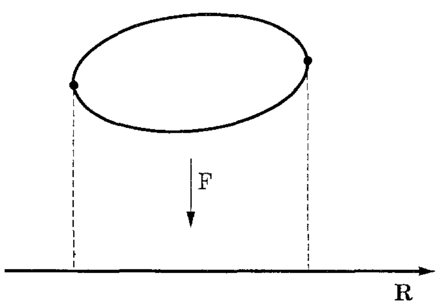

Fig. 3. A two-to-one map from a topological circle to the real line.

Theorem 2.5 (Fractal Delay Embedding Prevalence Theorem). Let Φ be a flow on an open subset U of R^k, and let A be a compact subset of U of box-counting dimension d. Let $n > 2d$ be an integer, and let $T > 0$. Assume that A contains at most a finite number of equilibria, no periodic orbits of Φ of period T or $2T$, at most finitely many periodic orbits of period $3T$, $4T$,..., nT, and that the linearizations of those periodic orbits have distinct eigenvalues. Then for almost every smooth function h on U, the delay coordinate map $F(h, \Phi, T)$: $U \to R^n$ is:

1. One-to-one on A.
2. An immersion on each compact subset C of a smooth manifold contained in A.

Where Takens[27] showed that the delay-coordinate maps generically (in the C^2-topology) give embeddings of smooth manifolds of dimension d, we substitute compact sets of box-counting dimension d, and replace generic with prevalent.

The assumption of Theorem 2.5 that there are no periodic orbits of period T or $2T$ can be satisfied by choosing the time delay T to be sufficiently small. In fact, if we assume that the vector field on A satisfies a Lipschitz condition, that is, $\dot{x} = V(x)$, where $|V(x) - V(y)| \leqslant L|x - y|$, then it is known[30] that each periodic orbit must have period at least $2\pi/L$. Hence, if $T < \pi/L$, there will be no periodic orbits of period T or $2T$.

Theorem 2.5 assumes $n > 2d$ to avoid self-intersection of the reconstructed image of A. To see that this requirement cannot be relaxed in general, consider the case $d = 1$, $n = 2d = 2$ shown in Fig. 4a. Let the observation function h be the coordinate function x_1, and consider the delay coordinate map $R^k \to R^2$ defined by

$$F(x_1, \Phi, T) = (x_1(x), x_1(\Phi_{-T}(x)))$$

In the situation illustrated in Fig. 4a, $x_1(\Phi_{-T}(b)) < x_1(\Phi_{-T}(a)) < x_1(a) = x_1(b)$, and $x_1(\Phi_{-T}(c)) < x_1(\Phi_{-T}(d)) < x_1(c) = x_1(d)$. Setting $F = F(x_1, \Phi, T)$, this means that in the reconstruction space R^2, $F(a)$ lies directly above $F(b)$, and $F(d)$ lies directly above $F(c)$. See Fig. 4b. The map F is continuous on the trajectory, so there is a continuous path, parametrized by x_1, connecting $F(a)$ and $F(c)$. There is also such a path connecting $F(b)$ and $F(d)$. According to Fig. 4b, there must be a value of x_1 in between where the curves meet, and two different points on the circle map together under F. Otherwise said, somewhere in between there is an x_1 coordinate such that the upper and lower parts of the trajectory advance the same amount in the x_1 direction during the time interval T, and thus have identical delay coordinates. The map $F(h, \Phi, T)$ is not an embedding.

If the observation function or flow is perturbed a small amount, the same topological argument can be made. Thus, this example is robust. No small perturbation of the map is an embedding.

Theorem 2.5 is a special case of a statement about diffeomorphisms. Before stating that version, we redefine delay coordinate maps for diffeomorphisms.

Definition 2.6. If g is a diffeomorphism of an open subset U of R^k, and h: $U \to R$ is a function, define the *delay coordinate map* $F(h, g)$: $U \to R^n$ by

$$F(h, g)x = (h(x), h(g(x)), h(g^2(x)),..., h(g^{n-1}(x)))$$

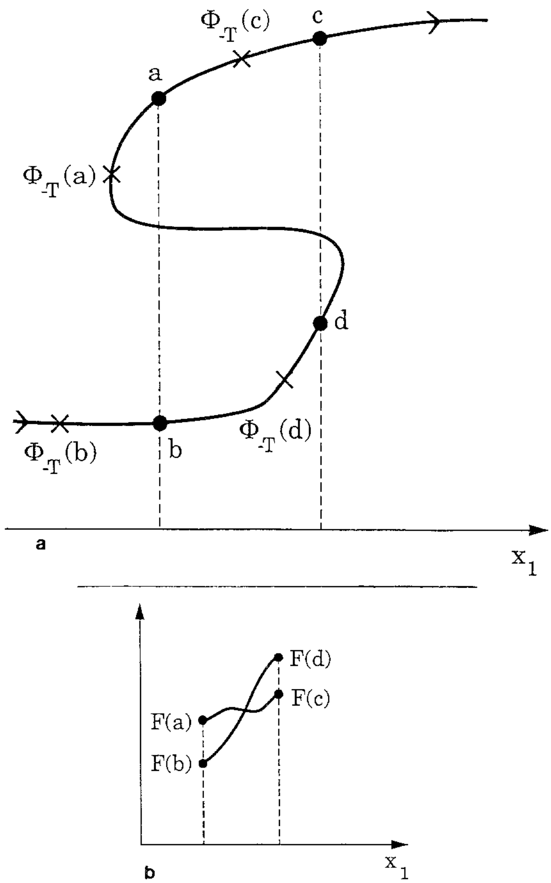

Fig. 4. (a) A trajectory of a flow that cannot be mapped using two delay coordinates in a one-to-one way. (b) The point at which the paths cross corresponds to a set of delay coordinates shared by two points on the trajectory.

We get the previous theorem by substituting $g = \Phi_{-T}$ in the following statement.

Theorem 2.7. Let g be a diffeomorphism on an open subset U of R^k, and let A be a compact subset of U, $\mathrm{boxdim}(A) = d$, and let $n > 2d$ be an integer. Assume that for every positive integer $p \leqslant n$, the set A_p of periodic points of period p satisfies $\mathrm{boxdim}(A_p) < p/2$, and that the linearization Dg^p for each of these orbits has distinct eigenvalues.

Then for almost every smooth function h on U, the delay coordinate map $F(h, g)\colon U \to R^n$ is:

1. One-to-one on A.
2. An immersion on each compact subset C of a smooth manifold contained in A.

Remark 2.8. The probe space for this prevalent set can be taken to be any set $h_1,..., h_t$ of polynomials in k variables which includes all polynomials of total degree up to $2n$. Given any smooth function h_0 on U, for almost all choices of $\alpha = (\alpha_1,..., \alpha_t)$ from R^t, the function $h_\alpha = h_0 + \sum_{k=1}^{t} \alpha_k h_k$ satisfies properties 1 and 2.

Remark 2.9. The proof of Theorem 2.7 is easily extended to the more general case where the reconstruction map F consists of a mixture of lagged observations. The more general result says that

$$F(x) = (h_1(x),..., h_1(g^{n_1 - 1}(x)),..., h_q(x),..., h_q(g^{n_q - 1}(x)))$$

satisfies the conclusions of Theorem 2.7 as long as $n_1 + \cdots + n_q > 2d$ and the corresponding conditions on the periodic points are satisfied. Those conditions are that $\mathrm{boxdim}(A_p) < p/2$ for $p \leqslant \max\{n_1,..., n_q\}$.

The reconstruction of chaotic attractors using independent coordinates from a time series was advocated in 1980 by Packard *et al.*[21] The delay-coordinate map is attributed in that work to a communication with D. Ruelle. The method actually illustrated in ref. 21 is somewhat different; namely, it is to use the value u_t of the time series and its time derivatives $\dot{u}_t$, $\ddot{u}_t$,... as independent coordinates.

In 1981, Takens[27] published the first mathematical results on the delay-coordinate map. Around the same time, Roux and Swinney[23] exhibited plots of delay-coordinate reconstructions of experimental data from the Belousov–Zhabotinski reaction.

In 1985, Eckmann and Ruelle[9] took the idea one step further and suggested examining not only the delay coordinates of a point, but also the relation between the delay coordinates of a point and the next point which occurs T time units later. In principle, one can then approximate not only the attractor, but the attractor together with its dynamics. Since ref. 9 it has become common practice to gather points that are close in reconstruction space, and use their next images to construct a low-order parametric model which approximates the dynamics in a small region. This idea has begun to be used for prediction and noise reduction applications. See, for example, refs. 1, 6, 12, 13, 15, 16, 18, and 28.

2.3. Self-Intersection

In the case that the reconstruction dimension n is not greater than twice the box-counting dimension d of the set A, the map F in the fractal Whitney embedding prevalence theorem (Theorem 2.3) will often not be an embedding. However, if $d<n$, most of A will still be embedded. In the case that A is a smooth manifold of dimension d, almost every F will be an embedding outside a subset of A of dimension at most $2d-n$. If $d<n$, then $2d-n<d$, and so this exceptional subset will have positive codimension in A.

If A is simply a compact set of box-counting dimension d, then the situation is slightly different. We will call the pair x, y of points *δ-distant* if the distance between them is at least δ. Then we define the *δ-distant self-intersection set* of F to be the subset of A consisting of all x such that there is a δ-distant point y with $F(x)=F(y)$; that is,

$$\Sigma(F,\delta)=\{x\in A\colon F(x)=F(y)\text{ for some }y\in A,\ |x-y|\geqslant\delta\}$$

Then the result is that for every $\delta>0$, the lower box-counting dimension of the δ-distant self-intersection set $\Sigma(F,\delta)$ is at most $2d-n$ for almost every F. A precise statement is given by the next theorem.

Theorem 2.10 (Self-Intersection Theorem). Let A be a compact subset of R^k of box-counting dimension d, let $n\leqslant 2d$ be an integer, and let $\delta>0$. For almost every smooth map $F\colon R^k\to R^n$:

1. The δ-distant self-intersection set $\Sigma(F,\delta)$ of F has lower box-counting dimension at most $2d-n$.
2. F is an immersion on each compact subset C of an m-manifold contained in A except on a subset of C of dimension at most $2m-n-1$.

For example, consider mapping a circle to the real line. In this case $d=m=n=1$, and Theorem 2.10 says that a prevalent set of F are immersions outside a zero-dimensional set. This is clear from Fig. 3, where the zero-dimensional set consists of a pair of points. The map is at least 2 to 1 outside this set, and hence nowhere an embedding.

On the other hand, setting $d = m = 1$ and $n = 2$ in the theorem, we see that a prevalent set of maps F from the circle to the plane are immersions, and are embeddings outside a zero-dimensional subset. Thus, the maps shown in Figs. 1a and 1b are of the prevalent type, immersions which are one-to-one except for at most a discrete (zero-dimensional) set of points. Figure 1c, on the other hand, is nonprevalent. Almost any map near F will perturb away the cusp.

There is also a self-intersection version of the fractal delay embedding prevalence theorem (Theorem 2.5) which one gets by making the obvious changes. Thus, if $n \leqslant 2d$, then for each $\delta > 0$ there exists a subset $\Sigma(F, \delta)$, whose box-counting dimension is at most $2d - n$, on which the delay-coordinate map fails to be one-to-one. Note that the result is independent of $\delta > 0$. If M is a closed subsed of an m-manifold contained in A, then there is a subset E_1 of M of dimension at most $2m - n - 1$ on which the map fails to be an immersion.

2.4. How Many Delay Coordinates Do You Need?

When using a delay coordinate map (or filtered delay coordinate map, described in the next section) to examine the image $F(A)$ in R^n of a set A in R^k, the choice of n depends on the objective of the investigation. Different choices of n suffice for the different goals of prediction, calculation of dimension and Lyapunov exponents, and the determination of the stability of periodic orbits.

To compute the dimension of A, all that is required is that

$$\dim F(A) = \dim A \tag{2.1}$$

whether the dimension being used is box-counting, Hausdorff, information, or correlation dimension. The latter two depend on a probability density on A and $F(A)$. It is shown in ref. 24 that for the case of Hausdorff dimension, the equality (2.1) holds for almost every measurable map F, in the sense of prevalence, as long as $n \geqslant \dim(A)$. The probe space of perturbations for this result is the space of all linear transformations from R^k to R^n. Mattila[19] proved that equality (2.1) holds for almost every orthogonal projection F.

It is somewhat surprising that there are examples for which (2.1) does not hold for *any* map F when box-counting dimension is used, even under the hypothesis $n > \text{boxdim}(A)$. An example of this type is given in ref. 25. However, in most cases of compact sets which arise in dynamical systems, we expect Hausdorff dimension to equal box-counting dimension.

In practical situations, if attempts to measure $\text{boxdim}(A)$ result in answers dependent on n, where $n > \text{boxdim}(A)$, then the variation would seem to be a numerical artifact, since there is no theoretical justification for which of the values of n greater than boxdim(A) gives the more accurate result. The usual technique is to increase n until the observed dimension of boxdim $F(A)$ reaches a plateau, and to use this result. The resulting number might be called the *plateau dimension*. While the plateau dimension may indeed give the best numerical estimate of the dimension of A, there does not seem to be theoretical or numerical justification of this bias, and the question needs further investigation. Notice that $n > \text{boxdim}(A)$ does not guarantee that almost every F is one-to-one, but that is not required for dimension calculation.

If the objective is to use $F(A)$ to predict the future behavior of trajectories, then it is sufficient to have the map F be one-to-one, in which case $n > 2 \cdot \text{boxdim}(A)$ is needed. Knowing the current state in $F(A)$ is sufficient to predict the future of the trajectory (at least in the short run). In the situation of Fig. 1b, on the other hand, prediction on the periodic orbit A would still be possible, except when the trajectory was at the midpoint of the "figure eight."

If the objective is to compute the Lyapunov exponents of the system, it is necessary to ask which exponents are to be computed. For a simple example, suppose the attractor A is a periodic orbit. Then the best possible result of the examination of $F(A)$ is to observe that 0 is a Lyapunov exponent. The other exponents, presumably all negative, cannot be observed without introducing perturbations. More generally, if an attractor A lies on a manifold of dimension m (as a 2.2-dimensional attractor might lie on a three-dimensional manifold), it will certainly be impossible to measure more than m true exponents from an embedding, even if the reconstructed image $F(A)$ lies in R^n with $n > m$. There are no criteria for determining the smallest manifold containing A.

Theorems 2.3 and 2.5 say that if $n > 2 \cdot \text{boxdim}(A)$, then almost every F is an embedding of all smooth manifolds that lie in A. The smooth manifolds we have in mind are the surface corresponding to the unstable directions on the attractor A, that is, the unstable manifolds. Under an embedding, the differential information is preserved along smooth directions, such as unstable manifolds, indicating that positive Lyapunov exponents should be computable from the image $F(A)$.

The stable manifolds, on the other hand, will be likely to intersect A in a Cantor set. The image of a Cantor set in $F(A)$ may be quite compressed. For example, a set which is the product of five Cantor sets whose dimensions sum to 0.5 might be mapped to a one-dimensional line in $F(A)$. It seems difficult to recover any exponents in these directions from knowledge of the reconstructed dynamics in $F(A)$.

The self-intersection results in Section 2.3 are aimed at another kind of question. A relevant experiment involving a vibrating ribbon is described in refs. 8 and 26. In this case, the Poincaré map has an attractor whose dimension was experimentally calculated to be 1.2. The investigators were interested in determining the eigenvalues of the linearization of a period-3 point on the attractor.

Using a delay-coordinate map of the attractor into R^2 did not result in a one-to-one map, which is consistent with our results in Section 2.2. Theorem 2.10 of Section 2.3, which deals with self-intersection, suggests that the subset Σ of A on which the map into R^2 fails to be one-to-one should have dimension at most $2d-n=2\times 1.2-2=0.4$. They found that the self-intersection set looked like a finite set. If Σ indeed has dimension 0.4 or less, as we would expect, then the set Σ would be unlikely to include the periodic point in question, and the delay-coordinate map would be expected to be one-to-one in a neighborhood of that orbit. Numerical investigations of the dynamics near the periodic orbit revealed that the dynamics did appear to be two-dimensional, and the researchers were able to estimate numerically the eigenvalues of the orbit at these points.

3. THE DELAY COORDINATE MAP AND FILTERS

3.1. Main Results

So far, we have defined the delay coordinate map $x \to F(h, g)x$ from the hidden phase space R^k to the reconstruction space R^n. Under suitable conditions on the diffeomorphism g, the delay coordinate map $F(h, g)$ is an embedding for almost all observation functions h. In this formulation, information from the previous n time steps is used to identify a state of the original dynamical system in R^k.

For purposes of measuring quantitative invariants of the dynamical systems, noise reduction, or prediction, it may be advantageous to create an embedding that identifies a state with information from a larger number of previous time steps. However, working with embeddings in R^n is difficult for large n. A way around this problem is to incorporate large numbers of previous data readings by "averaging" their contributions in some sense. This problem has also been treated in ref. 7.

To this end, generalize the delay-coordinate map $F(h, g)$: $R^k \to R^w$,

$$F(h, g)x = (h(x), h(g(x)),..., h(g^{w-1}(x)))^T$$

where the superscript T denotes transpose, by defining the *filtered delay-coordinate map* $F(B, h, g)$: $R^k \to R^n$ to be

$$F(B, h, g)x = BF(h, g)x \tag{3.1}$$

where B is an $n \times w$ constant matrix. Thus, each coordinate of $F(B, h, g)x$ is a linear combination of the w coordinates of $F(h, g)x$. Here we are considering the case where g is a diffeomorphism, for notational convenience. Everything we say applies to a flow Φ by setting g equal to the time $-T$ map of the flow. We will call w the *window length* of the reconstruction, since there are w evenly-spaced observations used. We call n the *reconstruction dimension*, since R^n is the range space of the map. We may as well assume that $n \leqslant w$ and that B has rank n; otherwise we could throw away some rows of B without losing information. Assuming that B is a fixed matrix restricts the filter to be a linear multidimensional moving average (MA) filter. Autoregressive (AR) filters in general can change the dimension of the attractor.[4,20]

If B is the identity matrix (denoted I), the map is the original Takens delay coordinate map. As stated in the previous section, in that case, $F(I, h, g) = F(h, g)$ is almost always an embedding as long as n is greater than twice the box-counting dimension of the attractor and the periodic points of period p less than n have distinct eigenvalues and make up a set of boxdim $< p/2$.

Under filtering, some complications are caused by the existence of periodic cycles. On the other hand, the next theorem states that in the absence of cycles of length smaller than the window length w, every moving average filter B gives a faithful representation of the attractor.

Theorem 3.1 (Filtered Delay Embedding Prevalence Theorem). Let U be an open subset of R^k, g be a smooth diffeomorphism on U, and let A be a compact subset of U, $\operatorname{boxdim}(A) = d$. For a positive integer $n > 2d$, let B be an $n \times w$ matrix of rank n. Assume g has no periodic points of period less than or equal to w. Then for almost every smooth function h, the delay coordinate map $F(B, h, g)$: $U \to R^n$ is:

1. One-to-one on A.
2. An immersion on each closed subset C of a smooth manifold contained in A.

The probe space for perturbing h can be taken to be any space of polynomials in k variables which includes all polynomials of total degree up to $2w$. Furthermore, in case $n \leqslant 2d$, the results of Theorem 3.1 hold outside exceptional subsets of A precisely as in Theorem 2.10.

For example, consider the 3×9 matrix

$$B = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix} \tag{3.2}$$

Then

$$\begin{aligned} F(B, h, g)x = (&\tfrac{1}{3}(h(x) + h(g(x)) + h(g^2(x))),\\ &\tfrac{1}{3}(h(g^3(x)) + h(g^4(x)) + h(g^5(x))),\\ &\tfrac{1}{3}(h(g^6(x)) + h(g^7(x)) + h(g^8(x)))) \end{aligned}$$

Although the map $F(B, h, g)$ uses information from 9 different lags, the "moving average" reconstruction space is only 3-dimensional. According to the theorem, if the dynamical system g has no periodic points of period less than $w = 9$, then $F(B, h, g)$ is an embedding for almost all observation functions h.

Remark 3.2. When the diffeomorphism g has periodic points, certain special choices of filters B will cause self-intersection to occur at the periodic points. However, under the genericity hypotheses on the dynamical system of Theorem 2.5, for example, almost all choices of an $n \times w$ matrix B imply the conclusions of Theorem 3.1. This follows from Remarks 3.4 and 3.6. A more detailed view of the effect of periodic points of the dynamical system is given in Sections 3.3 and 3.4.

3.2. Examples of Filters

In this section we will list some examples of filters that may be useful in given situations. The easiest example is a simple averaging filter. For any integers m, n, let B be a $n \times mn$ matrix of form

$$B = \begin{pmatrix} 1/m \cdots 1/m & & & \\ & 1/m \cdots 1/m & & \\ & & \ddots & \\ & & & 1/m \cdots 1/m \end{pmatrix} \tag{3.3}$$

where there are m nonzero entries in each row. In the presence of noise, this filter should perform well compared to the more standard delay-coordinate embedding which uses every mth reading and discards the rest.

A more sophisticated noise filter was suggested in ref. 5 for a slightly different purpose, and elaborated on in the very readable ref. 2, where it is used for dimension measurements. It is based on the singular value decomposition from matrix algebra, also known as principal component analysis. Let $y_1, \ldots, y_L$ be the reconstructed vectors in R^w, where L is the length of the data series. Following Broomhead and King,[5] define the $L \times w$ trajectory matrix

$$A = \frac{1}{\sqrt{L}} \begin{pmatrix} y_1^T \\ \vdots \\ y_L^T \end{pmatrix}$$

where the y_i^T are treated as row vectors. The *covariance matrix* of this multivariate distribution is $A^T A$. The off-diagonal entries of $A^T A$ measure the statistical dependence of the variables.

The singular value decomposition[14] of the $L \times w$ matrix A, where $L \geqslant w$, is

$$A = VSU^T \tag{3.4}$$

where V is an $L \times L$ orthogonal matrix, U is a $w \times w$ orthogonal matrix (this means that $V^T V = I$, $U^T U = I$), and S is an $L \times w$ diagonal matrix (meaning that the entries σ_{ij} of S are zero if $i \neq j$). By rearranging the rows and columns of U and V, we can arrange for the *singular values* of A to satisfy $\sigma_{11} \geqslant \sigma_{22} \geqslant \cdots \geqslant \sigma_{ww} \geqslant 0$. The bottom $L - w$ rows of S are zero.

The singular value decomposition suggests the use of the filter $B = U^T$. That is, instead of plotting the vectors $y_1,..., y_L$ in reconstruction space R^w, plot the vectors $U^T y_1,..., U^T y_L$. One immediate positive consequence of this change of variables is the statistical linear independence of the new variables. The covariance matrix of the new trajectory matrix

$$\frac{1}{\sqrt{L}} \begin{pmatrix} (U^T y_1)^T \\ \vdots \\ (U^T y_L)^T \end{pmatrix} = AU$$

is $(AU)^T AU = S^T S$, a diagonal matrix.

In practice, one can do better than $B = U^T$. This is because some of the nonzero singular values are dominated by noise. A rule of thumb is to ignore (by setting to zero) all singular values below the noise floor of the experimental data. Ignoring all but the largest k singular values is equivalent to letting the filter B in Eq. (3.1) be the top k rows of U^T. The rows of U^T are orthogonal, so B is still full rank. Theorem 3.1 implies that $F(B, h, g)$ will typically be one-to-one and immersive.

This program was followed in ref. 2, in the context of measuring the correlation dimension of chaotic attractors in a stable way. They used a filter B that consisted of the rows of U^T that corresponded to singular values above 10^{-4}.

3.3. Conditions on Periodic Orbits Which Imply One-to-One

For special filters B, conclusions 1 and 2 of Theorem 3.1 can fail, but only for periodic points. That is, some periodic points of period less than w may be mapped together under the map $F(B, h, g)$.

For example, assume

$$B = \begin{pmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 & 0 \\ 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 \\ 0 & 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{pmatrix} \tag{3.5}$$

and assume that g has a period-4 orbit, that is, $g^4(x) = x$. Then for any h, $F(B, h, g)$ maps all four points of the period-4 orbit to the same point in R^3, so $F(B, h, g)$ fails to be one-to-one. There is no way for any observation function to distinguish the four points, since their outputs are being averaged over the entire cycle. Thus, the filtered delay coordinate map fails, for *all* observation functions h, to be one-to-one.

A similar problem occurs with the filter

$$B = \begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix} \tag{3.6}$$

Now

$$\begin{aligned} F(B, h, g)x = (&\tfrac{1}{2}(h(x) + h(g^2(x))), \\ &\tfrac{1}{2}(h(g(x)) + h(g^3(x))), \\ &\tfrac{1}{2}(h(g^2(x)) + h(g^4(x)))) \end{aligned}$$

Assume that the period-four orbit of g consists of $x_0, x_1 = g(x_0)$, $x_2 = g^2(x_0)$, and $x_3 = g^3(x_0)$. Now x_0 and x_2 are mapped to the same point in the reconstruction space R^3 by $F(B, h, g)$, and the same goes for x_1 and x_3. Again, the map cannot be one-to-one for any h.

A second obvious problem can be illustrated when the dynamical system has more than one fixed point. No matter how h is chosen, the filter

$$B = \begin{pmatrix} \frac{1}{2} & -\frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & -\frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{2} & -\frac{1}{2} \end{pmatrix} \tag{3.7}$$

maps all fixed points to the origin in R^3, violating the one-to-one condition.

In each of these situations, the underlying dynamical system g may dictate that some periodic points will become identified under a particular filter B, no matter how "generic" the observation function h. On the other hand, these identifications occur only at periodic points. Further, even in the case of periodic points, it turns out that the restrictions on B exemplified by the three cases above are the *only* restrictions. That is, if these are avoided, then $F(B, h, g)$ is one-to-one for a prevalent set of observation functions h.

To be more precise about these restrictions, we need to make some definitions. For each positive integer p, denote by A_p the set of period-p points of g lying on A. That is, $A_p = \{x \in A\colon g^p(x) = x\}$. Let I_n denote the $n \times n$ identity matrix and $(\cdot, \cdot)$ denote greatest common divisor. We will use the convention that $(p, 0) = 0$. For integers $p > q \geqslant 0$, define the $p \times (p - (p, q))$ matrix

$$C_{pq} = \begin{pmatrix} & I_{p-(p,q)} & \\ -I_{(p,q)} & \cdots & -I_{(p,q)} \end{pmatrix} \tag{3.8}$$

Define C_{pq}^{∞} to be the $\infty \times (p - (p, q)))$ matrix formed by repeating the block C_{pq} vertically, and for a positive integer w, define C_{pq}^{w} to be the matrix formed by the top w rows of C_{pq}^{∞}.

Theorem 3.3. Let U be an open subset of R^k, let g be a smooth diffeomorphism on U, and let A be a compact subset of U of box-counting dimension d. Let w and n be integers satisfying $w \geqslant n > 2d$. Assume that B is an $n \times w$ matrix of rank n which satisfies:

A1. rank $BC_{pc}^{w} > 2 \cdot \text{boxdim}(A_p)$ for all $1 \leqslant p \leqslant w$.

A2. rank $BC_{pq}^{w} > \text{boxdim}(A_p)$ for all $1 \leqslant q < p \leqslant w$.

Then for almost every smooth function h, $F(B, h, g)$ is one-to-one on A.

Remark 3.4. Note that rank $C_{pq} = p - (p, q)$, and so rank $C_{pq}^{w} = \min\{w, p - (p, q)\}$. It follows that rank $BC_{p0}^{w} \geqslant \min\{n, p\}$ and rank $BC_{pq}^{w} \geqslant \min\{n, p/2\}$ for $B = I_n$, and also for almost every $n \times w$ matrix B.

To illustrate the restrictions that Theorem 3.3 puts on moving average filters, assume that B is the 3×6 matrix (3.5). In particular, the filter B must satisfy condition A2 for $p = 4$, $q = 1$, which means

$$\text{rank } B \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ -1 & -1 & -1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} > \text{boxdim } A_4$$

The rank on the left-hand side is zero, however, and if there exists any period-4 orbit, the filter (3.5) fails this condition. This is consistent with what we have already noticed: in the presence of a period-4 orbit, the map $F(B, h, g)$ is not one-to-one for any h.

The filter (3.6) satisfies the above condition as long as there are finitely many period-4 orbits. However, it fails condition A2 for $p=4$, $q=2$, which requires

$$\operatorname{rank} B\begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -1 & 0 \\ 0 & -1 \\ 1 & 0 \end{pmatrix} > \operatorname{boxdim} A_4$$

This is again consistent with our earlier observation.

Finally, if there exist fixed points, the filter (3.7) fails the condition A1 for $p=1$ if there exist fixed points. That is because condition A1 requires

$$\operatorname{rank} B\begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} > 2 \cdot \operatorname{boxdim} A_1$$

Since the rank on the left side is zero, the condition fails unless the set of fixed points is empty.

3.4. Conditions on Periodic Orbits Which Imply an Immersion

There are also rather obvious situations when certain filters cause $F(B, h, g)$ to fail to an an immersion. Assume that g is a diffeomorphism on a circle that has a fixed point x. Assume that the derivative of g at x is -2. Consider the filter

$$B=\begin{pmatrix} \frac{2}{3} & \frac{1}{3} & 0 & 0 \\ 0 & \frac{2}{3} & \frac{1}{3} & 0 \\ 0 & 0 & \frac{2}{3} & \frac{1}{3} \end{pmatrix} \tag{3.9}$$

In this case, the map $F(B, h, g)$ cannot be an immersion at x for any observation function h. For a tangent vector v in $T_x M = R^1$, the derivative map is

$$DF(B, h, g)(x)v = B\begin{pmatrix} \nabla h(x)^T v \\ \nabla h(g(x))^T Dg(x)v \\ \vdots \\ \nabla h(g^{w-1}(x))^T Dg^{w-1}(x)v \end{pmatrix}$$

$$= \begin{pmatrix} \frac{2}{3} & \frac{1}{3} & 0 & 0 \\ 0 & \frac{2}{3} & \frac{1}{3} & 0 \\ 0 & 0 & \frac{2}{3} & \frac{1}{3} \end{pmatrix}\begin{pmatrix} \nabla h(x)^T(v) \\ \nabla h(x)^T(-2v) \\ \nabla h(x)^T(4v) \\ \nabla h(x)^T(-8v) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

so the tangent map of $F(B, h, g)$ at x is the zero map.

In the case of an m-dimensional manifold M with a fixed point x, it can be checked that for a filter B of this type, $F(B, h, g)$ will fail to be an immersion for all h as long as the linearization of g at x has an eigenvalue of -2. As in the one-to-one case, the immersion will fail only for periodic points.

To be precise, given numbers $c_1,..., c_r$, define the $\infty \times rp$ matrix

$$D_p^{\infty}(c_1,..., c_r) = \begin{pmatrix} I_p & \cdots & I_p \\ c_1 I_p & \cdots & c_r I_p \\ c_1^2 I_p & \cdots & c_r^2 I_p \\ \vdots & & \vdots \end{pmatrix} \tag{3.10}$$

where I_p denotes the $p \times p$ identity matrix. For a positive integer w, let $D_p^w(c_1,..., c_r)$ be the matrix formed by the top w rows of $D_p^{\infty}(c_1,..., c_r)$. If the c_i are distinct, then rank $D_p^w(c_1,..., c_r) = \min\{w, rp\}$.

Theorem 3.5. Let U be an open subset of R^k, let g be a smooth diffeomorphism on U, and let A be a compact subset of a smooth m-manifold in U. Let w and n be integers satisfying $w \geqslant n \geqslant 2n$. Assume that the linearizations Dg^p of periodic orbits of period p less than or equal to w have distinct eigenvalues. Assume that B is an $n \times w$ matrix of rank n which satisfies:

A3. rank $BD_p^w(\lambda_{i_1},..., \lambda_{i_r}) > \text{boxdim}(A_p + r - 1)$ for all $1 \leqslant p < w$, $1 \leqslant r \leqslant m$, and for all subsets $\lambda_{i_1},..., \lambda_{i_r}$ of eigenvalues of the linearization at a point in A_p.

Then for almost every smooth function h, $F(B, h, g)$ is an immersion on A.

Remark 3.6. See Theorem 4.14 for a proof. Note that since rank $D_p^w(\lambda_1,..., \lambda_r) = \min\{w, rp\}$ for distinct eigenvalues λ_i, it follows that rank $BD_p^w = \min\{n, rp\}$ for the original delay coordinate case of $B = I_n$, and also for almost every $n \times w$ matrix B.

To illustrate, the condition A3 is not satisfied for filter (3.9) when g has a fixed point with an eigenvalue of -2. That condition requires that rank $BD_1^w(-2) > 0$, but

$$BD_1^4(-2) = B\begin{pmatrix} 1 \\ -2 \\ (-2)^2 \\ (-2)^3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

4. PROOFS

This section contains the proofs of the results stated above. After some fundamental lemmas, we give the proofs of the Whitney forms of the embedding theorems. These follow Lemma 4.11. The proofs of the delay-coordinate forms involving filters, Theorems 3.3 and 3.5, follow immediately from Theorems 4.13 and 4.14, respectively. This section concludes with the proof of Theorems 2.7 and 3.1, which are special cases of Theorems 3.3 and 3.5.

Lemma 4.1. Let n and k be positive integers, $y_1,..., y_n$ distinct points in R^k, and $u_1,...,u_n$ in R, $v_1,..., v_n$ in R^k.

1. There exists a polynomial h in k variables of degree at most $n-1$ such that for $i = 1,..., n$, $h(y_i) = u_i$.
2. There exists a polynomial h in k variables of degree at most n such that for $i = 1,..., n$, $\nabla h(y_i) = v_i$.

Proof. 1. We may assume, by linear change of coordinates, that the first coordinates of $y_1,..., y_n$ are distinct. Then ordinary one-variable interpolation guarantees such a polynomial.

2. First assume $k = 1$. There exists a polynomial of degree at most $n-1$ in one variable that interpolates the data. The antiderivative is the desired polynomial h.

In the general case, by a linear change of coordinates, we may assume that for each $j = 1,..., k$, the jth coordinates of $y_1,..., y_n$ are distinct. The above paragraph shows that for $j = 1,..., k$ there is a polynomial of degree at most n in the jth coordinate x_j whose derivative h_{x_j} interpolates the jth coordinate of u_i for $i = 1,..., n$. The sum of all k of these polynomials is a polynomial of degree at most n which satisfies the conclusion.

Lemma 4.2. Let $F(x) = Mx + b$ be a map from R^t to R^n, where M is an $n \times t$ matrix and $b \in R^n$. For a positive integer r, let $\sigma > 0$ be the rth largest singular value of M. Denote by B_ρ the ball centered at the origin of radius ρ in R^t, and by B_δ the ball centered at the origin of radius δ in R^n. Then

$$\frac{\mathrm{Vol}(B_\rho \cap F^{-1}(B_\delta))}{\mathrm{Vol}(B_\rho)} < 2^{t/2}(\delta/\sigma\rho)^r$$

Proof. Note that decreasing any singular value of M does not decrease the left-hand side. Thus we may assume that the singular values of M satisfy $\sigma_1 = \cdots = \sigma_r = \sigma$, and $0 = \sigma_{r+1} = \sigma_{r+2} = \cdots$. Let $M = VSU^T$ be the singular value decomposition of M. Here S is a diagonal matrix with entries $s_{11} = \cdots = s_{rr} = \sigma$ and all other entries zero, V is an $n \times n$ orthogonal matrix, and U is a $t \times t$ orthogonal matrix.

Since the columns of U and V each form an orthonormal set, we recognize $MB_\rho = VSU^T B_\rho$ as an r-dimensional ball of radius $\sigma\rho$ lying in R^n. In fact, the first r columns of V magnified by the factor $\sigma\rho$ are radii which span MB_ρ.

The set $F^{-1}(B_\delta) \cap B_\rho$ consists of the vectors in B_ρ whose image by M lands in a ball of radius δ in R^n. This is a cylindrical subset of B_ρ with base dimension r and base radius δ/σ. The subset thus has t-dimensional volume less than $(\delta/\sigma)^r C_r \rho^{t-r} C_{t-r}$, where $C_r = \pi^{r/2}/(r/2)!$ denotes the volume of the r-dimensional unit ball. The volume of B_ρ is $\rho^t C_t$, so

$$\frac{\mathrm{Vol}(B_\rho \cap F^{-1}(B_\delta))}{\mathrm{Vol}(B_\rho)} < \frac{(\delta/\sigma)^r \rho^{t-r} C_{t-r} C_n}{\rho^t C_t} < 2^{t/2}\left(\frac{\delta}{\sigma\rho}\right)^r$$

Lemma 4.3. Let S be a bounded subset of R^k, $\mathrm{boxdim}(\bar{S}) = d$, and let $G_0, G_1, \ldots, G_t$ be Lipschitz maps from S to R^n. Assume that for each x in S, the rth largest singular value of the $n \times t$ matrix

$$M_x = \{G_1(x), \ldots, G_t(x)\}$$

is at least $\sigma > 0$. For each $\alpha \in R^t$ define $G_\alpha = G_0 + \sum_{i=1}^t \alpha_i G_i$. Then for almost every α in R^t, the set $G_\alpha^{-1}(0)$ has lower box-counting dimension at most $d - r$. If $r > d$, then $G_\alpha^{-1}(0)$ is empty for almost every α.

Proof. For a positive number ρ, define the set B_ρ to be the ball of radius ρ centered at the origin in R^t. For the purposes of proving the theorem, we may replace R^t by B_ρ. For the remainder of the proof, we will say that G_α has some property with probability p to mean that the Lebesgue measure of the set of $\alpha \in B_\rho$ for which G_α has the property is p times the measure of B_ρ. For example, if $x \in S$, then Lemma 4.2 shows that $|G_\alpha(x)| = |G_0(x) + M_x(\alpha)| \leqslant \varepsilon$ for $\alpha \in B_\rho$ with probability at most $2^{t/2}(\varepsilon/\sigma\rho)^r$.

Let $D > d$, and let $\varepsilon_0 > 0$ be such that for $0 < \varepsilon < \varepsilon_0$, the following two facts hold. First, S can be covered by ε^{-D} k-dimensional balls $B(x, \varepsilon)$ of radius ε, centered at $x \in S$. Second, by the Lipschitz condition there exists a constant C such that the image under any G_α, $\alpha \in B_\rho$, of any ε-ball in R^k intersecting S is contained in a $C\varepsilon$-ball in R^n. For the remainder of the proof, we assume $\varepsilon < \varepsilon_0$.

The probability that the set $G_\alpha(B(x, \varepsilon))$ contains 0 is at most the probability that $|G_\alpha(x)| < C\varepsilon$, which is a constant times e^r, since ρ and σ are fixed. For any positive number M, the probability that at least M of the ε^{-D} images $G_\alpha(B(x, \varepsilon))$ contain 0 is at most $C_1 \varepsilon^{r-D}/M$. Therefore, $G_\alpha^{-1}(0)$ can be covered by fewer than $M = \varepsilon^{-b}$ of the ε-balls except with probability at most $C_1 e^{b-(D-r)}$. As long as $b > D - r$, this probability can be made as small as desired by decreasing ε.

Let $p > 0$. There is a sequence $\{\varepsilon_i\}_{i=1}^{\infty}$ approaching 0 such that $G_\alpha^{-1}(0)$ can be covered by fewer than ε_i^{-b} balls except for probability at most $p2^{-i}$. Thus, the lower box-counting dimension of $G_\alpha^{-1}(0)$ is at most b, except for a probability p subset of α. Since $p > 0$ was arbitrary, lower $\text{boxdim}(G_\alpha^{-1}(0)) \leqslant b$ for almost every α. Finally, since this holds for all $b > d - r$, lower $\text{boxdim}(G_\alpha^{-1}(0)) \leqslant d - r$. ∎

Remark 4.4. In case boxdim(S) does not exist, the hypotheses of the lemma can be slightly weakened by allowing d to be the lower box-counting dimension of S. A slight adaptation of the proof shows that boxdim can be replaced throughout Lemma 4.3 by Hausdorff dimension. In particular, if $r > \text{HD}(S)$, then $G_\alpha^{-1}(0)$ is empty for almost every α in R^t.

If in Lemma 4.3 we assume that $\text{rank}(M_x) \geqslant d$ for each $x \in S$ instead of the assumption on the singular values, then $G_\alpha^{-1}(0)$ is empty for almost every α. That is because one can apply Lemma 4.3 to the set $S_\sigma = \{x \in S$: rth largest singular value of $M_x > \sigma\}$ to get $G_\alpha^{-1}(0) \cap S_\sigma = \varnothing$. Then $S = \bigcup_{\sigma > 0} S_\sigma$ implies $G_\alpha^{-1}(0) = \varnothing$. We state this fact in the next lemma.

Lemma 4.5. Let S be a bounded subset of R^k, $\text{boxdim}(\bar{S}) = d$, and let $G_0, G_1, \ldots, G_t$ be Lipschitz maps from S to R^n. Assume that for each x in S, the rank of the $n \times t$ matrix

$$M_x = \{G_1(x), \ldots, G_t(x)\}$$

is at least r. For each $\alpha \in R^t$ define $G_\alpha = G_0 + \sum_{i=1}^{t} \alpha_i G_i$. Then for almost every α in R^t, the set $G_\alpha^{-1}(0)$ is the nested countable union of sets of lower box-counting dimension at most $d - r$. If $r > d$, then $G_\alpha^{-1}(0)$ is empty for almost every α.

Lemma 4.6. Let A be a compact subset of R^k. Let $F_0, F_1,..., F_t$ be Lipschitz maps from A to R^n. For each integer $r \geqslant 0$, let S_r be the set of pairs $x \neq y$ in A for which the $n \times t$ matrix

$$M_{xy} = \{F_1(x) - F_1(y),..., F_t(x) - F_t(y)\}$$

has rank r, and let $d_r = \text{lower boxdim}(\overline{S_r})$. Define $F_\alpha = F_0 + \sum_{i=1}^{t} \alpha_i F_i$: $A \to R^n$. Then for $\alpha = (\alpha_1,..., \alpha_k)$ outside a measure zero subset of R^t, the following hold:

1. If $d_r < r$ for all integers $r \geqslant 0$, then the map F_α is one-to-one.
2. If $d_r \geqslant r$ for some integer $r \geqslant 0$, then for every $\delta > 0$, the lower box-counting dimension of the δ-distant self-intersection set $\Sigma(F_\alpha, \delta)$ is at most $d_r - r$.

Proof. For $i = 0,..., t$, define $G_i(x, y) = F_i(x) - F_i(y)$. On the set S_r, the rank of the $n \times t$ matrix

$$M_{xy} = \{G_1(x, y),..., G_t(x, y)\}$$

is r.

If $r > d_r$, Lemma 4.5 shows that for almost every $\alpha \in R^t$, the origin is not in the image of S_r under the map $G_\alpha = G_0 + \sum \alpha_i G_i$, or equivalently, $F_\alpha(x) \neq F_\alpha(y)$ for $x \neq y$ in S_r. If $r > d_r$ for all r, then F_α is one-to-one, since each pair $x \neq y$ lies in some S_r.

If $r \leqslant d_r$, let $(A \times A)_\delta = \{(x, y) \in A \times A\colon |x - y| \geqslant \delta\}$ be the subset of δ-distant pairs of points in $A \times A$. Since $(A \times A)_\delta$ is compact for any $\delta > 0$, the minimum of the nth singular value of M_{xy} in $(A \times A)_\delta$ is greater than 0. Lemma 4.3 shows that for almost every α, the origin is in $G_\alpha((A \times A)_\delta)$ for a subset of $(A \times A)_\delta$ with lower box-counting dimension at most $d_r - r$. Therefore the δ-distant self-intersection subset $\Sigma(F_\alpha, \delta)$ of A, which is the image of this subset under the projection of $(A \times A)_\delta$ to A, has dimension at most $d_r - r$. ∎

Theorem 4.7. Let A be a compact subset of R^k, lower boxdim$(A) = d$. If $n > 2d$, then almost every linear transformation of R^k to R^n is one-to-one on A.

Proof. This follows immediately from Lemma 4.6 and the remark following it. Let $\{F_i\}$ be a basis for the nk-dimensional space of linear transformations. For each pair $x \neq y$, the vector $x - y$ can be moved to any direction in R^n by a linear transformation. In the terminology of Lemma 4.6, $S_n = A \times A - \Delta$ and S_r is empty for $r \neq n$. Since lower boxdim$(\overline{S_n}) = 2d < n$, almost every $F_\alpha = \sum \alpha_i F_i$ is one-to-one on A. ∎

Remark 4.8. It is interesting that no statement similar to Theorem 4.7 can be made if box-counting dimension is replaced by Hausdorff dimension. In an Appendix to this work provided by I. Kan, examples are constructed of compact subsets A of any Euclidean space R^k that have Hausdorff dimension $d=0$, and such that no projection to R^n for $n<k$ is one-to-one on A.

This striking difference between box-counting dimension and Hausdorff dimension is related to the fact that Hausdorff dimension does not work well with products. Extra hypotheses are needed on A, in particular on the Hausdorff dimension of the product $A \times A$, to prove an analogue to Theorem 4.7. For example, Mañé has shown (see ref. 17 and its correction in ref. 9, p. 627) that if $n > \mathrm{HD}(A \times A) + 1$, then the conclusion of Theorem 4.7 again holds. Of course, using Lemma 4.3 and Remark 4.4, it turns out that only $n > \mathrm{HD}(A \times A)$ is required:

Theorem 4.9. Let A be a compact subset of R^k, and let $n > \mathrm{HD}(A \times A)$. Then almost every linear transformation of R^k to R^n is one-to-one on A.

It was shown in ref. 10 that under the hypotheses of Theorem 4.7, almost every *orthogonal* projection is one-to-one (and in fact has a Hölder continuous inverse).

Definition 4.10. For a compact differentiable manifold M, let $T(M) = \{(x, v)\colon\ x \in M,\ v \in T_x M\}$ be the *tangent bundle* of M, and let $S(M) = \{(x, v) \in T(M)\colon |v| = 1\}$ denote the *unit tangent bundle* of M.

Lemma 4.11. Let A be a compact subset of a smooth manifold embedded in R^k. Let $F_0, F_1,..., F_t\colon R^k \to R^n$ be a set of smooth maps from an open neighborhood U of A to R^n. For each positive integer r, let S_r be the subset of the unit tangent bundle $S(A)$ such that the $n \times t$ matrix

$$\{DF_1(x)(v),..., DF_t(x)(v)\}$$

has rank r, and let $d_r = \text{lower boxdim}(\overline{S_r})$. Define $F_\alpha = F_0 + \sum_{i=1}^{t} \alpha_i F_i\colon U \to R^n$. Then the following hold:

1. If $d_r < r$ for all integers $r \geqslant 0$, then for almost every $\alpha \in R^t$, the map F_α is an immersion on A.
2. If $d_r \geqslant r$ for some $r \geqslant 0$, then for almost every $\alpha \in R^t$, F_α is an immersion outside a subset of A of lower boxdim $\leqslant d_r - r$.

Proof. For $i = 0,..., t$, define $G_i\colon S(A) \to R^n$ by $G_i(x, v) = DF_i(x)v$. If $r > d_r$ for all $r \geqslant 0$, then Lemma 4.5 applies to show that for almost every α, $G_\alpha^{-1}(0) \cap S_r$ is the empty set. Since $S(A)$ is the union of all S_r, $G_\alpha^{-1}(0)$

is empty. Thus, no unit tangent vector is mapped to the origin, and F_α is an immersion.

In case $d_r \geqslant r$ for some r, there is a positive lower bound on the singular values of the G_i on $S(A)$. Lemma 4.3 implies that there is a subset of unit tangent vectors of lower boxdim $\leqslant d_r - r$ that can map to zero. The projection of this subset into A has lower boxdim $\leqslant d_r - r$. ▮

Proof of Theorems 2.2, 2.3, and 2.10. Theorem 2.2 is a special case of Theorem 2.3. To prove the latter, we need to show that a prevalent set of maps are one-to-one and immersive.

Let $F_1,\ldots,F_t$ be a basis for the set of linear transformations from $R^k \to R^n$. In the notation of Lemma 4.6, the set $S_n = A \times A - \Delta$ and $S_r = \varnothing$ for $r \neq n$. Since $\operatorname{boxdim}(A \times A) = 2d < n$, F_α is one-to-one on A for almost every $\alpha \in R^t$. If any other maps $F_{t+1},\ldots,F_{t'}$ are added, the rank of M_{xy} cannot drop for any pair $x \neq y$, so almost every linear combination of $F_1,\ldots, F_{t'}$ is one-to-one on A.

The proof of the immersion half uses Lemma 4.11 instead of Lemma 4.6. Since $\operatorname{boxdim}(A) = d$, C is a subset of a smooth manifold of dimension at most d, and therefore boxdim $S(C) \leqslant 2d - 1$. In the notation of Lemma 4.11, $S_n = S(C)$ and $S_r = \varnothing$ for $r \neq n$. Since $n > 2d > 2d - 1 =$ boxdim S_n, the proof follows from Lemma 4.11.

The proof of Theorem 2.10 is similar, except that the second part of the conclusions of Lemmas 4.6 and 4.11 are used. For example, in the use of Lemma 4.6, $S_n = A \times A - \Delta$ and $S_r = \varnothing$ for $r \neq n$ as before, but now $\operatorname{boxdim}(A \times A) = 2d \geqslant n$. Thus for each $\delta > 0$, for almost every F_α, the δ-distant self-intersection set $\Sigma(F_\alpha, \delta)$ has lower box-counting dimension at most $2d - n$. The immersion half is again analogous. ▮

Definition 4.12. Let U be an open subset of R^k, let g: $U \to U$ be a map, and let h: $U \to R$ be a function. Let $w^- < w^+$ be integers and set $w = w^+ - w^- + 1$. For $1 \leqslant i \leqslant w$, set $g_i = g^{w^- + i - 1}$, so that $g_1 = g^{w^-}$ and $g_w = g^{w^+}$. Let B be an $n \times w$ matrix. Define the *filtered delay-coordinate map*

$$F_{w^-}^{w^+}(B, h, g)\colon \quad U \to R^n$$

by

$$\begin{aligned} F_{w^-}^{w^+}(B, h, g)(x) &= B(h(g_1(x)), h(g_2(x)),\ldots, h(g_w(x)))^T \\ &= B(h(g^{w^-}(x)),\ldots, h(g^{w^+}(x)))^T \end{aligned}$$

Theorems 2.7, 3.1, 3.3, and 3.5 are corollaries of the next two results, for which we will use the following notation. Let g denote a smooth diffeomorphism on an open neighborhood U in R^k. Let $h_1,\ldots,h_t$ be a basis for the polynomials in k variables of degree at most $2w$. For a smooth function h_0 on R^k and for $\alpha \in R^t$, define $h_\alpha = h_0 + \sum_{i=1}^t \alpha_i h_i$. For each positive integer p, denote by A_p the set of period-p points of g lying on A. That is, $A_p = \{x \in A: g^p(x) = x\}$. Let the matrices C_{pq}^w be as in Theorem 3.3.

Theorem 4.13. Let g be a smooth diffeomorphism on an open neighborhood U of R^k, and let A be a compact subset of U, $\operatorname{boxdim}(A) = d$. Let n and $w^- < w^+$ be integers, $n \leqslant w = w^+ - w^- + 1$. Assume that the $n \times w$ matrix B satisfies:

A1. rank $BC_{p0}^w > 2 \cdot \operatorname{boxdim}(A_p)$ for all $1 \leqslant p \leqslant w$.

A2. rank $BC_{pq}^w > \operatorname{boxdim}(A_p)$ for all $1 \leqslant q < p \leqslant w$.

Let $h_1, \ldots, h_t$ be a basis for the polynomials in k variables of degree at most $2w$. Then for any smooth function h_0 on R^k, and for almost every $\alpha \in R^t$, the following hold:

1. If $n > 2d$, then $F(B, h_\alpha, g)$: $U \to R^n$ is one-to-one on A.
2. If $n \leqslant 2d$, then for every $\delta > 0$, the δ-distant self-intersection set $\Sigma(F(B, h_\alpha, g), \delta)$ has lower box-counting dimension at most $2d - n$.

Proof. For $i = 1, \ldots, t$ define

$$F_i(x) = B \begin{pmatrix} h_i(g_1(x)) \\ \vdots \\ h_i(g_w(x)) \end{pmatrix}$$

By definition, $F(B, h_\alpha, g) = \sum_{i=1}^t F_i$. To use Lemma 4.6, we need to check for each $x \neq y$ the rank of the matrix

$$M_{xy} = (F_1(x) - F_1(y), \ldots, F_t(x) - F_t(y))$$

which can be written as

$$B \begin{pmatrix} h_1(g_1(x)) - h_1(g_1(y)) & \cdots & h_t(g_1(x)) - h_t(g_1(y)) \\ \vdots & & \vdots \\ h_1(g_w(x)) - h_1(g_w(y)) & \cdots & h_t(g_w(x)) - h_t(g_w(y)) \end{pmatrix} = BJH$$

where

$$H = \begin{pmatrix} h_1(z_1) & \cdots & h_t(z_1) \\ \vdots & & \vdots \\ h_1(z_q) & \cdots & h_t(z_q) \end{pmatrix}$$

$q \leqslant 2w$, the z_i are distinct, and $J = J_{xy}$ is a $w \times q$ matrix each of whose rows consists of zeros except for one 1 and one -1. By part 1 of Lemma 4.1, the rank of H is q. We divide the study of the rank of $M_{xy} = BJH$ into three cases.

Case 1: x and y are not both periodic with period $\leqslant w$.

In this case, J_{xy} is upper or lower triangular, and $\text{rank}(J_{xy}) = w$. Since B, J, and H are onto linear transformations, the product BJH is onto and has rank n. The set of pairs $x \neq y$ of case 1 has box-counting dimension at most $2d$, and $\text{rank}(M_{xy}) = n$. If g has no periodic points of period $\leqslant w$, we are done, and conclusion 1 (respectively, 2) of Lemma 4.6 implies conclusion 1 (resp., 2) of the theorem.

The remaining two cases are necessary to deal with periodic points of period $\leqslant w$. We show that conclusion 1 of Lemma 4.6 applies in both cases.

Case 2: x and y lie in distinct periodic orbits of period $\leqslant w$.

Assume p and q are minimal such that $g^p(x) = x$, $g^q(y) = y$, and that $1 \leqslant q \leqslant p \leqslant w$. In this case the matrix J_{xy} contains a copy of C^w_{p0}. Since H is onto, $\text{rank } M_{xy} = \text{rank } BJ_{xy}H = \text{rank } BJ_{xy}$. By hypothesis, $\text{rank } BJ_{xy} \geqslant \text{rank } BC^w_{p0} > 2 \cdot \text{boxdim } A_p$, which is the box-counting dimension of the set of pairs treated in case 2. By Lemma 4.6, for almost every $\alpha \in R^t$, $F_\alpha(x) \neq F_\alpha(y)$ for every such pair $x \neq y$.

Case 3: Both x and y lie in the same periodic orbit of period $\leqslant w$.

Assume p and q are minimal such that $g^p(x) = x$, $g^q(x) = y$, and that $1 \leqslant q < p \leqslant w$. Since x and y lie in the same periodic orbit, the column space of J_{xy} contains the column space of C^w_{pq}. Thus, $\text{rank } BJ_{xy}H = \text{rank } BJ_{xy} \geqslant \text{rank } BC^w_{pq} > \text{boxdim } A_p$, which is the dimension of the pairs $x \neq y$ of case 3. Now Lemma 4.6 applies to give the conclusion. ∎

Theorem 4.14. Let g be a smooth diffeomorphism on an open neighborhood U in R^k, and let A be a compact subset of a smooth m-manifold in U. Assume that the linearizations of periodic orbits of period less than w have distinct eigenvalues. Let $n \leqslant w$ be positive integers as in Theorem 4.13, and assume that the $n \times w$ matrix B satisfies:

A3. $\text{rank } BD^w_p(\lambda_{i_1}, \ldots, \lambda_{i_r}) > \text{boxdim}(A_p + r - 1)$ for all $1 \leqslant p < w$, $1 \leqslant r \leqslant m$, and for all subsets $\lambda_{i_1}, \ldots, \lambda_{i_r}$ of eigenvalues of the linearization Dg^p at a point in A_p.

Let $h_1, \ldots, h_t$ be a basis for the polynomials in k variables of degree at most $2w$. Then for any smooth function h_0 on R^k, and for almost every $\alpha \in R^t$, the following hold:

1. If $n \geqslant 2m$, then $F(B, h_\alpha, g)$: $U \to R^n$ is an immersion on A.
2. If $n < 2m$, then $F(B, h_\alpha, g)$ is an immersion outside an exceptional subset of A of dimension at most $2m - n - 1$.

Proof. To apply Lemma 4.11, we need to check the rank of the $n \times t$ matrix

$$(DF_1(x)(v),..., DF_t(x)(v)) \tag{4.1}$$

for each (x, v) in the unit tangent bundle $S(A)$. For a given observation function h, the derivative of $F(B, h, g)$ is

$$DF(B, h, g)(x)v = B \begin{pmatrix} \nabla h(g^{\omega^-}(x))^T Dg^{w^-}(x)v \\ \vdots \\ \nabla h(g^{w^+}(x))^T Dg^{w^+}(x)v \end{pmatrix}$$

If x is not a periodic point of period less than w, then $g^{w^-}(x),..., g^{w^+}(x)$ are distinct points. The facts that g is a diffeomorphism and $v \neq 0$ imply that $Dg^i(x)v \neq 0$ for all i. Therefore by Lemma 4.1, part 2, the set of vectors $\{DF(B, h_\alpha, g)(x)v\colon \alpha \in R^t\}$ spans R^n. In the notation of Lemma 4.11, the subset S_n contains all points of $S(A)$ that are not periodic with period less than w, and $d_n = \text{lower boxdim}(\overline{S_n}) \leqslant 2m - 1$. If g has no periodic points of period less than w, the proof is finished, by Lemma 4.11.

If x is a periodic point of period $p < w$, then

$$DF(B, h, g)(x)v = B \begin{pmatrix} H_1^T w_1 \\ \vdots \\ H_p^T w_p \\ H_1^T D_1 w_1 \\ \vdots \\ H_p^T D_p w_p \\ H_1^T D_1^2 w_1 \\ \vdots \end{pmatrix}$$

where

$$\begin{aligned} x_i &= g^{w^- + i}(x) = x_{p+i} \\ H_i &= \nabla h(x_i) \\ w_i &= Dg(x_{i-1}) \cdots Dg(x_1)\, Dg^{w^-}(x)v \\ D_i &= Dg(x_{i-1}) \cdots Dg(x_1)\, Dg(x_p) \cdots Dg(x_i) \end{aligned}$$

Each matrix D_i has the same set of eigenvalues $\lambda_1,..., \lambda_m$, and by hypothesis, they are distinct. If $u_1,..., u_m$ is a spanning set of eigenvectors for D_1, then it checks that $u_{ij} = Dg(x_{i-1}) \cdots Dg(x_1)u_j$ for $1 \leqslant i \leqslant p$, $1 \leqslant j \leqslant m$ defines a spanning set $\{u_{i1},..., u_{im}\}$ of eigenvectors for D_i. Thus, if

$w_1 = \sum_{j=1}^{m} a_j u_{1j}$ is the eigenvector expansion of w_1, then the eigenvector expansion of w_i is $\sum_{j=1}^{m} a_j u_{ij}$, which has the same coefficients.

Thus $DF(B, h, g)(x)v$ can be written as B times the w-vector

$$\begin{bmatrix} 1 & \cdots & 1 \\ 0 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & 0 \\ 0 & \cdots & 0 \\ \lambda_1 & \cdots & \lambda_m \\ 0 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & 0 \\ 0 & \cdots & 0 \\ \lambda_1^2 & \cdots & \lambda_m^2 \\ \vdots & & \vdots \end{bmatrix} \begin{pmatrix} a_1 u_{11}^T \\ \vdots \\ a_m u_{1m}^T \end{pmatrix} H_1 + \cdots + \begin{bmatrix} 0 & \cdots & 0 \\ 0 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & 0 \\ 1 & \cdots & 1 \\ 0 & \cdots & 0 \\ 0 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & 0 \\ \lambda_1 & \cdots & \lambda_m \\ 0 & \cdots & 0 \\ \vdots & & \vdots \end{bmatrix} \begin{pmatrix} a_1 u_{p1}^T \\ \vdots \\ a_m u_{pm}^T \end{pmatrix} H_p \tag{4.2}$$

To find the rank of the matrix (4.1) for (x, v) where x is periodic, we need to find the span of B times the vectors (4.2) for $h = h_\alpha = \sum \alpha_i h_i$, $\alpha \in R^t$. Assume that the eigenvector expansion of v has exactly r nonzero coefficients $a_{i_1}, ..., a_{i_r}$. By Lemma 4.1, part 2, the set of vectors $\{\nabla h_\alpha(x_i)$: $\alpha \in R^t\}$ spans R^k. Then because the u_{ij}, $1 \leqslant j \leqslant m$, are linearly independent, the vectors of form (4.2) span a space of dimension $\min\{w, rp\}$ as α spans R^t.

Therefore, for this (x, v), the span of the vectors (4.1) has dimension equal to the rank of $BD_p^w(\lambda_{i_1}, ..., \lambda_{i_r})$. By hypothesis, the boxdim of such pairs (x, v) in $S(A)$ is $\operatorname{boxdim}(A_p) + r - 1$. By hypothesis, the rank of the $n \times t$ matrix (4.1) is strictly larger, so that Lemma 4.11 applies to give the conclusion.

Proof of Theorem 2.7. Apply Theorems 3.3 and 3.5 with $B = I_n$. According to Remarks 3.4 and 3.6, the conditions A1–A3 translate to $p > 2 \cdot \operatorname{boxdim}(A_p)$, $p/2 > \operatorname{boxdim}(A_p)$, and $\min\{n, rp\} > \operatorname{boxdim}(A_p) + r - 1$, respectively, for $1 \leqslant p \leqslant n$ and $1 \leqslant r \leqslant m$. Thus, the hypothesis $\operatorname{boxdim}(A_p) < p/2$ guarantees that A1–A3 hold.

Proof of Theorem 3.1. Since A_p is empty for $1 \leqslant p \leqslant w$, the conditions A1–A3 of Theorems 3.3 and 3.5 are satisfied vacuously.

APPENDIX. HAUSDORFF DIMENSION-ZERO SETS WITH NO ONE-TO-ONE PROJECTIONS

Ittai Kan[4]

The purpose of this Appendix is to construct a Cantor set $C \subset R^m$ whose Hausdorff dimension is zero and which has the property that every projection of rank less than m is not one-to-one when restricted to C.

Definition A.1. The *Hausdorff s-dimensional outer measure* of a set K is

$$\mathscr{H}^s(K) = \lim_{\delta \downarrow 0} \inf_{|U_i| < \delta} \sum_{i=1}^{\infty} |U_i|^s$$

where the infimum is taken over all covers $\{U_i\}$ of K with the diameters of the U_i uniformly less than δ. The *Hausdorff dimension* of a nonempty set K is the unique value of s such that

$$\mathscr{H}^t(K) = \infty \text{ if } t < s \qquad \text{and} \qquad \mathscr{H}^t(K) = 0 \text{ if } t > s$$

Example A.2. We construct the subset C of R^m as the union of two sets $A = \bigcup_{n=1}^m A_n$ and $B = \bigcup_{n=1}^m B_n$ each of Hausdorff dimension zero, with the property that for any projection P of rank less than m the images under P of A and B intersect, and thus P is not injective when restricted to C.

The set A_n lies on a face of the unit m-cube and $a = (a_1, a_2, \ldots, a_m)$ is in A_n if it satisfies the following restrictions on the binary expansion $a_i = a_i^1 a_i^2 a_i^3 \ldots$ of its coordinates:

1. If $i = n$, then $a_i^l = 0$.
2. If $i \neq n$ and $k \geqslant 0$, then either (a) $a_i^l = 0$ for all $l \in (M_{2k}, M_{2k+1}]$; or (b) $a_i^l = 1$ for all $l \in (M_{2k}, M_{2k+1}]$.

Here the sequence $0 = M_0 < M_1 < M_2 \ldots$ increases sufficently rapidly so that $\lim(M_{j+1}/M_j) = \infty$. If $i \neq n$, then the orthogonal projection of A_n on the ith coordinate axis is a Cantor set which can be covered by 2^{r_k} intervals of length $2^{-M_{2k+1}}$, where $r_k = k + \sum_{j=1}^k (M_{2j} - M_{2j-1})$. Thus, A_n can be covered by $2^{(m-1)r_k}$ cubes with edges of length $2^{-M_{2k+1}}$. Since $r_k \leqslant M_{2k}$, we see that $\lim_{k \to \infty}(m-1)r_k/M_{2k+1} = 0$ and both the lower box-counting and Hausdorff dimensions of A_n are zero. Since A is the union of m copies of A_n, we see that both the lower box-counting and Hausdorff dimensions of A are zero.

[4] Department of Mathematical Sciences, George Mason University, Fairfax, Virginia 22030.

The set B_n lies on a face of the unit m-cube opposite A_n and b is in B_n if it satisfies the following restrictions on the binary expansion of its coordinates:

1. If $i = n$, then $b_i^l = 1$.
2. If $i \neq n$ and $k \geqslant 0$, then either (a) $b_i^l = 0$ for all $l \in (M_{2k+1}, M_{2k+2}]$; or (b) $b_i^l = 1$ for all $l \in (M_{2k+1}, M_{2k+2}]$.

Here $\{M_j\}$ is as above. The lower box-counting and Hausdorff dimensions of B are zero. The Hausdorff dimension of $C = A \cup B$ is zero.

Let P denote a projection of rank less than m. Let $v = (v_1, v_2, \ldots, v_m)$ in the null space of P be chosen so that $|v_i| \leqslant 1$ for all i and $v_n = 1$ for some particular n. We now show that P restricted to C is not injective by finding some $b \in B_n$ and $a \in A_n$ such that $v = b - a$. Using the binary expansion coordinate notation, we define a and b as follows:

1. If $i = n$, then $a_i^l = 0$ and $b_i^l = 1$.
2. If $i \neq n$ and $k \geqslant 0$, then (a) $a_i^l = 0$ and $b_i^l = v_i^l$ for all $l \in (M_{2k}, M_{2k+1}]$; and (b) $a_i^l = (v_i^l + 1) \bmod 2$ and $b_i^l = 1$ for all $l \in (M_{2k+1}, M_{2k+2}]$.

Clearly we have $v = b - a$ and by the definition of A_n and B_n we also have $a \in A_n$ and $b \in B_n$. ∎

ACKNOWLEDGMENTS

The research of T.S. and J.A.Y. was supported by the Applied and Computational Mathematics Program of DARPA, that of J.A.Y. additionally by AFOSR and the U. S. Department of Energy (Basic Energy Sciences), and that of M.C. by grants to the Santa Fe Institute, including core funding from the John D. and Catherine T. MacArthur Foundation, the National Science Foundation, and the U. S. Department of Energy.

REFERENCES

1. H. Abarbanel, R. Brown, and J. Kadtke, Prediction in chaotic nonlinear systems: Methods for time series with broadband Fourier spectra, preprint.
2. A. M. Albano, J. Muench, C. Schwartz, A. Mees, and P. Rapp, Singular value decomposition and the Grassberger–Procaccia algorithm, *Phys. Rev. A* **38**:3017–3026 (1988).
3. V. I. Arnold, *Geometrical Methods in the Theory of Ordinary Differential Equations* (Springer-Verlag, New York, 1983).
4. R. Badii, G. Broggi, B. Derighetti, M. Ravani, S. Ciliberto, A. Politi, and M. A. Rubio, Dimension increase in filtered chaotic signals, *Phys. Rev. Lett.* **60**:979–982 (1988).
5. D. S. Broomhead and G. P. King, Extracting qualitative dynamics from experimental data, *Physica* **20D**:217–236 (1986).

6. M. Casdagli, Nonlinear prediction of chaotic time series, *Physica* **35D**:335–356 (1989).
7. M. Casdagli, S. Eubank, D. Farmer, and J. Gibson, State-space reconstruction in the presence of noise, preprint.
8. W. Ditto, S. Rauseo, and M. Spano, Experimental control of chaos, *Phys. Rev. Lett.* **65**:3211–3214 (1990).
9. J.-P. Eckmann and D. Ruelle, Ergodic theory of chaos and strange attractors, *Rev. Mod. Phys.* **57**:617–656 (1985).
10. A. Eden, C. Foias, B. Nicolaenko, and R. Temam, Hölder continuity for the inverse of Mañé's projection, *Comptes Rendus*, to appear.
11. K. Falconer, *Fractal Geometry* (Wiley, New York, 1990).
12. J. D. Farmer and J. Sidorowich, Predicting chaotic time series, *Phys. Rev. Lett.* **59**:845–848 (1987).
13. J. D. Farmer and J. Sidorowich, Exploiting chaos to predict the future and reduce noise, Technical Report LA-UR-88-901, Los Alamos National Laboratory (1988).
14. G. Golub and C. Van Loan, *Matrix Computations*, 2nd ed. (Johns Hopkins University Press, Baltimore, Maryland, 1989).
15. E. Kostelich and J. Yorke, Noise reduction: Finding the simplest dynamical system consistent with the data, *Physica* **41D**:183–196 (1990).
16. E. Kostelich and J. Yorke, Noise reduction in dynamical systems, *Phys. Rev. A* **38**:1649–1652 (1988).
17. R. Mañé, On the dimension of the compact invariant sets of certain nonlinear maps, in *Lecture Notes in Mathematics*, No. 898 (Springer-Verlag, 1981).
18. P. Marteau and H. Abarbanel, Noise reduction in chaotic time series using scaled probabilistic methods, preprint.
19. P. Mattila, Hausdorff dimension, orthogonal projections and intersections with planes, *Ann. Acad. Sci. Fenn. Math.* **1**:227–224 (1975).
20. F. Mitschke, M. Möller, and W. Lange, Measuring filtered chaotic signals, *Phys. Rev. A* **37**:4518–4521 (1988).
21. N. Packard, J. Crutchfield, D. Farmer, and R. Shaw, Geometry from a time series, *Phys. Rev. Lett.* **45**:712 (1980).
22. W. Rudin, *Real and Complex Analysis*, 2nd ed. (McGraw-Hill, New York, 1974).
23. J.-C. Roux and H. Swinney, Topology of chaos in a chemical reaction, in *Nonlinear Phenomena in Chemical Dynamics*, C. Vidal and A. Pacault, eds. (Springer, Berlin, 1981).
24. B. Hunt, T. Sauer and J. Yorke, Prevalence: A translation-invariant "almost every" on infinite-dimensional spaces, preprint.
25. T. Sauer and J. Yorke, Statistically self-similar sets, preprint.
26. J. Sommerer, W. Ditto, C. Grebogi, E. Ott, and M. Spano, Experimental confirmation of the theory for critical exponents of crises, *Phys. Lett. A* **153**:105–109 (1991).
27. F. Takens, Detecting strange attractors in turbulence, in *Lecture Notes in Mathematics*, No. 898 (Springer-Verlag, 1981).
28. B. Townshend, Nonlinear prediction of speech signals, preprint.
29. H. Whitney, Differentiable manifolds, *Ann. Math.* **37**:645–680 (1936).
30. J. Yorke, Periods of periodic solutions and the Lipschitz constant, *Proc. Am. Math. Soc.* **22**:509–512 (1969).