Proof of Takens' Theorem

The goal of this note is to present a sketch of the proof of Takens' theorem [1]. This is heavily based on Huke's technical report [2].

1 Useful results

In this section are stated several basic results of differential topology that will be used in the following sections.
Let M be a *compact differential manifold* of dimension m.
Note that M can be seen as a *complete metric space* because M is Hausdorff, second-countable, and since it is locally euclidean, it is locally compact hence paracompact. So it admits a metric inducing its topology.
Recall that *Hausdorff* means $\forall x, x' \in M, x \neq x', \exists U, V$ open sets such that $x \in U, x' \in V, U \cap V = \varnothing$.
Recall that *second-countable* means that there exists a countable basis for the topology. A collection of open sets is a *basis* if every open set of the topology is the union of elements of the basis.
A topological space is *locally compact* if every point admits a compact neighbourhood i.e., $\forall x \in M, \exists U$ open and K compact such that $x \in U \subset K$.
A topological space is *paracompact* if every open cover admits a locally finite open refinement. An open *refinement* (*not* equivalent to a subcover) of an open cover $\{U_a\}$ is another open cover $\{V_b\}$ such that $\forall V_b, \exists U_a, V_b \subseteq U_a$. An open cover $\{V_b\}$ is *locally finite* if $\forall x, \exists V_x$ neighbourhood of x such that $\{V_b : V_b \cap V_x \neq \varnothing\}$ is finite.

Property 1: Good atlas

Let $\{U_\mu : \mu \in \Lambda\}$ be an open cover of M.
There exists an atlas $\{V_\nu : \nu \in \mathbb{N}, g_\nu : V_\nu \to V'_\nu\}$ called a *good atlas, subordinate to* $\{U_\mu\}$ such that:

- $\{V_\nu\}$ is a *locally finite refinement* of $\{U_\mu\}$, meaning that $\forall \nu, \exists \mu, V_\nu \subset U_\mu$ and $\forall x \in M, \exists O_x$ neighbourhood of x such that $O_x \cap V_\nu \neq \varnothing$ for finitely many ν.
- $V'_\nu = B(3)$ the open ball centered at 0 with radius 3 (or whatever strictly positive real number).
- Let $W_\nu = g_\nu^{-1} B(1)$ (or whatever radius strictly smaller than the radius of V'_ν), then $\{W_\nu, \nu \in \mathbb{N}\}$ is still an open cover of M.

Note that, since we shall be only concerned with *compact manifolds*, we can assume that any atlas has a finite number of charts.

Property 2: Differentiable bump functions

Let $r > 0$ and $\epsilon > 0$.
There exists a C^∞ function $\lambda : \mathbb{R}^m \to \mathbb{R}$ such that:

- $\forall x \in \mathbb{R}^m, 0 \leq \lambda(x) \leq 1$.
- $\lambda(x) = 1 \Leftrightarrow x \in \overline{B(r)}$.

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

Note that bump functions are used in conjunction with good atlases.

Property 3: Partition of unity

Let A be a closed subset of M and $\{U_i : i \in \Lambda\}$ be an open cover of A.
There exists a set of C^∞ functions $\lambda_i : M \to [0, 1], i \in \Lambda$ such that:

- $\forall i \in \Lambda, \text{support}(\lambda_i) \subset U_i$.
- $\{\text{support}(\lambda_i)\}_{i \in \Lambda}$ is locally finite.
- $\forall x \in A, \sum_{i \in \Lambda} \lambda_i(x) = 1$.

Lemma 1

Let M be a manifold with dimension m, and N a manifold with dimension n.
Assume that $m < n$.
Let $f : M \to N$ be a C^1 function.
Then, $N - f(M)$ is dense in N.

Lemma 2: Implicit function theorem

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

Definiton 1: Whitney C^s topology

Let $r \geq 0$ and $s \leq r$.
Let $\| \cdot \|$ be the Euclidean norm.
The C^s topology on $C^r(M, N)$ is generated by the sub-base $B^s = \{\mathcal{N}^s(f; (U, h), (V, g), K, \epsilon) : f \in C^r(M, N), (U, h)$ and (V, g) charts on M and $N, K \subset U$ compact such that $f(K) \subset V, 0 < \epsilon \leq \infty\}$.
Given $f, (U, h), (V, g), K, \epsilon$, define $\mathcal{N}^s(f; (U, h), (V, g), K, \epsilon) = \{\hat{f} \in C^r(M, N) : \hat{f}(K) \subset V$, and $\forall x \in h(K), \forall 0 \leq k \leq s, \|D^k g \hat{f} h^{-1}(x) - D^k g f h^{-1}(x)\| < \epsilon\}$

Recall that, given (X, T) a topological space:

$$\begin{aligned}
B \text{ is a sub-base of a topology } T \Leftrightarrow & B \text{ generates } T \\
\Leftrightarrow & T \text{ is the smallest topology containing } B \\
\Leftrightarrow & \{\cap_{i \in I} B_i : B_i \in B, I \text{ finite}\} \cup \{X\} \text{ is a basis of } T \\
\Leftrightarrow & \text{Every proper open set (i.e., different from } \varnothing \text{ and } X) \text{ is the union} \\
& \text{of finite intersections of elements of } B \\
\Leftrightarrow & \forall U \in T, x \in U, \exists B_1, \ldots, B_n \in B \text{ such that } x \in B_1 \cap \cdots \cap B_n \subset U.
\end{aligned}$$

Let $S \subset C^r(M,N)$ (e.g., the set of immersions, or the set of embeddings of $C^r(M,N)$).
Recall that: S is open $\Leftrightarrow \forall f \in S, \exists O_f$ neighbourhood of f such that $O_f \subset S$.
Let $f \in S$ and O_f a neighbourhood of f.
Since B^s is a sub-base, there is an open set such that $f \in \cap_i \mathcal{N}^s(f;(U_i,h_i),(V_i,g_i),K_i,\epsilon_i) \subset O_f$. Then:

$$S \text{ is open in } C^r(M,N) \Leftrightarrow \exists \cap_i \mathcal{N}^s(f;(U_i,h_i),(V_i,g_i),K_i,\epsilon_i) \subset S.$$

Now, recall that: S is dense $\Leftrightarrow \forall f \in C^r(M,N), \forall O_f$ neighbourhood of $f, \exists \hat{f} \in O_f$ such that $\hat{f} \in S$.
Let us only consider the case $N = \mathbb{R}^n$.
Since B^s is a sub-base, and given a good atlas $\{(U_\nu,h_\nu),W_\nu\}$ for M, we have that:

$$S \text{ is dense in } C^r(M,\mathbb{R}^n) \Leftrightarrow \forall f \in C^r(M,\mathbb{R}^n), \forall \epsilon > 0, \exists \hat{f} \in S \text{ such that } \hat{f} \in \cap_\nu \mathcal{N}^s(f;(U_\nu,h_\nu),(\mathbb{R}^n,\mathrm{id}),\overline{W}_\nu,\epsilon).$$

2 Takens' Theorem

Theorem 1: Takens' theorem

Let M be a compact manifold of dimension m.
Let $\phi \in Diff^1(M)$ and $y \in C^1(M,\mathbb{R})$.
Define the map $\Phi_{(\phi,y)} : M \to \mathbb{R}^{2m+1}$ by $\Phi_{(\phi,y)}(x) = (y(x), y(\phi(x)), \ldots, y(\phi^{2m}(x)))$.
Then, for pairs (ϕ,y), it is a generic property that $\Phi_{(\phi,y)}$ is an embedding.

The functions $y \in C^1(M,\mathbb{R})$ are called *measurement* functions, and the map $\Phi_{(\phi,y)}$ is called the *delay map*.

We will say that a property is *generic* if it holds for a set which is *open and dense* in the C^1 topology. The dual concept of an open dense set is a *nowhere dense* set (also called a *rare* set). A set is nowhere dense if and only if its closure has empty interior. Equivalently, a set is nowhere dense if and only if it is a subset of the boundary of an open set, or if its closure is the boundary of an open set.

A *meagre* set is a countable union of nowhere dense sets. A *residual* set (also called a *comeagre* set) is the complement of a meagre set. A residual set is a countable intersection of open dense sets. Since $C^1(M,N)$ is a Baire space (a space where countable unions of closed sets with empty interiors also have empty interiors), every residual set of $C^1(M,N)$ is dense. The notion of generic property can be extended to be any property that hold for a set containing a residual set. A generic property is 'typical' in the sense that if a point does not satisfy the property, any small perturbation of the point will land on the residual set. Hence, for example an algorithm can be designed to work only on generic points. Since a open dense set is a residual set, it is sufficient to prove that the property holds on an open dense set to be able to say that it is a generic property. This is the approach followed here.

There is two parts in the proof: one part to establish the openness of the embeddings, and the other to establish their denseness. The openness is the easiest part. Moreover, it is used several times during the proof of the denseness, so this is why the openness is addressed first.

3 Openness of the set of pairs (ϕ,y) that embed M

In this section, we show that the set $\{(\phi,y) \in Diff^1(M) \times C^1(M,\mathbb{R}) : \Phi_{(\phi,y)}$ is an embedding$\}$ is open in the C^1 topology.

Define the mapping $\mathcal{F}^1 : Diff^1(M) \times C^1(M,\mathbb{R}) \to C^1(M,\mathbb{R}^{2m+1})$ by $(\phi,y) \mapsto \Phi_{(\phi,y)}$. We will show that this mapping is continuous. But first, let us prove the following lemma.

Lemma 3

The function $F_1 : Diff^1(M) \times C^1(M,\mathbb{R}) \to C^1(M,\mathbb{R})$ defined by $(\phi,y) \mapsto y \circ \phi$ is continuous.

Proof: Lemma 3

Consider two finite good atlases for M. The first one is denoted by $\{(U_i, h_i)\}$ with $W_i = h_i^{-1}B(1)$. The second one is $\{(V_j, g_j)\}$ subordinate to the cover $\{\phi^{-1}W_i\}$ with $X_j = g_j^{-1}B(1)$. For each V_j, there is a W_i denoted by $W_{i(j)}$, such that $\phi V_j \subset W_{i(j)}$.
The derivatives $Dyh_i^{-1} : h_i\overline{W}_i \to \mathbb{R}^n$ and $Dh_i\phi g_j^{-1} : g_j\overline{X}_j \to \mathbb{R}^{m\times m}$ are uniformly continuous and have compact domains, hence their norms are bounded. There exist A and B such that $\forall i, \forall u \in h_i\overline{W}_i, \|Dyh_i^{-1}(u)\| < A$ and $\forall i, j, \forall u \in g_j\overline{X}_j, \|Dh_j\phi g_j^{-1}(u)\| < B$. Given $\epsilon > 0$, we can find $\delta > 0$ such that $\|u' - u\| < \delta \Rightarrow \|Dyh_i^{-1}(u') - Dyh_i^{-1}(u)\| < \epsilon$.
Given any neighbourhood of $y \circ \phi \in C^2(M, \mathbb{R})$, there is a neighbourhood of the form $\mathcal{N} = \cap_j \mathcal{N}^1(y \circ \phi; (V_j, g_j), (\mathbb{R}, id), \overline{X}_j, \epsilon')$ contained within it. Let δ and ϵ be chosen sufficiently small such that the following are satisfied:

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

(there are finitely many yh_i^{-1} all uniformly continuous using Heine's theorem)

$$\begin{aligned}
\forall i, \forall u, u' \in \overline{W}_i, \|u' - u\| < \delta \Rightarrow \|Dyh_i^{-1}(u') - Dyh_i^{-1}(u)\| &< \epsilon'/3B,\\
\delta &< \epsilon'/3A,\\
\delta &< B,\\
\epsilon &< \min\{\epsilon'/2, \epsilon'/6B\}.
\end{aligned}$$

Now, consider the open neighbourhood $\mathcal{N}^1(\delta, \epsilon) \subset Diff^1(M) \times C^1(M, \mathbb{R}$ of (ϕ, y) defined by:

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

For F_1 to be continuous, it is sufficient to show that $F_1\mathcal{N}^1(\delta, \epsilon) \subset \mathcal{N}$.
Let $(\hat{\phi}, \hat{y}) \in \mathcal{N}^1(\delta, \epsilon)$, $x \in \overline{X}_j$ and $u = g_jx$ for a given j. Then:

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

If we define $u' = h_{i(j)}\hat{\phi}g_j^{-1}(u) \in h_{i(j)}W_{i(j)}$, we can rewrite the left term:

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

using the fact that $y' \in \cap_i \mathcal{N}^1(y; (U_i, h_i), (\mathbb{R}, id), \overline{W}_i, \epsilon)$ and $\epsilon < \min\{\epsilon'/2, \epsilon'/6B\}$.
Similarly, we can rewrite the right term by defining $u'' = h_{i(j)}\phi g_j^{-1}(u)$ (since $x \in \overline{X}_j \subset V_j$ and $\phi V_j \subset W_{i(j)}$):

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

using the fact that $\|u' - u''\| = \|h_{i(j)}\hat{\phi}g_j^{-1}(u) - h_{i(j)}\phi g_j^{-1}(u)\| < \delta$
(since $\hat{\phi} \in \mathcal{N}^1(\delta, \epsilon) = \cap_j \mathcal{N}^1(\phi; (V_j, g_j), (W_{i(j)}, h_{i(j)}), \overline{X}_j, \delta)$).
Hence, we get:

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

Now, we can apply a similar argument for the derivatives (albeit long and boring, also the chain rule is needed at some point) to get:

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

using the assumptions imposed on the choice of δ and ϵ.
So we have showed that if $(\hat{\phi}, \hat{y} \in \mathcal{N}(\delta, \epsilon)$ then $y \circ \phi \in \mathcal{N}$. That is, F_1 is continuous. $\square$

Using the preceding lemma, we can show following result:

Lemma 4

The function $F_n : Diff^1(M) \times C^1(M,\mathbb{R}) \to C^1(M,\mathbb{R})$ defined by $(\phi, y) \mapsto y \circ \phi^n$ is continuous.

Proof: Lemma 4

By induction: using lemma 3 and the fact that $F_n = F_{n-1} \circ G$ where $G : Diff^1(M) \times C^1(M,\mathbb{R}) \to Diff^1(M) \times C^1(M,\mathbb{R})$ is defined by $G(\phi, y) = (\phi, F_1(\phi, y))$, we see that F_n is the composition of continuous functions. $\square$

Property 4

$\mathcal{F}^1$ is continuous.

Proof: Property 4

$\mathcal{F}^1 = T \circ F$ with $F : Diff^1(M) \times C^1(M,\mathbb{R}) \to [C^1(M,\mathbb{R}]^{2m+1}$ specified by its components: $F = (F_0, F_1, \ldots, F_{2m})$ where F_0 is defined as $(\phi, y) \mapsto y$, and $T : [C^1(M,\mathbb{R}]^{2m+1} \to C^1(M,\mathbb{R}^{2m+1})$ maps each $2m+1$-tuple of real-valued functions to the vector-valued function having these functions as its components. By lemma 4, each component of F is continuous so F is continuous. We only need to show that T is continuous. This can be done with the same method as used in the proof of lemma 3. $\square$

Property 5

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

Proof: Property 5

Using property 4 along with the fact that the set of immersions on K (respectively of embeddings on K) is open in $C^1(M,\mathbb{R}^{2m+1})$. $\square$

The openness part of theorem 1 is proven by using property 5 with $K = M$.
Note that property 5 is still true for ϕ differentiable from M to M (no need for ϕ to be a differomorphism), and for any number of delays. The fact that the number of delays has to be $2m+1$ is used in the denseness part.

4 A slightly different result

Consider the following result:

Theorem 2: Takens, unstated

Let M be a compact manifold of dimension m.
Let $\phi : M \to M$ be a diffeomorphism such that:

- The periodic points of ϕ with periods less or equal to $2m$ are finite in number,
- If x is a periodic point with period $k \leq 2m$, then the eigenvalues of the derivative of ϕ^k at x are all distinct.

Then, for generic $y \in C^1(M, \mathbb{R})$, the delay map $\Phi_{(\phi,y)} : M \to \mathbb{R}^{2m+1}$ is an embedding.

We will now prove theorem 2. The openness part is very similar to that of theorem 1, the main difference being that we only need to prove that the set of measurement functions y giving rise to embeddings $\Phi_{(\phi,y)}$ is open in $C^1(M, \mathbb{R})$ (and not the set of (ϕ, y) since ϕ is fixed). Hence, we obtain a result very similar to property 5:

Property 6

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

Note that, in theorem 1, we considered generic (open and dense) sets of diffeomorphisms, while in theorem 2, ϕ is fixed but with supplementary assumptions. The next sections are devoted to proving that the set of y giving rise to embeddings $\Phi_{(\phi,y)}$ is dense in $C^1(M, \mathbb{R})$. But before doing that, we have to prove that the denseness of theorem 2 implies the denseness of theorem 1, such that by proving theorem 2 we will also complete the proof of theorem 1. We will rely on the following result.

Theorem 3: Kupka-Smale (somehow)

Let M be a compact manifold, and n be a finite positive integer.
For generic $\phi \in Diff^1(M)$, the number of periodic points with period n or less is finite.
Moreover, the property still hold if we assume the periodic points to be *hyperbolic*.
Hence, for generic ϕ, the fixed points have distinct eigenvalues, and the periodic points have distinct eigenvalues when considered as fixed points of ϕ^k.

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

Now, assume the denseness part of theorem 2. Let A be the set of ϕ's with finite numbers of periodic points with period less or equal to $2m$ and having distinct eigenvalues. theorem 3 tells us that $A \subset Diff^1(M)$ is open and dense. Let X and Y be topological spaces, and let $V \subset X \times Y$ with the following property: $\exists A \subset X$ dense, $\forall x \in A$, $\exists O_x \subset Y$ open and dense, such that $\{(x, y) : y \in Ox\} \subset V$. Hence, V is dense in $X \times Y$. By interpreting X to be $Diff^1(M)$ and Y to be $C^1(M, \mathbb{R})$, then we obtain the denseness part of theorem 1.

5 Denseness of the set of measurement functions: general method

Given any $y \in C^1(M, \mathbb{R})$ and any neighbourhood $\mathcal{N}$ of y, we have to show that there exists $y' \in \mathcal{N}$ such that $\Phi_{(\phi,y')}$ is an embedding of M. This is done by showing how to construct y' explicitly:

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

where N is finite, $a_i \in \mathbb{R}$, and the $\psi_i : M \to \mathbb{R}$ are C^2.

The measurement function y is adjusted several times using equation (1). At each adjustment, the constructed function y' receives a new property (being immersive, or embedding some compact set). To ensure that the construction of y' does not destroy the previous property of y, recall that according to property 6, if $\mathcal{N}$ is small enough and if $y' \in \mathcal{N}$ then y' will share the properties of y (if $\mathcal{N}$ is not small enough, we can choose an open set included in $\mathcal{N}$). The new property of y' is also true for all functions in a neighbourhood $\mathcal{O}$ of y'. By taking a smaller neighbourhood if necessary, it is possible to choose $\mathcal{O}$ such that $\mathcal{O} \subset \mathcal{N}$. Then, we may find another function $y'' \in \mathcal{O}$ with yet another property, and since $y'' \in \mathcal{O}$, we also have $y'' \in \mathcal{N}$ such that y'' will share the properties of y' and y. By repeating this process a finite number of time, we may manage to construct an embedding of M which still lie in $\mathcal{N}$ hence proving the denseness part of theorem 2.

In the next arguments, we will require y and ϕ to be C^2. To do so, we use the following (supposedly classical) result:

Lemma 5

$C^2(M, \mathbb{R})$ is dense in $C^1(M, \mathbb{R})$ (with the C^1 topology).
Similarly, $Diff^2(M)$ is dense in $Diff^1(M)$.
It follows that $Diff^2(M) \times C^2(M, \mathbb{R})$ is dense in $Diff^1(M) \times C^1(M, \mathbb{R})$.

The C^1 topology on $C^2(M, \mathbb{R})$ is simply the induced topology when $C^2(M, \mathbb{R})$ is considered as a subset of $C^1(M, \mathbb{R})$ (and similarly for $Diff^2(M)$ and $Diff^2(M) \times C^2(M, \mathbb{R})$).

So, if we prove that the set of pairs (ϕ, y) giving rise to delay maps which are embeddings of M is dense in $Diff^2(M) \times C^2(M, \mathbb{R})$, then, since $Diff^2(M) \times C^2(M, \mathbb{R})$ is dense in $Diff^1(M) \times C^1(M, \mathbb{R})$, these pairs must be dense in $Diff^1(M) \times C^1(M, \mathbb{R})$ as well. From now on, we assume that ϕ is C^2, and that all the considered measurement function y are C^2. In particular, property 6 is still true if we replace C^1 by C^2.

In order to find a y' in a sufficiently small neighbourhood of y, we will need the following lemma:

Lemma 6

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

Proof: Lemma 6

Let $\{(U_j, h_j)\}$ be a finite good atlas for M. Then, there exists $\epsilon > 0$ such that the open set $\cap_j \mathcal{N}^1(y; (U_j, h_j), (\mathbb{R}, id), \overline{W}_i, \epsilon)$ is a subset of $\mathcal{N}$. $\forall 1 \leq i \leq N$ and $\forall j$, the function $\phi_i h_j^{-1} : h_j(\overline{W}_j) \to \mathbb{R}$ is continuous on a compact domain, hence it is bounded (in absolute value). Since there are finitely many such functions, there exists a B that is an upper bound of all these functions. Using equation 1, we have that if $x \in \overline{W}_j$ then:

$$\|y' h_j^{-1}(h_j x) - y h_j^{-1}(h_j x)\| = \| \sum_{i=1}^{N} a_i \phi_i h_j^{-1}(h_j x)\| \leq \sum_{i=1}^{N} |a_i| |\phi_i h_j^{-1}(h_j x)| \leq B \sum_{i=1}^{N} |a_i|.$$

Similarly, the derivatives $D\phi_i h_j^{-1}$ are continuous functions on a compact domain, so there exists a B' such that:

$$\|Dy' h_j^{-1}(h_j x) - Dy h_j^{-1}(h_j x)\| \leq B' \sum_{i=1}^{N} |a_i|.$$

Let $\delta < \epsilon/(\sqrt{n}\max\{B, B'\})$. Since $\|.\|_1 \leq \sqrt{n}\|.\|$, we find that for $\|a\| < \delta$, and for all j and $x \in \overline{W}_j$:

$$\|y' h_j^{-1}(h_j x) - y h_j^{-1}(h_j x)\| \leq \epsilon$$
$$\|Dy' h_j^{-1}(h_j x) - Dy h_j^{-1}(h_j x)\| \leq \epsilon.$$

Hence, $y' \in \mathcal{N}$. □

The procedure to build an embedding on M has four stages, each one producing a measurement function whose delay map is an injective immersion on successively larger parts of M. Denote by P_k the set of periodic points of ϕ with period less or equal than k. We will create functions which give a delay map which is

- immersive at every point in P_{2m}, then an injective immersion on some compact neighbourhood of P_{2m},
- an immersion on the whole M,
- an injective immersion on orbit segments,
- an injective immersion on M.

Each of this stage requires one or more adjustments as described in equation 1. For each one of these adjustments, it must be shown that $\|a\|$ can be made arbitrarily small and yet still give a y' having the desired property. To do so, we will rely on lemma 1.

6 Denseness of the set of measurement functions that embed a compact neighbourhood of the periodic points

Why do we focus on periodic points of ϕ? Because, for any $x \in M$ whose period is less or equal to $2m$, $\Phi_{(\phi,y)}(x)$ will have at least two identical coordinates. In particular, for fixed points, all the coordinates are equal (or said otherwise, all the images of fixed points lie on the diagonal of $\mathbb{R}^{2m+1}$). This degeneracy causes some delay maps to fail to be embeddings (consider the case where ϕ is the identity, then whatever choice of measurement function y, and whatever the number of delays, the delay map $\Phi_{(\phi,y)}$ will not be an embedding).

These difficulties are the reason why it is assumed in theorem 2 that the number of periodic points of ϕ with period less or equal to $2m$ are finite (i.e., P_{2m} is a finite set). Given this condition, there exists an open neighbourhood of each $x_i \in P_{2m}$ containing no other point in P_{2m}. By taking a smaller neighbourhood if necessary, we can assume that this neighbourhood lies in some chart domain U_i of a good atlas $\{(U_i, h_i)\}$, such that our neighbourhood is homeomorphic (under h_i) to an open ball B_i in $\mathbb{R}^m$ centred at $h_i(x_i)$. Using the Hausdorff property of M, we can choose these neighbourhoods so that they do not intersect each other.

As stated above, if x is a fixed point then $\Phi_{(\phi,y)}(x) = (y(x), y(x), \dots, y(x))$. To make sure that none of these fixed points have the same image, we must adjust y so that it takes a different value at every fixed point. In fact, we can even say that if y takes a different value for every $x_i \in P_{2m}$, then no two of these points map to the same image, since at least the first component will be different. If y has such property, then it is injective on the set P_{2m}, which is compact since it is finite. If we manage to prove that it is also an immersion on P_{2m}, then we can apply property 6 to say that the measurement functions y that embed P_{2m} form an open subset of $C^2(M, \mathbb{R})$.

6.1 Making an injection on P_{2m}

Now, we will show how to adjust y so that it becomes an injection on P_{2m}. Let $x_1, x_2 \in P_{2m}$ at which y takes the same value. Let (U_1, h_1) be the chart containing x_1 (so that $h_1 x_1$ is the centre of $B(3)$. Define the function $\psi : M \to \mathbb{R}$ by

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

where $\lambda : \mathbb{R}^m \to \mathbb{R}$ is a bump function having support in $B(3)$, and equal 1 on $B(1)$ (so ψ is a C^∞ function on M). Then define $y' = y + a\psi$ where a is a real number. For every $a > 0$, $y'(x_1)$ and $y'(x_2)$ are different (they differ by a). Using lemma 6, we can find a y' in any neighbourhood of y (by taking a sufficiently small a). If more than two x_i's have the same image, we can use a similar argument with several perturbations. So we have proved that, for generic y, $\Phi_{(\phi,y)}$ is injective when restricted to P_{2m}.

6.2 Making an immersion on fixed points of ϕ

Now, we need to show that, for generic y, $\Phi_{(\phi,y)}$ is an immersion on P_{2m}, meaning that $D\Phi_{(\phi,y)}h_i^{-1}$ has to be full rank at every $x_i \in P_{2m}$.

Consider first the fixed point of ϕ. Let x_1 be such a fixed point. Then, the k-th row of $D\Phi_{(\phi,y)}h_1^{-1}(h_1x_1)$ is $Dy\phi^{k-1}h_1^{-1}(h_1x_1)$. Using the chain rule we get:

$$\begin{aligned}
Dy\phi^{k-1}h_1^{-1}(h_1x_1) &= Dyh_1^{-1}h_1\phi^{k-1}h_1^{-1}(h_1x_1)\\
&= Dyh_1^{-1}(h_1\phi^{k-1}h_1^{-1}h_1x_1)Dh_1\phi^{k-1}h_1^{-1}(h_1x_1)\\
&= Dyh_1^{-1}(h_1\phi^{k-1}x_1)Dh_1\phi^{k-1}h_1^{-1}(h_1x_1)\\
&= Dyh_1^{-1}(h_1x_1)Dh_1\phi^{k-1}h_1^{-1}(h_1x_1) \text{ (since } x_1 \text{ is a fixed point of } \phi)
\end{aligned}$$

Let the row vector $v = Dyh_1^{-1}(h_1x_1)$ and $J = Dh_1\phi h_1^{-1}(h_1x_1)$, we have that the k-th row of the Jacobian matrix becomes vJ^{k-1}. Hence $\Phi_{(\phi,y)}$ is immersive at x_1 if and only if the set $\{v, vJ, \dots, vJ^{2m}\}$ contains m linearly independent vectors. By assumptions (in theorem 2), J has distinct eigenvalues λ_j and hence linearly independent eigenvectors. In terms of these eigenvectors, we have $v = \sum \alpha_j e_j$ so that $vJ^{k-1} = \sum \alpha_j \lambda_j^{k-1} e_j$ (assume that the α_i are real to avoid any more technicalities). In this basis, $\{v, vJ, \dots, vJ^{2m}\}$ takes the form:

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

We consider the first m rows, discarding the later ones to leave a square matrix which will be linearly independent if and only if its determinant is nonzero. As a function of the α_i's, the determinant is a polynomial from $\mathbb{R}^m$ to $\mathbb{R}$. Hence, the vectors v for which the matrix is not full rank correspond to the set of α's which are zeros of the polynomial. If a polynomial is not identically zero then its zeros form a closed nowhere dense set. To see that the determinant is not identically zero, consider $(\alpha_1, \dots, \alpha_m) = (1, \dots, 1)$. The matrix becomes a Vandermonde matrix which is full rank if and only if the λ_j's are distinct, which was assumed in the assumptions of theorem 2. Thus, the v's (transpose of the gradient of yh_1^{-1} at a fixed point of ϕ) for which $\{v, vJ, \dots, vJ^{2m}\}$ contains m linearly independent vectors form an open dense set (recall that the complement of a nowhere dense set contains an open dense set). Thus, given any v, we can find v' such that $\{v, vJ, \dots, vJ^{2m}\}$ span $\mathbb{R}^m$ and the difference $a = v' - v$ has arbitrarily small norm. Define C^∞ functions $\psi_j : M \to \mathbb{R}$ by

$$\psi_j(x) = \begin{cases} \mu_j(x)\lambda(h_1x) & \text{for } x \in h_1^{-1}B(3)\\ 0 & \text{otherwise}\end{cases}$$

where $\mu_j : U_1 \to \mathbb{R}$ is the j-th coordinate function ($\mu_j = u_j \circ h_1$). We will use this construction several times because it has the following property for any $u = h_1x \in \overline{B(1)}$:

$$\begin{aligned}
\frac{\partial\psi_j h_1^{-1}}{\partial u_k}(u) &= \frac{\partial(\mu_j h_1^{-1} \times \lambda)}{\partial u_k}(u)\\
&= \frac{\partial\mu_j h_1^{-1}}{\partial u_k}(u) \times 1 + \mu_j h_1^{-1}(u) \times \frac{\partial\lambda}{\partial u_k}(u) \text{ (since } \lambda(u) = 1)\\
&= \frac{\partial\mu_j h_1^{-1}}{\partial u_k}(u) \text{ (since the bump function has zero derivative on } \overline{B(1)})\\
&= \frac{\partial((u_j \circ h_1) \circ h_1^{-1})}{\partial u_k}(u) \text{ (definition of } \mu_j)\\
&= \delta_{kj}.
\end{aligned}$$

Using these ψ_j, we define y' using equation 1. Then $Dy'h_1^{-1}(h_1x_1) = Dyh1^{-1}(h_1x_1) + a$. As we have just shown, $\Phi_(\phi, y')$ can be made immersive at x_1 for arbitrarily small value of $\|a\|$. Hence, we can apply lemma 6 and say that a y' giving rise to an immersion at x_1 can be found in any neighbourhood of y. Since the number of fixed points of ϕ is finite, a finite number of such adjustments to y will give a y' for which $\Phi_(\phi, y')$ is immersive at all of them. Lemma 6 tells us that such a y' can be found in any neighbourhood of y.

6.3 Making an immersion on P_{2m}

The argument extends easily on to the periodic points of ϕ. For example, let x_1, x_2 be a pair of period 2 points such that $\phi(x_1) = x_2, \phi(x_2) = x_1$. Since these points are distinct, there exist disjoint open set containing x_1 and x_2 homeomorphic to open balls B_1 and B_2 centred at $h1(x_1)$ and $h_2(x_2)$. Then, we can use bump functions as above to perturb y independently at these two points. Let us focus on x_1. We have

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

We want to show that $D\Phi_{(\phi,y)}h_1^{-1}(h_1x_1)$ has full rank. Once gain, using the chain rule and the fact that $\phi^2(x_1) = x_1$, we can write the odd rows ($2i+1$-th row) of this matrix $Dy\phi^{2i}h_1^{-1}(h_1x_1) = vJ^i$ with $v = Dyh_1^{-1}(h_1x_1)$ and $J = Dh_1\phi^2h_1^{-1}(h_1x_1)$ and the even rows ($2i$-th row) $Dy\phi^{2i-1}h_1^{-1}(h_1x_1) = wJ^i$ with $w = Dy\phi^{-1}h_1^{-1}(h_1x_1)$. Taking as above an eigenbasis of J, we can write $D\Phi_{(\phi,y)}h_1^{-1}(h_1x_1)$ as a matrix whose determinant is a polynomial function of the coordinates of the v and w in the eigenbasis. This polynomial is not identically zero since we can choose the coordinates to obtain a Vandermonde matrix. It follows that the set of pairs (v, w) which make the derivative full rank is open and dense in $\mathbb{R}^m \times \mathbb{R}^m$. Say otherwise, given any v and w, we can find v' and w' such that $\|(v, w) - (v', w')\|$ is arbitrarily small, and the set $\{v', v'J, \ldots, v'J^{m+1}, w', w'J, \ldots, w'J^m\}$ spans $\mathbb{R}^m$.

Now, let y be a measurement function such that $\Phi_(\phi, y)$ is not immersive at x_1. Define $\psi_i : M \to \mathbb{R}$ and $\chi_i : M \to \mathbb{R}$ by

$$\psi_i(x) = \begin{cases} \mu_{1,i}(x)\lambda(h_1x) & \text{for } x \in h_1^{-1}B_1 \\ 0 & \text{otherwise} \end{cases}$$

and

$$\chi_i(x) = \begin{cases} \mu_{2,i}(x)\lambda(h_2x) & \text{for } x \in h_2^{-1}B_2 \\ 0 & \text{otherwise} \end{cases}$$

where $\mu_{j,i}$ is the i-th coordinate function of h_j, and B_1, B_2 are chosen so that there are disjoint. Then, we use equation (1) to construct y'. It can be shown that we can choose the coefficients of ψ_i and χ_i such that $Dyh_1^{-1}(h_1x_1) = v'$ and $Dy\phi^{-1}h_1^{-1}(h_1x_1) = w'$ such that $\Phi_(\phi, y')$ is immersive at x_1 and we can choose the perturbation to have arbitrarily small norm. The measurement functions giving rise to delay maps which are immersive at x_1 form an open set (by lemma 6), so we can make another adjustment such that the measurement function is immersive at both x_1 and x_2. We can extend this argument to cover all the points of period less than or equal to $2m$. Finally, by making a finite sequence of adjustments to y, we can construct a y' such that $\Phi_(\phi, y')$ is immersive on P_{2m}, and also injective on P_{2m} as discussed above. Since P_{2m} is compact, we can apply proposition 6 to say that these measurement functions form an open subset of $C^2(M, \mathbb{R})$.

6.4 Making an embedding on a compact neighbourhood of P_{2m}

Let y be a measurement function constructed such that $\Phi_{(\phi,y)}$ is an injective immersion of P_{2m}. By the Inverse Function Theorem, there is a neighbourhood of each $x_i \in P_{2m}$ which is embedded in $\mathbb{R}^{2m+1}$ by $\Phi_{(\phi,y)}$. We may take the neighbourhoods of the x_i's to be open balls $b_i(r_i, x_i)$ (since M is a metric space). $\Phi_{(\phi,y)}$ is an immersion on the union of these balls (since it is a local property) and an injection on each ball, but not on the union of the balls. However, by taking smaller radii r_i, it must be possible to find balls whose images do not intersect (since $\Phi_{(\phi,y)}$ is injective on P_{2m} and continuous on M). Then, $\Phi_{(\phi,y)}$ is an injective immersion on the union of these smaller balls.

Now, consider closed balls with a radius strictly lower than the open balls. The union of these closed balls is closed hence compact and is a subset of the union of the open balls, such that $\Phi_{(\phi,y)}$ is an injective immersion on this compact set. Let us call this set $V_y = \cup_i bi$ where b_i is the closed ball containing x_i. V_y is a compact neighbourhood of P_{2m} so we have just shown that there is a dense (and open) set of measurement functions such that for each element y of this set there exists a compact neighbourhood of the periodic points that is embedded by $\Phi_{(\phi,y)}$.

7 Denseness of the set of measurement functions that immerse M

Let y be such that there exists a compact neighbourhood V_y of P_{2m} such that $\Phi_{(\phi,y)}$ is an embedding of V_y. We show now that in every neighbourhood of y, we can find another measurement function which gives rise to an immersion of M. The general idea is to cover M with compact sets and to show that by making arbitrarily small perturbations of the measurement function, we can produce a delay map which is an immersion of one of these sets. Then, we can proceed similarly to obtain a delay map that immerse another compact set, and we use the openness of immersions to show that after this second adjustment, both the first and the second compact sets are immersed. By doing this for each compact set one by one, we obtain a delay map that immerses M with a measurement function arbitrarily close to our original y.

7.1 Construction of a suitable atlas

First, let us construct a suitable atlas. For every $x_i \in P_{2m}$, we can find a chart (U_i, h_i) such that $U_i \subset b_i$ and $U_i = h_i^{-1}B(3)$. Note that the chart domains are disjoint, that the sets $W_i = h_i^{-1}B(1)$ form an open cover for P_{2m}, and that $\overline{W}_i$ is a compact subset of b_i. Now, consider the complement of P_{2m}, denoted P_{2m}^c. P_{2m}^c is an open set and $\forall x \in P_{2m}^c$, the points $\{x, \phi x, \dots, \phi^{2m}x\}$ are all distinct. So we can find an open set $U_x \subset P_{2m}^c$ containing x such that $U_x, \phi U_x, \dots, \phi^{2m}U_x$ are disjoint. By taking smaller U_x if necessary, we can find h_x such that (U_x, h_x) is a chart with $U_x = h_x^{-1}B(3)$ and $W_x = h_x^{-1}B(1)$. The collection of sets $\{W_x\} \cup \{W_i\}$ forms an open cover of M. Since M is compact, we can extract a finite subcover, which must contains every W_i. Let us relabel the sets so that $W_i, 1 \le i \le k$ are the sets containing the periodic points, and $W_i, k < i \le l$ are the sets contained in P_{2m}^c. The corresponding charts (U_i, h_i) form our suitable atlas.

V_y is different for each y. But by proposition 6, delay embeddings of the compact set V_y are open. Hence, there exists a neighbourhood $\mathcal{U}_y$ of y in $C^2(M,\mathbb{R})$ such that $\forall \hat{y} \in \mathcal{U}_y$, $\Phi_{(\phi,\hat{y})}$ is an embedding of V_y. We will show that every neighbourhood of y contained in $\mathcal{U}_y$ contains an immersion of M. $\forall \hat{y} \in \mathcal{U}_y$, $\Phi_{(\phi,\hat{y})}$ is an immersion of the compact set $\cup_{i=1}^k \overline{W}_i$ (it is even an embedding). It remains to adjust the measurement function so as to make an immersion on the remaining $\overline{W}_i$'s.

7.2 Making an immersion on a compact set of the cover

Let i be the smallest index greater than k for which $\Phi_{(\phi,y)}$ fails to be an immersion of $\overline{W}_i$. Let $x \in U_i$, let $\mu_j : U_i \to \mathbb{R}$ be the j-th coordinate function of h_i, and let $u = h_i x$ and $u_j = \mu_j x$. For some $u \in h_i\overline{W}_i = \overline{B}(1)$, the Jacobian matrix of $\Phi_{(\phi,y)}$ does not have full rank. To make it full rank, we perturb y such that one a the column of the Jacobian matrix becomes linearly independent of the columns to its left. After at most m such perturbations, we obtain a full rank matrix. We will show that we can make these perturbations arbitrarily small. Suppose that the first s columns of the matrix are linearly independent for all $u \in \overline{B}(1)$. Define $\lambda : \mathbb{R}^m \to \mathbb{R}$ to be a bump function equal to 1 on $\overline{B}(1)$ and having support in $B(2)$ (so that its support is strictly included in $h_iU_i = B(3)$). Define $\psi : M \to \mathbb{R}$ by

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

so that $\psi(x) = \mu_{s+1}(x)$ if $x \in \overline{W}_i$. ψ has support in U_i, hence $\psi_j = \psi \circ \phi^{-j}$ has support in $\phi^j U_i$ which are all disjoint as explained above. Using equation 1, we construct

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

y' is C^2 and we have $y'(x) = y(x)$ whenever x in not in one of the sets $U_i, \phi U_i, \ldots, \phi^{2m} U_i$. Let $u \in h_i \overline{W}_i$. Then, $\phi^k h_i^{-1}(u) \in \phi^k U_i$ so

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

and hence

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

Thus, for all $u \in h_i \overline{W}_i$, the perturbation has added the vector a to the $s+1$-th column of the Jacobian matrix, the other column being unaffected. Now let $J_s(x)$ be the matrix formed from the first s columns of the Jacobi matrix of $\Phi_{(\phi,y)}$ at $x \in U_i$. By assumption, $J_s(x)$ is full rank on $\overline{W}_i$. Since J_s is a continuous function from U_i to the space of $(2m+1) \times s$ matrices, and since the full rank matrices form an open subset of this space, J_s is still full rank on some open set $X \subset U_i$ with $\overline{W}_i \subset X$. Now define the function $\mathcal{S} : \mathbb{R}^s \times X \to \mathbb{R}^{2m+1}$ by

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

with $u = h_i x$. The function $\mathcal{S}$ is C^1 because y and ϕ are assumed to be C^2 (this is where the C^2 assumption is actually required). Since $s \le m-1$ and X has dimension m, the dimension of $\mathbb{R}^s \times X$ is smaller or equal to $2m-1$, so in particular it is strictly smaller than that of $\mathbb{R}^{2m+1}$. Hence, we can apply lemma 1 and say that the complement of $\mathcal{S}(\mathbb{R}^s \times X)$ is dense in $\mathbb{R}^{2m+1}$. So there exists a vector $a \in \mathbb{R}^{2m+1}$ with arbitrarily small norm such that $a \notin \mathcal{S}(\mathbb{R}^s \times X)$. Say differently, there exists $a \in \mathbb{R}^{2m+1}$ with arbitrarily small norm such that $\forall x \in \overline{W}_i \subset X$, the first $s+1$-th columns of the Jacobian matrix of $\Phi_{(\phi,y')}$ are linearly independent (where y' is defined using the vector a). By repeating the argument for each column of the Jacobian matrix, we construct a measurement function y' arbitrarily close to y that gives rise to an immersion of $\overline{W}_i$. Note that to apply lemma 1, the number of delays must be at least $2m$. The condition that we must have $2m+1$ delays will be used later, to obtain the injectivity.

7.3 Making an immersion on M

Since we can find immersions of $\overline{W}_i$ with arbitrarily small perturbations, lemma 6 shows that we can find one in $\mathcal{U}_y$, which is already an immersion of any $\overline{W}_j$ with $j < i$ by assumption. By repeating the argument for $\overline{W}_{i+1}$ and so on, we finally obtain a y' such that $\Phi_{(\phi,y')}$ immerses the whole of M.

A final remark to conclude this section. Since $\Phi_{(\phi,y')}$ is an immersion of M, by the Inverse Function Theorem for each point $x \in M$ there exists an open neighbourhood N_x of x such that $\Phi_{(\phi,y')}$ is an embedding of N_x. We can find a closed ball $\overline{\beta}_x \subset N_x$ centred at x. The interiors of these balls form an open cover of M (since there is one for each x) from which we can extract a finite subcover. The corresponding finite collection of closed balls $\{\overline{\beta}_i\}$ forms a compact cover and each of the closed balls $\overline{\beta}_i$ is embedded by $\Phi_{(\phi,y')}$. Note that the balls are embedded individually and that $\Phi_{(\phi,y')}$ is not an embedding of $\cup_i \overline{\beta}_i$. If we select one of the (compact) balls, the set of measurement functions giving rise to embeddings of it is open (by proposition 6), and so is the set of measurement functions giving embeddings of all the balls (individually), since the intersection of a finite number of open sets is open. Let us call this set $\mathcal{U}'_y$. Since it is a neighbourhood of y' and since $y' \in \mathcal{U}_y$, we can take $\mathcal{U}'_y \subset \mathcal{U}_y$. Consider the following lemma:

Lemma 7: Lebesgue number

Let (M, ρ) be a compact metric space and $\{\beta_i\}$ an open cover of M.
There exists $\epsilon > 0$ such that a closed ball of radius ϵ centred at any point of M is contained in β_i for at least one i.
ϵ is called a Lebesgue number of the cover $\{\beta_i\}$.

It follows from this lemma that every closed ϵ-ball is embedded by any $\Phi_{(\phi,\hat{y})}$ where $\hat{y} \in \mathcal{U}'_y$. Hence, the conclusion of this section is the following: if $\hat{y} \in \mathcal{U}'_y$, then $\Phi_{(\phi,\hat{y})}$ is an immersion of M, an embedding of V_y, and $\forall x \neq x', \rho(x, x') \leq \epsilon \Rightarrow \Phi_{(\phi,\hat{y})}(x) \neq \Phi_{(\phi,\hat{y})}(x')$.

8 Denseness of the set of measurement functions that embed orbit segments

In the previous section, we have constructed an immersion of M. To complete the proof, we must make this immersion injective. For each $x \in M$, let us call the collection of points $\{x, \phi x, \ldots, \phi^{2m} x\}$ the *orbit segment* of x. Let $x \in M$ be a periodic point with period less or equal to $4m$ and let x' be a point belonging to the orbit segment of x. The segments $\{x, \phi x, \ldots, \phi^{2m} x\}$ and $\{x', \phi x', \ldots, \phi^{2m} x'\}$ overlap at both ends, so that we cannot change the coordinates of $\Phi_{(\phi,y)}(x)$ without also changing the coordinates of $\Phi_{(\phi,y)}(x')$. In this section, we create a delay map such that no $x \in P^c_{2m}$ shares an image under $\Phi_{(\phi,y)}$ with another point of its orbit segment. In the next section, we extend the injectivity on the whole of M.

Lemma 8

Let y' be such that $\Phi_{(\phi,y')}$ is an injective immersion on V_y.
In every neighbourhood of y' in $C^2(M, \mathbb{R})$, there is a function y'' such that $\forall x \in M, \forall 1 \leq j \leq 2m$,

$$(x \neq \phi^j x) \Rightarrow (\Phi_{(\phi,y'')}(x) \neq \Phi_{(\phi,y'')}(\phi^j x)).$$

Proof: Lemma 8

Let $x \in M$ and let j be the smallest value for which the lemma is not already true.
Let $S = \cap_{i=0}^{2m} \phi^{-i} V_y$. S is the set of points of V_y which orbit segment lies entirely in V_y. Since we assumed that $\Phi_{(\phi,y')}$ is already injective on V_y, the lemma is already true for S.
We now focus on points outside S. Let T be the closure of the complement of S. if $x \in T$, $x \notin P_{2m}$ and $\{x, \phi x, \ldots, \phi^{2m} x\}$ are all different, so we can find an open set U_x such that $U_x, \phi U_x, \ldots, \phi^{2m} U_x$ are all disjoint. Let us consider two cases.
<u>Case 1</u>: x is not a periodic point whose period lies between $2m+1$ and $4m$. Then, there is a U_x such that $U_x, \ldots, \phi^{4m} U_x$ are all disjoint. We also assume that U_x is the domain of a good chart, i.e., $U_x = h_x^{-1} B(3), W_x = h_x^{-1} B(1)$. We will use the notation $X_x = W_x$ to remain coherent with Case 2.
<u>Case 2</u>: x is a periodic point of period k where $2m+1 \leq k \leq 4m$. We find U_x such that $U_x, \ldots, \phi^{k-1} U_x$ are all disjoint and U_x is the domain of a good chart. Define $X_x = W_x \cap \phi^{-k} W_x$, which is an nonempty open set (since $x \in X_x$).
Note that in Case 1 and Case 2, if $2m + j < k$, none of the sets $\phi^{2m+1} \overline{X}_x, \ldots, \phi^{2m+j} \overline{X}_x$ intersect $\cup_{l=0}^{2m} \phi^l U_x$ (because $U_x, \ldots, \phi^{k-1} U_x$ are disjoint and $\overline{X}_x \subset U_x$). Also note that in Case 2, if $2m+j \geq k$, none of $\phi^{2m+1} \overline{X}_x, \ldots \phi^{k-1} \overline{X}_x$ intersect $\cup_{l=0}^{2m} \phi^l U_x$, and since $\overline{X}_x \subset \phi^{-k} \overline{W}_x$ then $\phi^k \overline{X}_x \subset \overline{W}_x, \phi^{k+1} \overline{X}_x \subset \phi \overline{W}_x, \ldots, \phi^{2m+j} \overline{X}_x \subset \phi^{2m+1} \overline{W}_x$.

From the collection $\{X_x : x \in T\}$, extract a finite cover of T, which we relabel $X_i, i = 1, \ldots, N$ and (U_i, h_i) are the corresponding charts. We use the same procedure as in section 7. y' is adjusted so that the property holds on $\overline{X}_i$ (for the given j), then on $\overline{X}_{i+1}$ with another adjustment and so on until the property holds on the whole of T after a finite number of adjustments. Moreover, we have to show that these adjustments can be made arbitrarily small, so that the property is preserved on the sets already dealt with.

Let X_i be the next set on which the property does not hold, i.e., $\forall 1 \leq i' < i, \forall x \in X_{i'}, \Phi_{(\phi,y')}(x) \neq \Phi_{(\phi,y')}(\phi^j x)$. Define $\psi : M \to \mathbb{R}$ by

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

and $\psi_l = \psi \circ \phi^{-l}$ so that the support of ψ_l lies in $\phi^l U_i$. We define $y'' = y' + \sum_{l=0}^{2m} a_l \psi_l$ using equation 1. Let l be taken between 0 and $2m$. Given any $x \in \overline{X}_i$, we have that $x \in \overline{W}_i$ and hence $\phi^l x \in \phi^l \overline{W}_i$. So the values of y'' over the orbit segment of x are $y''(\phi^l x) = y'(\phi^l x) + a_l$. The values of y'' over the orbit segment of $\phi^j x$ depend on whether X_i is Case 1 or Case 2. We only treat Case 2 with $2m+j \leq k$, since Case 1 and Case 2 with $2m+j < k$ are similar. We can see that

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

with A being the following $(2m+1) \times (2m+1)$ matrix:

$$A = \begin{array}{cc}
\begin{matrix} 0 & \cdots & \cdots & \cdots & {\scriptstyle 2m+j-k} & \cdots & \cdots & \cdots & \cdots & j & \cdots & 2m \end{matrix} & \text{indexes} \\
\begin{bmatrix}
1 & & & & & & & & & -1 & & \\
 & \ddots & & & & & & & & & \ddots & \\
 & & 1 & & & & & & & & & -1 \\
 & & & 1 & & & & & & & & \\
 & & & & \ddots & & & & & & & \\
 & & & & & \ddots & & & & & & \\
 & & & & & & 1 & & & & & \\
-1 & & & & & & & 1 & & & & \\
 & \ddots & & & & & & & \ddots & & & \\
 & & \ddots & & & & & & & \ddots & & \\
 & & & \ddots & & & & & & & \ddots & \\
 & & & & -1 & & & & & & & 1
\end{bmatrix} &
\begin{matrix} 0 \\ \vdots \\ 2m-j \\ 2m-j+1 \\ \vdots \\ \vdots \\ k-j-1 \\ k-j \\ \vdots \\ \vdots \\ \vdots \\ 2m \end{matrix}
\end{array}$$

The rank of A is a least[a] $m+1$. Let r be this rank and let L be the r-dimensional subspace of $\mathbb{R}^{2m+1}$ which is the image of A. Let $\pi : \mathbb{R}^{2m+1} \to L$ be the orthogonal projector onto L. Let $F : U_i \to L$ defined by $x \mapsto \pi(\Phi_{(\phi,y')}(x) - \Phi_{(\phi,y')}(\phi^j x))$. F is C^2 from the m-dimensional manifold U_i to the r-dimensional manifold L (with $r > m$). So we can apply lemma 1 to say that there are vectors of arbitrarily small norm in L which are not contained in the image of U_i. Let b be such a vector. If V is the orthogonal complement of the null space of A, there is a unique[b] $b' \in V$ such that $b = Ab'$. We can arrange for the norm of b' to be arbitrarily small by choosing a b with a sufficiently small norm.

Let $a = -b'$. Then $\forall x \in \overline{X}_i, \Phi_{(\phi,y'')}(x) \neq \Phi_{(\phi,y'')}(\phi^j x)$. To see that, assume that $\Phi_{(\phi,y'')}(x) - \Phi_{(\phi,y'')}(\phi^j x) = 0$. We have shown that $\Phi_{(\phi,y'')}(x) - \Phi_{(\phi,y'')}(\phi^j x) = \Phi_{(\phi,y')}(x) - \Phi_{(\phi,y')}(\phi^j x) - Ab' = \Phi_{(\phi,y')}(x) - \Phi_{(\phi,y')}(\phi^j x) - b$. Then, using that $\pi(b) = b$ (since $b \in L$), we obtain $\pi(\Phi_{(\phi,y')}(x) - \Phi_{(\phi,y')}(\phi^j x)) - b = 0$ which can be rewritten $b = F(x)$. Since $x \in \overline{X}_i \subset U_i$, this means that $b \in FU_i$ which contradict the definition of b.

Since $\overline{X}_i$ is compact, we can apply (a slightly different version of) proposition 6 to say that the functions y'' which verify $\forall x \in \overline{X}_i, \Phi_{(\phi,y'')}(x) \neq \Phi_{(\phi,y'')}(\phi^j x)$ form an open set of $C^2(M, \mathbb{R})$. Hence, we can make a series of adjustments as described in the previous paragraphs each of which establishes the lemma on one of the $\overline{X}_i$. After a finite number of adjustments, we generate a function y'' that satisfy $\forall x \in T, \Phi_{(\phi,y'')}(x) \neq \Phi_{(\phi,y'')}(\phi^j x)$. Since T is compact (as a closed subset of the compact M) we can make further adjustments for every $1 \leq j \leq 2m$ while retaining the property on previously treated j. We obtain a y'' that is injective on the orbit segment of any $x \in T$. These adjustments can be made sufficiently small so that $\Phi_{(\phi,y'')}$ is still an injective immersion on V_y. □

[a] why?

[b] again, why?

Lemma 9

Let y'' be a function satisfying lemma 8.
There exists a number $\delta > 0$ such that $\forall x, x' \in M, x \neq x', \forall 0 \leq i, j \leq k$

$$(\rho(\phi^i x, \phi^j x') < \delta) \Rightarrow (\Phi_{(\phi,y'')}(x) \neq \Phi_{(\phi,y'')}(x'')).$$

Proof: Lemma 9

By contradiction.
Let $\delta_n \to 0$ be a sequence of positive numbers tending to zero.
If the lemma were not true, we should be able to find for all n a pair of points $x_n, x'_n, x_n \neq x'_n$, and integers $i_n, j_n, 0 \leq i_n, j_n, \leq k$ such that $\rho(\phi^{i_n} x_n, \phi^{j_n} x'_n) < \delta_n$ and $\Phi_{(\phi,y'')}(x_n) = \Phi_{(\phi,y'')}(x'_n)$.
Since M is compact, there exist subsequences $\{x_n\}$ and $\{x'_n\}$ which have limits, say x and x'. Since the number of values that i_n can take is finite, we can find an infinite subsequence in which all the i_n's have the same value i, and similarly $j_n = j$ for all n. By continuity, $\phi^i x_n \to \phi^i x$ and $\phi^j x'_n \to \phi^j x'_n$.
Since $\rho(\phi^{i_n} x_n, \phi^{j_n} x'_n) \to 0$, we see that $\phi^i x = \phi^j x'$ meaning that one of the x, x' lies on the orbit segment of the other. By continuity, we also see that $\Phi_{(\phi,y'')}(x) = \Phi_{(\phi,y'')}(x')$ so by lemma 8, it follows that $x = x'$.
Since x_n and x'_n tend to the same limit, we can find n large enough so that $\rho(x_n, x'_n) < \epsilon$ (from lemma 7). But then, we have that $x_n \neq x'_n, \rho(x_n, x'_n) < \epsilon$ and $\Phi_{(\phi,y'')}(x_n) = \Phi_{(\phi,y'')}(x'_n)$ which is a contradiction. □

As usual, the property of y'' are shared by all the measurement functions in an open neighbourhood of y''. And y'' is constructed such that it lies in $\mathcal{U}'_y$. To prove injectivity for points separated by larger distances, we rely on a different approach presented in the next section.

9 Denseness of the set of measurement functions that embed M

The approach followed in [1] to prove the injectivity is to consider the product space $M \times M$ and to remove the diagonal $\Delta' = \{(x, x) : x \in M\}$. Then, $\Phi_{(\phi,y'')}$ is injective if and only if the image of $(M \times M) \setminus \Delta'$ under the function $(x, x') \mapsto (\Phi_{(\phi,y'')}(x), \Phi_{(\phi,y'')}(x'))$ does not intersect the diagonal of $\mathbb{R}^{2m+1} \times \mathbb{R}^{2m+1}$. This is proved using considerations of transversality[1].

[1] but what is transversality?

In this document, we follow a slightly different approach. We aim to show that the image of $(M \times M) \setminus \Delta'$ under the map $(x, x') \mapsto (\Phi_{(\phi,y'')}(x) - \Phi_{(\phi,y'')}(x'))$ does not contain 0, which is equivalent to $\Phi_{(\phi,y'')}$ being injective. The idea is to consider the image of $M \setminus \mathrm{int}V_y$ under the maps $\phi, \phi^2, \ldots, \phi^{2m}$. We take the union of these images, then find a covering of it and make a partition of unity, so that we can apply equation (1) and adjust the measurement function to obtain a injective delay map on M.

9.1 Constructing a suitable cover

What we have so far is a y'' arbitrarily close to y such that:

- $\Phi_{(\phi,y'')}$ is an injective immersion of V_y.
- $\forall x, x' \in M, x \neq x', \forall 0 \leq i, j \leq k, \rho(\phi^i x, \phi^j x') < \delta \Rightarrow \Phi_{(\phi,y'')}(x) \neq \Phi_{(\phi,y'')}(x')$.

The set $M \setminus \mathrm{int}V_y$ is compact (because it is closed in the compact set M), and so is the set Z defined by

$$Z = \cup_{j=0}^{2m} \phi^j (M \setminus \mathrm{int}V_y)$$

We now construct a cover of Z. Note that for every $x \in Z, x \notin P_{2m}$, so that the points $x, \phi^{-1}x, \ldots, \phi^{-2m}x$ are all distinct. Hence, we can find a open set U_x containing x such that $U_x, \phi^{-1}U_x, \ldots, \phi^{-2m}U_x$ are all disjoint. By taking smaller U_x if necessary, we can ensure that $U_x \subset b(x, \delta/2)$. From such cover of Z, $\{U_x : x \in Z\}$, we extract a finite subcover $\{U_l : l = 1, \ldots, N\}$.

9.2 A partition of unity

We build a partition of unity $\{(\psi_l : M \to \mathbb{R})\}$ on Z subordinate to this cover. For $\epsilon \in \mathbb{R}^N$, define

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

Our goal is to adjust y'' to shift the image of $(x, x') \mapsto (\Phi_{(\phi,y'')}(x) - \Phi_{(\phi,y'')}(x'))$ away from 0 if it contains it. To achieve that, we will show that the codomain of $\Phi_{(\phi,y'')}$ needs to be of dimension greater or equal to $2m + 1$. Define $W \subset M \times M$ by

$$W = \{(x, x') : \forall 0 \leq i, j \leq k, \rho(\phi^i x, \phi^j x') \geq \delta \text{ and } (x, x') \notin \mathrm{int}V_y \times \mathrm{int}V_y\}.$$

Let us show that W is closed. Define $W'_{i,j} = \{(x, x') : \rho(\phi^i x, \phi^j x) \geq \delta\}$. $W'_{i,j}$ is closed by continuity of $(x, x') \mapsto \rho(\phi^i x, \phi^j x)$. Then $W' = \cap W'_{i,j}$ is closed as intersection of closed sets. Then $W = W' \cap ((M \times M) \setminus (\mathrm{int}V_y \times \mathrm{int}V_y))$ is closed as intersection of a closed set with the complement of an open set. By working on W instead of $M \times M$ directly, we remove pathological points (in particular Δ') where the construction below fails to work. These points were dealt with in the three previous sections.

Consider the map $\Psi : M \times M \times \mathbb{R}^N \to \mathbb{R}^{2m+1}$ defined by

$$\Psi(x, x', \epsilon) = \Phi_{\phi,y_\epsilon}(x) - \Phi_{\phi,y_\epsilon}(x')$$

We want to find ϵ with arbitrarily small norm such that Ψ sends $W \times \{\epsilon\}$ into $\mathbb{R}^{2m+1} \setminus \{0\}$.

9.3 Showing that Ψ is submersive

First, we show that Ψ is submersive at all the points in the set $W \times \{0\}$. Let $\{(V_p, h_p)\}$ be an atlas for M. Then, $\{(V_p \times V_q \times \mathbb{R}^N, g_{p,q})\}$ where $g_{p,q}(x, x', \epsilon) = (h_p(x), h_q(x'), \epsilon)$ is an atlas for $M \times M \times \mathbb{R}^N$. Let us write $u = h_p(x)$ and $u' = h_q(x')$. The derivative $D\Psi g_{p,q}^{-1}(u, u', 0)$ can be written as a block matrix:

$$\begin{bmatrix} \text{dimensions} & m & m & N \\ 2m+1 & D\Phi_{(\phi,y'')} h_p^{-1}(u) & -D\Phi_{(\phi,y'')} h_q^{-1}(u') & A(x) - A(x') \end{bmatrix}$$

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

$$\begin{aligned}
A_{il}(x) &= \frac{\partial r_i \Psi h_p^{-1}}{\partial \epsilon_l}(h_p(x)), \\
&= \frac{\partial y_\epsilon \phi^{i-1} h_p^{-1}}{\partial \epsilon_l}(h_p(x)), \\
&= \psi_l \phi^{i-1}(x) \text{ which does not depend on the chart map } h_p \text{ (or } h_q \text{ for } x'),
\end{aligned}$$

where r_i is the i-th coordinate function of $(\mathbb{R}^{2m+1}, id)$ and ϵ_l is the l-th coordinate function of $(\mathbb{R}^N, id)$. Since we want to establish injectivity by adjusting ϵ, we focus on the submatrix $A(x) - A(x')$. First, note that each *column* of $A(x) - A(x')$ has *at most* one non-zero element. By contradiction, assume there are distinct i, j such that $A_{il}(x) - A_{il}(x') \neq 0$ and $A_{jl}(x) - A_{jl}(x') \neq 0$. Then, at least one of $A_{il}(x), A_{il}(x')$ is non-zero, and one one $A_{jl}(x), A_{jl}(x')$ is non-zero. If $A_{il}(x) \neq 0$ and $A_{jl}(x) \neq 0$, then $\psi_l \phi^{i-1}(x) \neq 0$ and $\psi_l \phi^{j-1}(x) \neq 0$ and both $\phi^{i-1}(x)$ and $\phi^{j-1}(x)$ are in the support of ψ_l which is itself a subset of U_l. Hence $\{x\} \subset (\phi^{-(i-1)} U_l) \cap (\phi^{-(j-1)} U_l)$ which is impossible since U_l has been constructed so that $(\phi^{-(i-1)} U_l) \cap (\phi^{-(j-1)} U_l) = \varnothing$. If $A_{il}(x) \neq 0$ and $A_{jl}(x') \neq 0$, then $\phi^{i-1}(x) \in U_l$ and $\phi^{j-1}(x') \in U_l$ implying that $\rho(\phi^{i-1}(x), \phi^{j-1}(x')) < \delta$ which is impossible since $(x, x') \in W$. We use the same arguments to show that the two remaining possibilities are impossible.

Now, we show that every *row* has *at least* one non-zero element. Since $(x, x') \in W$, then at least one of x, x' is in $M \setminus \text{int} V_y$. Without loss of generality, assume that $x \in M \setminus \text{int} V_y$. Then, $\forall 1 \leq i \leq 2m+1, \phi^{i-1}(x) \in Z$. Hence $\sum_{l=1}^N \psi_l \phi^{i-1}(x) = 1$ (since $\{\psi_l\}$ is a partition of unity), then $\forall 1 \leq i \leq 2m+1$ there exists an l such that $\psi_l \phi^{i-1}(x) = A_{il}(x) \neq 0$. But if $A_{il}(x) \neq 0$ then $A_{il}(x') = 0$, otherwise both $\phi^{i-1}(x)$ and $\phi^{i-1}(x')$ would lie in the support of ψ_l and hence in a ball of radius $\delta/2$ which would imply that $(x, x') \notin W$. Hence, $A_{il}(x) - A_{il}(x') \neq 0$.

Since $A(x) - A(x')$ has at least one non-zero element in every row, but at most one non-zero element in every column, the matrix must have at least as many columns as rows and it must be full rank. This last fact can be shown by developing according to a column with one non-zero element, then developing the sub-determinant according to another column with one non-zero element, and so on until there is no remaining column with a non-zero element. This will only happen after at least the number of rows since there is at least one non-zero element in each row. Hence, the rank is the number of rows, and since the number of rows is less or equal to the number of columns, the matrix is full rank. Hence, the rank of $D\Psi g_{p;q}^{-1}(u, u', 0)$ must be $2m+1$ and Ψ is submersive at $(x, x', 0)$. By the Constant Rank Theorem, there exists a neighbourhood of $(x, x', 0)$ on which the derivative is full rank. These open sets (for every (x, x')) form a cover of $W \times \{0\}$ and their union is an open set X such that the restriction of Ψ to X is a submersion. Applying lemma 7, there exists an $\eta > 0$ such that every closed ball of radius η or less, centered at any point in $W \times \{0\}$, is contained in X. In particular, $\forall \epsilon \in \mathbb{R}^N, \|\epsilon\| \leq \eta \Rightarrow W \times \{\epsilon\} \subset X$.

9.4 Injectivity on W

Since $\Psi|_X : X \to \mathbb{R}^{2m+1}$ is a submersion, we can apply lemma 2 and say that $\Psi|_X^{-1}(0)$ is a submanifold of X of dimension $2m + N - (2m+1) = N - 1$. Let $\pi : X \to \mathbb{R}^N$ be the projection defined by $(x, x', \epsilon) \mapsto \epsilon$, and let $\hat{\pi}$ be its restriction to $\Psi|_X^{-1}(0)$. Suppose that there is some ϵ with $\|\epsilon\| < \eta$ (so that $\forall (x, x') \in W, (x, x', \epsilon) \in X$) which is *not* is the range of $\hat{\pi}$. This would mean that $\forall (x, x') \in W, (x, x', \epsilon) \notin \Psi|_X^{-1}(0)$ or in other words $\forall (x, x') \in W, \Phi_{(\phi, y_\epsilon)}(x) \neq \Phi_{(\phi, y_\epsilon)}(x')$. Now, since $\hat{\pi} : \Psi|_X^{-1}(0) \to \mathbb{R}^N$ is a C^1 map from a manifold of dimension $N-1$ to a manifold of dimension N, we can apply lemma 1 to say that $\mathbb{R}^N - \hat{\pi}\Psi|_X^{-1}(0)$ is dense in $\mathbb{R}^N$. In other words, the ϵ's that are not in the range of $\hat{\pi}$ are dense in $\mathbb{R}^{\mathbb{N}}$. In particular, we can find such ϵ's with arbitrarily small norm. Note that for lemma 1 to apply, the dimension of $\Psi|_X^{-1}(0)$ must be smaller than N. In other words, the number of delays N' must verify $\dim(X) - N' < N$. Since $\dim(X) = 2m + N$, we must have $N' > 2m$, so at least $N' = 2m + 1$. This is the reason for the assumption on the number of delays in theorem 2 and theorem 1. We could imagine that, by relying on a different argument, Takens' theorem could proven for a smaller number of delays. However, some counterexamples[2] show that this number is indeed the minimum.

[2] which one?

9.5 Injectivity on M

Now, we prove that we can find ϵ such that $\Phi_{(\phi,y_\epsilon)}$ is injective on M. Let ϵ be such that the image of $W \times \{\epsilon\}$ under Ψ does not contain 0. The pairs (x,x') that are in $M \times M$ but not in W are those for which either $\rho(\phi^i x, \phi^j x') < \delta$ for some $0 \leq i,j \leq k$, or either $(x,x') \in \mathrm{int}V_y \times \mathrm{int}V_y$ (by definition of W). If $(x,x') \in \mathrm{int}V_y \times \mathrm{int}V_y$ then by lemma 6, if we have chosen a small enough ϵ (and we can since ϵ has arbitrarily small norm), then $y_\epsilon \in \mathcal{U}'_y$ and by sections 6 and 7, $\Phi_{(\phi,y_\epsilon)}$ is an immersion of M and an injection of V_y. If $\rho(\phi^i x, \phi^j x') < \delta$ for some $0 \leq i,j \leq k$, then again by lemma 6 (and choosing a smaller ϵ if necessary), the function y_ϵ verifies lemma 8 from section 8 and lemma 9 tell us that $\Phi_{(\phi,y_\epsilon)}$ is injective on the set $\{(x,x') : \rho(\phi^i x, \phi^j x') < \delta \text{ for some } 0 \leq i,j \leq k\}$.

9.6 Conclusion of the proof of Taken's theorem

Hence, for small enough ϵ, $\Phi_{(\phi,y_\epsilon)}$ is injective on M, and it is an immersion of M (by section 7). So $\Phi_{(\phi,y_\epsilon)}$ is an embedding of M. ϵ can be chosen sufficiently small to ensure that $y_\epsilon \in \mathcal{N}$ (where $\mathcal{N}$ is an arbitrary neighbourhood of our original (and arbitrary) measurement function y, both introduced in section 5). So, we an conclude than the set of measurement functions that give rise to embedding delay maps is dense in $C^2(M,\mathbb{R})$. This fact, together with proposition 6, allows us to say that for generic measurement functions y, $\Phi_{(\phi,y)}$ is an embedding of M, which complete the proof of theorem 2. Since, according to section 4, theorem 2 implies theorem 1, the proof of Takens' theorem is also complete.

References

[1] F. Takens, "Detecting strange attractors in turbulence," *Dynamical Systems and Turbulence*, vol. 898, pp. 366–381, 1981.

[2] J. P. Huke, "Embedding nonlinear dynamical systems: A guide to takens' theorem," tech. rep., Manchester Institute for Mathematical Sciences, 1993.