

*Embedding Theorems for Non-uniformly Sampled  
Dynamical Systems*

Huke, J. P. and Broomhead, D. S.

2006

MIMS EPrint: **2006.369**

Manchester Institute for Mathematical Sciences  
School of Mathematics

The University of Manchester

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

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

ISSN 1749-9097

# Embedding Theorems for Non-uniformly Sampled Dynamical Systems

J.P. Huke & D.S. Broomhead

September 14, 2006

## Abstract

The embedding theorem of Takens [23], and its extensions, have provided the theoretical underpinning for a wide range of investigations of time series derived from nonlinear dynamical systems. The theorem applies when the dynamical system is sampled uniformly in time. There has, however, been increasing interest in situations where observations on the system are not uniform in time, and in particular where they consist of a series of inter-event (‘inter-spike’) intervals. Sauer [17] has provided an embedding theorem for the case where these intervals are generated by an integrate-and-fire mechanism. Here we prove several embedding theorems pertaining to non-uniform sampling. We consider two situations: in the first, the observations consist of the values of some function on the state space of the system, with the times between successive observations being given by another function—the *sampling interval function*; in the second, the sampling times are generated in the same way, but now the observations consist only of the intersample intervals. We prove embedding theorems both when the sampling interval function is allowed to be fairly general, and when it corresponds to integrate-and-fire sampling. We point out that non-uniform sampling might lead to better reconstructions than uniform sampling, for certain kinds of time series.

## 1 Introduction

Devising a dynamical system to model some time-evolving physical (or biological or chemical) phenomenon involves specifying a suitable set to be the state space of the system, and devising rules that govern the dynamics—that is, how the state changes with time. If the physics of the process is well understood it may be possible to write down the dynamical system more or less from first principles, without much need for observational data (except perhaps to fix parameters). In other circumstances we may be much less clear about the underlying physics: we might suspect that the system is governed by (for example) a set of differential equations, but not know what these are; we may even be uncertain what dynamical variables the equations involve. On the other hand we may have available a quantity of experimental data relating to how the system is observed to change through time; can we use this data to build a model dynamical system? Approaches to doing this, particularly relevant to situations where the behaviour is nonlinear, were described in the early 1980’s [10, 23], and have since become widespread.

Many such constructions (often called ‘reconstructions’) of dynamical systems from experimental data employ the *method of delays*, in which a time series of observations (often of just one quantity) is used to make a sequence of *delay vectors*, which are then treated as substitutes for the states of the system under study. Takens [23] showed that the approach is valid for a broad class of dynamical systems, and opened the gate to a flood of applications, which have included dimension estimation, prediction, noise reduction, and control [7, 9]. Extensions to Takens’s theorem have also appeared [18, 20, 21].

In the basic situation considered by Takens, the system is finite dimensional and its dynamics governed by a smooth flow (essentially, by a set of ordinary differential equations). This clearly represents a wide class of systems (including some, such as certain fluid flows, where the state space is infinite-dimensional, but the motion is confined to a finite dimensional subset). The data are assumed to be generated by observing the system at regular intervals: at each of the times  $0, T, 2T, \dots$  a measurement is made on the current state; the measurement produces the value of some fixed function on the state space—the function corresponding to the measurement procedure. Takens allows this function to be arbitrary except that it must be smooth. Many experimental protocols do indeed fit this picture, but not all: one fairly common case that does not is where the data consist, not of a series of values measured at uniformly spaced times, but of a series of *time intervals* between events. One motivating example is neurobiological: in neuronal interactions information is apparently encoded in the frequencies of pulse trains, while the amplitude of the pulses seems to contain little or none of the information. Thus if we observe the neuronal circuit, the relevant measurements are a series of inter-pulse intervals. Can we use this kind of information to build a dynamical system model, and if so, how is it related to the original system?

These questions were addressed by Sauer [16, 17], who retained the finite dimensional flow assumption for the underlying dynamical system, but replaced the regular evaluation of a measurement function by a mechanism for generating a sequence of inter-event intervals (specifically an integrate-and-fire mechanism, described in more detail below). Sauer showed that delay vectors of the inter-event (or ‘interspike’) intervals can be used, in a similar way to the uniform sampling case, to construct a model of the system. Such interspike interval vectors have since been applied in a number of areas [2, 1, 3, 14, 12, 22].

Racicot and Longtin [13] studied numerically the predictability of time series of interspike intervals generated by a range of neuron models driven by deterministic systems—integrate-and-fire is the simplest of these models. The ability to predict is closely related to the existence of a reconstruction: Racicot and Longtin thus raise the question of how generally it is possible to recover dynamical information from interspike intervals, and more specifically, what sorts of interval generating mechanisms allow reconstructions to be made.

In this paper we shall consider a broad class of interspike interval systems, and show that they can be reconstructed using the method of delays. The basic assumption is that the intervals (like the dynamics) are deterministic: the length of each interval is determined by the state of the system when the interval begins. The integrate-and-fire mechanism is an example of this class, but there are many others, and in fact we shall not be explicit about the mechanism itself, concentrating instead on the relationship (mapping) between states and intervals. Assuming this relation-

ship is sufficiently well behaved, we shall prove a theorem analogous to Takens's, for the interspike interval case.

Also described in this paper is an intermediate type of system: here we retain the measurement function of the original picture, and evaluate it at various points along the trajectory to provide the time series; but now the evaluation points are not uniformly separated in time: instead they are separated by intervals generated in the same way as the interspike intervals of the previous paragraph (for example, the integrate-and-fire process could be used to generate them). We refer to this method for producing a time series as *state dependent sampling*; we would expect to encounter experimental data of this form whenever the system controls when measurements are made (we might, for example, have a setup that automatically measures the weight of a bird whenever it returns to its nesting box, or we might measure the elastic properties of a geological formation when earthquakes send seismic waves through it). We show that delay vectors of such non-uniformly sampled time series can be used for reconstruction in the usual way. This case is simpler theoretically than the interspike interval case of the last paragraph, and so we treat it first. But state dependent sampling has a further interest because it suggests we might purposely sample a continuous time signal in a non-uniform way to produce good reconstructions of systems for which uniform sampling is not well adapted. If, for example, the signal consists of periods of rapid activity interspersed with periods of slow variation then it might be better to sample more quickly during the active phases and less during the slow phases than to sample at a single rate all the time (and thus either undersample or oversample parts of the signal).

## 2 Uniform Sampling and Takens's Theorem

Let us briefly review the experimental scenario that leads to the hypotheses for Takens's theorem. The object of study is assumed to be a continuous time dynamical system, whose behaviour is governed by a set of autonomous ordinary differential equations. Thus the state space is a differential manifold  $M$ , of finite dimension  $m$ , which for convenience we take to be compact. The dynamics is described by a smooth flow, that is, a differentiable function  $\phi: M \times \mathbb{R} \rightarrow M$ ; (the initial state  $x \in M$  evolves, after time  $t$ , to state  $\phi(x, t)$ ). Measurements are made every  $T$  time units. In the time between measurements, state  $x$  moves to  $\phi(x, T)$ ; say  $\phi_T$  is the function  $\phi(\cdot, T)$ , then, since  $\phi$  arises as the solution of a set of differential equations,  $\phi_T$  is a diffeomorphism of  $M$  (the 'time  $T$  map'), and  $\phi(\cdot, nT) = \phi_T^n$  for integer  $n$ . The sequence of states on which measurements are made is (taking  $x$  as the initial state)

$$x_0 = x, \quad x_1 = \phi_T(x), \quad x_2 = \phi_T^2(x), \quad \dots, \quad x_n = \phi_T^n(x), \quad \dots$$

The measurements detect the value of some property of the system (such as the temperature, the voltage at some point in an electrical system, the velocity at some point in a fluid system, etc.). We suppose there is a function  $y: M \rightarrow \mathbb{R}$  such that, if the system is in state  $x$ , the result of the measurement is  $y(x)$ . Thus the sequence of measured values is  $\{y_0 = y(x_0), y_1 = y(x_1), \dots, y_n = y(x_n), \dots\}$ . This sequence of values is the information produced by the experiment; generally neither  $x$  nor  $\phi_T$  will be known, so the data consist of this sequence (and of course the sample times at which the measurements were made).

To process the time series  $\{y_i\}$  we form a sequence of *delay vectors* by associating with each time a vector in  $\mathbb{R}^d$ , whose coordinates are the value of the measurement at that time and the values at the  $d - 1$  succeeding times:

$$\mathbf{y}_i = (y_i, y_{i+1}, \dots, y_{i+d-1})$$

The states are related to measurements by the function  $y$ ; they are related to the delay vectors by the function  $\Phi_{y, \phi_T}: M \rightarrow \mathbb{R}^d$ , defined by

$$\Phi_{y, \phi_T}(x) = (y(x), y \circ \phi_T(x), \dots, y \circ \phi_T^{d-1}(x)) \quad (1)$$

(the *delay map*). In particular,  $\mathbf{y}_i$  is equal to  $\Phi_{y, \phi_T}(x_i)$ .

The usefulness of the delay vectors stems from the fact that the relationship between them and the states of the system is often one-to-one: more precisely, as we shall see below, Takens showed that (if  $d$  is large enough) the map  $\Phi_{y, \phi_T}$  is typically an *embedding* of the state space into  $\mathbb{R}^d$ , that is, a diffeomorphism onto its range. This means that the delay vectors lie on a set  $\Phi_{y, \phi_T}M \subset \mathbb{R}^d$  which is essentially the same as the state space of the system apart from a nonlinear (but smooth) change of coordinates. ( $\Phi_{y, \phi_T}M$  is sometimes called the ‘reconstruction’ of  $M$ .) Furthermore, as the system moves along a trajectory in its state space, the delay vectors evolve through a corresponding trajectory in  $\Phi_{y, \phi_T}M$ : the dynamics of the system give rise, through the coordinate change, to corresponding dynamics in the reconstruction.

Generally speaking, the delay map  $\Phi_{y, \phi_T}$  is not known, (and cannot be determined if all we have available is the time series of measurements), however, the mere existence of the embedding is enough to ensure that the system and its reconstruction share many properties: the topology of the state space, for example; the numbers and types of periodic orbits; the dimensions of invariant measures, and so on. The fact that the reconstruction is a dynamical system in its own right means that we can use it to make predictions about the evolution of the system, and for a range of other applications [7, 9].

It was said above that the delay map is ‘typically’ an embedding. This is an important qualification: not all pairs  $(y, \phi_T)$  lead to an embedding (this is obvious if, for example,  $y$  is a constant function and  $M$  contains more than one point). In this paper the word typical is equivalent to ‘generic’ used in its precise topological sense: given a class (in our case, of functions) a property that the elements may or may not possess is *generic* if all the members in an open and dense subset of the class have the property. (This is the sense used in [23]; one of the main extensions [18] to Takens’s theorem involves using a somewhat different notion of typicality.)

To state Takens’s result, recall  $M$  is a compact manifold of dimension  $m$ , let  $C^2(M, \mathbb{R})$  be the set of real-valued  $C^2$  functions on  $M$  and  $\mathcal{X}^2(M)$  the set of  $C^2$  vector fields on  $M$ . Given a fixed  $T > 0$ , there is, for each pair in  $C^2(M, \mathbb{R}) \times \mathcal{X}^2(M)$ , a delay map defined as in equation (1). Then (Theorem 2 of [23], slightly reworded):

**Theorem 1** *If  $d > 2m$  it is a generic property of  $C^2(M, \mathbb{R}) \times \mathcal{X}^2(M)$  that the delay map is an embedding.*

Below we shall prove analogous theorems for non-uniformly sampled systems.

The proof Takens gave for this result provides some more detail about which pairs  $(y, X)$  will actually lead to embeddings. In particular, if a given vector field  $X$  has certain properties (to be described shortly), then for that  $X$  there is an open and

dense subset of measurement functions that, together with  $X$ , produce embeddings.  $X$  influences the delay map through the map  $\phi_T$ , and the properties we require of  $X$  are chosen so that  $\phi_T$  has corresponding desirable traits. So what properties of  $\phi_T$  tend to produce embeddings? Takens proved (without stating explicitly) the following:

**Theorem 2 (Takens, unstated)** *Let  $M$  be a compact manifold of dimension  $m$ , say  $d > 2m$ , and let  $\phi: M \rightarrow M$  be a diffeomorphism such that:*

1. *the periodic points of  $\phi$  with periods less than or equal to  $d$  are finite in number,*
2. *any such periodic point has distinct characteristic values.*

*Then the functions  $y \in C^2(M, \mathbb{R})$ , for which the map  $\Phi_{y,\phi}: M \rightarrow \mathbb{R}^d$  defined by*

$$\Phi_{y,\phi} = (y, y \circ \phi, \dots, y \circ \phi^{d-1})$$

*is an embedding, form an open and dense subset of  $C^2(M, \mathbb{R})$ .*

What properties of  $X$  will ensure that  $\phi_T$  has the properties described in the above theorem? Naturally these are to do with its periodic orbits: specifically, we could insist there be no periodic orbits having period  $kT/n$  with  $k$  an integer in the range  $1 \leq k \leq d$  ( $n$  is any positive integer), and if  $X$  has any critical points (where  $X = 0$ ) these must be finite in number, and the characteristic values of such a point must all be distinct. In that case  $\phi_T$  actually has no periodic points with period less than or equal to  $d$  at all except for fixed points, and these fixed points are finite in number and have distinct characteristic values; so certainly the conditions of Theorem 2 are satisfied. Proving Theorem 1 then amounts to showing the properties we required of  $X$  are in fact generic in  $\mathcal{X}^2(M)$ , which is a straightforward extension of the Kupka-Smale theorem [8, 19, 11]. (In fact, the requirements placed on  $X$  are more stringent than strictly necessary [9].)

### 3 State Dependent Sampling

As we have mentioned, there are occasions when the measurements are not taken at uniform sample times. Can the observations in this case still be used to reconstruct the system? It is a reasonable hope that, so long as the times at which the samples are taken are determined completely by the states the system passes through, then the measurements might be used to reconstruct the states. We shall therefore consider the situation where there is a positive valued function  $\tau: M \rightarrow \mathbb{R}^+$  that tells us, for each state at which we sample, what the next sampling interval will be; thus if a sample is taken when the system is in state  $x$  the next sample will be taken  $\tau(x)$  time units later. For some systems, such as small networks of neurons, where we do not control the sampling, we consider  $\tau$  to be an intrinsic property of the system. In other cases, where we have access to a continuous time history and decide how to sample it ourselves,  $\tau$  represents the strategy we use for choosing the sampling times. (Of course, such a strategy must work in the absence of explicit knowledge of the system state, but may nevertheless be consistent with the existence of a function  $\tau$ ; for instance, in the uniform sampling case  $\tau$  is just a constant function; the integrate-and-fire procedure (see below) provides another example.) Any sampling procedure that can be described by a  $\tau$  function could be termed *state dependent sampling*.

The integrate-and-fire mechanism is one important procedure for generating non-uniform sampling: here, a positive function,  $v: M \rightarrow \mathbb{R}^+$ , is defined on the state space, and as the system moves along a trajectory this function is integrated with respect to time. When the integral reaches a preset threshold,  $\Theta$ , a ‘spike’ is generated (that is, a sample is taken) and the integration is restarted. A series of sampling times  $\{T_i\}$  is generated recursively (from an initial state  $x_0$ ) by

$$\int_{T_i}^{T_{i+1}} v \circ \phi(x_0, t) dt = \Theta$$

We may also think of this process as defining a function  $\tau$  on  $M$  by saying  $\tau(x)$  is the positive number  $T$  such that  $\int_0^T v \circ \phi(x, t) dt = \Theta$ .

Another mechanism [6] for producing non-uniform sampling again starts with a general picture of an ‘activity’ level rising until a threshold is reached (triggering a sample), but now the rate at which the activity rises is constant, and it is the threshold that varies with the state of the system (and so with time). So now there is a positive threshold function  $\Theta: M \rightarrow \mathbb{R}^+$ . Suppose a sample is taken at time  $t_0$ , while the system has state  $x_0$ : the threshold thereafter changes with time according to  $\Theta(\phi(x_0, t - t_0))$ . The activity starts from zero and increases according to  $\lambda(t - t_0)$ , where  $\lambda$  is a positive constant: the next sample is taken when these become equal. Thus the sampling interval  $\tau(x_0)$  is the solution  $T$  of

$$\lambda T = \Theta(\phi(x_0, T))$$

(or the smallest solution if there is more than one).

A further way in which intervals might be generated from the dynamical system is threshold crossing: again a function  $v: M \rightarrow \mathbb{R}$  (not now necessarily positive) is considered, but now a spike is generated whenever the value of this function crosses a fixed threshold  $v_c$ , that is, at times  $t$  that satisfy  $v \circ \phi(x_0, t) = v_c$ . (One often distinguishes between upward crossings, where the time derivative of  $v$  is positive at the crossing, and downward crossings where it is negative; frequently only upward crossings are considered to generate spikes.) This mechanism is somewhat different in character to the previous two: sampling intervals are not automatically defined for all states, since samples are only ever taken when the system state satisfies  $v(x) = v_c$  (typically, for smooth  $v$ , this set  $S$  will be a codimension 1 submanifold of  $M$ ). The obvious approach of defining  $\tau(x)$  to be the time until the next (upward) crossing of  $v_c$ , starting from  $x$ , produces a  $\tau$  which is discontinuous at points in  $S$ , and so is not likely to be useful for embedding. It is more natural to consider this  $\tau$  to be a function on  $S$  only, and consider embeddings of  $S$  rather than  $M$ —this is the approach taken, for example, in [4]; we shall not consider this sort of situation further here: we assume for the remainder of this paper that  $\tau$  is defined on the whole of  $M$ .

Once the sampling intervals have been specified, the states at which samples are taken are also determined. If we sample at  $x$  the next sample will be taken at the state that  $x$  evolves to after a time  $\tau(x)$ , namely  $\phi(x, \tau(x))$ . Let us define the *sampling state map*  $g: M \rightarrow M$  by  $g(x) = \phi(x, \tau(x))$ ; then sampling at  $x$  implies the next sample will be taken at  $g(x)$ , the sample after that at  $g(g(x)) = g^2(x)$  and so on. The sequence of sampled states is thus  $\{x, g(x), \dots, g^n(x), \dots\}$  and the sampling intervals are  $\{\tau(x), \tau(g(x)), \dots, \tau(g^n(x)), \dots\}$ . The sequence of measurements is  $\{y(x), y(g(x)), \dots, y(g^n(x)), \dots\}$  which has the form of a time series of observations made on a dynamical system  $(g, M)$ . If the observations are collected into delay vectors as before then to each state  $x$  of the system there corresponds a delay vector

$\Phi_{y,g}(x)$ , given by

$$\Phi_{y,g}(x) = (y(x), y \circ g(x), \dots, y \circ g^{d-1}(x)) \quad (2)$$

### 3.1 When is the sampling state map a diffeomorphism?

The similarity between delay maps (2) and (1) is clear, with  $g$  in (2) playing the role of  $\phi_T$  in (1). This suggests there should be an embedding theorem analogous to that for the uniform sampling case, if  $g$  is a diffeomorphism (the main requirement on  $\phi_T$  in (1)). A major difference between state dependent and uniform sampling is that  $g$  is not guaranteed to be a diffeomorphism just because it arises from a smooth flow. We can see this in a simple example: take  $M$  to be the circle  $S^1$ , with  $\theta \in S^1$  evolving according to  $\dot{\theta} = \omega$ , where  $\omega$  is a positive constant; then  $\phi(\theta, t) = \theta + \omega t \bmod 2\pi$ . We will say the sampling interval function  $\tau$  is given by  $\tau(\theta) = a + b \cos \theta$ , with  $a$  and  $b$  non-negative constants, and  $a > b$  to ensure  $\tau$  is positive. In that case

$$g(\theta) = \theta + \omega a + \omega b \cos \theta \quad \bmod 2\pi$$

It is clear that  $g$  is a diffeomorphism if  $b = 0$ , and hence also for all sufficiently small  $b$  (for given  $a$  and  $\omega$ ). On the other hand, if  $b = \pi/(2\omega)$  then  $g(\pi/2) = a\omega + \pi/2 = g(\pi)$ , so that  $g$  is not injective. For this system it is simple enough to show that  $g$  is a diffeomorphism if and only if  $b < 1/\omega$ ; but generally we shall need to identify conditions that ensure that  $g$  is a diffeomorphism.

Recalling that  $g(x) = \phi(x, \tau(x))$  we see that  $g$  is differentiable whenever  $\tau$  and  $\phi$  (or equivalently  $X$ ) are, so we shall insist that  $\tau$  and  $X$  (and so  $g$ ) are  $C^2$  from now on.

A clear picture of how  $g$  can fail to be injective emerges once we note that each trajectory of  $X$  is invariant under  $g$ , so we need only consider whether  $g$  is injective on individual trajectories. Consider one such trajectory  $L$ , taken for the sake of convenience to be non-periodic, so that the points of  $L$  have a well-defined time ordering. It is apparent that the only way  $g$  can fail to be injective on  $L$  is for it to ‘fold’  $L$  onto itself, which means that the time order of (at least) two points is reversed by  $g$ . ( $g$  cannot reverse the order of *all* the points because of the boundedness of  $\tau$  and  $X$ .) Say the points are  $x_1$  and  $x_2$ , with  $x_2$  in the forward orbit of  $x_1$ , and say it takes time  $T$  to flow from  $x_1$  to  $x_2$ . For  $g$  to reverse the order of these points  $\tau(x_1)$  must be large enough for  $x_1$  to flow past  $x_2$  and  $g(x_2)$ ; that is,  $\tau(x_1) > \tau(x_2) + T$  so  $(\tau(x_2) - \tau(x_1))/T < -1$ . The mean value theorem now indicates the presence of a point on the trajectory at which the *time* derivative of  $\tau$  is less than  $-1$ ; that is, there is some  $x \in M$  such that  $d(\tau \circ \phi(x, t))/dt|_{t=0} = 0$ . To express this condition in terms of  $\tau$  and  $X$ , we choose a coordinate system  $h: U \subset M \rightarrow \mathbb{R}^m$  and say  $\hat{\tau}$  and  $\hat{X}$  are the local representations of  $\tau$  and  $X$ : then the condition is  $D\hat{\tau}(u)\hat{X}(u) < -1$ , where  $u = h(x)$ . However, from now on we shall suppress explicit reference to coordinates, and abbreviate this to  $D\tau(x)X(x) < -1$ .

This condition turns out to be the crucial one in deciding whether  $g$  is a diffeomorphism; indeed, we have the following:

**Lemma 1** *Let  $X$  be a  $C^r$  ( $r \geq 1$ ) vector field on a compact connected manifold  $M$ , and  $\tau: M \rightarrow \mathbb{R}$  a  $C^r$  function. The mapping  $g: M \rightarrow M$  defined by  $g(x) = \phi(x, \tau(x))$ , (where  $\phi: M \times \mathbb{R} \rightarrow M$  is the flow arising from  $X$ ), is a  $C^r$  diffeomorphism if and only if  $D\tau(x)X(x) > -1$  for all  $x \in M$ .*



This is proved in section 7.1.

As a simple illustration we may take the circle example given above: there  $X(\theta) = \omega$  and  $\tau(\theta) = a + b \cos \theta$  so  $D\tau(\theta)X(\theta) = -\omega b \sin \theta$ . Thus the condition in the lemma becomes  $\omega b \sin \theta < 1$ , which is satisfied for all  $\theta$  if and only if  $\omega b < 1$ .

For a more important example, consider again the integrate-and-fire mechanism. Here  $\tau$  is defined through

$$\int_0^{\tau(x)} v \circ \phi(x, t) dt = \Theta \quad (3)$$

Differentiating this gives  $0 = v(g(x))D\tau(x) + \int_0^{\tau(x)} Dv(\phi(x, t))J_t(x)dt$  (where  $J_t$  is the Jacobian of the time  $t$  map). It follows that  $0 = v(g(x))D\tau(x)X(x) + v(g(x)) - v(x)$ , and hence the condition  $D\tau(x)X(x) > -1$  is automatically satisfied if  $v(x) > 0$  for all  $x$ . It is thus a feature of the integrate-and-fire procedure that the sampling state maps that it gives rise to are always diffeomorphisms.

For an example where a diffeomorphism is not guaranteed we return to the variable threshold mechanism. Recall that given a positive valued threshold function  $\Theta(x)$ ,  $\tau(x)$  satisfies  $\lambda\tau(x) = \Theta(\phi(x, \tau(x)))$ . Differentiating this with respect to  $x$  we find

$$D\tau(x)X(x) = \frac{D\Theta(g(x))X(g(x))}{\lambda - D\Theta(g(x))X(g(x))}$$

which implies (recalling  $\lambda > 0$ ) that  $g$  will be a diffeomorphism if  $\lambda > D\Theta(x)X(x)$  for all  $x \in M$ . For smaller values of  $\lambda$  a diffeomorphism is not obtained; as mentioned in [13], this condition corresponds to there being one or more points on  $M$  at which a sample can never be taken.

### 3.2 Embedding theorems for state dependent sampling

Now that we can decide whether  $g$  is a diffeomorphism, we immediately have access to an embedding theorem relevant to state dependent sampling: Theorem 2 applies, and shows that the state dependent sampling delay map,  $\Phi_{y,g}$  will generically be an embedding if  $g$  satisfies the relevant conditions. Of course, as in the uniform sampling case, checking the conditions for individual diffeomorphisms may not be particularly easy, so we would prefer a theorem that concerns a broad class of  $g$ 's: that is, a theorem corresponding to Theorem 1. Such a theorem can be proved, but to state it precisely we need to specify the sets of functions that we are considering.

Embedding theorems of the kind we are looking for always assert that the property of giving rise to an embedding is generic for a certain set of functions. Here our primary focus will be on the sampling intervals (which will sometimes be the only observations we make) and so we shall consider sets of  $\tau$  functions. (An alternative might be to consider sets of  $g$  functions, but this does not suit our purposes so well: see the last paragraph of this section.)

The condition  $D\tau(x)X(x) > -1$  tells us whether  $g$  is a diffeomorphism and hence whether we are within the purview of Theorem 2. To make a statement about a broad class of state dependent sampled systems, rather than a single one, we need to produce an analogue of Theorem 1, and this involves showing that the conditions of Theorem 2 are generically satisfied by diffeomorphisms arising from some suitable set of vector fields and sampling interval functions. So we first discard the cases that fail the condition for  $g$  to be a diffeomorphism: let  $\mathcal{T} \subset C^2(M, \mathbb{R}^+) \times \mathcal{X}^2(M)$  be the subset  $\{(\tau, X) : D\tau(x)X(x) > -1, \forall x \in M\}$ ; then it is clear that  $\mathcal{T}$  is an open

subset of  $C^2(M, \mathbb{R}^+) \times \mathcal{X}^2(M)$ . Only pairs within  $\mathcal{T}$  can be expected to give rise to embeddings.

We can now state the embedding theorem for state dependent sampled systems:

**Theorem 3** *If  $d > 2m$  then the triples  $(y, \tau, X)$ , for which the map  $\Phi_{y,g}$  defined in equation (2) is an embedding, form an open and dense subset of  $C^2(M, \mathbb{R}) \times \mathcal{T}$ .*

This result follows from the fact that for generic  $(\tau, X)$  in  $\mathcal{T}$  the diffeomorphism  $g$  satisfies the conditions on periodic orbits required by Theorem 2. Details of the proof are given in section 7.2.

Theorem 3 refers to sets of  $\tau$  functions; to arrive at it we have imagined the diffeomorphism in Theorem 2 to be derived from a sampling interval map  $\tau$ . However we might alternatively postulate a mechanism that specified—for each state  $x$  at which a sample is taken—what the next sampled *state* will be, rather than the time interval; that is,  $g$  is given rather than  $\tau$ . At one level these are actually equivalent, since the sampling is assumed not to affect the continuous time dynamics, and  $g$  must therefore respect the underlying vector field  $X$  in the sense that  $g(x)$  must be a point on the (forward)  $X$ -trajectory of  $x$ . This further means that for every  $x$  there must be a positive time  $\tau(x)$  such that  $g(x) = \phi(x, \tau(x))$ . Note however that if  $x$  lies on a periodic orbit, this time is not unique. Thus specifying  $g$  gives a less complete description of the experimental situation than does specifying  $\tau$ , since it may be ambiguous about when the samples are taken (and hence about the interspike intervals). This is why we chose to state Theorem 3 in terms of sets of  $\tau$  functions.

### 3.3 Sampling continuous time records

In some experimental situations, state dependent sampling is forced upon us because the system under observation produces data at non-uniform time intervals which are not under our control. There are other occasions when we may choose to adopt this kind of sampling because it is advantageous to do so. Consider for example a time series of the kind shown in figure 1.

This figure represents the variation of one of the components of a three-dimensional system investigated by Rössler [15]; (the system is given in equation (1) in [15], with parameters given in Fig. 3 of that reference; we use the  $z$  component). The time series consists of sections showing relatively slow, low amplitude variation near a baseline value, with occasional spikes in which the signal makes brief excursions to large values. This could be viewed as a candidate for interspike interval sampling, but it is clear that the spikes have different heights and that these heights could contain useful information about the state of the system. It seems clear that if this time series is to be sampled uniformly, the sampling interval will need to be quite short to capture the spikes, especially if fairly accurate information about their heights and shapes is wanted. On the other hand a short interval will tend to oversample the stretches of baseline value, producing a lot of data without much useful information. It would seem better therefore to vary the sampling interval: to choose small intervals when sampling the spikes and larger ones between the spikes. But we also want to preserve if we can the embedding relationship between the sampled series and the original system; the preceding discussion suggests we may be able to do this if we can arrange for the sampling intervals to be determined by the states of the system. Since the states generate the time series (through the function  $y$ ), using the time

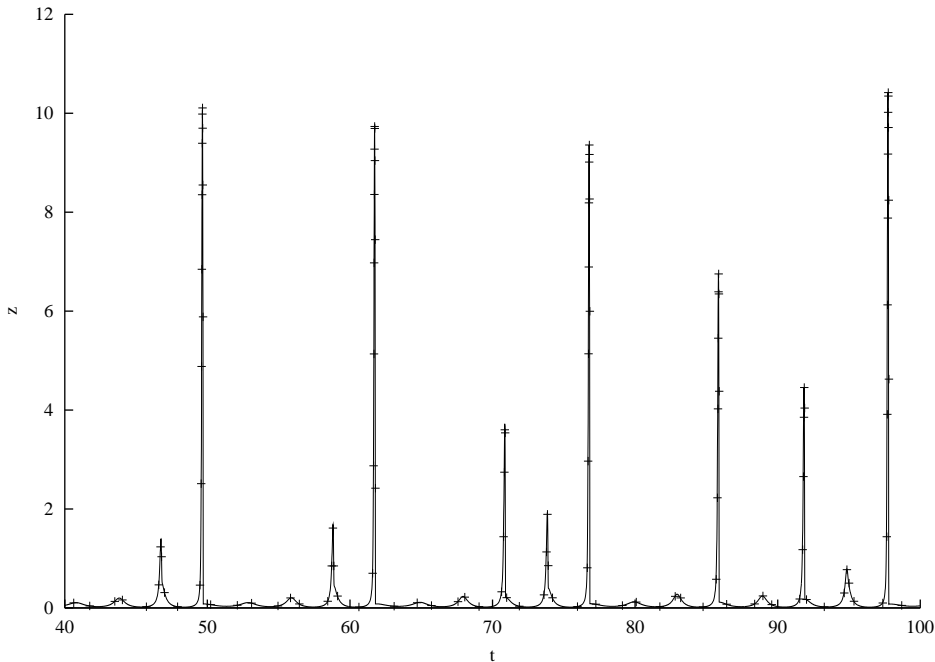


Figure 1: Time series from a three dimensional system specified in Ref. [15]. The points marked with symbols indicate the samples generated by integrate-and-fire sampling; see text for details.

series values to determine the sampling should mean that the intervals are indeed a function of the state (though not necessarily a function we could easily write down).

There are various approaches that could be taken: the sampling interval could be made to depend inversely on the amplitude of the signal, or on the magnitude of the slope of the signal so that the sampling is rapid when the signal varies rapidly. (Clearly, however, we cannot be too naive about this: the slope, for example, is low between the spikes and high at most points within a spike, but if we happened to sample close to the maximum of the peak we would find a small slope that could cause us to make the next interval too large.) An approach that works quite well for the time series in figure 1 is integrate-and-fire: during the sections between spikes the input to the integrator is small and firing (sampling) is infrequent; increase in the signal amplitude increases the frequency of firing so that the spikes are well sampled. Figure 1 shows the samples generated by this process; in this example the function to be integrated is chosen to be  $0.01 + z(t)$ , and the threshold is 0.1.

We should highlight a subtlety related to the approach to sampling just described. Although each sampling interval depends on the system state, we are not exactly in the situation covered by Theorem 3 because in that theorem the sampling interval function  $\tau$  and the measurement function  $y$  are regarded as independent: both are needed to specify one of the triples with which the theorem is concerned. But in the sampling strategy we used above  $\tau$  is clearly not independent of  $y$ , and so Theorem 3 does not give us any direct information about how likely we are to produce an embedding. In the next section we shall consider whether or not we get embeddings when we consider only  $\tau$  values, that is, making no use of  $y$  values at all; it will turn out that, with some restrictions, embeddings are still generic.

## 4 Interspike intervals

In the previous section we described how the sampling of a continuous time dynamical system could be controlled by a sampling interval function  $\tau$ . The observations consisted of values of a measurement function  $y: M \rightarrow \mathbb{R}$ . But since the sampling is non-uniform, and the intervals depend on the states of the system, it is conceivable that we may derive information about the states from the intervals themselves even in the absence of the measurement function  $y$ ; indeed we might try to use the intervals to embed the system. Another way of looking at this is to consider  $\tau$  to be the measurement function: when each sample is taken, the value of  $y$  at the corresponding state is just the time taken to the next sample. As the system moves along a trajectory it generates a sequence of samples (or spikes) and a corresponding sequence of interspike intervals: as before, spikes occur at  $x, g(x), g^2(x), \dots$ , and the intervals are  $\tau(x), \tau(g(x)), \tau(g^2(x)), \dots$ . If we collect these intervals into delay vectors then to each state there corresponds a vector

$$\Phi_{\tau,g}(x) = (\tau(x), \tau \circ g(x), \tau \circ g^2(x), \dots, \tau \circ g^{d-1}(x)) \quad (4)$$

Note this is identical to (2) if  $y$  is replaced by  $\tau$ .

Is the delay map (4) (generically) an embedding? Note that Theorem 3 gives no assurance that it is, since that theorem says nothing specifically about the case where  $y$  is constrained to be the same as  $\tau$ ; a new theorem is needed to cover this. Furthermore Theorem 2, on which the proof of Theorem 3 was based, is not applicable here, since considering a range of different  $y$  functions would entail considering a corresponding range of  $\tau$ 's, and hence  $g$ 's. Nevertheless, the techniques used to prove Theorems 2 and 3 can be adapted to produce an embedding theorem of the kind we require.

Again we must be specific about the class of functions  $\tau$ , and vector fields, that we consider. We have already introduced the set  $\mathcal{T} \subset C^2(M, \mathbb{R}^+) \times \mathcal{X}^2(M)$  as the subset of pairs  $(\tau, X)$  such that  $D\tau(x)X(x) > -1$  for all  $x$ , and we continue to restrict our attention to this, since we still require  $g$  to be a diffeomorphism. However, since there is now less flexibility about the functions that go to make up the delay map we shall need to consider subsets of  $\mathcal{T}$  to find a set for which the embedding property is generic. As usual, (see for example the conditions to Theorem 2), it is the periodic points that cause the most serious barriers to embedding, and we shall need to discard some  $(\tau, X)$  pairs where  $\tau$  takes particularly troublesome values at the periodic points. Especially awkward are those periodic (though not equilibrium) points of  $X$  that give rise to fixed points of  $g$ : if there is such a point, the periodicity of  $\tau$  on the periodic orbit of  $X$  implies there must be a second such point, and that the two have the same image under  $\Phi_{\tau,g}$ . Furthermore this situation is likely to be shared by all sufficiently close  $(\tau, X)$  pairs, ruling out the possibility of generic embeddings; this is the situation we need to exclude.

The embedding theorem thus produced is:

**Theorem 4** *Let  $\mathcal{T}' \subset C^2(M, \mathbb{R}^+) \times \mathcal{X}^2(M)$  be the set of pairs  $(\tau, X)$  with the following properties:*

1.  $D\tau(x)X(x) > -1$  for every  $x \in M$ .
2. *There are no points  $x$  such that:  $x$  lies on a periodic orbit of  $X$ , with period  $T$ , and  $\tau(x) = nT$  for some integer  $n$ .*

3. There are no points  $x$  such that:  $x$  is an equilibrium point of  $X$ , with eigenvalue  $i\omega$ , and  $\tau(x) = 2\pi n/\omega$  for some integer  $n$ .

then  $\mathcal{T}'$  is an open subset of  $C^2(M, \mathbb{R}^+) \times \mathcal{X}^2$ .

For  $d > 2m$  the pairs  $(\tau, X)$  for which  $\Phi_{\tau, X}$  is an embedding of  $M$  form an open and dense subset of  $\mathcal{T}'$ .

(The third condition is needed to ensure that  $\mathcal{T}'$  is an open subset.) This is proved in section 7.3.

Consider again the circle example where  $X(\theta) = \omega$  and  $\tau(\theta) = a + b \cos \theta$ , ( $a > b > 0$ ). We have seen that for  $g$  to be a diffeomorphism we must have  $\omega b < 1$ . For this system every point is periodic with period  $T = 2\pi/\omega$ , so the second condition in the above theorem is not satisfied if there is some  $\theta$  such that  $a + b \cos \theta = nT$ ; that is, if  $a$  is in the range  $nT - b \leq a \leq nT + b$  for some positive integer  $n$ . Thus for a given  $b$  there are ‘forbidden intervals’ for  $a$ , centred on the values  $nT$  and of width  $2b$ . (Since  $2b < T/\pi$  these intervals do not overlap.) For a given  $a$  on the other hand, the condition fails if  $b$  is larger than the distance of  $a$  from the nearest multiple of  $T$ .

Returning to integrate-and-fire systems, recall that these automatically satisfy the first condition above. Suppose an integrate-and-fire system fails to satisfy the second condition, so that there is a point  $x$  on a periodic orbit  $C$  such that  $\tau(x) = nT$  ( $T$  is the period of  $C$ ). Then the integral defining  $\tau(x)$ ,  $\Theta = \int_0^{\tau(x)} v \circ \phi(x, t) dt$ , is taken  $n$  times around  $C$ , and it follows that every  $x \in C$  gives the same value of  $\tau$ , and every point is a fixed point of  $g$ . Hence none of these fixed points is hyperbolic. The implication of this is that although it is possible for integrate-and-fire systems to fail the second condition, generically they do not fail it, (and a similar remark applies to the third condition). So for integrate-and-fire systems none of the conditions proves a serious barrier to embedding. We shall investigate these systems in more detail later.

It is quite possible for the variable threshold mechanism described above (where  $\tau$  is defined through  $\lambda\tau(x) = \Theta \circ \phi(x, \tau(x))$ ) to have hyperbolic fixed points of  $g$ , and so fail condition 2 in a robust way. (It is easy to check that if  $g(x) = x$  then  $x$  will be hyperbolic so long as  $D\Theta(x)X(x) \neq 0$ , that is, so long as  $\Theta$  does not pass through a stationary point at  $x$  as we pass along the periodic orbit of the flow. Fixed points of  $g$  are studied in [6] for the case of uniform rotation in the circle, where  $\Theta$  is sinusoidal.)

A given vector field  $X$  may well have a great number of periodic orbits, and each gives rise to a set of values forbidden by condition 2.  $\tau$  must of course be bounded above on  $M$ , and orbits with period greater than this bound will be of no concern: it is the low period orbits that are most likely to cause problems with embedding. Commonly there will be some minimum period orbit, and if  $\tau(x)$  is everywhere less than this minimum period then condition 2 will be satisfied; that is, it is sufficient for condition 2 to hold that every periodic orbit is sampled at least once before a complete circuit of it is made.

## 5 Integrate-and-Fire Systems

Given a continuous time dynamical system on  $M$ , we convert it to an interspike interval system by supplying the sampling interval function  $\tau$ . We have mentioned several times that one convenient way of constructing  $\tau$  is by the integrate-and-fire procedure, and we have seen a number of examples of this. However we have not yet

provided embedding theorems specifically concerned with integrate-and-fire systems. Note that Theorems 3 and 4 refer to general interspike interval systems, and we cannot infer from them that integrate-and-fire systems will (generically) provide embeddings because the latter do not form an open subset of the former. Thus we shall need separate theorems for the integrate-and-fire class.

The special nature of the integrate-and-fire systems is seen most clearly from their behaviour on periodic orbits of the vector field. The sampling state map  $g$  is a diffeomorphism of this orbit (recall that integrate-and-fire systems automatically satisfy the condition in Lemma 1). Suppose first that  $g$  has periodic points on the orbit: we have already remarked that if there is a fixed point it is not hyperbolic, and the same applies to periodic points; however, arbitrarily close to  $\tau$  we can find another sampling interval function  $\tau'$  whose periodic points *are* hyperbolic—and which cannot therefore come from an integrate-and-fire system. (The construction of  $\tau'$  is described in the proof of Lemma 4 in section 7.2.) If none of the orbits of  $g$  is periodic then each has an accumulation point and it is clear that a periodic point can be created by an arbitrarily small change to the integrated function  $v$ ; then the remarks of the previous two sentences apply. Thus, at least for vector fields with periodic orbits, the collection of  $\tau$ 's that are *not* derivable from integrate-and-fire mechanisms contains an open and dense subset of the set  $C^2(M, \mathbb{R}^+)$  of general  $\tau$ 's.

In fact, there is a close connection between integrate-and-fire systems and uniformly sampled ones. For suppose we have a vector field  $X$  and produce an integrate-and-fire system using the positive valued function  $v$ . Then the resulting diffeomorphism  $g$  is given by  $g(x) = \int_0^{\tau(x)} X(\phi(x, t)) dt$ . For fixed  $x$ ,  $\hat{t} = \int_0^t v(\phi(x, t')) dt'$  is a differentiable, strictly monotone increasing function of  $t$ . Changing the variable in the integral for  $g(x)$  from  $t$  to  $\hat{t}$  we find

$$g(x) = \int_0^\Theta \hat{X}(\hat{\phi}(x, \hat{t})) d\hat{t} = \hat{\phi}(x, \Theta)$$

where  $\hat{X}(y) = \frac{1}{v(y)}X(y)$  for all  $y \in M$ , and  $\hat{\phi}$  is the corresponding flow. Thus  $g$  results from sampling the flow  $\hat{\phi}$  *uniformly*, with sampling interval  $\Theta$ .

We can prove a ‘state dependent sampling’ embedding theorem, analogous to Theorem 3, for integrate-and-fire systems; as usual, we must first identify the class of systems within which we hope the embedding property will be generic. It is natural to characterize integrate-and-fire systems by the function  $v$  rather than  $\tau$ , since the latter is defined implicitly from the former; thus we specify an integrate-and-fire system by a pair  $(v, X) \in C^2(M, \mathbb{R}^+) \times \mathcal{X}^2(M)$ . For the state dependent sampling case, where we also have a measurement function, we consider triples  $(y, v, X) \in C^2(M, \mathbb{R}) \times C^2(M, \mathbb{R}^+) \times \mathcal{X}^2(M)$ . Then

**Theorem 5** *Let  $\Theta > 0$  be fixed. For each  $(y, v, X) \in C^2(M, \mathbb{R}) \times C^2(M, \mathbb{R}^+) \times \mathcal{X}^2(M)$  define  $\tau$  by equation (3),  $g$  by  $g(x) = \phi(x, \tau(x))$ , and the delay map by equation (2).*

*For  $d > 2m$  the triples  $(y, v, X)$  for which the delay map is an embedding form an open and dense subset of  $C^2(M, \mathbb{R}) \times C^2(M, \mathbb{R}^+) \times \mathcal{X}^2(M)$ .*

This is proved in section 7.4. (In fact, this can be proved from Theorem 1 using the connection between integrate-and-fire and uniform sampling mentioned above. However the proof in section 7.4 parallels that of Theorem 3, mainly so that we produce Lemma 5, which is also needed for the following theorem.)

Finally we consider an analogue of Theorem 4 (in which  $\tau$  plays the roles both of sampling interval function and measurement function) but specialized to the case of integrate-and-fire systems. This is the situation treated by Sauer [16, 17], who has already supplied an embedding theorem for this case. There are however some differences between the theorem proved by Sauer and the theorems described here: the main two being that the manifold  $M$  is replaced with a compact set (with a corresponding adjustment of the notion of embedding), and a somewhat different notion ('prevalence') is used in place of genericity. For uniformity we state a theorem that is directly analogous to Theorem 4 in the same way that Theorem 5 is analogous to Theorem 3:

**Theorem 6** *Let  $\Theta > 0$  be fixed. For each  $(v, X) \in C^2(M, \mathbb{R}^+) \times \mathcal{X}^2(M)$  define  $\tau$  by equation (3),  $g$  by  $g(x) = \phi(x, \tau(x))$ , and the delay map by equation (4).*

*For  $d > 2m$  the pairs  $(v, X)$  for which the delay map is an embedding form an open and dense subset of  $C^2(M, \mathbb{R}^+) \times \mathcal{X}^2(M)$ .*

As with Theorem 4 there seems no very simple way to derive this theorem as a straightforward consequence of the theorems we already have to hand, but a proof can be constructed along similar lines to those used for Theorem 4; this is described in section 7.5.

## 6 Concluding Remarks

We have attempted to give embedding theorems for non-uniformly sampled systems that parallel as closely as possible Takens's theorem for uniformly sampled systems—hopefully this will illuminate the similarities and differences between the two, and help to make clear the relationship these results have to the treatment of experimental observations. But given that our theorems are slightly more complex in their statement than Takens's theorem, this relationship might be worth commenting on.

Embedding theorems always take a well defined class of functions (for example,  $C^2$  vector fields together with  $C^2$  measurement functions) and assert that, for a 'large subset' of this class, an embedding is obtained (by the method of delays or some other specified procedure). Since the theorems do not cover the *entirety* of the class, they give no assurance that any particular member of the class will produce an embedding—this would have to be verified for the particular case. Thus the theorems do not prove that applying the method of delays (say) to a particular experimental data set will result in an embedding. What they do is provide some assurance that the method of delays is a reasonable strategy to adopt and is likely (in a sense that is not mathematically precise) to achieve some success in analysing or modelling the data. However, to conclude even this we must be confident that the system does fall within the original class of functions, otherwise assertions about large subsets of the class are not of much value. The original class should therefore be chosen as far as possible so that membership of it is either easy to check or, more usually, follows from reasonable assumptions about the system under study. In the case of Takens's theorem the main assumptions we need are fairly straightforward: they are that the system is finite dimensional and that the dynamics and measurement are smooth. But in addition to the system lying within the original class, the theorem also requires that the number of delays  $d$  be sufficiently large, and we may not be sure of this *a priori* because we may not know the dimension  $m$  of the underlying system. Thus

even here there may be some investigating to do to establish the applicability of the theorem.

In any case we still have the problem of deciding whether or not we have an embedding in any particular case, and for this we must apply suitable tests to the data. This is an important part of analysing the data using the geometrical approach [7]: we need it to check that  $d$  is large enough, and ultimately it is our reassurance that, despite the existence of a small subset of problematic systems, our analysis of the data is not running into trouble.

In addition to the assumptions made for Takens's theorem, Theorem 3 for state dependent sampling further requires that our system belongs to the class  $\mathcal{T}$ , that is, that it satisfies  $D\tau(x)X(x) > -1$  for all  $x \in M$ . This condition is straightforward to check if we have a continuous time record such as described in section 3.3, or if we can monitor the value of  $\tau$  continuously, since the condition states that the *time derivative* of  $\tau$  shall be greater than  $-1$ . However, if all we have available is the sampled data then we must initially take the condition as a further assumption, and then check for an embedding as in the previous paragraph. Similar remarks apply, of course, to Theorem 4

Another aspect of the relationship between embedding theorems and experiments concerns the regions of state space that the experiment actually investigates. The embedding theorems set up a correspondence (the delay map  $\Phi$ ) between *every* point in the state space  $M$  and a corresponding point in (a subset of) delay space  $\mathbb{R}^d$ . Usually however experiments do not attempt to cover all of state space: often what is of interest are the attractors of the system. To investigate such a set, an initial condition  $x_0$  is chosen within the basin of attraction of the set, and the trajectory starting from  $x_0$  is observed; we assume that the  $\omega$ -limit set of this trajectory is the attractor of interest. Samples from the trajectory, uniform or non-uniform in time, then constitute the experimental data. As we have seen, these samples are governed by a discrete time map: in the uniform sampling case this is  $\phi_T$ , the time- $T$  map ( $T$  is the sampling interval); for non-uniform sampling it is the map  $g$  described in section 3. Thus what we actually observe is a trajectory of the discrete time map, and so it is the  $\omega$ -limit set of this map, rather than the underlying flow, about which the experiment gives us information. (This was noted by Takens [23] for the uniform sampling case, and in [22] and [5] for the non-uniform case.)

It is obvious that the  $\omega$ -limit set of  $x_0$  under the discrete time map is a subset of the continuous time  $\omega$ -limit set. It is quite possible for the former to be a rather small subset of the latter: indeed, in the simple example of uniform rotation on the circle, with  $\tau(\theta) = a + b \cos \theta$ , described in section 3.1, it is easy to find values of the parameters for which  $g$  has attracting periodic points, which form the  $\omega$ -limit sets for most of the points on the circle—whereas plainly the continuous time limit set is always the entire circle. Thus a single trajectory may be insufficient for investigating an attractor of the continuous time system, however long the trajectory may be. A more significant example of this has been observed by Suzuki *et al.* [22], who applied leaky integrate-and-fire sampling to the standard Rössler system. This arrangement was further investigated in [5].

(In principle this problem can even occur in the uniform sampling case: the  $\omega$ -limit set of  $x_0$  under  $\phi_T$  may be a proper subset of that under the flow: uniform rotation on the circle again provides an obvious example. However, Takens [23] showed that for a residual set of sampling times  $T$  these  $\omega$ -limit sets will be the



same. The relationship between integrate-and-fire sampling and uniform sampling described in section 5 shows that a corresponding result holds for integrate-and-fire systems.) Elucidation of the general circumstances under which the attractors of  $g$  are proper subsets of those of the flow for the non-uniform sampling case awaits further investigation.

Finally we return to more theoretical implications of Theorems 3 and 4. One way of viewing these is as assertions that delay maps are likely to produce embeddings if certain conditions are met; the conditions then give information about the significant *barriers* to embedding. Aside from the smoothness conditions (which we often just assume are met), much the most important condition is that  $D\tau(x)X(x) > -1$  for all  $x$ . We can use this to investigate whether we should expect embeddings (and hence whether dynamical information is transferred) in various scenarios. For example, Racicot and Longtin [13] describe four interval generating mechanisms (regarded as neuron models) and investigate whether the interval sequences reflect the characteristics of the underlying deterministic system. The four mechanisms are: the integrate-and-fire and variable threshold mechanisms described previously; a ‘leaky’ version of integrate-and-fire (in which the integral is modified by progressively reducing the contribution of earlier parts of the integrand); and a mechanism based on integration of the FitzHugh-Nagumo equations. The last of these is more immediately akin to the threshold crossing approach mentioned in section 3 than to the  $\tau$  function approach, so we shall not consider it here. As we have seen, the  $D\tau(x)X(x) > -1$  condition is always satisfied by integrate-and-fire, but for the variable threshold procedure it is only satisfied if the activity rise rate (and hence the sampling rate) is sufficiently high—if it is lower we cannot expect an embedding. Thus these two mechanisms are qualitatively different in a way which may not be particularly easy to detect in a numerical investigation, unless we know what we are looking for. For the leaky integrate-and-fire we find, briefly, that the condition is satisfied (and so embeddings are generic) when the integrand  $v$  everywhere exceeds  $k\Theta$ , where  $k$  is the leakage rate constant. We shall however postpone a detailed analysis of this case to a subsequent publication; the main point here is that the condition gives us a powerful way of investigating when embeddings can (and cannot) be expected.

## 7 Proofs

In this section we provide proofs for Lemma 1 and Theorems 3–6. Naturally we are concerned with various functions on the manifold  $M$ , such as the sampling interval function  $\tau$ , the sampling state map  $g$ , and the delay maps. Most of the time we work with coordinate representations of these functions, but we generally suppress explicit mention of charts—thus in the following subsection we write  $D\tau(x)X(x)$ ,  $x \in M$ , rather than (say)  $D\tau \circ h^{-1}(u)\hat{X}(u)$ , where  $u = h(x) \in U \subset M$ , (and  $\hat{X}$  is the local representation of  $X$ ). We only mention chart maps explicitly if we need to choose some particular form for them.

### 7.1 Proof of Lemma 1: the condition for $g$ to be a diffeomorphism

Lemma 1 follows as a straightforward corollary to the following:

**Lemma 2** *Let  $X$  be a  $C^1$  vector field on a compact manifold  $M$ , and  $\tau: M \rightarrow \mathbb{R}$  a  $C^1$  function. The mapping  $g: M \rightarrow M$  defined by  $g(x) = \phi(x, \tau(x))$ , where  $\phi: M \times \mathbb{R} \rightarrow M$  is the flow arising from  $X$ , fails to be a diffeomorphism if and only if there is a point  $x \in M$  at which  $D\tau(x)X(x) = -1$ .*

**Proof:** Let  $J_t(x)$  be the derivative of the time- $t$  map at  $x$ . By partial differentiation and the chain rule we find

$$Dg(x) = J_{\tau(x)}(x) + X(g(x))D\tau(x) \quad (5)$$

Using the fact that  $J_t(x)X(x) = X(\phi(x, t))$  we see

$$Dg(x)X(x) = J_{\tau(x)}(x)X(x) + X(g(x))D\tau(x)X(x) = (1 + D\tau(x)X(x))X(g(x))$$

Thus if  $D\tau(x)X(x) = -1$  then  $Dg(x)X(x) = 0$  (clearly  $X(x) \neq 0$ ) and  $g$  is not immersive at  $x$ . This proves the ‘if’ part of the lemma.

Now assume that  $g$  is not a diffeomorphism. Since  $g$  is certainly  $C^1$  (and  $M$  is compact),  $g$  must either fail to be immersive or fail to be injective. Suppose first that  $g$  is not immersive: then there is some  $x \in M$ , and some non-zero vector  $v$  tangent to  $M$  at  $x$  such that  $Dg(x)v = 0$ . From (5) we have

$$Dg(x)v = J_{\tau(x)}(x)v + X(g(x))D\tau(x)v$$

$J_t(x)$  is an isomorphism for any  $t$  (the time- $t$  map being a diffeomorphism) so  $J_{\tau(x)}(x)v \neq 0$ . Thus  $Dg(x)v = 0$  implies  $X(g(x)) \neq 0$  and  $J_{\tau(x)}(x)v$  is collinear with  $X(g(x))$ . Moreover  $J_{\tau(x)}(x)X(x) = X(g(x))$  so  $X(x) \neq 0$  and  $v$  is collinear with  $X(x)$ ; let us say  $v = \alpha X(x)$  ( $\alpha \neq 0$ ). Then

$$Dg(x)v = \alpha J_{\tau(x)}(x)X(x) + \alpha X(g(x))D\tau(x)X(x) = \alpha(1 + D\tau(x)X(x))X(g(x))$$

Hence  $Dg(x)v = 0$  implies  $D\tau(x)X(x) = -1$  as required.

The remaining possibility is that  $g$  fails to be injective. It follows immediately from the definition of  $g$  that every trajectory of the vector field  $X$  is an invariant set of  $g$ , so if  $g$  is not injective then there must be two distinct points,  $x_1, x_2 \in M$  such

that  $g(x_1) = g(x_2)$ , where  $x_1$  and  $x_2$  are on the same trajectory, so that there is some  $t \in \mathbb{R}$  such that  $x_2 = \phi(x_1, t)$ . Clearly we can take  $t > 0$ , and if the trajectory of  $X$  is periodic with period  $T$  we can further take  $t < T$ .

Consider the function  $f: \mathbb{R} \rightarrow \mathbb{R}$  defined by  $f(t') = t' + \tau(\phi(x_1, t'))$ ; clearly this is a  $C^1$  function on  $\mathbb{R}$ , with  $f(0) = \tau(x_1)$ . Note that

$$\begin{aligned} \phi(x_1, f(t)) &= \phi(x_1, t + \tau(\phi(x_1, t))) = \phi(x_1, t + \tau(x_2)) = \phi(\phi(x_1, t), \tau(x_2)) \\ &= \phi(x_2, \tau(x_2)) = g(x_2) = g(x_1) = \phi(x_1, \tau(x_1)) = \phi(x_1, f(0)) \end{aligned}$$

so  $\phi(x_1, f(t)) = \phi(x_1, f(0))$ . If the trajectory containing  $x_1$  and  $x_2$  is not periodic, this implies that  $f(t) = f(0)$ . It immediately follows from Rolle's Theorem that there is some  $\hat{t}$  between 0 and  $t$  at which  $f'(\hat{t}) = 0$ .

If the trajectory containing  $x_1$  and  $x_2$  is periodic with period  $T$ ,  $\phi(x_1, f(t)) = \phi(x_1, f(0))$  implies that  $f(t) = f(0) + nT$  for some integer  $n$ . Recall that  $0 < t < T$ , and note from the definition of  $f$  that  $f(T) = T + \tau(x_1) = f(0) + T$ . Now if  $n \leq 0$  the Intermediate Value Theorem (IVT) shows that  $\exists \tilde{t}$ , with  $t \leq \tilde{t} \leq T$ , such that  $f(\tilde{t}) = f(0)$ ; and Rolle's Theorem now implies  $\exists \hat{t}$ ,  $0 < \hat{t} < \tilde{t}$ , at which  $f'(\hat{t}) = 0$ . If  $n \geq 1$  the IVT shows  $\exists \tilde{t}$ ,  $0 \leq \tilde{t} \leq t$ , such that  $f(\tilde{t}) = f(0) + T = f(T)$ ; and Rolle's Theorem now implies  $\exists \hat{t}$ ,  $\tilde{t} < \hat{t} < T$ , at which  $f'(\hat{t}) = 0$ .

Hence  $g(x_1) = g(x_2)$  implies there is a  $\hat{t} > 0$  such that  $f'(\hat{t}) = 0$ . From the definition of  $f$  we get  $f'(t') = 1 + D\tau(\phi(x_1, t'))X(\phi(x_1, t'))$ . Setting  $x = \phi(x_1, \hat{t})$ , the condition  $f'(\hat{t}) = 0$  becomes  $1 + D\tau(x)X(x) = 0$ , as required.

**Corollary (Lemma 1)** *If the manifold  $M$  of Lemma 2 is connected, then  $g$  is a diffeomorphism if and only if  $D\tau(x)X(x) > -1$  for all  $x \in M$ . If  $X$  and  $\tau$  are  $C^r$ ,  $r > 1$ , then  $g$  is a  $C^r$  diffeomorphism.*

**Proof:** If  $D\tau(x)X(x) > -1$  then  $g$  is a diffeomorphism by Lemma 2. To prove the converse, note that if  $M$  is connected then it is path-connected; so if there were  $x \in M$  such that  $D\tau(x)X(x) > -1$ , and  $x' \in M$  such that  $D\tau(x')X(x') < -1$ , then by continuity of  $D\tau$  and  $X$  (and the IVT) there would be some  $x''$  on a path joining  $x$  to  $x'$  at which  $D\tau(x'')X(x'') = -1$ , and so by Lemma 2  $g$  would fail to be a diffeomorphism. Hence  $g$  is a diffeomorphism only if either  $D\tau(x)X(x) > -1 \forall x \in M$ , or  $D\tau(x)X(x) < -1 \forall x \in M$ . However, if we choose a point  $x_0$ , let  $x(t) = \phi(x_0, t)$  and define  $\hat{\tau}: \mathbb{R} \rightarrow \mathbb{R}$  by  $\hat{\tau}(t) = \tau(x(t))$  we see that  $\hat{\tau}'(t) = D\tau(x(t))X(x(t))$  so  $\hat{\tau}(t) = \hat{\tau}(0) + \int_0^t D\tau(x(t))X(x(t)) dt$ . If  $D\tau(x)X(x) < -1 \forall x$  then  $\hat{\tau}(t) < \hat{\tau}(0) - t$  for all  $t$  and so  $\hat{\tau}$ , and hence  $\tau$ , is unbounded, contradicting the continuity of  $\tau$ . The only remaining possibility, if  $g$  is a diffeomorphism, is  $D\tau(x)X(x) > -1 \forall x \in M$ .

If  $X$  and  $\tau$  are  $C^r$ ,  $r > 1$ , then  $g$  is plainly  $C^r$ , and is a  $C^r$  diffeomorphism by the Inverse Function Theorem.

## 7.2 Proof of Theorem 3: State Dependent Sampling Embedding Theorem

Roughly speaking an embedding theorem for dynamical systems states that a delay map, (or some other analogous map) constitutes an embedding of the state space of a dynamical system, for most of the systems in some suitable class. (Different theorems differ in the kind of delay map they consider, the class of systems they consider, or the meaning of 'most'.) A large part of proving such a theorem is concerned with showing

that if, for a particular system, the corresponding delay map is not an embedding, then arbitrarily small changes to the system (including its measurement function) can be found that result in a system that *is* embedded by its corresponding delay map. For the kinds of system we are considering in this paper these changes can mostly be found (or at least their existence demonstrated) by standard transversality arguments, but these arguments are impeded by the presence of certain problematic points in the state space. If we use  $\Phi$  to stand for whatever delay map we are considering,  $x$  is a problematic point if some of the components of  $\Phi(x)$  are necessarily the same as other components: for standard delay maps this happens when  $x$  is a periodic point with period less than the number of components of  $\Phi(x)$ . Also problematic are pairs  $x$  and  $y$  for which some components of  $\Phi(x)$  are necessarily the same as some (not necessarily the corresponding ones) of  $\Phi(y)$ ; this happens when  $x$  and  $y$  lie close together on the same trajectory. Typically the proof of the embedding theorem consists of stages, in which the problematic points are dealt with first, and all the remaining points are dealt with subsequently using transversality arguments. Thus in the proofs below we begin by ensuring that the most problematic points—those periodic points with a period less than or equal to some suitable integer  $l_p$ —are finite in number, and an embedding is established on a neighbourhood of these points.

We shall need the following result:

**Lemma 3** *Let  $A$  be a complex  $n \times n$  matrix having distinct eigenvalues  $\lambda_i$ ,  $i = 1, \dots, n$  (so that the corresponding eigenvectors  $v_i$  form a basis for  $\mathcal{C}^n$ ). The matrix  $A + v_1 w^T$ , where  $w$  is an arbitrary but fixed vector, has the same eigenvalues as  $A$ , except that  $\lambda_1$  is replaced by  $\lambda_1 + w^T v_1$ .*

**Proof:** Let  $V$  be the matrix of eigenvectors of  $A$ , so that  $V^{-1}AV = \Lambda$  where  $\Lambda = \text{diag}\{\lambda_1, \dots, \lambda_n\}$ . Then  $V^{-1}(A + v_1 w^T)V = \Lambda + e_1 w^T V$  where  $e_1^T = (1, 0, \dots, 0)$ . The first row of  $e_1 w^T V$  is  $(w^T v_1, w^T v_2, \dots, w^T v_n)$ , and all the other elements of this matrix are zero. Hence  $\Lambda + e_1 w^T V$  is an upper triangular matrix whose diagonal elements are  $\lambda_1 + w^T v_1, \lambda_2, \dots, \lambda_n$ .

To prove the state dependent sampling theorem we need to show that, for an open and dense set of pairs  $(\tau, X) \in \mathcal{T}$ , the periodic points of the  $g$  that arise from  $(\tau, X)$  have the conditions required by Takens's theorem.

**Lemma 4** *Let  $l_p$  be a given positive integer. Let  $\mathcal{S} \subset \mathcal{T}$  be the set of pairs  $(\tau, X)$  satisfying the following conditions:*

1. *The periodic points of  $g_{\tau, X}$  having period  $l_p$  or less are finite in number, and these points are hyperbolic.*
2. *For each of these periodic points the characteristic values are all distinct.*

*Then  $\mathcal{S}$  is open and dense in  $\mathcal{T}$ .*

(We refer to the periodic points of  $g$  whose period is  $l_p$  or less as the *low period* points of  $g$ . The value of  $l_p$  that we must use is dictated by the embedding theorem we want to prove—this is discussed further below.)

**Proof:** The ‘openness’ part of the lemma is straightforward. The Kupka-Smale theorem shows that the diffeomorphisms satisfying the first condition form an open subset of  $D^2(M)$ , (the set of  $C^2$  diffeomorphisms on  $M$ ). Furthermore, if  $f \in D^2(M)$

has a hyperbolic periodic point with distinct characteristic values, then every diffeomorphism  $f'$  in a sufficiently small neighbourhood of  $f$  has a corresponding periodic point, also hyperbolic, whose characteristic values are also distinct, since these depend continuously on  $f'$  within the neighbourhood. Thus the diffeomorphisms satisfying both conditions form an open set of  $D^2(M)$ . Since  $g$  depends continuously on  $(\tau, X)$ , the pairs producing diffeomorphisms which satisfy the conditions also form an open set.

To prove that  $\mathcal{S}$  is dense in  $\mathcal{T}$  requires rather more work. We need to show that for any  $(\tau, X) \in \mathcal{T}$ , every neighbourhood  $\mathcal{N}$  of  $(\tau, X)$  contains a pair  $(\tau', X')$  such that  $g_{\tau', X'}$  satisfies the conditions. (We may assume of course that  $\mathcal{N}$  is of the form  $\mathcal{N}_1 \times \mathcal{N}_2$  where  $\mathcal{N}_1 \subset C^2(M, \mathbb{R})$  is an open neighbourhood of  $\tau$  and  $\mathcal{N}_2 \subset \mathcal{X}^2(M)$  is an open neighbourhood of  $X$ .) Note that if  $x$  is a periodic point of  $g_{\tau, X}$  then  $x$  must lie on a periodic trajectory of  $X$ . It is well-known ([11]) that given a positive constant  $K$  and a vector field  $X \in \mathcal{X}^2(M)$ , every neighbourhood of  $X$  contains a vector field whose periodic trajectories of period less than  $K$  (including equilibrium points) are finite in number and are hyperbolic. Let  $B = \max_{x \in M} |\tau(x)|$  and choose  $A > B$ . We select  $X' \in \mathcal{N}_2$  to have finitely many hyperbolic periodic trajectories with period less than  $l_p A$ , and to satisfy certain other conditions on the characteristic values of these orbits: namely, that the characteristic values of each of the orbits are distinct, and further, that if  $\lambda$  and  $\mu$  are (possibly complex) distinct characteristic values of an orbit whose period is  $T$ , then no equation of the form  $\lambda^n = \mu^n$  is satisfied for an integer  $n$  in the range  $0 < n < l_p A/T$ .  $X'$  is considered fixed from now on. We may assume (by using a smaller neighbourhood if necessary) that  $|\tau'(x)| < A$  for all  $\tau' \in \mathcal{N}_1$  and  $x \in M$ .

Given  $\tau' \in \mathcal{N}_1$  (and using our fixed  $X'$ ) define, for integer  $r > 0$ :

$$\sigma_r(x) = \tau'(x) + \tau'(g'(x)) + \tau'(g'^2(x)) + \dots + \tau'(g'^{r-1}(x)) \quad (6)$$

(where  $g' = g_{\tau', X'}$ ). Then  $g'^r(x) = \phi'(x, \sigma_r(x))$  ( $\phi'$  is the flow corresponding to  $X'$ .) If  $x$  is a period- $r$  point of  $g'$ , then  $x$  lies on a periodic trajectory of  $X'$ , and if the period of this trajectory is  $T$  we must have  $\sigma_r(x) = nT$  for some positive integer  $n$ . Since  $|\sigma_r(x)| < rA$  it follows that  $T < rA/n \leq rA$ . Thus every periodic point of  $g'$  whose period is  $l_p$  or less lies on a periodic trajectory of  $X'$  whose period is less than  $l_p A$ ; we have seen that there are only finitely many such trajectories. (We refer to these trajectories as the *low period* trajectories of  $X'$ .)

For convenience relabel  $\tau'$  as  $\tau$ , and  $g'$  as  $g$ . The fact that the low period points of  $g$  are confined to a finite collection of periodic trajectories of  $X'$  does not mean that these points are themselves finite in number—indeed it is quite possible for all the points on such a trajectory to be periodic points of  $g$ . To deal with this situation we make a series of perturbations to  $\tau$  to produce a new function that satisfies the conditions of the lemma. We need to show that such perturbations can be found and that they can be taken to be arbitrarily small, so that the resulting function lies in  $\mathcal{N}_1$ . If all the low period points of  $g$  on a given low period trajectory of  $X'$  have the properties described in the lemma, these properties will persist under all sufficiently small perturbations. We may thus deal with the low period trajectories one at a time.

Consider first the equilibrium points of  $X'$ : each of these gives rise to a fixed point of  $g$  (irrespective of  $\tau$ ). From equation (5) we find  $Dg(x) = J_{\tau(x)}(x)$  so it follows directly from the properties of  $X'$  that these points are hyperbolic and have distinct eigenvalues.

Now consider a low period trajectory,  $C$ , of  $X'$ . Restricting  $X'$  to  $C$  gives a one-dimensional system which is  $C^2$ -conjugate to rotation on the circle  $S^1$  at uniform speed  $\omega = 2\pi/T$ , where  $T$  is the period of  $C$ , the conjugacy being  $F: S^1 \rightarrow C$ ,  $F(\theta) = \phi(x_0, \theta T/(2\pi))$  ( $x_0$  is any arbitrarily chosen point of  $C$ ). Note that if  $\tau$  is a sampling interval function on  $C$  there is a corresponding function  $\hat{\tau} = \tau \circ F$  on  $S^1$ , and the corresponding diffeomorphisms  $g(x) = \phi(x, \tau(x))$  and  $\hat{g}(\theta) = \theta + \omega\hat{\tau}(\theta) \pmod{2\pi}$ , are conjugate under  $F$ ;  $g = F \circ \hat{g} \circ F^{-1}$ . Suppose that given  $\hat{\tau}: S^1 \rightarrow \mathbb{R}$  we can always find, arbitrarily close to  $\hat{\tau}$ , a  $\bar{\tau}$  such that the diffeomorphism  $\bar{g}: S^1 \rightarrow S^1$ ,  $\bar{g}(\theta) = \theta + \omega\bar{\tau}(\theta) \pmod{2\pi}$ , has finitely many low period points all of which are hyperbolic; then given any  $\tau$  on  $C$  we can find, arbitrarily close to  $\tau$ , a  $\tilde{\tau}$  such that  $\tilde{g}(x) = \phi(x, \tilde{\tau}(x))$  has finitely many low period points (which are hyperbolic). (We choose  $\tilde{\tau}$  to be  $\bar{\tau} \circ F^{-1}$ , where  $\bar{\tau}$  is sufficiently close to  $\tau \circ F$ .) So we only need to confirm that we can find a suitable  $\bar{\tau}$  on  $S^1$ .

Given  $\hat{\tau}$  on  $S^1$  we can define  $f: \mathbb{R} \rightarrow \mathbb{R}$  by setting  $f(x) = \hat{\tau}(2\pi x)$  for  $x \in [0, 1]$ , and letting  $f$  be a periodic function with unit period. Then the function  $G(x) = x + \frac{\omega}{2\pi}f(x)$  is a lift of  $\hat{g}(\theta) = \theta + \omega\hat{\tau}(\theta) \pmod{2\pi}$ . There is a diffeomorphism  $\bar{g}$  arbitrarily close to  $\hat{g}$  such that  $\bar{g}$  has only finitely many periodic points (and they are hyperbolic), ( $\bar{g}$  is a Morse-Smale diffeomorphism [11]). Say  $\bar{G}$  is the lift of  $\bar{g}$  closest to  $G$ ; clearly we can arrange for  $\bar{G}$  to be as close as we please to  $G$  by choosing  $\bar{g}$  sufficiently close to  $\hat{g}$ . Define  $\bar{f}(x)$  by  $(2\pi/\omega)(\bar{G}(x) - x)$  and  $\bar{\tau}(\theta)$  by  $\bar{f}(\theta/(2\pi))$ ; then  $\bar{g}(\theta) = \theta + \omega\bar{\tau}(\theta) \pmod{2\pi}$ , and we can arrange for  $\bar{\tau}$  to be arbitrarily close to  $\hat{\tau}$ . This is what we needed for the argument of the previous paragraph.

We have thus seen that for any  $\tau$  on  $C$  there is an arbitrarily small function  $p$  on  $C$  such that  $g_{\tau+p, X'}$  on  $C$  has only finitely many periodic points. We now only need to extend  $p$  to  $M$  to produce a new  $\tau$  with the properties we require—this can be done using a partition of unity on  $C$ .

We now need to consider the characteristic values of the low period points on  $C$ . Let  $x$  be a period- $r$  point of  $g$ : then  $\sigma_r(x) = nT$  for some positive integer  $n$ , implying  $nT < rA \leq l_p A$ , so that  $n < l_p A/T$ . From  $g^r(x) = \phi(x, \sigma_r(x))$  we get

$$Dg^r(x) = J_{\sigma_r(x)}(x) + X'(g^r(x))D\sigma_r(x) = J_{nT}(x) + X'(x)D\sigma_r(x)$$

Since  $J_{nT}(x)X'(x) = X'(x)$  Lemma 3 shows that the eigenvalues of  $Dg^r$  are just those of  $J_{nT}$  except that the unit eigenvalue of  $J_{nT}$  is replaced by  $1 + D\sigma_r(x)X'(x)$ . The eigenvalues of  $J_{nT}$  are of course just those of  $J_T$  raised to the power  $n$ , and our original choice of the vector field  $X'$  means that they are all distinct. Thus the eigenvalues of  $Dg^r(x)$  are all distinct so long as  $1 + D\sigma_r(x)X'(x)$  does not equal any of the non-unit eigenvalues of  $J_{nT}$ . The periodic points are isolated so we can construct a perturbation function  $p$  to  $\tau$  whose support contains no periodic point except  $g^{r-1}(x)$ ; from equation (6) we see that  $\sigma_r(x)$  is perturbed to  $\sigma_r(x) + p(g^{r-1}(x))$ . We can arrange for  $p(g^{r-1}(x)) = 0$  (so that  $x$  remains periodic) while  $Dp(g^{r-1}(x))$  can be chosen arbitrarily; hence we can make arbitrary adjustments to  $D\sigma_r(x)$ , and so to  $1 + D\sigma_r(x)X'(x)$ . Clearly even arbitrarily small adjustments make this value different to the other eigenvalues of  $J_{nT}$ .

We have thus shown that arbitrarily close to any pair  $(\tau, X) \in \mathcal{T}$  is a pair  $(\tau', X')$  for which the low period points of  $g_{\tau', X'}$  are finite in number, are hyperbolic, and have distinct characteristic values. This completes the proof of Lemma 4.

We are now in a position to prove the state dependent sampling theorem. For

$y \in C^2(M, \mathbb{R})$  and  $(\tau, X) \in \mathcal{T}$  we define  $\Phi_{y,\tau,X}: M \rightarrow \mathbb{R}^d$  by

$$\Phi_{y,\tau,X}(x) = (y(x), y(g_{\tau,X}(x)), \dots, y(g_{\tau,X}^{d-1}(x))) \quad (7)$$

Then we have

**Theorem 3** *For  $d > 2m$  the triples  $(y, \tau, X) \in C^2(M, \mathbb{R}) \times \mathcal{T}$  for which  $\Phi_{y,\tau,X}$  is an embedding of  $M$  form an open and dense subset of  $C^2(M, \mathbb{R}) \times \mathcal{T}$*

**Proof:** Let  $(\tau, X) \in \mathcal{S}$ , where  $\mathcal{S}$  is as in Lemma 4 with  $l_p$  set to  $d$ ,  $d > 2m$ ; then by Theorem 2 there is a dense subset of  $C^2(M, \mathbb{R})$  such that, for all  $y$  in this subset,  $\Phi_{y,\tau,X}$  is an embedding of  $M$ . Since  $\mathcal{S}$  is dense in  $\mathcal{T}$  by Lemma 4, the set of triples  $(y, \tau, X) \in C^2(M, \mathbb{R}) \times \mathcal{T}$  for which  $\Phi_{y,\tau,X}$  is an embedding, is dense.

Clearly  $\Phi_{y,\tau,X}$  depends continuously on  $(y, g_{\tau,X})$ , so it depends continuously on  $(y, \tau, X)$ ; that is, the mapping  $C^2(M, \mathbb{R}) \times \mathcal{T} \rightarrow C^2(M, \mathbb{R}^d)$  given by  $(y, \tau, X) \mapsto \Phi_{y,\tau,X}$  is continuous.  $C^2$  embeddings of  $M$  form an open subset of  $C^2(M, \mathbb{R}^d)$ , so the triples  $(y, \tau, X)$  giving rise to embeddings form an open subset of  $C^2(M, \mathbb{R}) \times \mathcal{T}$ .

### 7.3 Proof of Theorem 4: Interspike Intervals Embedding Theorem

To prove the interspike intervals theorem we must of course again restrict the class of pairs  $(\tau, X)$  that we consider to those satisfying the conditions of Lemma 1—i.e. to the set  $\mathcal{T}$ . But this is not enough:  $\mathcal{T}$  may still contain an open set of pairs for which interspike intervals do not provide an embedding, so we need to restrict attention further; this is attended to in the following paragraph.

We first introduce a new notation for the delay map: for each pair  $(\tau, X)$  we say  $\Phi_{\tau,X}: M \rightarrow \mathbb{R}^d$  is defined by

$$\Phi_{\tau,X}(x) = (\tau(x), \tau(g_{\tau,X}(x)), \dots, \tau(g_{\tau,X}^{d-1}(x))) \quad (8)$$

( $\Phi_{\tau,X}$  is exactly the same as  $\Phi_{\tau,g}$  of equation (4), but it is convenient to have a slightly more explicit notation for it at this stage.)

Let  $\hat{\mathcal{T}} \subset \mathcal{T}$  be the set of pairs  $(\tau, X)$  such that there is a periodic orbit  $C$  of  $X$  and a point  $x \in C$  for which  $\tau(x)$  is an integer multiple of the period of  $C$ . Clearly  $g(x) = x$ , and all the components of  $\Phi_{\tau,X}(x)$  are equal to  $\tau(x)$ . If  $\tau|C$  has neither a maximum nor a minimum at  $x$  then there must be another point  $x'$  at which  $\tau(x') = \tau(x)$  (by the IVT), and we see that  $g(x') = x'$ , and  $\Phi_{\tau,X}(x') = \Phi_{\tau,X}(x)$ . If  $\tau|C$  has a maximum or a minimum at  $x$  then  $D\tau(x)X(x) = 0$  and  $D\Phi_{\tau,X}(x)X(x) = 0$ . Thus  $\Phi_{\tau,X}$  fails to be an embedding for every  $(\tau, X) \in \hat{\mathcal{T}}$ . Furthermore if  $D\tau(x)X(x) \neq 0$  then  $x$  is a hyperbolic fixed point of  $g|C$ , and all sufficiently close  $(\tau', X')$  will have a corresponding fixed point, and so will belong to  $\hat{\mathcal{T}}$ . Hence the pairs  $(\tau, X)$  giving rise to embeddings cannot be dense in  $\mathcal{T}$ . To prove an embedding theorem in this situation we must eliminate  $\hat{\mathcal{T}}$  from consideration. However, simply considering  $\mathcal{T} \setminus \hat{\mathcal{T}}$  runs into the problem that  $\hat{\mathcal{T}}$  may not be a closed subset of  $\mathcal{T}$ , so that  $\mathcal{T} \setminus \hat{\mathcal{T}}$  is not open: this is unsatisfactory since small perturbations of a pair in  $\mathcal{T} \setminus \hat{\mathcal{T}}$  are not guaranteed to be in  $\mathcal{T} \setminus \hat{\mathcal{T}}$ , preventing us from showing that embeddings are dense. Thus we need to discard the closure of  $\hat{\mathcal{T}}$  rather than  $\hat{\mathcal{T}}$ . Let  $\hat{\mathcal{S}}$  be the set of pairs  $(\tau, X) \in \mathcal{T}$  such that  $X$  has a non-hyperbolic equilibrium point  $x$  with an eigenvalue  $i\omega$  where  $\omega = 2\pi n/\tau(x)$  for some positive integer  $n$ . It is not difficult to see that any

limit point of  $\hat{\mathcal{T}}$  not in  $\hat{\mathcal{T}}$  must be in  $\hat{S}$  (and that  $\hat{S}$  is closed in  $\mathcal{T}$ ). So we consider the (open) set  $\mathcal{T}' = \mathcal{T} \setminus (\hat{\mathcal{T}} \cup \hat{S})$ .

We want to know if  $\Phi_{\tau, X}$  is likely to be an embedding. The theorem we shall prove is the following:

**Theorem 4** *For  $d > 2m$  the pairs  $(\tau, X) \in \mathcal{T}'$  for which  $\Phi_{\tau, X}$  is an embedding of  $M$  form an open and dense subset of  $\mathcal{T}'$ .*

The main difference between this and the state dependent sampling case of Theorem 3 is that since  $\tau$  now plays the role of the measurement function, the measurement function and the diffeomorphism  $g$  are inextricably linked—they cannot be perturbed independently, and this means there is no straightforward way of deriving Theorem 4 by an application of Takens’s Theorem. Thus we must prove Theorem 4 from scratch; we do this by a strategy that closely parallels the proof of Takens’s Theorem itself [23].

As usual, the difficult part of the theorem is the denseness part. Given any pair  $(\tau, X) \in \mathcal{T}'$  we must show that arbitrarily small perturbations to  $\tau$  and  $X$  can be found that produce an embedding. A series of perturbations is constructed, each of which endows the map  $\Phi_{\tau', X'}$  with some desired property, culminating in  $\Phi_{\tau', X'}$  being an embedding. The first perturbations produce a new pair in the set  $\mathcal{S}$  described in Lemma 4; (note that since  $\mathcal{S}$  is open and dense in  $\mathcal{T}$  (by Lemma 4), and  $\mathcal{T}'$  is open in  $\mathcal{T}$ ,  $\mathcal{S} \cap \mathcal{T}'$  is open and dense in  $\mathcal{T}'$ , so the required perturbations have already been described in the proof to the lemma). The vector field is fixed from now on—all the subsequent perturbations are made to  $\tau$  (and have, of course, a corresponding effect on  $g$ ).

Perturbations are made so that  $\Phi$  is injective on the set of low period points, and further perturbations ensure  $\Phi$  is immersive at these points, so is an embedding of a neighbourhood of them. The remainder of  $M$  is covered with coordinate patches, and perturbations are made on each patch to ensure  $\Phi$  is immersive there; once this has been done for each patch,  $\Phi$  is an immersion of the whole manifold. Finally, another system of patches is chosen and a perturbation constructed based on it that makes  $\Phi$  injective on  $M$ , and so an embedding.

Of course, the hard work comes in showing that the perturbations described above can actually be found: those dealing with the low period points can largely be written down directly, while for the remainder each perturbation is expressed as a linear combination of suitably chosen localized functions, and transversality arguments used to show that the coefficients in these combinations can be chosen so that  $\Phi$  acquires the right properties.

### Making $\Phi_{\tau, X}$ injective on the low period points

Let us begin with the low period points. We may assume that our pair  $(\tau, X)$  lies in  $\mathcal{S}$ , so the low period points of  $g$  are finite in number, hyperbolic, and have distinct eigenvalues (and these properties are shared by all sufficiently close pairs), (see Lemma 4). The first task is to ensure that  $\tau$  takes on distinct values at all the low period points. Since we are working in  $\mathcal{T}'$ , any fixed point  $x$  of  $g$  must be an equilibrium point of  $X$ , and so will remain a fixed point regardless of how we change  $\tau$ . We find an open neighbourhood  $U$  of  $x$  containing none of the other low period points, and a  $C^2$  function  $\psi: M \rightarrow \mathbb{R}$  such that  $\psi(y) = 0 \forall y \notin U$  and  $\psi(x) \neq 0$ . Then it is clear that for all except a finite number of values of  $a$  the function  $\tau' = \tau + a\psi$  takes a



different value at  $x$  to those at the other low period points. If necessary we make such a perturbation at each fixed point.

Periodic points other than fixed points are rather different: changing the value of  $\tau$  at one of these will generally change the *positions* of others. Let  $x$  be a low period point of  $g$ , of period at least 2, so that  $x_1 = g(x) \neq x$ . Perturbing  $\tau$  will generally shift  $x_1$ , and may well result in  $x$  no longer being a periodic point. To overcome this we shall introduce perturbations both at  $x$  and at  $x_1$  such that the latter counteracts the former and  $x$  remains a periodic point. The Hausdorff property means we can find an open neighbourhood  $U$  of  $x$  such that  $U$  and  $gU$  are disjoint, and such that  $U \cup gU$  contains none of the low period points except  $x$  and  $x_1$ . Letting  $B_r \subset \mathbb{R}^m$  be the open ball of radius  $r$  centred at the origin of  $\mathbb{R}^m$ , we may further assume that there is a chart map  $h$  of  $M$  such that  $hU = B_1$  (and  $h(x) = (0, \dots, 0)$ ). Let  $\lambda: \mathbb{R}^m \rightarrow [0, 1]$  be a  $C^2$  function such that  $\lambda(u) = 0 \forall u \notin B_{3/4}$  and  $\lambda(u) = 1 \forall u \in B_{1/2}$ ; also let  $\psi: M \rightarrow \mathbb{R}$  be defined by  $\psi(y) = \lambda \circ h(y) \forall y \in U$  and  $\psi(y) = 0$  otherwise: then  $\psi$  is  $C^2$ . Let  $\tau_a = \tau + a\psi$ , and  $g_a$  be the corresponding diffeomorphism (recall that  $X$  is a fixed vector field); note that  $g_a(y) = g(y) \forall y \notin h^{-1}B_{3/4}$ . For sufficiently small  $a$  we must have  $g_a h^{-1}B_{3/4} \subset gU$  and hence  $g_a U \subset gU$ . For such  $a$  we define  $\chi_a: M \rightarrow \mathbb{R}^m$  by  $\chi_a = \psi \circ g_a^{-1}$ ;  $\chi_a$  has support in  $gU$ . Now let  $\tau_{a,b} = \tau_a + b\chi_a$  (with  $g_{a,b}$  the corresponding diffeomorphism—note that  $g_{a,b}(y) = g_a(y)$  for  $y \in U$ , and  $g_{a,b}(y) = g_a(y) = g(y) \forall y \notin U \cup gU$ ). We choose  $b$  so that, at the periodic point  $x$ ,  $g_{a,b}^2(x) = g^2(x)$ , that is, so that  $\tau_{a,b}(x) + \tau_{a,b}(g_{a,b}(x)) = \tau(x) + \tau(g(x))$ ; it is easy to see that the  $b$  we require for this is  $b = \tau(g(x)) - \tau(g_a(x)) - a$ . Since  $g_{a,b}$  equals  $g$  at all other points (if any) on the orbit of  $x$  it is clear that  $x$  is a periodic point of  $g_{a,b}$  (with the same period as under  $g$ ). By taking  $a$  sufficiently small,  $\tau_{a,b}$  can be made arbitrarily close to  $\tau$ . With the above choice for  $b$  we have  $\tau_{a,b}(x) = \tau(x) + a$  and  $\tau_{a,b}(g_{a,b}(x)) = \tau(g(x)) - a$ , so it is clear that for all but finitely many values of  $a$ ,  $\tau_{a,b}(x)$  differs from the value of  $\tau_{a,b}$  at all the other low period points.

Having established the desired property at one low period point,  $x$ , it may turn out that a subsequent perturbation changes the position and  $\tau$  value of  $x$ ; however, sufficiently small perturbations will not make the new  $\tau$  value the same as at any of the other low period points. Hence this kind of perturbation can be applied to each of the points in turn until they all have different values.

### Making $\Phi_{\tau, X}$ immersive at the low period points

We now move on to ensuring  $\Phi_{\tau, X}$  is an immersion at the low period points. Suppose that  $\{x_0, x_1, \dots, x_{p-1}\}$  is a (low) period  $p$  orbit of  $g$ . For each  $i = 0, \dots, p-1$  we may find an open neighbourhood  $U_i$  of  $x_i$ , such that the neighbourhoods are disjoint, and there are chart maps  $h_i$ , with  $h_i U_i = B_1$  (and  $h_i(x_i) = (0, \dots, 0)$ ). We let  $u_{ji}: U_i \rightarrow \mathbb{R}$  be the  $j$ -th component of  $h_i$ , and  $\psi_i: M \rightarrow \mathbb{R}$  be  $\psi_i(y) = \lambda \circ h_i(y) \forall y \in U_i$  and  $\psi_i(y) = 0$  otherwise ( $\lambda$  is the same bump function as above). We now consider the perturbation

$$\tau' = \tau + \sum_{j=1}^m \sum_{i=0}^{p-1} a_{ji} u_{ji} \cdot \psi_i \quad (9)$$

for arbitrary coefficients  $a_{ji}$ . Note that for all  $y \in U_i$  we have  $\tau'(y) = \tau(y) + \sum_{j=1}^m a_{ji} u_{ji}(y) \psi_i(y)$ , and in particular  $\tau'(x_i) = \tau(x_i)$  for all  $i$ . This means that  $\{x_0, x_1, \dots, x_{p-1}\}$  is still a periodic orbit of the new diffeomorphism  $g'$ . We also have  $D\tau'(x_i) = D\tau(x_i) + a_i$ , where we have let  $a_i = (a_{1i}, a_{2i}, \dots, a_{mi})$ , so this construction

allows us to perturb  $D\tau(x_i)$  to any vector we please.

We wish to ensure that  $D\Phi_{\tau',X}(x_0)$  is full rank (obviously similar arguments will work for the other points on the orbit).  $D\Phi_{\tau',X}(x_0)$  has the form:

$$\begin{bmatrix} D\tau'(x_0) \\ D\tau'(x_1)Dg'(x_0) \\ D\tau'(x_2)Dg'(x_1)Dg'(x_0) \\ \vdots \\ D\tau'(x_{p-1})Dg'(x_{p-2}) \dots Dg'(x_0) \\ D\tau'(x_0)Dg'(x_{p-1}) \dots Dg'(x_0) \\ D\tau'(x_1)Dg'(x_0)Dg'(x_{p-1}) \dots Dg'(x_0) \\ \vdots \end{bmatrix}$$

which we write in a more convenient form by letting  $D\tau'(x_i) = w_i^T$  and  $Dg'(x_i) = A_i$ :

$$\begin{bmatrix} w_0^T \\ w_1^T A_0 \\ w_2^T A_1 A_0 \\ \vdots \\ w_{p-1}^T A_{p-2} \dots A_0 \\ w_0^T A_{p-1} \dots A_0 \\ w_1^T A_0 A_{p-1} \dots A_0 \\ \vdots \end{bmatrix} \quad (10)$$

From (5) we find that  $Dg'(x_i) = J_{\tau'(x_i)}(x_i) + X(g'(x_i))D\tau'(x_i)$ , or (writing  $J_i$  for  $J_{\tau'(x_i)}(x_i)$  and  $X_i$  for  $X(g'(x_i))$ ),  $A_i = J_i + X_i w_i^T$ . Both  $w_i^T$  and  $A_i$  are affected by the perturbation to  $\tau$ , but  $J_i$  is unaffected, since the value  $\tau(x_i)$  is unchanged. We can clarify the effect of the perturbation on  $D\Phi_{\tau,X}(x_0)$  by noting that the matrix (10) has the same rank as

$$\begin{bmatrix} w_0^T \\ w_1^T J_0 \\ w_2^T J_1 J_0 \\ \vdots \\ w_{p-1}^T J_{p-2} \dots J_0 \\ w_0^T J_{p-1} \dots J_0 \\ w_1^T J_0 J_{p-1} \dots J_0 \\ \vdots \end{bmatrix} \quad (11)$$

This can be established by showing that each row of (10) is the sum of the corresponding row of (11) and a linear combination of earlier rows of (11) (this shows that (10) is obtained from (11) by multiplication by a lower triangular matrix whose diagonal elements are all 1); for this it is enough to show that each row of (10) is the

sum of the corresponding row of (11) and a linear combination of earlier rows of (10). Assume then that this is true for the first  $k$  rows of (10). The  $k + 1$ -th row has the form  $w_i^T A_{i-1} A_{i-2} \dots A_0 A_{p-1} \dots A_0 \dots A_0$ , and

$$\begin{aligned} w_i^T A_{i-1} A_{i-2} \dots A_0 A_{p-1} \dots A_0 \dots A_0 &= \\ &= w_i^T (J_{i-1} + X_{i-1} w_{i-1}^T) A_{i-2} \dots A_0 A_{p-1} \dots A_0 \dots A_0 \\ &= w_i^T J_{i-1} A_{i-2} \dots A_0 A_{p-1} \dots A_0 \dots A_0 + (w_i^T X_{i-1}) w_{i-1}^T A_{i-2} \dots A_0 A_{p-1} \dots A_0 \dots A_0 \end{aligned}$$

The second of these terms is a scalar (namely  $w_i^T X_{i-1}$ ) times the  $k$ -th row of (10); we can treat the first term in the same way:

$$\begin{aligned} w_i^T J_{i-1} A_{i-2} \dots A_0 A_{p-1} \dots A_0 \dots A_0 &= \\ &= w_i^T J_{i-1} (J_{i-2} + X_{i-2} w_{i-2}^T) \dots A_0 A_{p-1} \dots A_0 \dots A_0 \\ &= w_i^T J_{i-1} J_{i-2} \dots A_0 A_{p-1} \dots A_0 \dots A_0 + (w_i^T J_{i-1} X_{i-2}) w_{i-2}^T \dots A_0 A_{p-1} \dots A_0 \dots A_0 \end{aligned}$$

in which the second term is a scalar ( $w_i^T J_{i-1} X_{i-2}$ ) times the  $k - 1$ -th row of (10). Proceeding in this way produces the required result for the  $k + 1$ -th row of (10).

Writing  $v_i^T = w_i^T J_{i-1} \dots J_0$ ,  $i = 0, \dots, p - 1$ , and  $M = J_{p-1} \dots J_0$ , the matrix (11) becomes

$$\begin{bmatrix} v_0^T \\ v_1^T \\ \vdots \\ v_{p-1}^T \\ v_0^T M \\ \vdots \end{bmatrix} \quad (12)$$

We have already seen that we can arrange for  $w_i^T$  to be any vector in  $\mathbb{R}^m$  by a suitable choice of the vector of coefficients  $a_i$ , and since the matrices  $J_j$ ,  $j = 0, \dots, p - 1$ , are all full rank the same is true of  $v_i^T$ . A standard result from linear systems theory (in which (12) is referred to as an ‘observability matrix’) tells us that if the eigenvalues of  $M$  are all distinct then the set of  $p$ -tuples  $(v_0^T, v_1^T, \dots, v_{p-1}^T)$  such that (12) is full rank is open and dense in  $\mathbb{R}^{pm}$ . The coefficients  $(a_0, a_1, \dots, a_{p-1})$  which make (12) full rank are correspondingly open and dense. Thus to show that  $D\Phi_{\tau, X}(x_0)$  can be made full rank by an arbitrarily small perturbation of the form (9) we need only confirm that  $M$  has distinct eigenvalues. If  $T$  is the period of the orbit of the vector field  $X$  on which  $x_0$  lies, then  $\tau(x_0) + \tau(x_1) + \dots + \tau(x_{p-1}) = rT$  for some positive integer  $r$ , and  $M$  is just the matrix  $D\phi_{rT}(x_0)$  (recall  $\phi_{rT}$  is the time- $rT$  map of the flow). We arranged for this to have distinct eigenvalues by performing perturbations to the vector field in the proof of Lemma 4.

Let  $P$  be the set of low period points of  $g'$ . (Note we have not yet set a value for  $l_p$ —the above discussion applies whatever value we choose.) Though we shall make further perturbations to  $g'$  these will not alter the set of low period points, so  $P$  is a fixed set from now on. It was established above that the images of all the low period points are distinct under  $\Phi_{\tau', X}$ , and so we can find a neighbourhood of each point such that the images under  $\Phi_{\tau', X}$  of no two of the neighbourhoods overlap. Further, since  $\Phi_{\tau', X}$  is immersive at each low period point, each  $x_i \in P$  has a neighbourhood  $U_i$  such that  $\Phi_{\tau', X}|_{U_i}$  is an embedding, and by the previous sentence we may take

the  $U_i$ 's to have disjoint images, so that  $\Phi_{\tau', X}$  is in fact an embedding of the union of the neighbourhoods. Since  $M$  is locally compact, we may further take the  $U_i$ 's to be compact neighbourhoods; their union is then also compact since there are only finitely many of them. Let  $V$  be the union of the neighbourhoods; it is clear that for all sufficiently small perturbations to  $\tau'$ , the resulting delay map will be an injective immersion on  $V$ .

### Making $\Phi_{\tau, X}$ immersive on the whole of $M$

Attention now turns to producing an immersion on the (closed) set  $M' = M \setminus \text{int}(V)$ . For this we use a construction that we shall need again several times below, so to avoid some repetition we set it out in general terms here. For convenience we relabel the current measurement function  $\tau$ , and the corresponding diffeomorphism  $g$  (*i.e.* we drop the primes). Let  $l_p$  be  $d - 1$  (or larger). Let  $C$  be any closed subset of  $M$  such that  $C \cap P = \emptyset$ , and let  $x \in C$ ; since  $x$  is not a low period point the points  $x, g(x), \dots, g^{d-1}(x)$  are all distinct, and we can find an open neighbourhood  $U_x$  of  $x$ , with compact closure  $\overline{U}_x$ , such that  $h_x \overline{U}_x = \overline{B}_1$  for some chart map  $h_x$  (and  $h_x x = (0, \dots, 0)$ ), the sets  $\overline{U}_x, g\overline{U}_x, \dots, g^{d-1}\overline{U}_x$  are all disjoint, and  $\overline{U}_x$  contains no points in  $P$ . Letting  $W_x = h_x^{-1}\overline{B}_{1/4}$  we see that  $\{W_x | x \in C\}$  is a collection of compact sets whose interiors cover  $C$ : we select a finite subset, which we relabel  $W_1, \dots, W_K$  (with corresponding points  $x_1, \dots, x_K$ , open neighbourhoods  $U_1, \dots, U_K$ , and chart maps  $h_1, \dots, h_K$ ), so that the interiors of these sets cover  $C$ . We call such a collection a ‘good cover’ for  $C$ . It would be more complete to refer to a ‘good cover with respect to  $g$ ’ since one of the properties of the good cover (the fact that  $\overline{U}_i, g\overline{U}_i, \dots, g^{d-1}\overline{U}_i$  are all disjoint) depends on  $g$ . The crucial point to note is that, although the sets  $gU_i$  etc. depend on  $g$ , the collection  $W_1, \dots, W_K$  (with the corresponding  $U_1, \dots, U_K$ ) still form a good cover of  $C$  for all  $g'$  sufficiently close to  $g$ . We can always find a good cover so long as  $C$  contains no low period points; sometimes we may need the cover to have other properties as well, in which case we will need to show how these can be arranged.

We now return to the particular case where  $C$  is the set  $M'$ ; we find a good cover for  $M'$ . Each  $W_i$  is dealt with in turn: a perturbation is made (with support in  $U_i \cup gU_i \cup \dots \cup g^{d-1}U_i$ ) which establishes an immersion at every point in  $W_i$ ; since this set is compact, subsequent perturbations will preserve the immersivity property.

Consider then the set  $U_1$ ; similar arguments will apply to all the other  $U_i$ 's and so for convenience we will drop the ‘1’ subscript from  $U_1$ ,  $h_1$  and  $W_1$ . To deal with this set we define  $\psi_{ji}: M \rightarrow \mathbb{R}$  by

$$\psi_{ji}(y) = \begin{cases} u_j(g^{-i}y) \times \lambda \circ h(g^{-i}y) & \text{for } y \in g^i U \\ 0 & \text{for } y \notin g^i U \end{cases}$$

where  $u_j$  is the  $j$ -th component of  $h$  and  $\lambda$  is the bump function used above. Now define a perturbed measurement function  $\tau'$  by

$$\tau' = \tau + \sum_{i=0}^{d-1} \sum_{j=1}^m a_{ij} \psi_{ji}$$

with corresponding mapping  $g'$ . Of course, we are only interested in cases where the coefficients are small enough that  $g'$  is actually a diffeomorphism. Let  $a \in \mathbb{R}^{dm}$  be the vector  $(a_{01}, a_{02}, \dots, a_{d-1, m})$ , then there is some  $\delta > 0$  such that if  $a \in B_\delta$  then  $g'$

is a diffeomorphism. ( $B_\delta$  is the open ball of radius  $\delta$  in  $\mathbb{R}^{dm}$ .) Indeed, for sufficiently small  $a$  we must have  $g^i U \subseteq g^i U$  for  $i = 1, \dots, d-1$ , so we will take  $\delta$  small enough to ensure this. This means that for  $y \in U$  the points  $y, g'(y), \dots, g^{d-1}(y)$  lie in the sets  $U, gU, \dots, g^{d-1}U$  respectively, so  $\psi_{ji}(g^{lk}(y)) = 0$  unless  $i = k$  and hence  $g^{lk}(y)$  does not depend on the coefficient  $a_{ij}$  if  $i \geq k$ .

Define the function  $F: \mathbb{R}^{dm} \times B_1 \times S^{m-1} \rightarrow \mathbb{R}^d$  by

$$F(a, x, v) = D(\Phi_{\tau', X} \circ h^{-1})(x)v$$

where  $S^{m-1} = \{v \in \mathbb{R}^m : \|v\| = 1\}$  is the  $m-1$  dimensional sphere. (Then  $D\Phi_{\tau', X}(y)$  is full rank precisely when  $F(a, hy, \cdot): S^{m-1} \rightarrow \mathbb{R}^d$  does not contain 0 in its image.) To investigate the rank of the Jacobian matrix  $DF(a, x, v)$  we concentrate on that part of it that corresponds to partial differentiation with respect to the coefficients  $a_{ij}$ . Say  $F_k$  is the  $k$ -th component of  $F$ , then

$$\begin{aligned} F_k(a, x, v) &= \sum_{i=1}^m \left[ D(\Phi_{\tau', X} \circ h^{-1})(x) \right]_{ki} v_i = \sum_{i=1}^m D_i(\tau' \circ g^{lk-1} \circ h^{-1})(x)v_i \\ &= \sum_{i=1}^m D_i(\tau \circ g^{lk-1} \circ h^{-1})(x)v_i + \sum_{i=1}^m \sum_{p=1}^m \sum_{q=0}^{d-1} a_{qp} D_i(\psi_{pq} \circ g^{lk-1} \circ h^{-1})(x)v_i \end{aligned}$$

We saw in the previous paragraph that  $\psi_{pq} \circ g^{lk-1} \circ h^{-1}(x) = 0$  for  $x \in B_1$  if  $q \neq k-1$ , and so  $D_i(\psi_{pq} \circ g^{lk-1} \circ h^{-1})(x) = 0$  in this case. Hence

$$F_k(a, x, v) = \sum_{i=1}^m D_i(\tau \circ g^{lk-1} \circ h^{-1})(x)v_i + \sum_{i=1}^m \sum_{p=1}^m a_{k-1,p} D_i(\psi_{p,k-1} \circ g^{lk-1} \circ h^{-1})(x)v_i$$

Now consider  $\partial F_k / \partial a_{rs}$  where  $r > k-1$ . Since  $g^{lk-1}$  (and hence  $\tau \circ g^{lk-1} \circ h^{-1}$  and  $D_i(\psi_{pq} \circ g^{lk-1} \circ h^{-1})$ ) have no dependence on  $a_{rs}$  we see that  $\partial F_k / \partial a_{rs} = 0$ . For the case  $r = k-1$ ,  $g^{lk-1}$  still has no dependence on  $a_{rs}$ , and so

$$\frac{\partial F_k}{\partial a_{k-1,s}} = \sum_{i=1}^m D_i(\psi_{s,k-1} \circ g^{lk-1} \circ h^{-1})(x)v_i$$

Taking  $a = 0$ , so that  $g'$  becomes  $g$ , and taking  $x \in \overline{B}_{1/2}$  we have

$$D_i(\psi_{s,k-1} \circ g^{lk-1} \circ h^{-1})(x) = D_i u_s \circ h^{-1}(x) = \delta_{is}$$

hence  $\partial F_k / \partial a_{k-1,s}(0, x, v) = v_s$ . The first  $dm$  columns of the matrix  $DF(0, x, v)$  take the form

$$\begin{bmatrix} v^T & 0 & 0 & \dots & 0 \\ * & v^T & 0 & \dots & 0 \\ * & * & v^T & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ * & * & * & \dots & v^T \end{bmatrix}$$

where  $v^T$  and 0 are (row)  $m$ -vectors, and  $*$  stands for elements about which we have no information. Since  $v$  cannot be 0,  $DF(0, x, v)$  has at least  $d$  linearly independent columns and so is full rank.

$F$  is  $C^1$ , so the fact that  $DF$  is full rank everywhere on the compact set  $\{0\} \times \overline{B}_{1/2} \times S^{m-1}$  implies it is full rank on a neighbourhood of this set; in particular there is some  $\epsilon > 0$  such that  $DF$  is full rank on all points of the manifold  $B_\epsilon \times B_{1/2} \times S^{m-1}$ .

The argument now follows the standard pattern. The function  $F|_{B_\epsilon \times B_{1/2} \times S^{m-1}}$  is transversal to  $0 \in \mathbb{R}^d$  and so the inverse image of  $\{0\}$  is a co-dimension  $d$  (hence dimension  $(d+1)(m-1)+m$ ) submanifold. The projection of this set into  $\mathbb{R}^{dm}$  is precisely the set of coefficient vectors  $a$  for which  $\Phi_{\tau',X} \circ h^{-1}$  fails to be immersive on  $B_{1/2}$ . Projection is differentiable, so the projected set has dense complement in  $\mathbb{R}^{dm}$  if  $F^{-1}\{0\}$  has dimension smaller than  $dm$ , that is, if  $d \geq 2m$ . In that case we can clearly find  $\tau'$  arbitrarily close to the unperturbed  $\tau$  such that  $\Phi_{\tau',X}$  is immersive at all points of  $h^{-1}B_{1/2}$ , and in particular, all points of  $W$ .

This deals with  $W_1$ . The argument for  $W_2$  etc. is the same; the only extra point is that the perturbations for  $W_2$  will have support in  $U_2 \cup g'U_2 \cup \dots \cup g'^{d-1}U_2$  where  $g'$  is the new diffeomorphism constructed when we dealt with  $W_1$ , and we need to check that these sets are disjoint. Recall that  $U_2$  was selected so that  $\overline{U_2}, g\overline{U_2}, \dots, g^{d-1}\overline{U_2}$  were disjoint, and since these sets are compact the disjointness will be preserved by all sufficiently small changes to  $g$ .

### Making $\Phi_{\tau,X}$ injective: pairs $(x, y)$ where $y$ is a low-period point

We have now constructed a function  $\tau$  (again dropping the primes for convenience) such that  $\Phi_{\tau,X}$  is an immersion of  $M$ , and injective when restricted to  $V$ . The task now is to extend injectivity to the whole of  $M$ . This is done in several stages. The initial stages deal as usual with those points that present particular difficulties: pairs  $x, y \in M$  for which some of the components of  $\Phi_{\tau,X}(x)$  and  $\Phi_{\tau,X}(y)$  are necessarily the same, either because  $x$  or  $y$  is periodic, or because  $y = g^j(x)$  for some  $j$ . Once these points have been disposed of the rest of  $M$  can be dealt with fairly easily. To facilitate the discussion we introduce the function  $\Psi: M \times M \rightarrow \mathbb{R}^d$ , defined by  $\Psi(x, y) = \Phi_{\tau,X}(x) - \Phi_{\tau,X}(y)$ ;  $\Phi_{\tau,X}$  is then injective if  $\Psi$  has the property:  $\Psi(x, y) = 0$  implies  $x = y$ .  $\Psi$  already has this property when restricted to certain subsets of  $M \times M$ . In particular, since  $\Phi_{\tau,X}$  is locally an embedding, there is a neighbourhood  $N$  of  $\Delta = \{(x, x) | x \in M\}$ , the diagonal of  $M \times M$ , such that  $\Psi|_N(x, y) = 0 \Rightarrow x = y$ .

The first stage in creating an injection is to arrange for the low period points not to share an image under  $\Phi_{\tau,X}$  with any other point in  $M$ ; that is,  $[\Psi|_{M \times P}](x, y) = 0 \Rightarrow x = y$ . Since  $\Phi_{\tau,X}$  is an embedding of  $V$ , we already know that  $[\Psi|_{V \times P}](x, y) = 0 \Rightarrow x = y$ . Let  $M' = M \setminus \text{int}(V)$ ; since  $M' \cap P = \emptyset$  we need to ensure that  $\Psi(M' \times P) \not\ni 0$ . Find a good cover for  $M'$ . For each of the sets  $W_1, W_2, \dots, W_K$  in this cover we arrange for  $\Psi(W_i \times P) \not\ni 0$ . We show how to do this for  $W_1$  (the other sets are similar of course), and for convenience we drop the '1' subscript (from  $W_1, U_1$  and  $h_1$ ). Let  $\psi_j: M \rightarrow \mathbb{R}$  be given, for each  $0 \leq j \leq d-1$ , by

$$\psi_j(y) = \begin{cases} \lambda \circ h(g^{-j}y) & \text{for } y \in g^jU \\ 0 & \text{for } y \notin g^jU \end{cases} \quad (13)$$

and for each  $a \in \mathbb{R}^d$  let  $\tau_a = \tau + \sum_{j=0}^{d-1} a_j \psi_j$ . As usual we can consider  $a$  sufficiently small that the corresponding  $g_a$  is a diffeomorphism, and  $g_a^i U = g^i U$  for  $i = 0, \dots, d-1$ . Consider  $F: \mathbb{R}^d \times B_1 \times P$ , defined by  $F(a, x, p) = \Phi_{\tau_a, X}(h^{-1}x) - \Phi_{\tau_a, X}(p)$ . The  $k$ -th component is  $F_k(a, x, p) = \tau_a(g_a^{k-1}y) - \tau_a(g_a^{k-1}p)$  where  $y = h^{-1}x$ . Since  $\tau_a(p') = \tau(p')$  for all  $p' \in P$  we see that  $\tau_a(g_a^{k-1}p) = \tau(g^{k-1}p)$ . Further, since  $y, g_a y, \dots, g_a^{d-1}y$  lie in  $U, gU, \dots, g^{d-1}U$  respectively,  $\tau_a(g_a^l y)$  depends on  $a_m$  only for  $m \leq l$ , and so  $g_a^l y$  depends on  $a_m$  only for  $m \leq l-1$ ; thus  $g_a^l y$  does not depend on  $a_m$  if  $m \geq l$ .

Now

$$\begin{aligned} F_k(a, x, p) &= \tau(g_a^{k-1}y) + \sum_{j=0}^{d-1} a_j \psi_j(g_a^{k-1}y) - \tau(g^{k-1}p) \\ &= \tau(g_a^{k-1}y) + a_{k-1} \psi_{k-1}(g_a^{k-1}y) - \tau(g^{k-1}p) \end{aligned}$$

For  $m \geq k-1$ ,  $\partial F_k / \partial a_m(a, x, p) = \delta_{m, k-1} \psi_{k-1}(g_a^{k-1}y)$ , so if  $x \in \overline{B}_{1/2}$  then  $\partial F_k / \partial a_m(0, x, p) = \delta_{m, k-1}$ . Thus the first  $d$  columns of the matrix  $DF(0, x, p)$  form a lower triangular matrix whose diagonal elements are all 1, and so  $DF(0, x, p)$  is full rank. The argument now follows the same course as before: the fact that  $DF$  is full rank on the compact set  $\{0\} \times \overline{B}_{1/2} \times P$  means there is some  $\epsilon > 0$  such that  $DF$  is full rank on  $B_\epsilon \times B_{1/2} \times P$ . The inverse image of  $0 \in \mathbb{R}^d$  under the restriction of  $F$  to this latter set is a codim  $d$ , hence dim  $m$ , submanifold; its projection into  $\mathbb{R}^d$  has dense complement if  $d > m$ . But this projection is the set of  $a \in \mathbb{R}^d$  such that  $\Psi(h^{-1}B_{1/2} \times P) \ni 0$ , and so for any  $a$  in the complement  $\Psi(W \times P) \not\ni 0$ . Repeating this for all the sets in the good cover we find  $\Psi(W_1 \cup W_2 \cup \dots \cup W_K \times P) \supset \Psi(M' \times P) \not\ni 0$ . Hence if  $x \in M$ ,  $y \in P$  and  $\Psi(x, y) = 0$  then  $x \in V$  and  $x = y$ , so  $\Psi$  has the desired property on  $M \times P$ .

In fact, without making more adjustments to  $\tau$  we can say that there is a neighbourhood  $L$  of  $P$ , which we may as well take to be compact, such that  $[\Psi|_{M \times L}](x, y) = 0 \Rightarrow x = y$ .

### Making $\Phi_{\tau, X}$ injective: pairs $(x, y)$ where $y = g^j x$

Obviously, for  $0 \leq j \leq d-1$  the  $j+1$ -th component of  $\Phi_{\tau, X}(x)$  is identical to the first component of  $\Phi_{\tau, X}(g^j x)$ ; so pairs of the form  $(x, g^j x)$  need special treatment. This is the situation we deal with next: we want to ensure that if  $\Phi_{\tau, X}(x) = \Phi_{\tau, X}(g^j x)$  then  $x = g^j x$ .

To arrange this we shall need  $P$  to contain points of period  $2d-2$  or less (that is, we take  $l_p = 2d-2$ ). Certainly there is a compact neighbourhood  $N$  of  $P$  such that  $N \cup gN \cup \dots \cup g^{d-1}N \subset V$ .  $\Phi_{\tau, X}$  is injective on  $V$ , so for every  $x \in N$ ,  $x$  and  $g^j(x)$  (where  $1 \leq j \leq d-1$ ) have the same image under  $\Phi_{\tau, X}$  only if  $g^j(x) = x$ , (which means that  $x \in P$ ). We now extend this property to the (closed) set  $M'' = M \setminus \text{int}(N)$ .

For every point  $x \in M''$  the points  $x, g(x), \dots, g^{2d-2}(x)$  are distinct. We perform the standard good cover construction on  $M''$  with the small modification that we make sure that the sets  $\overline{U}_x, g\overline{U}_x, \dots, g^{2d-2}\overline{U}_x$  are all disjoint, rather than just the first  $d$  of these. The resulting sets  $W_1, \dots, W_K$  are dealt with in turn, and for each of these sets the values of  $j$ ,  $1 \leq j \leq d-1$  are dealt with in turn: for each  $j$  a perturbation is made (with support in  $U_i \cup gU_i \cup \dots \cup g^{2d-2}U_i$ ) which establishes  $\Phi_{\tau, X}(x) \neq \Phi_{\tau, X}(g^j(x))$  for every point  $x$  in  $W_i$ ; since this set is compact, subsequent perturbations will preserve this property.

Consider then the set  $U_1$ ; similar arguments will apply to all the other  $U_i$ 's and so we will drop the '1' subscript. Take  $j$  to be some fixed number in the range  $1 \leq j \leq d-1$ . Define  $\psi_i: M \rightarrow \mathbb{R}$  ( $0 \leq i \leq d-1$ ) by

$$\psi_i(x) = \begin{cases} \lambda \circ h(g^{-(j+i)}x) & \text{for } x \in g^{j+i}U \\ 0 & \text{for } x \notin g^{j+i}U \end{cases}$$

where  $\lambda$  is the usual bump function. Now define a perturbed measurement function  $\tau'$  by

$$\tau' = \tau + \sum_{i=0}^{d-1} a_i \psi_i$$

with corresponding mapping  $g'$ . As always, we are only interested in cases where the coefficients are small enough that  $g'$  is a diffeomorphism: letting  $a$  be the vector  $(a_0, \dots, a_{d-1})$ , there is some  $\delta > 0$  such that if  $a \in B_\delta \subset \mathbb{R}^d$  then  $g'$  is a diffeomorphism. Indeed, for sufficiently small  $a$  we must have  $g'^i U \subseteq g^i U$  for  $i = 1, \dots, 2d-2$ . Then for  $x \in U$  the points  $x, g'(x), \dots, g'^{2d-2}(x)$  lie in the sets  $U, gU, \dots, g^{2d-2}U$  respectively, so  $\psi_i(g'^k(x)) = 0$  unless  $j+i=k$  and hence  $g'^k(x)$  does not depend on the coefficient  $a_i$  if  $j+i \geq k$ .

Define  $F: \mathbb{R}^d \times B_1 \rightarrow \mathbb{R}^d$  by  $F(a, x) = \Phi_{\tau', X}(y) - \Phi_{\tau', X}(g'^j y)$  where  $y = h^{-1}x$ . Then in particular the  $k$ -th component of  $F$  is given by

$$\begin{aligned} F_k(a, x) &= \tau'(g'^{(k-1)}y) - \tau'(g'^{(k-1+j)}y) \\ &= \tau(g'^{(k-1)}y) + \sum_{i=0}^{d-1} a_i \psi_i(g'^{(k-1)}y) - \tau(g'^{(k-1+j)}y) - \sum_{i=0}^{d-1} a_i \psi_i(g'^{(k-1+j)}y) \\ &= \begin{cases} \tau(g'^{(k-1)}y) - \tau(g'^{(k-1+j)}y) - a_{k-1} \psi_{k-1}(g'^{(k-1+j)}y) & \text{if } k < j+1 \\ \tau(g'^{(k-1)}y) + a_{k-1-j} \psi_{k-1-j}(g'^{(k-1)}y) \\ \quad - \tau(g'^{(k-1+j)}y) - a_{k-1} \psi_{k-1}(g'^{(k-1+j)}y) & \text{if } k \geq j+1 \end{cases} \end{aligned}$$

If  $r \geq k-1$  then obviously  $r > k-1-j$  so neither  $g'^{(k-1+j)}y$  nor  $g'^{(k-1)}y$  depends on  $a_r$ , and from the above expressions we see that  $\frac{\partial F_k}{\partial a_r}(a, x) = -\delta_{r, k-1} \psi_{k-1}(g'^{(k-1+j)}y)$ ; and in particular if  $x \in \overline{B}_{1/2}$  then  $\frac{\partial F_k}{\partial a_r}(0, x) = -\delta_{r, k-1}$ . Hence the first  $d$  columns of  $DF(0, x)$  (where  $x \in \overline{B}_{1/2}$ ) form a lower triangular  $d \times d$  matrix whose diagonal elements are all  $-1$ .  $DF$  is thus full rank on the compact set  $\{0\} \times \overline{B}_{1/2}$ , and hence on some neighbourhood of this set: in particular there must be some  $\epsilon > 0$  such that  $DF$  is full rank on  $B_\epsilon \times B_{1/2}$ . Now we use the standard argument:  $F|_{B_\epsilon \times B_{1/2}}$  is transversal to  $0$ , so  $F^{-1}(0)$  is a codim  $d$ , (i.e. dimension  $d+m-d=m$ ) submanifold of  $B_\epsilon \times B_{1/2}$ , and the projection of this set onto its first  $d$  components has dense complement in  $\mathbb{R}^d$  so long as  $d > m$ . But this projected set is precisely the set of  $a$  vectors such that there is  $x \in B_{1/2}$  for which  $\Phi_{\tau', X}(y) = \Phi_{\tau', X}(g'^j y)$ . Thus we can find  $\tau'$  arbitrarily close to the unperturbed  $\tau$  such that  $\Phi_{\tau', X}(y) \neq \Phi_{\tau', X}(g'^j y)$  at all points of  $h^{-1}B_{1/2}$ , and in particular, all points of  $W$ .

We have now established that  $\Psi(x, g^j x) = 0$  implies  $x = g^j x$  for all  $x \in M$ , and  $j$  in the range  $1 \leq j \leq d-1$ . It is clear that we can immediately extend the range of  $j$  to include  $-(d-1) \leq j \leq -1$ , (we need only let  $y = g^j x$  and rewrite the implication in terms of  $y$ ). Letting  $T_j \subset M \times M$  be the set  $\{(x, g^j x) | x \in M\}$ , and  $T = \bigcup_{j=-(d-1)}^{d-1} T_j$  we have that if  $(x, y) \in T$  and  $\Psi(x, y) = 0$  then  $x = y$ . Furthermore there is a neighbourhood  $S$  of  $T$  (which we may as well take to be compact) such that  $\Psi|_S(x, y) = 0 \Rightarrow x = y$ .

### Making $\Phi_{\tau, X}$ injective on the whole of $M$

Having established the property we want on  $M \times L$  and on  $S$  we have dealt with all the problematic points; it is now straightforward to adjust  $\tau$  so that  $\Phi_{\tau, X}$  becomes an injection on  $M$ . Recall that we wish to arrange for  $\Psi(x, y) = 0 \Rightarrow x = y$  for all  $(x, y) \in M \times M$ . Since we already know that the implication holds for all pairs in



$L \times M$  we need only arrange it for pairs in  $(M \setminus L) \times M$ . So consider the (compact) set  $M' = M \setminus \text{int}(L)$ ; we construct a good cover by the usual construction, with the following modifications: in choosing  $U_x$  we arrange for  $g^j \overline{U}_x$ ,  $-(d-1) \leq j \leq d-1$  to be disjoint (this is possible since  $x \notin P$  and so the points  $g^{-(d-1)}x, \dots, g^{d-1}x$  are all distinct); further we arrange for the sets  $\overline{U}_x \times g^j \overline{U}_x$ ,  $-(d-1) \leq j \leq d-1$ , all to be subsets of  $S$ , (this is possible since all the pairs  $(x, g^j x)$  lie in  $T$ .) As usual, the interiors of the compact sets  $W_1, \dots, W_K$  cover  $M'$ . Each  $W_i$  is considered in turn, and we arrange for  $\Psi$  to have the desired property on  $W_1 \times M, W_2 \times M$  etc.; once we have done this for all  $K$  sets the property holds on all of  $M \times M$ . So consider first  $W_1$  (and the corresponding open set  $U_1$ ), and drop the subscripts since the same argument applies to each of the other sets. We shall perturb  $\tau$  so that the desired property holds on  $W \times M$ . We already know that it holds on  $W \times (g^{-(d-1)}\overline{U} \cup \dots \cup g^{d-1}\overline{U})$  since this is a subset of  $\overline{U} \times (g^{-(d-1)}\overline{U} \cup \dots \cup g^{d-1}\overline{U})$  and hence of  $S$ ; thus it is sufficient to arrange for it to hold on  $W \times M \setminus (g^{-(d-1)}\overline{U} \cup \dots \cup g^{d-1}\overline{U})$ .

The argument is now very similar to that used for  $M \times P$  above. Let  $\psi_j: M \rightarrow \mathbb{R}$  be given, for each  $0 \leq j \leq d-1$ , by equation (13), and for each  $a \in \mathbb{R}^d$  let  $\tau_a = \tau + \sum_{j=0}^{d-1} a_j \psi_j$ . As usual we consider  $a$  sufficiently small that the corresponding  $g_a$  is a diffeomorphism, and  $g_a^i U = g^i U$  for  $i = 0, \dots, d-1$ . Consider  $F: \mathbb{R}^d \times B_1 \times M \setminus (g^{-(d-1)}U \cup \dots \cup g^{d-1}U)$ , defined by  $F(a, x, z) = \Phi_{\tau_a, X}(h^{-1}x) - \Phi_{\tau_a, X}(z)$ . The  $k$ -th component is  $F_k(a, x, z) = \tau_a(g_a^{k-1}y) - \tau_a(g_a^{k-1}z)$  where  $y = h^{-1}x$ . Since  $z$  lies outside  $g^{-(d-1)}U \cup \dots \cup g^{d-1}U$  we see that  $\tau_a(g_a^{k-1}z) = \tau(g^{k-1}z)$  for  $k = 1, \dots, d$ , and so the second term of  $F_k(a, x, z)$  is independent of  $a$ . By the same argument as follows equation (13) we see that  $DF$  is full rank on  $\{0\} \times \overline{B}_{1/2} \times M \setminus (g^{-(d-1)}U \cup \dots \cup g^{d-1}U)$ , and hence, for some  $\epsilon > 0$ , on  $B_\epsilon \times B_{1/2} \times M \setminus (g^{-(d-1)}\overline{U} \cup \dots \cup g^{d-1}\overline{U})$ . And following the argument further we see that the  $a$  vectors for which  $\Psi$  has the desired property on  $h^{-1}B_{1/2} \times M \setminus (g^{-(d-1)}\overline{U} \cup \dots \cup g^{d-1}\overline{U})$  are dense in  $B_\epsilon$ , so long as  $d > 2m$ . Since we already know the property holds on  $h^{-1}B_{1/2} \times (g^{-(d-1)}\overline{U} \cup \dots \cup g^{d-1}\overline{U})$ , it holds on  $h^{-1}B_{1/2} \times M$ , hence, of course, on the compact set  $W \times M$ . Dealing with the other  $W$  sets in the same way, we finally create a  $\tau$  such that  $\Phi_{\tau, X}$  is injective on  $M$ . This completes the proof.

## 7.4 Proof of Theorem 5: State Dependent Sampling Embedding Theorem for Integrate-and-Fire Systems

Recall that we specify an integrate-and-fire system by a pair  $(v, X) \in C^2(M, \mathbb{R}^+) \times \mathcal{X}^2(M)$ , where  $C^2(M, \mathbb{R}^+) = \{v \in C^2(M, \mathbb{R}) \mid v(x) > 0 \forall x \in M\}$ . Each pair gives rise to a sampling interval function  $\tau$  through the implicit relation equation (3), where  $\phi$  is the flow arising from  $X$ . (We could have included the threshold  $\Theta$  along with  $v$  and  $X$  in a triple that specified the integrate and fire system; the reason we have not done so is that in proving our embedding theorems we shall have no occasion to perturb  $\Theta$ . Thus we consider  $\Theta$  chosen in advance, and the integrate and fire systems referred to in the theorems below use this  $\Theta$  for the threshold; the value chosen for  $\Theta$  is arbitrary.) We have already seen that the interspike interval system is a member of  $\mathcal{T}$ .

For the state dependent sampling embedding theorem we shall also of course introduce a measurement function  $y \in C^2(M, \mathbb{R})$  to produce a triple  $(y, v, X)$ . The corresponding delay map  $\Phi_{y, \tau, X}: M \rightarrow \mathbb{R}^d$  is defined as in equation (7). The theorem we want is then

**Theorem 5** For  $d > 2m$  the triples  $(y, v, X)$  for which  $\Phi_{y,\tau,X}$  is an embedding of  $M$  form an open and dense subset of  $C^2(M, \mathbb{R}) \times C^2(M, \mathbb{R}^+) \times \mathcal{X}^2(M)$ .

The openness part of this theorem is proved in much the same way as in Theorem 3: it follows from the observation that  $\Phi_{y,\tau,X}$  depends continuously on  $(y, v, X)$ . The denseness part, again as for Theorem 3, is essentially an application of Theorem 2, and requires an analogue for integrate-and-fire systems of Lemma 4: the lemma we use is:

**Lemma 5** Let  $l_p$  be a given positive integer. Let  $\mathcal{S} \subset C^2(M, \mathbb{R}^+) \times \mathcal{X}^2(M)$  be the set of pairs  $(v, X)$  satisfying the following conditions:

1. The only periodic points of  $g_{\tau,X}$  having period  $l_p$  or less are fixed points; these are finite in number, and hyperbolic.
2. For each of these fixed points the characteristic values are all distinct.

Then  $\mathcal{S}$  is open and dense in  $C^2(M, \mathbb{R}^+) \times \mathcal{X}^2$ .

**Proof:** The openness part of this lemma is proved in the same way as in Lemma 4: it follows from the observation that  $g_{\tau,X}$  depends continuously on  $(v, X)$ .

The denseness part is rather different to that in Lemma 4. Given a pair  $(v, X)$ , let  $\mu$  be the minimum value  $v$  takes on  $M$ , and note from (3) that  $\tau(x) \leq \Theta/\mu$ . We can find, arbitrarily close to  $X$ , another vector field  $X'$  whose periodic trajectories having period less than or equal to  $l_p\Theta/\mu$  are finite in number, are hyperbolic and have distinct eigenvalues. Any periodic point of  $g_{\tau,X'}$  with period  $l_p$  or less must lie on one of these periodic trajectories. That is, if  $g_{\tau,X'}^r(x) = x$  for some  $x$  then there is a low period trajectory  $C$  of  $X'$  such that  $x \in C$ ; furthermore if  $T$  is the period of  $C$  then  $\sigma_\tau(x) = nT$  for some positive integer  $n$ , (recall  $\sigma_\tau$  is defined in equation (6)), and

$$r\Theta = \int_0^{nT} v(\phi(x, t))dt = n \int_0^T v(\phi(x, t))dt$$

Note this means every point on  $C$  is a period  $r$  point of  $g_{\tau,X'}$ . Make a new integrand  $v'$  by multiplying  $v$  by a real-valued function taking the values 1 outside a neighbourhood of  $C$  and  $\alpha \neq 1$  on  $C$ ; the above equality will not be satisfied if  $v'$  replaces  $v$ . For  $\alpha$  sufficiently close to 1, no condition of the form  $q\Theta = n' \int_0^T v'(\phi(x, t))dt$  (where  $n'$  is a positive integer and  $q$  is an integer in the range  $1 \leq q \leq l_p$ ) will be satisfied. Hence there are now no low period points on  $C$ . We can treat the other low period trajectories of the vector field in the same way, with the exception of equilibrium points, and so generate a new vector field, say  $\tilde{X}$ , for which  $g_{\tau,\tilde{X}}$  has no low period points, except for fixed points arising from equilibrium points of  $\tilde{X}$ . By our original choice of  $X'$  these fixed points are hyperbolic and have distinct eigenvalues. This completes the denseness part of the lemma.

We can now complete the proof of Theorem 5. Given any pair  $(v, X) \in \mathcal{S}$  (where  $\mathcal{S}$  is as in Lemma 5 with  $l_p = d - 1$  and  $d > 2m$ ) by Theorem 2 there is a dense subset of  $C^2(M, \mathbb{R})$  such that, for all  $y$  in this subset,  $\Phi_{y,\tau,X}$  is an embedding of  $M$ . Since  $\mathcal{S}$  is dense in  $C^2(M, \mathbb{R}^+) \times \mathcal{X}^2(M)$  by Lemma 5, the set of triples  $(y, v, X)$  for which  $\Phi_{y,\tau,X}$  is an embedding, is dense in  $C^2(M, \mathbb{R}) \times C^2(M, \mathbb{R}^+) \times \mathcal{X}^2(M)$ .

## 7.5 Proof of Theorem 6: Interspike Interval Embedding Theorem for Integrate-and-Fire Systems

To find a theorem for integrate and fire systems analogous to Theorem 4 we consider delay maps where the sampling interval function  $\tau$  also supplies the measurement function—but where  $\tau$  is derived from  $v$  through relation (3). The delay map  $\Phi_{\tau,X}$  is defined as in equation (8); the theorem we want is then:

**Theorem 6** *For  $d > 2m$  the pairs  $(v, X) \in C^2(M, \mathbb{R}^+) \times \mathcal{X}^2$ , for which  $\Phi_{\tau,X}$  is an embedding of  $M$ , form an open and dense subset of  $C^2(M, \mathbb{R}^+) \times \mathcal{X}^2(M)$ .*

As with Theorem 4 we will have to prove this from scratch; we use exactly the same strategy that we used for Theorem 4, so rather than go through this proof again here, we will just set out the modifications that must be made to the proof of Theorem 4; mainly these modifications arise from the fact that now we are perturbing  $v$  rather than  $\tau$  (directly).

The openness part is as usual straightforward, and follows from the observation that  $\Phi_{\tau,X}$  depends continuously on  $(v, X)$ . The proof of the denseness part begins by using Lemma 5 to show that arbitrarily close to any  $(v, X)$  is a pair  $(v, \tilde{X})$  for which the corresponding  $g$  has no periodic points of period  $l_p$  or less, except for fixed points arising from the equilibrium points of  $\tilde{X}$  (and these fixed points are hyperbolic and have distinct eigenvalues). The vector field is fixed from now on (and renamed  $X$ ).

Following the pattern of the proof of Theorem 4, we now ensure that  $\Phi_{\tau,X}$  is injective on the low period points—for which it is sufficient to ensure  $\tau$  takes on distinct values at these points. Since such a point (say  $x_0$ ) is in fact an equilibrium point of  $X$  relation (3) gives  $\tau(x_0) = \Theta/v(x_0)$ , so it is enough that  $v$  take distinct values at these points; and since the points are finite in number this is easy to arrange by perturbations local to each point.

The next step is to make  $\Phi_{\tau,X}$  immersive at the low period points. Here again matters are simplified by the fact that we are only concerned with equilibrium points. The same argument as used for Theorem 4 shows here that sufficient conditions for  $D\Phi_{\tau,X}(x_0)$  to be full rank are that  $J_{\tau(x_0)}(x_0)$  should have distinct eigenvalues (which has already been arranged by our choice of  $X$ ), and that  $D\tau(x_0)$  should belong to an open dense set depending on  $J_{\tau(x_0)}(x_0)$ . Thus we need only confirm that we can make arbitrary perturbations to  $D\tau$  by making suitable changes to  $v$ . A straightforward calculation beginning with (3) shows that

$$D\tau(x_0) = -\frac{Dv(x_0)}{v(x_0)} \left( \exp \left[ \frac{\Theta}{v(x_0)} DX(x_0) \right] - I \right)$$

and since  $x_0$  is a hyperbolic equilibrium point of  $X$  the matrix in parentheses on the right hand side of this expression is full rank. Thus we can make arbitrary changes to  $D\tau(x_0)$  if we can make arbitrary changes to  $Dv(x_0)$  without altering  $v(x_0)$ : we can do this by making perturbations of the form (9), (with  $v$  replacing  $\tau$ ;  $p = 1$  in this case). As in Theorem 4 we can conclude that there is a compact neighbourhood  $V$  of the set of low period points such that  $\Phi_{\tau,X}|_V$  is an injective immersion, and delay maps derived from all  $v'$  sufficiently close to  $v$  are also injective immersions of  $V$ .

The remainder of the proof relies on the construction of ‘good covers’—and perturbation functions to put on them—for certain subsets of  $M$ . We use perturbation

functions to establish the properties we want for our integrate and fire systems somewhat in the way we did for interspike interval systems; there are, however, some significant differences. For interspike interval systems we typically chose an open set  $U$  and considered its images  $g^i U$ ,  $i = 0, \dots, d-1$ , installing a perturbation function  $\psi_i$  on each of these and then forming new sampling interval functions  $\tau_a = \tau + \sum_{i=0}^{d-1} a_i \psi_i$ . For integrate and fire systems the approach is similar, but now we are perturbing  $v$ , so we have  $v_a = v + \sum_{i=0}^{d-1} a_i \psi_i$ . The effect this has on  $g$  is more complicated than perturbing  $\tau$ : in particular, for a point  $x \in g^i U$  the effect on  $g(x)$  of perturbing  $\tau$  is determined solely by  $\psi_i(x)$ , but the effect of perturbing  $v$  can depend upon all the values of the perturbations along the (continuous time) trajectory between  $x$  and  $g(x)$ —and this trajectory may pass through any or all the  $g^i U$  sets. To deal with this we shall use more sophisticated perturbation functions; one of the things we shall arrange is that the integral (w.r.t. time) of each  $\psi_i$  along a trajectory from the point at which it enters  $g^i U$ , to the point at which it leaves, is zero. This will ensure that the only relevant perturbations are those at the beginning and end of the trajectory between  $x$  and  $g(x)$ .

To simplify things further it is convenient to consider a slightly modified delay map. We define

$$\hat{\Phi}_X^v(x) = (\tau(x), \sigma_2(x), \dots, \sigma_d(x)) \quad (14)$$

where  $\sigma_r$  is as in (6); note that for integrate and fire systems  $\sigma_r$  is related to  $v$  by

$$r\Theta = \int_0^{\sigma_r(x)} v(\phi(x, t)) dt \quad (15)$$

Since  $\hat{\Phi}_X^v = A\Phi_{\tau, X}$  where  $A$  is the constant matrix

$$\begin{pmatrix} 1 & 0 & \dots & 0 \\ 1 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \dots & 1 \end{pmatrix}$$

$\hat{\Phi}_X^v$  will be an embedding (or immersion) precisely when  $\Phi_{\tau, X}$  is. Note that the  $r$ -th component of  $\hat{\Phi}_X^v(x)$  is determined by the integral (15) along the trajectory between  $x$  and  $g^r(x)$ ; by using the kinds of perturbation functions described in the last paragraph we ensure that this component is affected only by perturbations in the regions of  $x$  and  $g^r(x)$ . In fact we can make things even simpler by not installing a perturbation function on the set  $U$  containing  $x$ : then only the perturbation in the region of  $g^r(x)$  will be relevant. Thus we typically place perturbation functions on the sets  $g^i U$ ,  $i = 1, \dots, d$  (whereas for the interspike interval case we had  $i = 0, \dots, d-1$ ; this means we need to choose  $l_p$  correspondingly larger than before).

We describe now how to construct the ‘good covers’ (the corresponding perturbation functions are described later.) We take  $l_p = d$  or larger (one greater than we used in Theorem 4), let  $P$  be the set of low period points, and consider a closed subset  $C \subset M$  such that  $C \cap P = \emptyset$ . Let  $x \in C$ , then  $x, gx, \dots, g^d x$  are all distinct, and we can find an open neighbourhood  $U_x$  of  $x$  such that  $\overline{U}_x, g\overline{U}_x, \dots, g^d \overline{U}_x$  are all disjoint,  $\overline{U}_x$  contains no points in  $P$ , and such that  $\overline{U}_x$  is a *flow box* ([11]): that is, there is a chart map  $h_x: O \supset \overline{U}_x \rightarrow \mathbb{R}^m$  such that  $h_x \overline{U}_x = [-\gamma_x, \gamma_x]^m$  (with  $h_x(x) = (0, \dots, 0)$ ), and  $Dh_x(y)X(y) = (1, 0, \dots, 0)^T$  for all  $y \in \overline{U}_x$  (so that the induced vector field in  $[-\gamma_x, \gamma_x]^m$  is a constant). Letting  $W_x = h_x^{-1}[-\gamma_x/4, \gamma_x/4]^m$

we see that  $\{W_x|x \in C\}$  is a collection of compact sets whose interiors cover  $C$ : we select a finite subset,  $W_1, \dots, W_K$  (with corresponding neighbourhoods  $U_1, \dots, U_K$ , and chart maps  $h_1, \dots, h_K$ ), so that the interiors of these sets cover  $C$ ; this is our ‘good cover’ for  $C$ .

We are now ready to move on to the next stage in the proof of Theorem 6, which is to establish an immersion on the whole of  $M$ . We construct a good cover on the closed set  $M' = M \setminus \text{int}(V)$ , and make an immersion on each  $W_i$  in turn; we show how to do this for  $W_1$  (the others are similar) and drop the subscript for convenience.

To define suitable perturbation functions, let  $\mu: (-1, 1) \rightarrow \mathbb{R}$  be a  $C^\infty$  function such that  $0 \leq \mu(y) \leq 1$  for all  $y$ ,  $\mu(y) = 1$  for  $y \in (-1/2, 1/2)$  and  $\mu(y) = 0$  for  $y \notin (-3/4, 3/4)$ . Let  $\lambda: (-1, 1) \rightarrow \mathbb{R}$  be a  $C^\infty$  function such that:  $\lambda(y) = 0$  for  $y \in (-1/2, 1/2)$  and for  $y \notin (-3/4, 3/4)$ ;  $\int_{-1}^1 \lambda(y) dy = 0$ ; and  $\int_0^1 \lambda(y) dy \neq 0$  (which implies that  $\int_t^1 \lambda(y) dy \neq 0$  for every  $t \in (-1/2, 1/2)$ ). Also let  $\lambda_1: (-1, 1) \rightarrow \mathbb{R}$  be  $C^\infty$ , with  $\lambda_1(y) = 1$  for  $y \in (-1/2, 1/2)$ ,  $\lambda_1(y) = 0$  for  $y \notin (-3/4, 3/4)$ ,  $\int_{-1}^1 \lambda_1(y) dy = 0$ , and  $\int_t^1 \lambda_1(y) dy \neq 0$  for every  $t \in (-1/2, 1/2)$ . Let  $x = (x_1, \dots, x_m) \in [-\gamma, \gamma]^m$  and define  $\hat{\mu}(x) = \mu(x_2/\gamma) \times \mu(x_3/\gamma) \times \dots \times \mu(x_m/\gamma)$ . Also define the following  $m + 1$  functions on  $[-\gamma, \gamma]^m$ :

$$\begin{aligned}\chi_1(x) &= \lambda_1(x_1/\gamma) \times \hat{\mu}(x) \\ \chi_j(x) &= x_j \lambda(x_1/\gamma) \times \hat{\mu}(x) \quad \text{for } j = 2, \dots, m \\ \chi_{m+1}(x) &= \lambda(x_1/\gamma) \times \hat{\mu}(x)\end{aligned}$$

The perturbation functions we require are  $\psi_{ji}: M \rightarrow \mathbb{R}$

$$\psi_{ji}(y) = \begin{cases} \chi_j \circ h(g^{-i}y) \times [1 + D\sigma_i(g^{-i}y)X(g^{-i}y)]^{-1} & \text{for } y \in g^i U \\ 0 & \text{for } y \notin g^i U \end{cases} \quad (16)$$

(where  $0 \leq i \leq d$ ,  $1 \leq j \leq m + 1$ ). These functions are arranged so that if the trajectory of  $y$  enters  $g^k \bar{U}$  at time  $t_1$  and subsequently leaves it at  $t_2$  then  $\int_{t_1}^{t_2} \psi_{jk}(\phi(y, t)) dt = 0$  for  $j = 1, \dots, m + 1$ . Thus if  $y \in U$  (so that  $g^i y \in g^i U$ ) then  $\int_0^{\sigma_i(y)} \psi_{jk}(\phi(y, t)) dt$  is nonzero only if  $k = 0$  or  $i$ . Further,  $\int_0^{\sigma_i(y)} \psi_{j0}(\phi(y, t)) dt = \int_0^{t_1} \psi_{j0}(\phi(y, t)) dt$  where  $t_1$  is the time at which the trajectory from  $y$  first leaves  $\bar{U}$ , and  $\int_0^{\sigma_i(y)} \psi_{ji}(\phi(y, t)) dt = \int_{t_2}^{\sigma_i(y)} \psi_{ji}(\phi(y, t)) dt$  where  $t_2$  is the last time before  $\sigma_i(y)$  that the trajectory enters  $g^i \bar{U}$ . Also, if  $h(y) = x = (x_1, \dots, x_m)$  then

$$\int_0^{t_1} \psi_{j0}(\phi(y, t)) dt = \gamma \hat{\mu}(x) \begin{cases} \int_{x_1/\gamma}^1 \lambda_1(s) ds & \text{for } j = 1 \\ x_j \int_{x_1/\gamma}^1 \lambda(s) ds & \text{for } j = 2, \dots, m \\ \int_{x_1/\gamma}^1 \lambda(s) ds & \text{for } j = m + 1 \end{cases} \quad (17)$$

and  $\int_0^{t_1} \psi_{j0}(\phi(y, t)) dt + \int_{t_2}^{\sigma_i(y)} \psi_{ji}(\phi(y, t)) dt = 0$  (which further means that  $\int_0^{\sigma_i(y)} \psi_{j0}(\phi(y, t)) dt + \int_0^{\sigma_i(y)} \psi_{ji}(\phi(y, t)) dt = 0$ ). Now define a new function

$$v' = v + \sum_{i=1}^d \sum_{j=1}^{m+1} a_{ij} \psi_{ji} \quad (18)$$

with corresponding functions  $\tau'$ ,  $\sigma'_i$  and  $g'$  (and let  $a = (a_{11}, a_{12}, \dots, a_{d, m+1})$ ); note that  $i = 0$  is *not* included in the sum. As in the proof of Theorem 4, for sufficiently small  $a_{ij}$ 's,  $g'$  is a diffeomorphism and  $g^i U \subseteq g^i U$  for  $i = 1, \dots, d$ . Define  $F: \mathbb{R}^{(m+1)d} \times (-\gamma, \gamma)^m \times S^{m-1} \rightarrow \mathbb{R}^d$  by  $F(a, x, u) = D(\hat{\Phi}'_X \circ h^{-1})(x)u$ . The  $k$ -th component of  $F$

is  $F_k(a, x, u) = D(\sigma'_k \circ h^{-1})(x)u$ . An expression for  $\partial F_k/\partial a_{rs}$  can be found from (15) and (18): letting  $\hat{v}_k \equiv v' \circ g^{rk} \circ h^{-1}$  and  $G_{ji}(x) = \int_0^{\sigma'_k \circ h^{-1}(x)} \psi_{ji}(\phi(h^{-1}(x), t))dt$  we find by differentiating (15) w.r.t.  $a_{rs}$

$$0 = \hat{v}_k(x) \frac{\partial \sigma'_k \circ h^{-1}}{\partial a_{rs}}(x) + G_{sr}(x) \quad (19)$$

Differentiating this w.r.t.  $x$ , and using (19) to replace  $\partial \sigma'_k \circ h^{-1}(x)/\partial a_{rs}$  gives

$$\hat{v}_k(x)^2 \frac{\partial F_k}{\partial a_{rs}} = G_{sr}(x) D\hat{v}_k(x)u - \hat{v}_k(x) DG_{sr}(x)u$$

From the remarks following (16) we see that (at  $a = 0$ ) for all  $x \in (-\gamma, \gamma)^m$ ,  $G_{sr}(x) = 0$  unless  $r = k$ ; differentiating this gives  $DG_{sr}(x) = 0$ . The relation  $\int_0^{\sigma_k(y)} \psi_{s0}(\phi(y, t))dt + \int_0^{\sigma_k(y)} \psi_{sk}(\phi(y, t))dt = 0$  implies that (at  $a = 0$ ),  $G_{s0}(x) = -G_{sk}(x)$ , and differentiating this gives  $DG_{s0}(x) = -DG_{sk}(x)$ . Thus the first  $(m+1)d$  columns of the matrix  $DF(0, x, u)$  take the form

$$\begin{bmatrix} -w_1^T & 0 & 0 & \dots & 0 \\ 0 & -w_2^T & 0 & \dots & 0 \\ 0 & 0 & -w_3^T & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \dots & -w_d^T \end{bmatrix} \quad (20)$$

This matrix will have full rank so long as none of the vectors  $w_k^T$  is zero. The  $s$ -th component,  $[w_k^T]_s$ , is given by

$$\hat{v}_k(x)^2 [w_k^T]_s = G_{s0}(x) D\hat{v}_k(x)u - \hat{v}_k(x) DG_{s0}(x)u$$

(evaluated at  $a = 0$ ). Consider first  $s = m + 1$ . From (17) we see that  $G_{m+1,0}(x) = \gamma \hat{\mu}(x) \int_{x_1/\gamma}^1 \lambda(t)dt$ , so if  $x \in (-\gamma/2, \gamma/2)^m$  then  $G_{m+1,0}(x) \neq 0$ . On the other hand, (again if  $x \in (-\gamma/2, \gamma/2)^m$ , so that  $\hat{\mu}(x) = 1$ ),  $DG_{m+1,0}(x) = (-\lambda(x_1/\gamma), 0, \dots, 0)^T$ , which is 0 since  $x_1/\gamma \in (-1/2, 1/2)$ . Hence, if  $D\hat{v}_k(x)u \neq 0$ , then  $[w_k^T]_{m+1} \neq 0$ .

Suppose then that  $D\hat{v}_k(x)u = 0$ ; then  $\hat{v}_k(x)[w_k^T]_s = -DG_{s0}(x)u$ . From (17) we see that, if  $x \in (-\gamma/2, \gamma/2)^m$ ,  $D_1 G_{1,0}(x) = -\hat{\mu}(x)\lambda_1(x_1/\gamma) = -1$ , and  $D_s G_{1,0}(x) = 0$  for  $s = 2, \dots, m$ . Hence  $-DG_{1,0}(x)u = u_1$ . For  $s = 2, \dots, m$  we have  $D_1 G_{s0}(x) = -\hat{\mu}(x)x_s \lambda(x_1/\gamma) = 0$ ;  $D_s G_{s0}(x) = \gamma \hat{\mu}(x) \int_{x_1/\gamma}^1 \lambda(t)dt = c$ , say, where  $c \neq 0$ ; and  $D_i G_{s0}(x) = 0$  otherwise. Hence  $-DG_{s0}(x)u = -cu_s$ . Since  $u$  is not the zero vector, not all of the quantities  $-DG_{s0}(x)u$  can be zero, and hence  $[w_k^T]_s \neq 0$  for some  $s$ . Hence  $w_k^T \neq 0$  and  $DF(0, x, u)$  is full rank. The same argument as in Theorem 4 then shows that arbitrarily close to  $v$  there is a  $v'$  such that  $\hat{\Phi}_X^{v'}$  is immersive at all points of  $h^{-1}(-\gamma/2, \gamma/2)$ . We conclude that we can make  $\hat{\Phi}_X^v$  immersive on the whole of  $M$  using arbitrarily small perturbations.

To complete the embedding we must make  $\hat{\Phi}_X^v$  injective. Consider the function  $\Psi: M \times M \rightarrow \mathbb{R}^d$  defined by  $\Psi(x, y) = \hat{\Phi}_X^v(x) - \hat{\Phi}_X^v(y)$ ;  $\hat{\Phi}_X^v$  is injective if  $\Psi(x, y) = 0 \Rightarrow x = y$ . As before we establish this in stages, beginning with  $[\Psi|M \times P](x, y) = 0 \Rightarrow x = y$ . It is already true that  $[\Psi|V \times P](x, y) = 0 \Rightarrow x = y$ , so we need  $\Psi(M' \times P) \not\ni 0$ , where  $M' = M \setminus \text{int}(V)$ . We find a good cover of  $M'$  and arrange for  $\Psi(W_i \times P) \not\ni 0$  for each of the sets in it; we deal with  $W_1$  (and drop the subscripts). The perturbation functions we require are among those we used to make

an immersion: in particular, we set  $\psi_i = \psi_{m+1,i}$  where  $\psi_{m+1,i}$  is as in equation (16). For each  $a \in \mathbb{R}^d$  let  $v' = v + \sum_{i=1}^d a_i \psi_i$  (but we are only interested in  $a$  sufficiently small that  $g'$  is a diffeomorphism, and  $g^i U = g'^i U$  for  $i = 1, \dots, d$ ). Consider  $F: \mathbb{R}^d \times (-\gamma, \gamma)^m \times P \rightarrow \mathbb{R}^d$  defined by  $F(a, x, p) = \hat{\Phi}_X^{v'}(h^{-1}x) - \hat{\Phi}_X^{v'}(p)$ ; the  $k$ -th component is  $F_k(a, x, p) = \sigma'_k(h^{-1}x) - \sigma'_k(p)$ . However,  $\bar{U} \cap P = \emptyset$  so neither  $p$  nor  $g^k p$  lies in one of the sets  $g^i \bar{U}$ ; if the trajectory from  $p$  enters one of these sets before reaching  $g^k p$  it must also leave it before reaching  $g^k p$ , and so from the discussion following (16),  $\sigma'_k(p) = \sigma_k(p)$ . Hence (using (19))

$$\frac{\partial F_k}{\partial a_i}(a, x, p) = \frac{\partial \sigma'_k(h^{-1}x)}{\partial a_i} = -\frac{1}{\hat{v}_k(x)} \int_0^{\sigma'_k(h^{-1}x)} \psi_i(\phi(h^{-1}x, t)) dt \quad (21)$$

Evaluating this at  $a = 0$ , using the discussion after (16), gives:  $\partial F_k / \partial a_i(0, x, p) = 0$  unless  $i = k$ ;  $\partial F_k / \partial a_k(0, x, p) = \gamma(\hat{\mu}(x) / \hat{v}_k(x)) \int_{x_1/\gamma}^1 \lambda(s) ds = c_k(x)$ , say, where  $c_k(x) \neq 0$  if  $x \in (-\gamma/2, \gamma/2)^m$ . Thus the first  $d$  columns of  $DF(0, x, p)$  have the form (20) except that each  $-w_i^T$  is replaced by the single number  $c_k(x)$ . Exactly the same argument as in Theorem 4 now shows that arbitrarily small perturbations lead to  $\Psi(M' \times P) \not\cong 0$ . Indeed, there is a neighbourhood  $L$  of  $P$  such that  $[\Psi|M \times L](x, y) = 0 \Rightarrow x = y$ .

The next stage of the injectivity construction is to deal with pairs of the form  $(y, g^j y)$  where  $1 \leq j \leq d$ . We take  $l_p = 2d$ , and find a compact neighbourhood  $N$  of  $P$  such that  $N \cup gN \cup \dots \cup g^d N \subset V$ ;  $\hat{\Phi}_X^v$  already has the required property on  $N$  so we extend it to  $M'' = M \setminus \text{int}(N)$ . We form a good cover on  $M''$ , now making sure that the sets  $\bar{U}_x, g\bar{U}_x, \dots, g^{2d}\bar{U}_x$  are all disjoint. The resulting sets  $W_1, \dots, W_K$  are dealt with in turn, and for each of these sets the values of  $j$ ,  $1 \leq j \leq d$  are dealt with in turn. Consider  $U_1$  (and drop subscripts as usual); take  $j$  to be some fixed number in the range  $1 \leq j \leq d$ . Set  $\psi_i = \psi_{m+1,i+j}$  where  $\psi_{m+1,i+j}$  is as in equation (16). For each  $a \in \mathbb{R}^d$  let  $v' = v + \sum_{i=1}^d a_i \psi_i$  ( $a$  is sufficiently small that  $g'$  is a diffeomorphism and  $g^i U = g'^i U$  for  $i = 1, \dots, d$ ). Now say  $F: \mathbb{R}^d \times (-\gamma, \gamma)^m \rightarrow \mathbb{R}^d$  is given by  $F(a, x) = \hat{\Phi}_X^{v'}(h^{-1}x) - \hat{\Phi}_X^{v'}(g'^j(h^{-1}x))$  so that  $F_k(a, x) = \sigma'_k(h^{-1}x) - \sigma'_k(g'^j(h^{-1}x))$ . Note that  $\sigma'_{k+j}(h^{-1}x) = \sigma'_j(h^{-1}x) + \sigma'_k(g'^j(h^{-1}x))$  so

$$F_k(a, x) = \sigma'_j(h^{-1}x) + \sigma'_k(h^{-1}x) - \sigma'_{k+j}(h^{-1}x)$$

$\sigma'_j(h^{-1}x)$  does not depend on  $a$ , since there is no perturbation on  $g^j U$ . From (21)

$$\left. \frac{\partial \sigma'_k(h^{-1}x)}{\partial a_i} \right|_{a=0} = -\frac{1}{\hat{v}_k(x)} \int_0^{\sigma_k(h^{-1}x)} \psi_{m+1,i+j}(\phi(h^{-1}x, t)) dt$$

and this is zero, unless  $i = k - j$ ; otherwise it is  $c_k(x)$  (this quantity is described just after (21)) and is non-zero for  $x \in (-\gamma/2, \gamma/2)^m$ . Similarly  $\partial \sigma'_{k+j}(h^{-1}x) / \partial a_i|_{a=0}$  is zero unless  $i = k$ , when it is  $c_{k+j}(x)$ . Hence the first  $d$  columns of  $DF(0, x)$  form a matrix whose only non-zero entries are on the diagonal, or are  $j$  places to the left of a diagonal element; this matrix is lower triangular, and full rank for  $x \in (-\gamma/2, \gamma/2)^m$ . The same argument as in Theorem 4 now shows we can find arbitrarily small  $a$  such that  $\hat{\Phi}_X^{v'}(h^{-1}x) \neq \hat{\Phi}_X^{v'}(g'^j(h^{-1}x))$  for all  $x \in W$ , which we extend to  $x \in M''$  by treating each  $W_i$  in turn. Following the argument of Theorem 4 we conclude that there is a neighbourhood  $S$  of the set  $T = \cup_{j=-d}^d \{(x, g^j x) | x \in M\}$  such that  $\Psi|_S(x, y) = 0 \Rightarrow x = y$ .

The final stage of the construction closely parallels the corresponding stage in Theorem 4. As there, we need to establish the desired property of  $\Psi$  on the set

$(M \setminus L) \times M$ , so we consider the set  $M' = M \setminus \text{int}(L)$ , and for each  $x$  in this set find  $U_x$  such that  $g^j \overline{U}_x$ ,  $-d \leq j \leq d$  are all disjoint and so that  $\overline{U}_x \times g^j \overline{U}_x$ ,  $-d \leq j \leq d$  are all subsets of  $S$ ; we extract a good cover in the usual way. We need to perturb  $v$  so the desired property holds on  $W \times M$ ; it already holds on  $W \times (g^{-d} \overline{U} \cup \dots \cup g^d \overline{U})$  so we arrange it for  $W \times M \setminus (g^{-d} U \cup \dots \cup g^d U)$ .

Set  $\psi_i = \psi_{m+1,i}$  where  $\psi_{m+1,i}$  is as in equation (16), and let  $v' = v + \sum_{i=1}^d a_i \psi_i$  ( $a$  sufficiently small that  $g'$  is a diffeomorphism, and  $g^i U = g^i U$  for  $i = 1, \dots, d$ ). Consider  $F: \mathbb{R}^d \times (-\gamma, \gamma)^m \times M \setminus (g^{-d} U \cup \dots \cup g^d U) \rightarrow \mathbb{R}^d$  defined by  $F(a, x, z) = \hat{\Phi}_X^{v'}(h^{-1}x) - \hat{\Phi}_X^{v'}(z)$ , so that  $F_k(a, x, z) = \sigma'_k(h^{-1}x) - \sigma'_k(z)$ . Since  $z$  and  $g^k z$  lie outside  $g^j U$  for  $j = 1, \dots, d$ ,  $\sigma'_k(z)$  does not depend on  $a$ , and  $\partial F_k / \partial a_i = \partial \sigma'_k(h^{-1}x) / \partial a_i$ . From (21) and the discussion following it we get  $\partial F_k / \partial a_i(0, x, z) = 0$  unless  $i = k$ , and  $\partial F_k / \partial a_k(0, x, z) = c_k(x)$ , where  $c_k(x) \neq 0$  if  $x \in (-\gamma/2, \gamma/2)^m$ . Thus the first  $d$  columns of  $DF(0, x, z)$  form a diagonal matrix, which is full rank if  $x \in (-\gamma/2, \gamma/2)^m$ . The argument used for Theorem 4 now applies, and completes the proof for Theorem 6.



# Acknowledgements

The authors have benefited from useful discussions with Mark Muldoon and Jaroslav Stark. This work was partially supported by the Royal Signals and Radar Establishment, Malvern, UK.

# References

- [1] R. Castro and T. Sauer. Chaotic stochastic resonance: Noise-enhanced reconstruction of attractors. *Phys. Rev. Lett.*, 79(6):1030–1033, 1997.
- [2] R. Castro and T. Sauer. Correlation dimension of attractors through interspike intervals. *Phys. Rev. E*, 55(1):287–290, 1997.
- [3] R. Castro and T. Sauer. Reconstructing chaotic dynamics through spike filters. *Phys. Rev. E*, 59(3):2911–2916, 1999.
- [4] M. Ding and W. Yang. Deterministic point processes generated by threshold crossings: Dynamics reconstruction and chaos control. *Phys. Rev. E*, 55(3):2397–2402, 1997.
- [5] T. Gedeon, M. Holzer, and M. Pernarowski. Attractor reconstruction from interspike intervals is incomplete. *Physica D*, 178(3–4):149–172, 2003.
- [6] L. Glass and M. C. Mackey. A simple model for phase locking of biological oscillators. *J. Math. Biol.*, 7:339–352, 1979.
- [7] H. Kantz and T. Schreiber. *Nonlinear Time Series Analysis*. Cambridge University Press, Cambridge, second edition, 2003.
- [8] I. Kupka. Contribution à la théorie des champs génériques. *Contributions to Differential Equations*, 2:457–484, 1963.
- [9] E. Ott, T. Sauer, and J. A. Yorke, editors. *Coping with Chaos*. John Wiley and Sons Inc., New York, 1994.
- [10] N. H. Packard, J. P. Crutchfield, J. D. Farmer, and R. S. Shaw. Geometry from a time series. *Phys. Rev. Lett.*, 45(9):712–716, 1980.
- [11] J. Palis and W. de Melo. *Geometric Theory of Dynamical Systems*. Springer-Verlag, New York, NY, first edition, 1982.
- [12] A. N. Pavlov, O. V. Sosnovtseva, E. Mosekilde, and V. S. Anishchenko. Chaotic dynamics from interspike intervals. *Phys. Rev. E*, 63(3):036205, 2001.
- [13] D. M. Racicot and A. Longtin. Interspike interval attractors from chaotically driven neuron models. *Physica D*, 104:184–204, 1997.
- [14] K. A. Richardson, T. T. Imhoff, P. Grigg, and J. J. Collins. Encoding chaos in neural spike trains. *Phys. Rev. Lett.*, 80(11):2485–2488, 1998.
- [15] O.E. Rössler. Chaotic behaviour in simple reaction systems. *Zeitschrift für Naturforschung Section A*, 31a:259–264, 1976.

- [16] T. Sauer. Reconstruction of dynamical systems from interspike intervals. *Phys. Rev. Lett.*, 72(24):3811–3814, 1994.
- [17] T. Sauer. Reconstruction of integrate-and-fire dynamics. In C. Cutler and D. Kaplan, editors, *Nonlinear Dynamics and Time Series*, volume 11 of *Fields Institute Publications*, pages 63–75. American Mathematical Society, Providence, RI, 1997.
- [18] T. Sauer, J. A. Yorke, and M. Casdagli. Embedology. *J. Stat. Phys.*, 65:579–616, 1991.
- [19] S. Smale. Stable manifolds for differential equations and diffeomorphisms. *Annali della Scuola normale superiore di Pisa. Scienze fisiche e matematiche.*, 18:97–116, 1963.
- [20] J. Stark. Delay embeddings for forced systems: I, deterministic forcing. *J. Nonlinear Sci.*, 9:255–332, 1999.
- [21] J. Stark, D. S. Broomhead, M. E. Davies, and J. Huke. Delay embeddings for forced systems: II, stochastic forcing. *J. Nonlinear Sci.*, 13:519–577, 2003.
- [22] H. Suzuki, K. Aihara, J. Murakami, and T. Shimozawa. Analysis of neural spike trains with interspike interval reconstruction. *Biol. Cybern.*, 82(4):305–311, 2000.
- [23] F. Takens. Detecting strange attractors in turbulence. In David A. Rand and Lai-Sang Young, editors, *Dynamical Systems and Turbulence*, volume 898 of *Lecture Notes in Mathematics*, pages 366–381, Berlin, 1981. Springer-Verlag.