Nonlinearity

PAPER

Learning strange attractors with reservoir systems

To cite this article: Lyudmila Grigoryeva *et al* 2023 *Nonlinearity* **36** 4674

IOP Publishing | London Mathematical Society Nonlinearity

Nonlinearity **36** (2023) 4674–4708 https://doi.org/10.1088/1361-6544/ace492

Learning strange attractors with reservoir systems

Lyudmila Grigoryeva[1,2], **Allen Hart**[3,5] **and Juan-Pablo Ortega**[4,*]

[1] Faculty of Mathematics and Statistics, University of St. Gallen, Bodanstrasse 6, CH-9000 St. Gallen, Switzerland
[2] Department of Statistics, University of Warwick, Coventry CV4 7AL, United Kingdom
[3] Department of Mathematical Sciences, University of Bath, Bath BA2 7AY, United Kingdom
[4] Division of Mathematical Sciences, School of Physical and Mathematical Sciences, Nanyang Technological University, 21 Nanyang Link, Singapore 637371, Singapore
[5] Department of Engineering, University of Exeter, Exeter, EX4 4PY, United Kingdom

E-mail: Juan-Pablo.Ortega@ntu.edu.sg

Received 8 May 2022; revised 9 June 2023
Accepted for publication 5 July 2023
Published 27 July 2023

Recommended by Dr Jean-Luc Thiffeault

Abstract

This paper shows that the celebrated embedding theorem of Takens is a particular case of a much more general statement according to which, randomly generated linear state-space representations of generic observations of an invertible dynamical system carry in their wake an embedding of the phase space dynamics into the chosen Euclidean state space. This embedding coincides with a natural generalized synchronization that arises in this setup and that yields a topological conjugacy between the state-space dynamics driven by the generic observations of the dynamical system and the dynamical system itself. This result provides additional tools for the representation, learning, and analysis of chaotic attractors and sheds additional light on the reservoir computing phenomenon that appears in the context of recurrent neural networks.

* Author to whom any correspondence should be addressed.

Keywords: dynamical systems, reservoir computing, generalized synchronization, attractor, Takens embedding, echo state property, fading memory property

Mathematics Subject Classification numbers: 37D45, 68T05, 57R40, 62M45

(Some figures may appear in colour only in the online journal)

1. Introduction

Takens' theorem [Take 81] and the associated method of delays have been used and studied for decades as they are powerful tools in the reconstruction of qualitative features of a dynamical system out of time series of low dimensional observations. This result is also at the origin of the development of powerful forecasting tools [Saue 91, Kant 03].

In order to put these results in context and to better motivate the contributions in this paper, we start by recalling Huke's formulation [Huke 06] of Takens' Theorem.

Theorem 1.1 (Takens). *Let M be a compact manifold of dimension $q \in \mathbb{N}$ and let $\phi \in \mathit{Diff}^2(M)$ be a twice-differentiable diffeomorphism that satisfies the following two properties:*

(i) *ϕ has only finitely many periodic points with periods less than or equal to $2q$.*
(ii) *If $m \in M$ is any periodic point of ϕ with period $k < 2q$, then the eigenvalues of the linear map $T_m\phi^k : T_mM \longrightarrow T_mM$ are distinct.*

Then for any generic scalar observation function $\omega \in C^2(M, \mathbb{R})$, the $(2q+1)$-delay map $\Phi_{(\phi,\omega)} : M \longrightarrow \mathbb{R}^{2q+1}$ defined by

$$\Phi_{(\phi,\omega)}(m) := \left(\omega(m), \omega \circ \phi(m), \omega \circ \phi^2(m), \ldots, \omega \circ \phi^{2q}(m)\right) \tag{1.1}$$

is an embedding in $C^1(M, \mathbb{R}^{2q+1})$.

The first consequence of this result is that, since the map $\Phi_{(\phi,\omega)}$ is an embedding, then it is necessarily injective and hence it can be used to represent in $\Phi_{(\phi,\omega)}(M) \subset \mathbb{R}^{2q+1}$ the dynamics induced by ϕ on M via the differentiable map $\varphi_{(\phi,\omega)} := \Phi_{(\phi,\omega)} \circ \phi \circ \Phi_{(\phi,\omega)}^{-1} : \Phi_{(\phi,\omega)}(M) \subset \mathbb{R}^{2q+1} \longrightarrow \Phi_{(\phi,\omega)}(M) \subset \mathbb{R}^{2q+1}$ (we recall that the inverse function theorem guarantees that the map $\Phi_{(\phi,\omega)}^{-1} : \Phi_{(\phi,\omega)}(M) \longrightarrow M$ is differentiable). In view of the expression (1.1), this map takes necessarily the form $\varphi_{(\phi,\omega)}(z_1, \ldots, z_{2q+1}) := (z_2, z_3, \ldots, h(z_1, \ldots, z_{2q+1}))$, for some differentiable map $h : \Phi_{(\phi,\omega)}(M) \subset \mathbb{R}^{2q+1} \longrightarrow \mathbb{R}$. In this situation, we say that the dynamical systems (M, ϕ) and $(\Phi_{(\phi,\omega)}(M), \varphi_{(\phi,\omega)})$ are topologically conjugate by the map $\Phi_{(\phi,\omega)}$. The importance of this representation is that the two systems (M, ϕ) and $(\Phi_{(\phi,\omega)}(M), \varphi_{(\phi,\omega)})$ have the same C^1 invariants like Lyapunov exponents, eigenvalues of linearizations, or dimensions of attractors and their computation may be more efficiently carried out in $\Phi_{(\phi,\omega)}(M) \subset \mathbb{R}^{2q+1}$.

More recently, the remarkable success of recurrent neural networks and reservoir computing (RC) [Luko 09, Tana 19] in the learning, forecasting [Jaeg 04, Path 17, Path 18, Lu 18], and classification [Carr 18] of chaotic attractors of complex nonlinear high-dimensional dynamical systems strongly suggests that these machine learning paradigms have Takens embedding-type properties. This fact has been rigorously established in [Hart 20, Hart 21] where the so-called *Echo State Networks* [Matt 92, Matt 93, Jaeg 04, Grig 18, Gono 20b, Gono 21] driven by one-dimensional observations of a given dynamical system on a compact manifold have been shown, under certain hypotheses, to produce dynamics that are topologically conjugate to that of the original system.

A concept that unifies the recurrent networks and the Takens approaches to the representation of dynamical systems is that of ***generalized synchronization*** (GS), as introduced in [Rulk 95] (see [Peco 97, Ott 02, Bocc 02, Erog 17] for self-contained presentations and many references). GSs represent dynamical systems in the space of states $\mathbb{R}^N$ of a state-space map $F:\mathbb{R}^N\times\mathbb{R}^d\longrightarrow\mathbb{R}^N$, $N,d\in\mathbb{N}$. More specifically, let (M,ϕ) be the same dynamical systems as above, with M compact and $\phi\in\mathrm{Diff}^1(M)$. Let $\omega\in C^1(M,\mathbb{R}^d)$, $d\in\mathbb{N}$, be a map that encodes d-dimensional observation function of the dynamical system (in different parts of our presentation we consider purely scalar observation functions ω as in theorem 1.1, which we shall make clear in the discussion), and define the (ϕ,ω)-*delay map* $S_{(\phi,\omega)}:M\longrightarrow\ell^\infty(\mathbb{R}^d)$ as $S_{(\phi,\omega)}(m):=\{\omega(\phi^t(m))\}_{t\in\mathbb{Z}}$. Consider now the drive-response system associated to the ω-observations of ϕ and determined by the recursions:

$$\mathbf{x}_t=F\left(\mathbf{x}_{t-1},S_{(\phi,\omega)}(m)_t\right),\quad t\in\mathbb{Z},m\in M. \tag{1.2}$$

We say that a map $f_{(\phi,\omega,F)}:M\longrightarrow\mathbb{R}^N$ is a GS between the dynamical system (M,ϕ) and its representation in the space of states of the state map $F:\mathbb{R}^N\times\mathbb{R}^d\longrightarrow\mathbb{R}^N$ driven by the observations $\omega\in C^1(M,\mathbb{R}^d)$ of $\phi\in\mathrm{Diff}^1(M)$, when for any $\mathbf{x}_t$ given by (1.2)

$$\mathbf{x}_t=f_{(\phi,\omega,F)}(\phi^t(m)),\quad t\in\mathbb{Z}. \tag{1.3}$$

When a GS exists, the time evolution of the dynamical system in phase space (not just its observations) drives the response in (1.2). We emphasize that the definition (1.3) presupposes that the recursions (1.2) have a (unique) solution, that is, that there exists a sequence $\mathbf{x}\in\ell^\infty(\mathbb{R}^N)$ such that (1.2) holds true. When that existence property holds and, additionally, the solution sequence $\mathbf{x}$ is unique, we say that F has the (ϕ,ω)-*Echo State Property* (ESP) (see [Jaeg 10, Manj 13, Manj 20] for in-depth descriptions of this property). Moreover, in the presence of the (ϕ,ω)-ESP, the state map F determines a unique causal and time-invariant filter $U^F:S_{(\phi,\omega)}(M)\longrightarrow(\mathbb{R}^N)^{\mathbb{Z}}$ that associates to each orbit $S_{(\phi,\omega)}(m)$ the unique solution sequence $\mathbf{x}\in(\mathbb{R}^N)^{\mathbb{Z}}$ of (1.2). The existence, continuity, and differentiability of GSs have been established in [Grig 21a] for a rich class of systems that exhibit the so-called fading memory property and that are generated by locally state-contracting maps F.

The relevance of these concepts in relation to the embedding of dynamical systems lays in the fact that Takens' Theorem can be easily reformulated in the language of GSs. Indeed, we first note that the map $\Phi_{(\phi,\omega)}$ for the dynamical system (M,ϕ) with $\phi\in\mathrm{Diff}^2(M)$ and $\omega\in C^2(M,\mathbb{R})$, and introduced in (1.1) is the GS $f_{(\phi,\omega,F)}:M\longrightarrow\mathbb{R}^{2q+1}$ between ϕ and its representation in $\mathbb{R}^{2q+1}$ corresponding to the linear state map $F(\mathbf{x},z):=A\mathbf{x}+\mathbf{C}z$, with A the lower shift matrix in dimension $2q+1$ and $\mathbf{C}=(1,0,\ldots,0)^\top\in\mathbb{R}^{2q+1}$. Takens' Theorem can now be stated by saying the GS $\Phi_{(\phi,\omega)}$ is an embedding for any generic scalar observation function $\omega\in C^2(M,\mathbb{R})$.

The main result in this paper shows that Takens' Theorem is a particular case of a more general statement that ensures that the GSs associated to generic randomly generated linear state-space systems of the type $F(\mathbf{x},z):=A\mathbf{x}+\mathbf{C}z$, *with* $A\in\mathbb{M}_{N,N}$, $\mathbf{C}\in\mathbb{R}^N$, *and* $N\geqslant 2q+1$, *and driven by generic observations* $\omega\in C^2(M,\mathbb{R})$ *are embeddings.*

The term *generic* is used in the previous statement with two different meanings. First, when we talk about *generic* randomly generated linear state-space systems, we mean that the embedding condition holds almost surely when $A\in\mathbb{M}_{N,N}$ and $\mathbf{C}\in\mathbb{R}^N$ are randomly drawn with respect to some probability distribution in a subset of the spaces where those elements are defined. Second, when we write *generic* observations $\omega \in C^2(M,\mathbb{R})$, we mean that they belong to an open and dense subset of $C^2(M,\mathbb{R})$ with respect to a Banach topology in that space that we define later on in the paper.

An important consequence of this result is that it sheds light on the good performance of RC [Jaeg 04, Luko 09, Tana 19] in the forecasting of dynamical systems. We recall that RC (also found in the literature under other denominations like *Liquid State Machines* [Maas 00, Maas 02, Nats 02, Maas 04, Maas 07]) capitalizes on the idea that there are randomly generated systems that attain universal approximation properties without the need to estimate all their parameters. RC has shown unprecedented abilities in the learning of the attractors of complex nonlinear infinite dimensional dynamical systems [Jaeg 04, Path 17, Path 18, Lu 18] and has given rise to forecasting techniques that outperform standard Takens-based strategies.

Our results explicitly contribute in relation to the RC phenomenon by showing that the dynamics of generic observations of invertible dynamical systems is almost surely learnable using randomly generated linear reservoir systems with nonlinear readouts (unlike what is common practice in the RC literature, where readouts are linear). Indeed, let $f_{(\phi,\omega,F)}: M \longrightarrow \mathbb{R}^N$ be a GS associated with a randomly generated linear state-space system that, as above, is driven by generic scalar observations $\omega \in C^2(M,\mathbb{R})$ of $\phi \in \mathrm{Diff}^2(M)$. Since our results show that $f_{(\phi,\omega,F)}$ is an embedding, it then has an inverse and we can hence construct the readout $h := \omega \circ \phi \circ f^{-1}_{(\phi,\omega,F)} : f(M) \subset \mathbb{R}^N \longrightarrow \mathbb{R}$ that, applied to the states $\mathbf{x}_t$ determined by (1.3) fully characterize the dynamics of the ω-observations $\{\omega(\phi^t(m))\}_{t\in\mathbb{Z}}$ of ϕ because $h(\mathbf{x}_t) = \omega \circ \phi(f^{-1}_{(\phi,\omega,F)}(\mathbf{x}_t)) = \omega(\phi^{t+1}(m))$. This observation implies that this dynamics can be captured via the learning of the function $h := \omega \circ \phi \circ f^{-1}_{(\phi,\omega,F)}$. This is what we call *learnability* (see, for instance, [Lu 20, Verz 21, Gaut 21]). We emphasize that the regularity properties of the map $f_{(\phi,\omega,F)}$ that we establish later on in the paper guarantee that the readout $h := \omega \circ \phi \circ f^{-1}_{(\phi,\omega,F)}$ can be efficiently approximated by a universal family (for instance neural networks or polynomials) and explains the good performance of this methodology in the applications cited above.

The paper is organized as follows. Section 2 contains a first introduction to the connection between GSs and embeddings and provides existence and regularity statements in the linear case (mainly proposition 2.3) that are used later on in the paper. Section 3 introduces and proves theorem 3.1, which establishes sufficient conditions for a linear system to yield immersive GSs for generic observation maps. In section 4 we show first (theorem 4.1) that basically without additional hypotheses, the globally immersive GSs whose existence was proved in theorem 3.1 are injective and hence are necessarily embeddings due to the compactness of M. Finally, it is also shown (theorem 4.2) in this section that randomly generated linear systems (linear reservoirs) yield synchronization maps $f_{(\phi,\omega,F)} \in C^2(M,\mathbb{R}^N)$ that are almost surely embeddings and are hence amenable to learnability from data. Section 5 contains a series of numerical illustrations that show the pertinence of the proposed results for attractor reconstruction, filtering in the presence of noise, and forecasting.

2. Definitions and preliminary discussion

All along this paper we consider an invertible and discrete-time dynamical system determined by a map ϕ that belongs to the set of diffeomorphisms $\mathrm{Diff}^1(M)$ of a finite-dimensional compact manifold M. Since later on we need to ensure that M can be endowed with a Riemannian metric g, we additionally assume that M is connected, Hausdorff, and second-countable (see [Carm 92, proposition 2.10]). The d-dimensional observations of the dynamical system are realized by maps ω that belong, most of the time, to $C^1(M,\mathbb{R}^d)$. The symbol TM denotes the tangent bundle of M, $T\phi: TM \longrightarrow TM$ the tangent map of ϕ, and $D\omega: TM \longrightarrow \mathbb{R}^d$ the differential of the observation map ω. Now, for any $f \in C^1(M,\mathbb{R}^N)$, define

$$\|Df\|_\infty = \sup_{m\in M}\{\|Df(m)\|\} \quad \text{with} \quad \|Df(m)\| = \sup_{\substack{\mathbf{v}\in T_mM\\ \mathbf{v}\neq \mathbf{0}}}\left\{\frac{\|Df(m)\cdot\mathbf{v}\|}{(g(m)(\mathbf{v},\mathbf{v}))^{1/2}}\right\},$$

where g is a Riemannian metric tensor on M which defines at each point $m\in M$ a positive-definite inner product $g(m): T_mM\times T_mM \to \mathbb{R}$ on each of the tangent spaces T_mM of M.

Analogously, if $\phi: M\to M$ is a C^1 map, we can define:

$$\|T\phi\|_\infty = \sup_{m\in M}\{\|T_m\phi\|\} \quad \text{with} \quad \|T_m\phi\| = \sup_{\substack{\mathbf{v}\in T_mM\\ \mathbf{v}\neq \mathbf{0}}}\left\{\frac{(g(\phi(m))(T_m\phi\cdot\mathbf{v},T_m\phi\cdot\mathbf{v}))^{1/2}}{(g(m)(\mathbf{v},\mathbf{v}))^{1/2}}\right\}.$$

It can be proved by using the results in chapter 2 of [Abra 67] that the norm $\|\cdot\|_{C^1}$ defined by

$$\|f\|_{C^1} := \|f\|_\infty + \|Df\|_\infty \tag{2.1}$$

endows $C^1(M,\mathbb{R}^N)$ with a Banach space structure. Additionally, (see [Abra 67, theorem 11.2 **(ii)**]) this norm generates a topology in $C^1(M,\mathbb{R}^N)$ that is independent of the choice of Riemannian metric and coincides with the weak and strong topologies introduced in chapter 2 of [Hirs 76]. These notions can be extended to higher-order differentiable maps in a straightforward manner. We emphasize that the embedding results obtained later on in this paper do not depend on the choice of g on M.

The embeddings that are at the core of this paper will be constructed using GSs associated with linear systems. That is why we start by recalling a result proved in [Grig 21a] in relation to the existence of these objects in a rich variety of situations. The statement requires the following constants defined with respect to a subset $V\subset\mathbb{R}^N$:

$$\begin{aligned} L_{F_x} &:= \sup_{(\mathbf{x},\mathbf{z})\in V\times\omega(M)}\{\|D_xF(\mathbf{x},\mathbf{z})\|\}, & L_{F_z} &:= \sup_{(\mathbf{x},\mathbf{z})\in V\times\omega(M)}\{\|D_zF(\mathbf{x},\mathbf{z})\|\},\\ L_{F_{xx}} &:= \sup_{(\mathbf{x},\mathbf{z})\in V\times\omega(M)}\{\|D_{xx}F(\mathbf{x},\mathbf{z})\|\}, & L_{F_{xz}} &:= \sup_{(\mathbf{x},\mathbf{z})\in V\times\omega(M)}\{\|D_{xz}F(\mathbf{x},\mathbf{z})\|\}, \end{aligned} \tag{2.2}$$

Theorem 2.1 (Existence and uniqueness of differentiable GSs). *Let $\phi\in \mathrm{Diff}^1(M)$ be a dynamical system on the compact manifold M and consider the observation $\omega\in C^1(M,\mathbb{R}^d)$ and state $F\in C^2(D_N\times D_d, D_N)$ maps, with $D_N\subset\mathbb{R}^N$ and $D_d\subset\mathbb{R}^d$ open subsets such that $\omega(M)\subset D_d$. Let $V\subset D_N$ be a closed convex subset and suppose that $F(V\times\omega(M))\subset V$. Suppose that the bounds for the partial derivatives of F introduced in (2.2) are all finite and that, additionally,*

$$L_{F_x} < \min\left\{1, 1/\left\|T\phi^{-1}\right\|_\infty\right\}. \tag{2.3}$$

Then there exists a compact and convex subset $W\subset V$ such that $F(W\times\omega(M))\subset W$ and:

(i) *The system determined by $F: W\times\omega(M)\longrightarrow W$ and driven by the ω-observations of ϕ has the (ϕ,ω)-ESP and a GS $f_{(\phi,\omega,F)}: M\longrightarrow W$ exists and is well-defined by the relation $U^F(S_{(\phi,\omega)}(m))_t = f_{(\phi,\omega,F)}(\phi^t(m))$, for any $t\in\mathbb{Z}$, $m\in M$.*

(ii) *The map $f_{(\phi,\omega,F)}$ belongs to $C^1(M,W)$ and it is the only one that satisfies the identity:*
$$f_{(\phi,\omega,F)}(m) = F\left(f_{(\phi,\omega,F)}(\phi^{-1}(m)),\omega(m)\right), \quad \text{for all} \quad m\in M.$$

If we now consider the linear system

$$F(\mathbf{x}, z) := A\mathbf{x} + \mathbf{C}z, \quad \text{with } A \in \mathbb{M}_{N,N},\ \mathbf{C} \in \mathbb{R}^N,\ N \in \mathbb{N}, \tag{2.4}$$

in the context of the previous theorem, we obtain the following corollary that is a straightforward consequence of the fact that, in this case, $A = D_x F(\mathbf{x}, z)$, for all $\mathbf{x} \in \mathbb{R}^N$ and $z \in \mathbb{R}$, and hence $L_{F_x} = \|A\|$. We shall refer to A as the *connectivity matrix* and to the vector $\mathbf{C}$ as the *input mask*.

Corollary 2.2. *Let $\phi \in \mathrm{Diff}^1(M)$ be a dynamical system on the compact manifold M and consider the observation map $\omega \in C^1(M, \mathbb{R})$. Let $F : \mathbb{R}^N \times \mathbb{R} \longrightarrow \mathbb{R}^N$ be the linear state map given by $F(\mathbf{x}, z) := A\mathbf{x} + \mathbf{C}z$ with $A \in \mathbb{M}_{N,N}$, $\mathbf{C} \in \mathbb{R}^N$, $N \in \mathbb{N}$, such that*

$$\|A\| < \min\left\{1, 1/\left\|T\phi^{-1}\right\|_\infty\right\}. \tag{2.5}$$

(i) *The system determined by $F : \mathbb{R}^N \times \omega(M) \longrightarrow \mathbb{R}^N$ and driven by the ω-observations of ϕ has the (ϕ, ω)-ESP and a GS $f_{(\phi,\omega,F)} : M \longrightarrow \mathbb{R}^N$ given by*

$$f_{(\phi,\omega,F)}(m) = \sum_{j=0}^{\infty} A^j \mathbf{C}\omega(\phi^{-j}(m)). \tag{2.6}$$

(ii) *The map $f_{(\phi,\omega,F)}$ belongs to $C^1(M, \mathbb{R}^N)$ and it is the only one that satisfies the identity:*
$$f_{(\phi,\omega,F)}(m) = Af_{(\phi,\omega,F)}(\phi^{-1}(m)) + \mathbf{C}\omega(m), \quad \textit{for all} \quad m \in M.$$

The features of the linear case allow us to prove the existence of GSs in situations that go beyond those spelled out in theorem 2.1 and corollary 2.2. More specifically, an argument similar to what can be found in proposition 4.2 in [Grig 21b] allows us to drop the compactness condition on the manifold M and to replace the hypotheses on the design matrix A with more general ones based on its spectral radius $\rho(A)$.

Proposition 2.3. *Let $\phi \in \mathrm{Diff}^1(M)$ be a dynamical system on the manifold M (not necessarily compact) and consider the observation map $\omega \in C^1(M, \mathbb{R})$. Let $F : \mathbb{R}^N \times \mathbb{R} \longrightarrow \mathbb{R}^N$ be a linear state map given by $F(\mathbf{x}, z) := A\mathbf{x} + \mathbf{C}z$ with $A \in \mathbb{M}_{N,N}$, $\mathbf{C} \in \mathbb{R}^N$, $N \in \mathbb{N}$.*

(i) *If the spectral radius of A satisfies that $\rho(A) < 1$ and ω maps into a bounded set of $\mathbb{R}$ then the GS $f_{(\phi,\omega,F)} : M \longrightarrow \mathbb{R}^N$ introduced in (2.6) exists and it is a continuous map.*

(ii) *Additionally, let $r \in \mathbb{N}$ and suppose that $\phi \in \mathrm{Diff}^r(M)$ and that there exist constants $k_1, \ldots, k_r \in \mathbb{N}$ such that $\left\|A^{k_i}\right\| \left\|T^i \phi^{-k_i}\right\|_\infty < 1$, $\left\|T^i \phi^{-1}\right\|_\infty < \infty$ for all $i \in \{1, \ldots, r\}$. Then for any $\omega \in C^r(M, \mathbb{R})$ such that $\left\|D^i \omega\right\|_\infty < \infty$, for all $i \in \{1, \ldots, r\}$, the map $f_{(\phi,\omega,F)}$ belongs to $C^r(M, \mathbb{R}^N)$ and the higher order derivatives are given by:*

$$D^i f_{(\phi,\omega,F)}(m) = \sum_{j=0}^{\infty} A^j \mathbf{C} D^i \left(\omega \circ \phi^{-j}\right)(m), \textit{ for all } i \in \{1, \ldots, r\}. \tag{2.7}$$

(iii) *Suppose now that M is compact. In the hypotheses of points **(i)** and **(ii)** above, the map*

$$\begin{array}{cccc} \Theta_{(\phi,F)}: & C^r(M, \mathbb{R}) & \longrightarrow & C^r(M, \mathbb{R}^N) \\ & \omega & \longmapsto & f_{(\phi,\omega,F)} \end{array} \tag{2.8}$$

is continuous. Moreover, the subsets Ω_i and Ω_e of $C^r(M, \mathbb{R})$ for which the corresponding GS are immersions and embeddings, respectively, are open.

Proof. (i) This statement is obtained out of a combination of Weierstrass M-test (see [Apos 74, theorem 9.6]) and Gelfand's formula for the spectral radius (see [Lax 02]), that is, $\lim_{k\to\infty}\left\|A^k\right\|^{1/k}=\rho(A)$. Since by hypothesis $\rho(A)<1$, we can guarantee the existence of a number $k_0\in\mathbb{N}$ such that $\left\|A^{k_0}\right\|<1$, for all $k\geqslant k_0$. Consider now the series $\sum_{j=0}^{\infty}A^j\mathbf{C}\omega(\phi^{-j}(m))$ in (2.6) that defines $f_{(\phi,\omega,F)}(m)$. Given that for any $j\in\mathbb{N}$ there exist $l\in\mathbb{N}$ and $i\in\{0,\ldots,k_0-1\}$ such that $A^j=A^{lk_0+i}$, we then have that,

$$\left\|A^j\mathbf{C}\omega(\phi^{-j}(m))\right\|\leqslant\left\|A^{k_0}\right\|^l\|\mathbf{C}\|K_AK_\omega, \tag{2.9}$$

with $K_A=\max\left\{1,\|A\|,\ldots,\left\|A^{k_0-1}\right\|\right\}$ and $K_\omega\in\mathbb{R}$ a constant that satisfies that $|\omega(m)|\leqslant K_\omega$ for any $m\in M$ and that is available by the boundedness hypothesis on $\omega(M)$.
The inequality (2.9) and the Weierstrass M-test guarantee that the series $\sum_{j=0}^{\infty}A^j\mathbf{C}\omega(\phi^{-j}(m))$ converges absolutely and uniformly on M and that

$$\left\|f_{(\phi,\omega,F)}(m)\right\|=\left\|\sum_{j=0}^{\infty}A^j\mathbf{C}\omega(\phi^{-j}(m))\right\|\leqslant\sum_{l=0}^{\infty}\left\|A^{k_0}\right\|^l\|\mathbf{C}\|K_AK_\omega=\frac{\|\mathbf{C}\|K_AK_\omega}{1-\|A^{k_0}\|}.$$

Finally, since each of the summands in the series is a continuous function then so is $f_{(\phi,\omega,F)}$.

(ii) The result that we just proved guarantees that if the differentials $D^if_{(\phi,\omega,F)}(m)$, $i\in\{1,\ldots,r\}$, exist then they are given by the series $\sum_{j=0}^{\infty}A^j\mathbf{C}D^i\left(\omega\circ\phi^{-j}\right)(m)$ that, using again the Weierstrass M-test and the hypotheses in the statement, will be now shown to uniformly converge to a continuous map. Indeed, using again the decomposition $A^j=A^{lk_i+s}$ in terms of the element $k_i\in\mathbb{N}$ such that $\left\|A^{k_i}\right\|\left\|T^i\phi^{-k_i}\right\|_\infty<1$ we can conclude that each summand of this series satisfies that

$$\left\|A^j\mathbf{C}D^i\left(\omega\circ\phi^{-j}\right)(m)\right\|\leqslant\left(\left\|A^{k_i}\right\|\left\|T\phi^{-k_i}\right\|_\infty\right)^l\|\mathbf{C}\|K_A^iK_{T^i\phi^{-1}}\left\|D^i\omega\right\|_\infty, \tag{2.10}$$

with $K_A^i:=\max\left\{1,\|A\|,\ldots,\left\|A^{k_i-1}\right\|\right\}$ and $K_{T^i\phi^{-1}}:=\max\left\{1,\left\|T^i\phi^{-1}\right\|_\infty,\ldots,\left\|T^i\phi^{-1}\right\|_\infty^{k_i-1}\right\}$, which proves the desired convergence and that $D^if_{(\phi,\omega,F)}(m)=\sum_{j=0}^{\infty}A^j\mathbf{C}D^i\left(\omega\circ\phi^{-j}\right)(m)$. Moreover,

$$\left\|D^if_{(\phi,\omega,F)}(m)\right\|=\left\|\sum_{j=0}^{\infty}A^j\mathbf{C}D^i\left(\omega\circ\phi^{-j}\right)(m)\right\|\leqslant\frac{\|\mathbf{C}\|K_A^iK_{T^i\phi^{-1}}\left\|D^i\omega\right\|_\infty}{1-\|A^{k_i}\|\left\|T^i\phi^{-k_i}\right\|_\infty}. \tag{2.11}$$

(iii) We start by noting that if the map (2.8) is continuous then the subsets Ω_i and Ω_e are indeed open because by theorems 1.1 and 1.4 in [Hirs 76] the immersions and the embeddings in $C^r(M,\mathbb{R}^N)$ are open and hence Ω_i and Ω_e are the preimages of those open sets by the continuous map $\Theta_{(\phi,F)}$. We establish now the continuity of $\Theta_{(\phi,F)}$ by showing that if the sequence $\{\omega_n\}_{n\in\mathbb{N}}$ in $C^r(M,\mathbb{R})$ converges to some element $\omega\in C^r(M,\mathbb{R})$ then so does $\left\{\Theta_{(\phi,F)}(\omega_n)\right\}_{n\in\mathbb{N}}\subset C^r(M,\mathbb{R}^N)$ with respect to $\Theta_{(\phi,F)}(\omega)\in C^r(M,\mathbb{R}^N)$. Indeed, if $\omega_n\longrightarrow\omega$ then, using the notation introduced in (2.11), we have that for a given $\epsilon>0$ and for n sufficiently large

$$\frac{\|\mathbf{C}\|K_A^iK_{T^i\phi^{-1}}}{1-\|A^{k_i}\|\left\|T^i\phi^{-k_i}\right\|_\infty}\left\|D^i\omega_n-D^i\omega\right\|_\infty<\epsilon/r.$$

Then,

$$\left\|\Theta_{(\phi,F)}(\omega_n)-\Theta_{(\phi,F)}(\omega)\right\|_{C^r(M,\mathbb{R}^N)}=\sum_{i=0}^{r}\left\|D^if_{(\phi,\omega_n,F)}-D^if_{(\phi,\omega,F)}\right\|_{\infty}$$
$$=\sum_{i=0}^{r}\left\|\sum_{j=0}^{\infty}A^j\mathbf{C}D^i\left((\omega_n-\omega)\circ\phi^{-j}\right)(m)\right\|\leqslant\sum_{i=0}^{r}\frac{\|\mathbf{C}\|\,K_A^i\,K_{T^i\phi^{-1}}}{1-\|A^{k_i}\|\,\|T^i\phi^{-k_i}\|_{\infty}}\left\|D^i\omega_n-D^i\omega\right\|_{\infty}$$
$$<\frac{\epsilon}{r}+\cdots+\frac{\epsilon}{r}=\epsilon,$$

as required. □

3. Immersive GSs

As we discussed in the introduction, the fact that the Takens delay map is an embedding under certain circumstances guarantees that the representation of the dynamical system associated with it can be used to learn the dynamics of its observations. In this section, we take the first steps to show that similar results can be achieved by using the GSs introduced in proposition 2.3. More specifically, we shall spell out conditions on linear state-space systems that guarantee that the resulting GSs are immersions for generic scalar observations $\omega\in C^2(M,\mathbb{R})$. All along this section, the phase space manifold M of the dynamical system $\phi\in\mathrm{Diff}^2(M)$ is compact and hence genericity in $C^2(M,\mathbb{R})$ is stated with respect to the topology associated to the extension to second-order differentiable functions of the Banach structure introduced in (2.1). The next theorem is the main statement of this section.

Theorem 3.1. *Let $\phi\in\mathrm{Diff}^2(M)$ be a dynamical system on a compact manifold M of dimension q that exhibits finitely many periodic orbits. Let $F:\mathbb{R}^N\times\mathbb{R}\longrightarrow\mathbb{R}^N$ be a linear state map as in (2.4) with $N\geqslant 2q$ whose connectivity matrix satisfies that $\rho(A)<1$ and such that for any observation map $\omega\in C^2(M,\mathbb{R})$ the corresponding GS $f_{(\phi,\omega,F)}\in C^2(M,\mathbb{R}^N)$ and, moreover, the map $\Theta_{(\phi,F)}:C^2(M,\mathbb{R})\longrightarrow C^2(M,\mathbb{R}^N)$ introduced in (2.8) is continuous. Suppose also that the two following conditions hold:*

(i) *For each periodic orbit m of ϕ with period $n\in\mathbb{N}$, the derivative $T_m\phi^{-n}$ has q distinct eigenvalues $\lambda_1,\lambda_2,\ldots,\lambda_q$. Let λ_{max} be the eigenvalue with the highest absolute value among the eigenvalues of all those linear maps and let n_{min} be the smallest period. Suppose that $\rho(\lambda_{max}A^{n_{min}})<1$ and that for any periodic point m, the vectors*

$$\left\{(\mathbb{I}-\lambda_jA^n)^{-1}(\mathbb{I}-A)^{-1}(\mathbb{I}-A^n)\mathbf{C}\right\}_{j\in\{1,\ldots,q\}},\ \text{with } \lambda_j \text{ eigenvalue of } T_m\phi^{-n}\tag{3.1}$$

form a linearly independent set.

(ii) *The vectors $\left\{A^j\mathbf{C}\right\}_{j\in\{0,1,\ldots,N-1\}}$ form a linearly independent set.*

Then, for generic $\omega\in C^2(M,\mathbb{R}^N)$ the GS $f_{(\phi,\omega,F)}\in C^2(M,\mathbb{R})$ is an immersion.

3.1. About the hypotheses of the theorem

All the hypotheses in this statement can be either easily guaranteed or, even better, they generically hold. More specifically, the condition on the linear state map F to produce GS maps

$f_{(\phi,\omega,F)} \in C^2(M,\mathbb{R}^N)$ for any observation map $\omega \in C^2(M,\mathbb{R})$ and the continuity of $\Theta_{(\phi,F)}$ can be enforced by using the first and second parts of proposition 2.3. The condition on ϕ exhibiting finitely many periodic orbits holds generically due to the Kupka–Smale Theorem [Kupk 63, Smal 63].

As to the condition (3.1), we shall see later on (see proposition 4.1) that it holds almost surely in a very specific sense. Regarding the hypothesis in point **(ii)**, this is a very important condition that amounts to reachability in a control theoretical sense (see [Kalm 10, Sont 98]). It has been shown in [Gono 20a] that if A is diagonalizable then this condition holds if and only if all the eigenvalues in the spectrum $\sigma(A)$ of A are distinct and in the linear decomposition $\mathbf{C} = \sum_{i=1}^{N} c_i \mathbf{v}_i$, with $\{\mathbf{v}_1,\ldots,\mathbf{v}_N\}$ a basis of eigenvectors of A, all the coefficients c_i, with $i \in \{1,\ldots,N\}$, are non-zero. This condition can be equivalently reformulated by saying that the Krylov space [Kryl 31] generated by A and $\mathbf{C}$ has maximal dimension.

3.2. Relation with Takens' theorem

The system spelled out in the introduction that allows us to see Takens's delay embedding $\Phi_{(\phi,\omega)}$ as the GS corresponding to a linear state map trivially satisfies the hypotheses of the theorem. Indeed, since in that case A is the lower shift matrix in dimension $2q+1$ and $\mathbf{C} = (1,0,\ldots,0)^\top \in \mathbb{R}^{2q+1}$, the set in condition **(ii)** coincides with the canonical basis in $\mathbb{R}^N$ which is a trivially linearly independent set. Regarding the conditions in **(i)**, as A is nilpotent then all its eigenvalues are zero and hence the hypotheses are trivially satisfied.

Based on this observation, we can formulate a more general statement by saying that any linear system with nilpotent connectivity matrix A that has an input mask $\mathbf{C}$ for which the vectors $\left\{A^j\mathbf{C}\right\}_{j\in\{0,1,\ldots,N-1\}}$ form a linearly independent set also satisfies the hypotheses of the theorem. Equivalently, with the terminology of the previous paragraph, we can rephrase this by writing that any reachable linear system with nilpotent connectivity matrix A satisfies the hypotheses of the theorem.

3.3. System isomorphisms

Given the linear state map introduced in (2.4) and a linear isomorphism of $\mathbb{R}^N$ with associated matrix $P \in \mathbb{M}_{N,N}$, consider the new map $\overline{F}(\mathbf{x},z) := PAP^{-1}\mathbf{x} + P\mathbf{C}z$. Let now $h : \mathbb{R}^M \longrightarrow \mathbb{R}^m$ be a readout for the state map F. In this setup, it is easy to see that the state-space systems (F,h) and $(\overline{F},\overline{h} := h\circ P^{-1})$ are isomorphic in the sense that, in the presence of the ESP, they determine identical input/output systems.

In view of this observation, it is important to emphasize that the hypotheses of theorem 3.1 are invariant under linear system isomorphisms. More explicitly, if we replace in the statement A and $\mathbf{C}$ by $\overline{A} := PAP^{-1}$ and $\overline{\mathbf{C}} := P\mathbf{C}$, respectively, then $\rho(A) = \rho(\overline{A})$ and the validity of the hypotheses **(i)** and **(ii)** is not altered. Indeed, regarding **(i)**, it suffices to notice that

$$
\begin{aligned}
\left(\mathbb{I} - \lambda_j\overline{A}^n\right)^{-1}\left(\mathbb{I} - \overline{A}\right)^{-1}\left(\mathbb{I} - \overline{A}^n\right)\overline{\mathbf{C}} &= \left(\sum_{i=0}^{\infty}\lambda_j^i\left(\overline{A}^n\right)^i\right)\left(\sum_{i=0}^{\infty}\lambda_j^i\overline{A}^i\right)\left(\mathbb{I} - \overline{A}^n\right)\overline{\mathbf{C}} \\
&= P\left(\sum_{i=0}^{\infty}\lambda_j^i\left(A^n\right)^i\right)P^{-1}P\left(\sum_{i=0}^{\infty}\lambda_j^i\overline{A}^i\right)P^{-1} \\
&\quad\times\left(PP^{-1} - PA^nP^{-1}\right)P\mathbf{C} \\
&= P\left(\left(\mathbb{I} - \lambda_jA^n\right)^{-1}\left(\mathbb{I} - A\right)^{-1}\left(\mathbb{I} - A^n\right)\mathbf{C}\right).
\end{aligned}
$$

As to **(ii)**, notice that $\left\{\overline{A}^j\overline{\mathbf{C}}\right\}_{j\in\{0,1,\ldots,N-1\}} = \left\{PA^j\mathbf{C}\right\}_{j\in\{0,1,\ldots,N-1\}}$. Since P is an invertible matrix, in both cases the linear independence is preserved.

Another observation that is worth pointing out is that the class of linear systems for which theorem 3.1 hold is strictly larger than the one determined (up to linear isomorphisms) by Takens' Theorem. As it was mentioned in the previous paragraph, Takens' result is associated to a linear system with nilpotent connectivity matrix A (whose eigenvalues are hence all zero). It is easy to see that when the entries of $\mathbf{C}$ are all non-zero then one can always find a non-singular diagonal matrix A for which the hypotheses of theorem 3.1 hold. Such system is not in the same isomorphism class as Takens' system.

3.4. Proof of the theorem

We proceed in two steps. In the first one, we show that $f_{(\phi,\omega,F)} \in C^2(M,\mathbb{R})$ is an immersion at periodic points, and in the second one we take care of the remaining points. We emphasize that equilibria can be seen as periodic points with period 1.

Step 1. Immersion at periodic points. We start this part with two preparatory lemmas.

Lemma 3.2. *Consider a connectivity matrix that satisfies the conditions $\rho(A) < 1$ and also that $\rho(\lambda_{max}A^{n_{min}}) < 1$ as in part **(i)** of the statement of the theorem. Then, for any periodic point m with period n and any eigenvalue λ_j of $T_m\phi^{-n}$, we have that $\rho(\lambda_j A^n) < 1$ and*

$$(\mathbb{I} - \lambda_j A^n)^{-1} = \sum_{k=0}^{\infty} \lambda_j^k A^{nk}. \tag{3.2}$$

Proof. Firstly, recall the general fact already used in the proof of proposition 2.3 (see also proposition 4.2 in [Grig 21b]) that for any square matrix B such that $\rho(B) < 1$ then $(\mathbb{I} - B)^{-1} = \sum_{j=0}^{\infty} B^j$. Let now m be a periodic point with period n and let λ_j be an eigenvalue of $T_m\phi^{-n}$. This implies that in order for (3.2) to hold we just need to show that $\rho(\lambda_j A^n) < 1$. This is indeed true since any element in the spectrum of $\lambda_j A^n$ can be written as $\lambda_j \mu_k^n$ with $\mu_k \in \mathbb{C}$ an eigenvalue of A. Moreover, let $c < 1$ such that $|\lambda_j| = c\,|\lambda_{\max}|$. Then

$$|\lambda_j \mu_k^n| = c\,|\lambda_{\max}\mu_k^{n_{\min}}|\,\left|\mu_k^{n-n_{\min}}\right| < 1,$$

as required. Notice that in the last inequality we used that $\rho(\lambda_{\max}A^{n_{\min}}) < 1$ and that $\rho(A) < 1$. □

Lemma 3.3. *In the hypotheses of the statement of the theorem, let $m \in M$ be a periodic point of $\phi \in \mathrm{Diff}^2(M)$ with period $n \in \mathbb{N}$. Let $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ be a basis of eigenvectors associated to the distinct eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_q$. Suppose that the set*

$$\left\{(\mathbb{I} - \lambda_j A^n)^{-1}\sum_{k=0}^{n-1} A^k\mathbf{C}D(\omega\circ\phi^{-k})(m)\mathbf{v}_j\right\}_{j\in\{1,\ldots,q\}} \tag{3.3}$$

is linearly independent. Then $f_{(\phi,\omega,F)}$ is an immersion at the periodic point m for generic $\omega \in C^2(M,\mathbb{R}^N)$.

Proof. Since the eigenvalues λ_j are distinct and the eigenvectors $\mathbf{v}_j$ are hence linearly independent, it, therefore, suffices to show that the set $\left\{Df_{(\phi,\omega,F)}(m)\mathbf{v}_j\right\}_{j\in\{1,\ldots,q\}}$ is linearly independent to conclude that $Df_{(\phi,\omega,F)}(m)$ is injective. Then, by the expression (2.7):

$$
\begin{aligned}
Df_{(\phi,\omega,F)}(m)\mathbf{v}_j &= \sum_{l=0}^{\infty} A^l \mathbf{C} D(\omega\circ\phi^{-l})(m)\mathbf{v}_j = \sum_{l=0}^{\infty}\sum_{k=0}^{n-1} A^{ln+k}\mathbf{C}D(\omega\circ\phi^{-(ln+k)})(m)\mathbf{v}_j \\
&= \sum_{l=0}^{\infty}\sum_{k=0}^{n-1} A^{ln+k}\mathbf{C}D(\omega\circ\phi^{-k})(m)[D\phi^{-n}(m)]^l\mathbf{v}_j \\
&= \sum_{l=0}^{\infty}\sum_{k=0}^{n-1} A^{ln+k}\mathbf{C}D(\omega\circ\phi^{-k})(m)\lambda_j^l\mathbf{v}_j \\
&= \sum_{l=0}^{\infty}(\lambda_j A^n)^l \sum_{k=0}^{n-1} A^{k}\mathbf{C}D(\omega\circ\phi^{-k})(m)\mathbf{v}_j \\
&= (\mathbb{I}-\lambda_j A^n)^{-1} \sum_{k=0}^{n-1} A^{k}\mathbf{C}D(\omega\circ\phi^{-k})(m)\mathbf{v}_j,
\end{aligned}
$$

which proves the statement. □

We now use this result to show that, for generic $\omega \in C^2(M,\mathbb{R}^N)$, the GS $f_{(\phi,\omega,F)} \in C^2(M,\mathbb{R})$ is an immersion at the periodic points of ϕ. Let $m_1,\ldots,m_P \in M$ be the distinct periodic points of ϕ, each of which have periods $n_1,\ldots,n_P \in \mathbb{N}$, respectively (the equilibria of ϕ are on this list with periods equal to one). The term *distinct* means that none of those points are in the orbits of the others. We now choose P disjoint open neighbourhoods B_i that contain each of the distinct periodic points m_i. Since there is a finite number of periodic points, the open sets B_i can be chosen small enough so that, additionally, all the open sets

$$
\phi^{-t}(B_i) \quad \text{for all } t\in\{0,\ldots,n_i\} \text{ and } i\in\{1,\ldots,P\}
$$

are disjoint.

Now, given any of the distinct periodic points $m_i \in M$ on the list, we show that $f_{(\phi,\omega,F)}$ for generic $\omega \in C^2(M,\mathbb{R}^N)$, that is, the set of observation maps ω for which $f_{(\phi,\omega,F)}$ is an immersion at m_i is open and dense in $C^2(M,\mathbb{R}^N)$. The openness is a consequence of the hypothesis on the continuity of the map $\Theta_{(\phi,F)}$ and of an argument identical to the beginning of the proof of part **(iii)** of proposition 2.3. Regarding the density, we show that if $f_{(\phi,\omega,F)}$ is not an immersion at m_i, then there is a perturbation ω' of ω in $C^2(M,\mathbb{R})$ for which $f_{(\phi,\omega',F)}$ is an immersion at m_i. Indeed, set

$$
\omega' = \omega + \sum_{l=0}^{n_i-1}\psi_l^i \tag{3.4}
$$

where $\psi_l^i \in C^\infty(M,\mathbb{R})$ are bump functions whose supports are contained in $\phi^{-l}(B_i)$ and, additionally, are chosen to satisfy

$$
D(\psi_l^i\circ\phi^{-l})(m_i) = \varepsilon\mathbf{v}^\top, \quad l\in\{0,\ldots,n_i-1\},
$$

for some small constant $\varepsilon > 0$ and $\mathbf{v}\in\mathbb{R}^p$ the unique vector that solves the linear system

$$
\begin{pmatrix} \mathbf{v}_1^\top \\ \vdots \\ \mathbf{v}_p^\top \end{pmatrix}\mathbf{v} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}, \tag{3.5}
$$

with $\{\mathbf{v}_1,\ldots,\mathbf{v}_p\}$ a basis of eigenvectors of $T_{m_i}\phi^{-n_i}$. Note that by construction and for any $l\in\{0,\ldots,n_i-1\}$,

$$D(\omega'\circ\phi^{-l})(m_i)=D(\omega\circ\phi^{-l})(m_i)+D(\psi_l^i\circ\phi^{-l})=D(\omega\circ\phi^{-l})(m_i)+\varepsilon\mathbf{v}^\top. \tag{3.6}$$

We now consider the vectors (3.3) in lemma 3.3 with respect to the perturbed observation map in (3.4). Indeed, by (3.6) and the way in which the vector $\mathbf{v}$ has been chosen in (3.5):

$$\begin{aligned}(\mathbb{I}-\lambda_jA^{n_i})^{-1}&\sum_{k=0}^{n_i-1}A^k\mathbf{C}D(\omega'\circ\phi^{-k})(m_i)\mathbf{v}_j\\ &=(\mathbb{I}-\lambda_jA^{n_i})^{-1}\sum_{k=0}^{n_i-1}A^k\mathbf{C}D(\omega\circ\phi^{-k})(m_i)\mathbf{v}_j+\varepsilon(\mathbb{I}-\lambda_jA^{n_i})^{-1}\sum_{k=0}^{n_i-1}A^k\mathbf{C}\mathbf{v}^\top\mathbf{v}_j\\ &=(\mathbb{I}-\lambda_jA^{n_i})^{-1}\sum_{k=0}^{n_i-1}A^k\mathbf{C}D(\omega\circ\phi^{-k})(m_i)\mathbf{v}_j+\varepsilon(\mathbb{I}-\lambda_jA^{n_i})^{-1}(\mathbb{I}-A)^{-1}(I-A^{n_i})\mathbf{C}.\end{aligned}$$

Given that when we vary $j\in\{1,\ldots,p\}$ in the previous expression the vectors in the second summand form by hypothesis a linearly independent set, we can use lemma 6.1 to choose $\varepsilon>0$ so that the family $\left\{(\mathbb{I}-\lambda_jA^{n_i})^{-1}\sum_{k=0}^{n_i-1}A^k\mathbf{C}D(\omega'\circ\phi^{-k})(m_i)\mathbf{v}_j\right\}_{j\in\{1,\ldots,p\}}$ forms a linearly independent set and, at the same time, ω' is as close to ω in $C^2(M,\mathbb{R})$ as desired. This shows by lemma 3.3 that $f_{(\phi,\omega',F)}$ is an immersion at m_i.

The choice of the open sets B_i implies that we can keep perturbing ω to make $f_{(\phi,\omega',F)}$ immersive at the other periodic points without spoiling that condition for the previous ones. This shows in particular that a perturbation of the type

$$\omega'=\omega+\sum_{i=1}^{P}\sum_{l=0}^{n_i-1}\psi_l^i \tag{3.7}$$

can be constructed so that $f_{(\phi,\omega',F)}$ is immersive at all the periodic points of ϕ, as required.

Step 2. Immersion at the remaining points. Having just proved that for generic $\omega\in C^2(M,\mathbb{R}^N)$ the GS $f_{(\phi,\omega,F)}\in C^2(M,\mathbb{R})$ is an immersion at the periodic points of ϕ, the Immersion Theorem (see [Abra 88, theorem 3.5.7]) guarantees that the same holds for the open set formed by the union of certain open neighbourhoods around those points. Let $\mathcal{M}\subset M$ be the compact subset of M obtained by removing that immersed open set. Our goal is now to show that $f_{(\phi,\omega,F)}\in C^2(M,\mathbb{R})$ is also an immersion at $\mathcal{M}$ for generic $\omega\in C^2(M,\mathbb{R}^N)$.

Recall first that the hypotheses that we imposed on M in the beginning of section 2 imply that it can be endowed with a Riemannian metric which makes it into a complete metric space by the Hopf and Rinow Theorem (see [Boot 03, theorem 7.7]). This implies in turn that the compact subset $\mathcal{M}\subset M$ is also a complete metric space which allows us to define open balls $B_r(m)$ of radius $r>0$ around each point $m\in\mathcal{M}$. Using this notation, in the next paragraphs we show that for any $\omega\in C^2(M,\mathbb{R}^N)$ and $m\in\mathcal{M}$ we can find a $n(m)\in\mathbb{N}$ and a perturbation $\omega'\in C^2(M,\mathbb{R}^N)$ as close to ω as desired such that the restriction of $f_{(\phi,\omega',F)}$ to $B_{2^{-n(m)}}(m)$ is an immersion.

Indeed, take an arbitrary $m\in\mathcal{M}$ and define a collection of balls $B_{2^{-n}}(m)$ centred at m with radius 2^{-n}, $n\in\mathbb{N}$. For a fixed n consider the infinite trajectory $\phi^{-t}(B_{2^{-n}}(m))$, $t\in\mathbb{N}$. Choose now $n_1\in\mathbb{N}$ large enough so that, for any $n>n_1$ the balls $\phi^{-t}(B_{2^{-n}}(m))$ are disjoint for $t=0,\ldots,N-1$ and $B_{2^{-n}}(m)\subset U$ where (U,h) is an admissible chart of M. Given that $\phi\in\mathrm{Diff}^2(M)$, we note that the family (U_t,h_t), $t\in\mathbb{N}$, defined by $U_t=\phi^{-t}(U)$ and $h_t:=h\circ\phi^t$

is made of admissible charts and that $\phi^{-t}(B_{2^{-n}}(m)) \subset U_t$, for all $n > n_1$. Let $T(n)$ denote the largest integer such that $\phi^{-t}(B_{2^{-n}}(m))$ are disjoint for $t = 0, \ldots, T(n) - 1$.

Now, for each $n > n_1$ and $t = 0, \ldots, N-1$ we define functions $\psi_{tn} \in C^\infty(M, \mathbb{R})$ that have their support included in $\phi^{-t}(B_{2^{-n}}(m))$ and satisfy

$$\frac{\partial(\psi_{tn} h_t^{-1})}{\partial u_j} = 1 \tag{3.8}$$

on $h_t(\phi^{-t}(B_{2^{-(n+1)}}(m))) = h(B_{2^{-(n+1)}}(m))$. We impose further that $\psi_{tn} = \psi_{t(n+1)}$ on $\phi^{-t}(B_{2^{-(n+2)}}(m))$ for all $n > n_1$, and that there is some $\kappa > 0$ independent of n and t such that $\|\psi_{tn} h_t^{-1}\|_{C^1} \leqslant \kappa$. These functions can be constructed by setting

$$\psi_{tn}(m) = \lambda_{tn}(m) \sum_{j=1}^{q} \xi_j(m)$$

where ξ_j is the jth coordinate map for the chart h_t and $\lambda_{tn} \in C^\infty(M, \mathbb{R})$ are bump functions that have support included in $\phi^{-t}(B_{2^{-n}}(m))$ and satisfy $\lambda_{tn}|_{\phi^{-t}(B_{2^{-(n+1)}}(m))} = 1$. Define now the perturbation ω_n of ω by

$$\omega_n = \omega + \sum_{t=0}^{N-1} \varepsilon_t \psi_{tn}, \tag{3.9}$$

where ε_t are the components of a vector $\boldsymbol{\varepsilon} \in \mathbb{R}^N$ with positive entries. By construction, for any $m' \in B_{2^{-n}}(m)$ and $t = 0, \ldots, N-1$, we have that

$$\omega_n \phi^{-t}(m') = \omega \phi^{-t}(m') + \varepsilon_t \psi_{tn}(m')$$

and moreover by (3.8) and for any $m' \in B_{2^{-(n+1)}}(m)$:

$$\frac{\partial(\omega_n \phi^{-t} h^{-1})}{\partial u_j}(h(m')) = \frac{\partial(\omega \phi^{-t} h^{-1})}{\partial u_j}(h(m')) + \varepsilon_t. \tag{3.10}$$

Let $\Phi : M \to \mathbb{R}^N$ be a backward version of the Takens delay map introduced in (1.1), that is,

$$\Phi(m) := \left(\omega(m), \omega \circ \phi^{-1}(m), \ldots, \omega \circ \phi^{-(N-1)}(m)\right)^\top,$$

and let $\Phi_n : M \to \mathbb{R}^N$ be its perturbed version defined by

$$\Phi_n(m) := \left(\omega_n(m), \omega_n \circ \phi^{-1}(m), \ldots, \omega_n \circ \phi^{-(N-1)}(m)\right)^\top.$$

Using these objects, we can rewrite (3.10) in vector form as

$$\frac{\partial(\Phi_n h^{-1})}{\partial u_j}(h(m')) = \frac{\partial(\Phi h^{-1})}{\partial u_j}(h(m')) + \boldsymbol{\varepsilon}, \tag{3.11}$$

for any $m' \in B_{2^{-(n+1)}}(m)$ and where $\boldsymbol{\varepsilon} \in \mathbb{R}^N$. Next, for any $t = N, \ldots, T(n) - 1$ notice that

$$\omega_n \phi^{-t}(m) = \omega \phi^{-t}(m). \tag{3.12}$$

Finally, if $t \geqslant T(n)$ then

$$\omega_n \phi^{-t}(m) = \omega \phi^{-t}(m) + \sum_{\tau=0}^{N-1} \varepsilon_\tau \psi_{\tau n} \phi^{-t}(m). \tag{3.13}$$

We now consider the perturbed GS $f_{(\phi,\omega_n,F)}: M \longrightarrow \mathbb{R}^N$ given by

$$
\begin{aligned}
f_{(\phi,\omega_n,F)} &= \sum_{t=0}^{\infty} A^t \mathbf{C} \omega_n \phi^{-t} \\
&= \sum_{t=0}^{N-1} A^t \mathbf{C} \omega_n \phi^{-t} + \sum_{t=N}^{\infty} A^t \mathbf{C} \omega_n \phi^{-t} = Q\Phi_n + \sum_{t=N}^{\infty} A^t \mathbf{C} \omega_n \phi^{-t}
\end{aligned}
$$

where Q is the $N \times N$ real matrix with $(t+1)$th column $A^t\mathbf{C}$. Now we take the partial derivatives with respect to u_j at points in $h(B_{2^{-(n+1)}}(m))$ and observe that by (3.11)–(3.13):

$$
\begin{aligned}
\frac{\partial(f_{(\phi,\omega_n,F)} h^{-1})}{\partial u_j} &= Q\frac{\partial(\Phi_n h^{-1})}{\partial u_j} + \sum_{t=N}^{\infty} A^t \mathbf{C} \frac{\partial(\omega_n \phi^{-t} h^{-1})}{\partial u_j} \\
&= Q\frac{\partial(\Phi h^{-1})}{\partial u_j} + Q\varepsilon + \sum_{t=N}^{\infty} A^t \mathbf{C} \frac{\partial(\omega_n \phi^{-t} h^{-1})}{\partial u_j} \\
&= Q\frac{\partial(\Phi h^{-1})}{\partial u_j} + Q\varepsilon + \sum_{t=N}^{T(n)-1} A^t \mathbf{C} \frac{\partial(\omega_n \phi^{-t} h^{-1})}{\partial u_j} + \sum_{t=T(n)}^{\infty} A^t \mathbf{C} \frac{\partial(\omega_n \phi^{-t} h^{-1})}{\partial u_j} \\
&= Q\frac{\partial(\Phi h^{-1})}{\partial u_j} + Q\varepsilon + \sum_{t=N}^{T(n)-1} A^t \mathbf{C} \frac{\partial(\omega \phi^{-t} h^{-1})}{\partial u_j} + \sum_{t=T(n)}^{\infty} A^t \mathbf{C} \frac{\partial(\omega_n \phi^{-t} h^{-1})}{\partial u_j} \\
&= Q\frac{\partial(\Phi h^{-1})}{\partial u_j} + Q\varepsilon + \sum_{t=N}^{T(n)-1} A^t \mathbf{C} \frac{\partial(\omega \phi^{-t} h^{-1})}{\partial u_j} + \sum_{t=T(n)}^{\infty} A^t \mathbf{C} \frac{\partial(\omega \phi^{-t} h^{-1})}{\partial u_j} \\
&\quad + \sum_{t=T(n)}^{\infty} A^t \mathbf{C} \left(\sum_{\tau=0}^{N-1} \varepsilon_\tau \frac{\partial(\psi_{\tau n} \phi^{-t} h^{-1})}{\partial u_j} \right) \\
&= Q\frac{\partial(\Phi h^{-1})}{\partial u_j} + Q\varepsilon + \sum_{t=N}^{\infty} A^t \mathbf{C} \frac{\partial(\omega \phi^{-t} h^{-1})}{\partial u_j} \\
&\quad + \sum_{t=T(n)}^{\infty} A^t \mathbf{C} \left(\sum_{\tau=0}^{N-1} \varepsilon_\tau \frac{\partial(\psi_{\tau n} \phi^{-t} h^{-1})}{\partial u_j} \right) \\
&= \sum_{t=0}^{N-1} A^t \mathbf{C} \frac{\partial(\omega \phi^{-t} h^{-1})}{\partial u_j} + Q\varepsilon + \sum_{t=N}^{\infty} A^t \mathbf{C} \frac{\partial(\omega \phi^{-t} h^{-1})}{\partial u_j} \\
&\quad + \sum_{t=T(n)}^{\infty} A^t \mathbf{C} \left(\sum_{\tau=0}^{N-1} \varepsilon_\tau \frac{\partial(\psi_{\tau n} \phi^{-t} h^{-1})}{\partial u_j} \right) \\
&= \sum_{t=0}^{\infty} A^t \mathbf{C} \frac{\partial(\omega \phi^{-t} h^{-1})}{\partial u_j} + Q\varepsilon + \sum_{t=T(n)}^{\infty} A^t \mathbf{C} \left(\sum_{\tau=0}^{N-1} \varepsilon_\tau \frac{\partial(\psi_{\tau n} \phi^{-t} h^{-1})}{\partial u_j} \right) \\
&= \frac{\partial(f_{(\phi,\omega,F)} h^{-1})}{\partial u_j} + Q\varepsilon + \sum_{t=T(n)}^{\infty} A^t \mathbf{C} \left(\sum_{\tau=0}^{N-1} \varepsilon_\tau \frac{\partial(\psi_{\tau n} \phi^{-t} h^{-1})}{\partial u_j} \right).
\end{aligned} \tag{3.14}
$$

In order to prove that $f_{(\phi,\omega,F)}$ is an immersion at the points in $h(B_{2^{-(n+1)}}(m))$ for a generic observation ω, we shall find an arbitrarily small vector ε for which the vectors corresponding to the ε-perturbed observation ω_n

$$\left\{\frac{\partial\left(f_{(\phi,\omega_n,F)}h^{-1}\right)}{\partial u_j}\right\}_{j\in\{1,\dots,q\}}$$

are a linearly independent family. We will proceed inductively by showing that if we assume for some s satisfying $1\leqslant s<q$ that the vectors

$$\left\{\frac{\partial\left(f_{(\phi,\omega,F)}h^{-1}\right)}{\partial u_j}\right\}_{j\in\{1,\dots,s\}} \tag{3.15}$$

are linearly independent, then we can choose an arbitrarily small vector $\boldsymbol{\varepsilon}$ such that the family corresponding to the perturbed observation ω_n defined in (3.9) satisfies that

$$\left\{\frac{\partial\left(f_{(\phi,\omega_n,F)}h^{-1}\right)}{\partial u_j}\right\}_{j\in\{1,\dots,s+1\}}$$

is a linearly independent family. To this end, we define the map $\Psi:\mathbb{R}^s\times h(U)\to\mathbb{R}^N$ as

$$\Psi(\boldsymbol{\alpha},\mathbf{u})=\sum_{j=1}^{s}\alpha_j\frac{\partial\left(f_{(\phi,\omega,F)}h^{-1}\right)}{\partial u_j}-\frac{\partial\left(f_{(\phi,\omega,F)}h^{-1}\right)}{\partial u_{s+1}}.$$

The hypothesis on the statement of the theorem about $f_{(\phi,\omega,F)}\in C^2(M,\mathbb{R}^N)$ for any observation map $\omega\in C^2(M,\mathbb{R})$ implies that Ψ is of class C^1 and maps a manifold of dimension $s+q$ to a manifold of dimension N. Since by hypothesis $s+q<2q\leqslant N$, then the set $\mathbb{R}^N\setminus\Psi(\mathbb{R}^s\times h(U))$ is dense in $\mathbb{R}^N$ (see [Hirs 76, chapter 3, proposition 1.2]). This implies that we can choose an arbitrarily small vector $\boldsymbol{\delta}\in\left(\mathbb{R}^N\setminus\Psi(\mathbb{R}^s\times h(U))\right)$ such that if we set $\boldsymbol{\varepsilon}:=Q^{-1}\boldsymbol{\delta}$ then we have that the vector

$$\frac{\partial(f_{(\phi,\omega,F)}h^{-1})}{\partial u_{s+1}}+Q\boldsymbol{\varepsilon}$$

is independent of the vectors in (3.15) when evaluated at any point in $h(U)$. Since the linear independence is stable under small perturbations we can choose $\boldsymbol{\varepsilon}$ small enough so that it is actually the family

$$\left\{\frac{\partial\left(f_{(\phi,\omega,F)}h^{-1}\right)}{\partial u_j}+Q\boldsymbol{\varepsilon}\right\}_{j\in\{1,\dots,s+1\}}$$

that is linearly independent. Now, in view of the identity (3.14) we note that the value $n\in\mathbb{N}$ can be chosen large enough so that the residual terms

$$\sum_{t=T(n)}^{\infty}A^t\mathbf{C}\left(\sum_{\tau=0}^{N-1}\varepsilon_\tau\frac{\partial(\psi_{\tau n}\phi^{-t}h^{-1})}{\partial u_j}\right),\quad j\in\{1,\dots,s+1\},$$

are small enough so that the family

$$\left\{\frac{\partial\left(f_{(\phi,\omega,F)}h^{-1}\right)}{\partial u_j}+Q\boldsymbol{\varepsilon}+\sum_{t=T(n)}^{\infty}A^t\mathbf{C}\left(\sum_{\tau=0}^{N-1}\varepsilon_\tau\frac{\partial(\psi_{\tau n}\phi^{-t}h^{-1})}{\partial u_j}\right)\right\}_{j\in\{1,\dots,s+1\}}$$
$$=\left\{\frac{\partial\left(f_{(\phi,\omega_n,F)}h^{-1}\right)}{\partial u_j}\right\}_{j\in\{1,\dots,s+1\}}$$

is linearly independent, as required. Notice that this equality is a consequence of (3.14). The possibility to shrink the residual term comes from the convergence of the series

$$\sum_{t=0}^{\infty} A^t \mathbf{C}\left(\sum_{\tau=0}^{N-1} \varepsilon_\tau \frac{\partial(\psi_{\tau n}\phi^{-t}h^{-1})}{\partial u_j}\right), \quad j \in \{1,\ldots,s+1\},$$

which is guaranteed by the hypothesis on the differentiability of $f_{(\phi,\omega,F)}$ for any observation map $\omega \in C^2(M,\mathbb{R})$ and the expression (2.7). In this case, the bump functions play the role of the observations for which we assumed the existence of a uniform bound κ over τ and n such that $\|\psi_{\tau n}\|_{C^1} < \kappa$.

If we recursively apply this procedure, we can conclude the existence of a small perturbation ω_n of ω obtained as a sequence of perturbations of the type (3.9) for which the family

$$\left\{\frac{\partial\left(f_{(\phi,\omega_n,F)}h^{-1}\right)}{\partial u_j}\right\}_{j\in\{1,\ldots,q\}}$$

is linearly independent when evaluated at $m \in \mathcal{M}$, which proves that $f_{(\phi,\omega_n,F)} \in C^2(M,\mathbb{R}^N)$ is an immersion at $m \in \mathcal{M}$.

Finally, observe that we just showed that for any $m \in \mathcal{M}$, there exists an $n(m) \in \mathbb{N}$ such that the restriction of the perturbation $f_{(\phi,\omega_{n(m)},F)}$ to $B_{2^{-n(m)}}(m)$ is an immersion. We note that the union

$$\bigcup_{m\in\mathcal{M}} B_{2^{-n}}(m)$$

is clearly an open cover of $\mathcal{M}$. Since $\mathcal{M}$ is compact, it admits a finite subcover. The finite subcover comprises sets for which, one at a time, we can construct an immersion using the procedure described earlier in this proof. For each set, we ensure that the perturbation is sufficiently small not to spoil the immersion on any other set.

This argument completes the proof of the immersion of the GS at the points of $\mathcal{M}$ and therefore, together with Step 1, shows that there exists a small perturbation of $f_{(\phi,\omega_n,F)} \in C^2(M,\mathbb{R}^N)$ of $f_{(\phi,\omega,F)}$ that is an immersion at all the points in M, as required. □

4. Linear reservoir embeddings

We continue in this section by showing two important facts. Firstly, we prove that without additional hypotheses, the globally immersive GSs whose existence we proved in theorem 3.1 are injective and hence are necessarily embeddings due to the compactness of M (see [Hirs 76]). As we already pointed out in the introduction this is very important in relation to the *learnability question*, that is, at the time of using the embedded state representation of the dynamical system to learn from data the dynamics of its observations. The second fact is related with the RC phenomenon as we show that randomly generated linear systems yield synchronization maps $f_{(\phi,\omega_n,F)} \in C^2(M,\mathbb{R}^N)$ that are almost surely embeddings and are hence amenable to learnability from data.

Theorem 4.1. *Assume that the hypotheses of theorem 3.1 hold true and that, additionally, $N > \max\{2q,\ell\}$ with $\ell \in \mathbb{N}$ the lowest common multiple of all the periods of the finite periodic points of ϕ. Then, for generic $\omega \in C^2(M,\mathbb{R}^N)$, the GS $f_{(\phi,\omega,F)} \in C^2(M,\mathbb{R}^N)$ is an embedding.*

Proof. As in the previous theorem, we proceed in two steps.

Step 1. Injectivity around the periodic set. We start by showing that the observations corresponding to the globally immersive GSs whose existence we proved in theorem 3.1 can be slightly perturbed in $C^2(M,\mathbb{R})$ so that the resulting GS is injective in an open subset V_P that includes all the periodic points of ϕ. We start this part of the discussion with a preparatory lemma.

Lemma 4.1. *In the hypotheses of the theorem, let $m_1,\ldots,m_P \in M$ be the distinct periodic points of ϕ, each of which has periods $n_1,\ldots,n_P \in \mathbb{N}$, respectively. Let $\ell \in \mathbb{N}$ be the lowest common multiple of all the periods and denote by M_P the set of all periodic points of ϕ (that is, the set that comprises $\{m_1,\ldots,m_P\}$ and all the corresponding orbits). Then, the restriction $f_{(\phi,\omega,F)}\big|_{M_P}$ of a GS $f_{(\phi,\omega,F)} \in C^2(M,\mathbb{R}^N)$ to M_P is injective if and only if the map $g_{(\phi,\omega,F)} : M_P \longrightarrow \mathbb{R}^N$ defined by*

$$g_{(\phi,\omega,F)} = \sum_{k=0}^{\ell-1} A^k \mathbf{C}\left(\omega \circ \phi^{-k}\right)$$

is injective.

Proof of the lemma. Let $m_1, m_2 \in M_P$ be such that $f_{(\phi,\omega,F)}(m_1) = f_{(\phi,\omega,F)}(m_2)$. This equality is equivalent to the following expressions:

$$\sum_{t=0}^{\infty} A^t \mathbf{C}\omega\phi^{-t}(m_1) = \sum_{t=0}^{\infty} A^t \mathbf{C}\omega\phi^{-t}(m_2),$$

$$\sum_{t=0}^{\infty}\sum_{k=0}^{\ell-1} A^{t\ell+k}\mathbf{C}\omega\phi^{-(t\ell+k)}(m_1) = \sum_{t=0}^{\infty}\sum_{k=0}^{\ell-1} A^{t\ell+k}\mathbf{C}\omega\phi^{-(t\ell+k)}(m_2),$$

$$\sum_{t=0}^{\infty}(A^\ell)^t \sum_{k=0}^{\ell-1} A^k\mathbf{C}\omega\phi^{-k}(m_1) = \sum_{t=0}^{\infty}(A^\ell)^t \sum_{k=0}^{\ell-1} A^k\mathbf{C}\omega\phi^{-k}(m_2).$$

Given that $\rho(A) < 1$ then $\rho(A^\ell) < 1$ necessarily and hence this equality can be rewritten as

$$(I - A^\ell)^{-1}\sum_{k=0}^{\ell-1} A^k\mathbf{C}\omega\phi^{-k}(m_1) = (I - A^\ell)^{-1}\sum_{k=0}^{\ell-1} A^k\mathbf{C}\omega\phi^{-k}(m_2),$$

which is equivalent to $\sum_{k=0}^{\ell-1} A^k\mathbf{C}\omega\phi^{-k}(m_1) = \sum_{k=0}^{\ell-1} A^k\mathbf{C}\omega\phi^{-k}(m_2)$ and hence, by definition, to $g_{(\phi,\omega,F)}(m_1) = g_{(\phi,\omega,F)}(m_2)$, which proves the statement. □

If we now define $\Phi_{(\ell,\omega)} : M \to \mathbb{R}^l$ as

$$\Phi_{(\ell,\omega)}(m) := \left(\omega(m), \omega\circ\phi^{-1}(m), \ldots, \omega\circ\phi^{-(\ell-1)}(m)\right)^\top,$$

we note that the map $g_{(\phi,\omega,F)}$ can be rewritten as $g_{(\phi,\omega,F)} = Q\Phi_{(\ell,\omega)}$, where $Q \in \mathbb{M}_{N,\ell}$ is a matrix whose $(k+1)$th-column is set to the vector $A^k\mathbf{C}$. The hypotheses on the vectors $\left\{A^j\mathbf{C}\right\}_{j\in\{0,1,\ldots,N-1\}}$ forming a linearly independent set and that $N > \ell$ guarantee that rank $Q = N$ and hence that the associated linear map $Q : \mathbb{R}^\ell \longrightarrow \mathbb{R}^N$ is injective. With this notation we now show that if $g_{(\phi,\omega,F)}$ is not injective in M_P then a perturbation $\omega' \in C^2(M,\mathbb{R}^N)$ of ω can be chosen so that $g_{(\phi,\omega',F)}$ is. More specifically, define

$$\omega' := \omega + \sum_{i=1}^{P}\sum_{j=1}^{n_i} \varepsilon_{ij}\Psi_{ij}, \tag{4.1}$$

where Ψ_{ij} are bump functions with non-intersecting supports U_{ij} such that $m_{ij} := \phi^{-(j-1)}(m_i) \in U_{ij}$ and, moreover, $\Psi_{ij}\left(\phi^{-(j-1)}(m_i)\right) = \Psi_{ij}(m_{ij}) = 1/\mathcal{L}(i,j)$. The symbol $\mathcal{L}(i,j) \in \mathbb{N}$ denotes the ordinal of the pair (i,j) in lexicographic order.

We now show that the constants ε_{ij} can be chosen so that ω' is as close as we want to ω and, at the same time, $g_{(\phi,\omega',F)}$ is injective. Firstly, it is easy to see that, by construction,

$$\Phi_{(\ell,\omega')}(m) := \Phi_{(\ell,\omega)}(m) + \sum_{i=1}^{P}\sum_{j=1}^{n_i} \varepsilon_{ij}\left(\Psi_{ij}(m), \Psi_{ij}\circ\phi^{-1}(m), \ldots, \Psi_{ij}\circ\phi^{-(\ell-1)}(m)\right)^{\top}.$$

Second, if $m_{i_1j_1}$ and $m_{i_2j_2}$ are two different periodic points then

$$\begin{aligned} g_{(\phi,\omega',F)}(m_{i_1j_1}) - g_{(\phi,\omega',F)}(m_{i_2j_2}) = g_{(\phi,\omega,F)}(m_{i_1j_1}) - g_{(\phi,\omega,F)}(m_{i_2j_2}) \\ + Q\left(\varepsilon_{i_1j_1}\mathbf{v}_{i_1j_1} - \varepsilon_{i_2j_2}\mathbf{v}_{i_2j_2}\right), \end{aligned} \tag{4.2}$$

where the vectors $\mathbf{v}_{i_1j_1} \in \mathbb{R}^{\mathrm{Card}M_P}$ have entries equal to zero except at the slots that are multiples of the period of the corresponding periodic point. More specifically, if the periodic point m_{ij} has period n_{ij}, then

$$(\mathbf{v}_{ij})_i := \begin{cases} 1/\mathcal{L}(i,j) & \text{when } i = 1 \text{ or } i-1 \text{ is a multiple of } n_{ij}, \\ 0 & \text{otherwise.} \end{cases} \tag{4.3}$$

Using the injectivity of Q and lemma 4.1, we now show that we can choose the perturbation constants ε_{ij} so that the restriction of $g_{(\phi,\omega',F)}$ to M_P is injective. Let

$$\varepsilon_{ij} := \epsilon\|g_{(\phi,\omega,F)}(m_{ij})\|, \quad \text{for some constant } \epsilon > 0. \tag{4.4}$$

We now show that if $\epsilon > 0$ is chosen so that

$$\begin{aligned} \epsilon \max_{(i_1,j_1),(i_2,j_2)} \left\{\left\|Q\left(\left\|g_{(\phi,\omega,F)}(m_{i_1j_1})\right\|\mathbf{v}_{i_1j_1} - \left\|g_{(\phi,\omega,F)}(m_{i_2j_2})\right\|\mathbf{v}_{i_2j_2}\right)\right\|\right\} \\ < \min_{(i_1,j_1),(i_2,j_2)} \left\{\left\|g_{(\phi,\omega,F)}(m_{i_1j_1}) - g_{(\phi,\omega,F)}(m_{i_2j_2})\right\|\right\} \end{aligned} \tag{4.5}$$

then the injectivity of $g_{(\phi,\omega',F)}|_{M_P}$ is guaranteed. Indeed, consider first the case of two distinct periodic points $m_{i_1j_1}$ and $m_{i_2j_2}$ for which $g_{(\phi,\omega,F)}$ fails to be injective, that is, $g_{(\phi,\omega,F)}(m_{i_1j_1}) = g_{(\phi,\omega,F)}(m_{i_2j_2})$. In that case, by (4.2) and (4.4) we have that

$$g_{(\phi,\omega',F)}(m_{i_1j_1}) - g_{(\phi,\omega',F)}(m_{i_2j_2}) = \epsilon\left\|g_{(\phi,\omega,F)}(m_{i_1j_1})\right\| Q\left(\mathbf{v}_{i_1j_1} - \mathbf{v}_{i_2j_2}\right). \tag{4.6}$$

Given that $\mathbf{v}_{i_1j_1} - \mathbf{v}_{i_2j_2} \neq \mathbf{0}$ (notice, for instance, that $(\mathbf{v}_{i_1j_1} - \mathbf{v}_{i_2j_2})_1 = 1/\mathcal{L}(i_1,j_1) - 1/\mathcal{L}(i_2,j_2) \neq 0$) and Q is injective then $Q(\mathbf{v}_{i_1j_1} - \mathbf{v}_{i_2j_2}) \neq \mathbf{0}$ and hence $g_{(\phi,\omega',F)}(m_{i_1j_1}) \neq g_{(\phi,\omega',F)}(m_{i_2j_2})$ necessarily by (4.6). In the case $g_{(\phi,\omega,F)}(m_{i_1j_1}) \neq g_{(\phi,\omega,F)}(m_{i_2j_2})$ the same conclusion can be drawn because the choice of $\epsilon > 0$ in (4.5) guarantees that

$$\|Q\left(\varepsilon_{i_1j_1}\mathbf{v}_{i_1j_1} - \varepsilon_{i_2j_2}\mathbf{v}_{i_2j_2}\right)\| < \left\|g_{(\phi,\omega,F)}(m_{i_1j_1}) - g_{(\phi,\omega,F)}(m_{i_2j_2})\right\|$$

which by (4.2) ensures that, again, $g_{(\phi,\omega',F)}(m_{i_1j_1}) \neq g_{(\phi,\omega',F)}(m_{i_2j_2})$, as required.

We now show that if $f_{(\phi,\omega,F)}\big|_{M_P}$ is injective then there exists an open set V_P such that $M_P \subset V_P$ and $f_{(\phi,\omega,F)}\big|_{\overline{V_P}}$ is also injective. By the Immersion Theorem [Abra 88, theorem 3.5.7] we know that there exists $n \in \mathbb{N}$ such that the balls $B_{2^{-n}}(m_{ij})$ do not intersect and that the restriction of $f_{(\phi,\omega,F)}$ to each of them is a collection of injective maps. It could still be, however, that the images of different balls intersect. The continuity of $f_{(\phi,\omega,F)}$ and the fact that $f_{(\phi,\omega,F)}\big|_{M_P}$ is injective implies that n can be chosen sufficiently high so that this does not happen. Indeed, if this was not the case for the balls around the periodic points, say, $m_{i_1j_1}$ and $m_{i_2j_2}$, then it would be possible to construct two sequences $\{m_{i_1j_1,l}\}_{l\in\mathbb{N}}$ and $\{m_{i_2j_2,l}\}_{l\in\mathbb{N}}$ with limits $m_{i_1j_1}$ and $m_{i_2j_2}$ for which $f_{(\phi,\omega,F)}(m_{i_1j_1,l}) = f_{(\phi,\omega,F)}(m_{i_2j_2,l})$ for each $l \in \mathbb{N}$. By continuity this implies that

$f_{(\phi,\omega,F)}(m_{i_1j_1}) = f_{(\phi,\omega,F)}(m_{i_2j_2})$ which is in contradiction with the injectivity of $f_{(\phi,\omega,F)}\big|_{M_P}$ and hence proves the injectivity of $f_{(\phi,\omega,F)}$ restricted to $V_P = \bigcup_{ij} B_{2^{-n}}(m_{ij})$, with n chosen so that the properties of the corresponding balls designed above are satisfied. Notice that by doubling n, if necessary, it is also easy to ensure the injectivity of $f_{(\phi,\omega,F)}\big|_{\overline{V_P}}$.

Step 2: Global injectivity. We firstly prove an important local intermediate result.

Lemma 4.2. *If M is a compact differentiable manifold endowed with a metric d and $f : M \longrightarrow \mathbb{R}^N$ is an immersion, then there exists a constant $r > 0$ such that for any $m \in M$ the restriction $f|_{B_r(m)}$ of f to the open ball $B_r(m) \subset M$ of radius r and centre m is injective.*

Proof. The Immersion Theorem ([Abra 88, theorem 3.5.7]) implies that each $m \in M$ has an open neighbourhood $U_m \subset M$ such that $f|_{U_m}$ is injective. The collection of sets $\{U_m\}_{m\in M}$ forms an open cover of M. Then, by Lebesgue's number lemma [Munk 14, lemma 27.5], there exists a $\delta > 0$ such that every set of diameter δ is contained in some set in the family $\{U_m\}_{m\in M}$. The lemma is proved by choosing $r = \delta/2$. □

Since M is compact and $f_{(\phi,\omega,F)} : M \longrightarrow \mathbb{R}^N$ is an immersion, this lemma implies the existence of a constant $r > 0$ such that for any $m \in M$ the restriction $f_{(\phi,\omega,F)}|_{B_r(m)}$ of $f_{(\phi,\omega,F)}$ to the open ball $B_r(m)$ is injective. We now define the set $W \subset M \times M$ as follows using the open set V_P whose existence we proved in Step 1.

$$W := \{(m_1, m_2) \in (M \times M) \setminus (V_P \times V_P) \mid \ d(m_1, m_2) \geqslant r\}.$$

The set W comprises pairs $(m_1, m_2) \in M$ whose entries satisfy one of two conditions:

1. Neither m_1 nor m_2 are in V_P.
2. One of m_1 and m_2 is in V_P and the other is not.

In view of this, the injectivity of $f_{(\phi,\omega,F)}|_{V_P}$ proved in the Step 1 together with lemma 4.2 imply that if we show that $f_{(\phi,\omega,F)}(m_1) \neq f_{(\phi,\omega,F)}(m_2)$ for all $(m_1, m_2) \in W$ then $f_{(\phi,\omega,F)}$ is globally injective and the proof is concluded.

We start the proof of this fact by first defining, for each $m \in M$, a collection of nested balls $\{B_{2^{-n}}(m) \mid n \in \mathbb{N}\}$ centred at m with radius 2^{-n}. Let $(m_1, m_2) \in W$, and assume from now on without loss of generality that $m_1 \in W \setminus V_p$. Let $T(n, m_1, m_2)$ denote the largest integer such that the following two properties hold. Firstly, the sets

$$\{B_{2^{-n}}(\phi^{-t}(m_1))\}_{t=0,\ldots,T(n,m_1,m_2)-1}$$

are disjoint and secondly

$$B_{2^{-n}}(\phi^{-t}(m_1)) \cap B_{2^{-n}}(\phi^{-s}(m_2)) = \emptyset \quad \text{for all} \quad t, s \in \{0, \ldots, T(n, m_1, m_2) - 1\}.$$

Notice now that by the continuity of ϕ, for each $n \in \mathbb{N}$ and pair $(m_1, m_2) \in W$ there is an open neighbourhood $U_{(m_1,m_2)} \subset M \times M$ of (m_1, m_2) such that $T(n, m_1', m_2') = T(n, m_1'', m_2'')$ for all $(m_1', m_2'), (m_1'', m_2'') \in U_{(m_1,m_2)}$. The collection $\{U_{(m_1,m_2)} \mid (m_1, m_2) \in W\}$ covers W and since it is a compact set we can extract a finite subcover $\{U_a \mid a \in \mathcal{A}\}$, where $\mathcal{A}$ is a finite set. Then we can choose one pair $(m_1^a, m_2^a) \in U_a$ for each $a \in \mathcal{A}$ and notice that

$$\min_{(m_1,m_2)\in W}\{T(n, m_1, m_2)\} = \min_{\{(m_1^a,m_2^a)\ \mid\ a\in A\}}\{T(n, m_1^a, m_2^a)\}.$$

The importance of this equality is that, since $\mathcal{A}$ is a finite set, the minimum on the right-hand side is realized by a pair $(m_1^*, m_2^*) \in W$. Let $T(n) = T(n, m_1^*, m_2^*) = \min_{(m_1,m_2)\in W} T(n, m_1, m_2)$. Observe that as $n \to \infty$ the families $\{B_{2^{-n}}(\phi^{-t}(m_1^*))\}_{t\in\mathbb{N}}$ and $\{B_{2^{-n}}(\phi^{-t}(m_2^*))\}_{t\in\mathbb{N}}$ converge to $\{\phi^{-t}(m_1^*)\}_{t\in\mathbb{N}}$ and $\{\phi^{-t}(m_2^*)\}_{t\in\mathbb{N}}$ respectively. The point m_1^* is not periodic so the infinite orbit $\{\phi^{-t}(m_1^*)\}_{t\in\mathbb{N}}$ of singletons is disjoint, and furthermore does not intersect any point in $\{\phi^{-t}(m_2^*)\}_{t\in\mathbb{N}}$. This allows us to conclude that $T(n)\to\infty$ as $n\to\infty$.

The fact that we just proved guarantees the existence of a $\nu\in\mathbb{N}$ such that $T(\nu)=N$. Thus for all pairs $(m_1,m_2)\in W$, the collection

$$\{B_{2^{-\nu}}(\phi^{-t}(m_1))\}_{t=0,\dots,N-1}$$

is disjoint and

$$B_{2^{-n}}(\phi^{-t}(m_1))\cap B_{2^{-n}}(\phi^{-s}(m_2))=\emptyset \quad \text{for all} \quad t,s\in\{0,\dots,N-1\}.$$

Now for any $n>\nu$ the collection

$$\mathcal{C}_n=\{B_{2^{-(n+1)}}(m)\mid m\in M\}$$

forms on open cover of M from which we can extract a finite subcover $\{B_i\mid i\in J_n\}$ for J a finite set with cardinality $\ell(n)\in\mathbb{N}$. Now define a partition of unity $\{\lambda_i\mid i\in J_n\}$ subordinate to $\{B_i\mid i\in J_n\}$. We impose on this partition of unity the special property that for each $m\in M$ there exists an $i\in J_n$ such that $\lambda_i(m)\geqslant 1/2$. Now we define the perturbed observation function

$$\omega_n=\omega+\sum_{i=1}^{\ell(n)}\epsilon_i\lambda_i$$

where $\epsilon_i\in\mathbb{R}$ is the ith component of a vector $\boldsymbol{\epsilon}\in\mathbb{R}^{\ell(n)}$ with positive entries. Then we define $\Psi_n:M\times M\times\mathbb{R}^{\ell(n)}\to\mathbb{R}^N$ by

$$\Psi_n(m_1,m_2,\epsilon)=f_{(\phi,\omega_n,F)}(m_1)-f_{(\phi,\omega_n,F)}(m_2). \tag{4.7}$$

Let $\Delta=\{(m,m)\in M\times M\mid m\in M\}$ be the diagonal set. Given an arbitrary open neighbourhood $\mathcal{N}\subset C^1(M,\mathbb{R})$ of the observation function $\omega\in C^2(M,\mathbb{R})$ our goal is to find $\boldsymbol{\epsilon}\in\mathbb{R}^{\ell(n)}$ such that $\omega_n\in\mathcal{N}$ and that for all $(m_1,m_2)\in(M\times M)\setminus\Delta$ we have that $\Psi_n(m_1,m_2,\boldsymbol{\epsilon})\neq\mathbf{0}$.

First of all, we observe that for any pair $(m_1,m_2)\in(M\times M)\setminus W$ either $d(m_1,m_2)<r$ or both $m_1,m_2\in V_P$. In the former case, $\Psi_n(m_1,m_2,\mathbf{0})\neq\mathbf{0}$ unless $(m_1,m_2)\in\Delta$ by lemma 4.2, and in the latter case, $\Psi_n(m_1,m_2,\mathbf{0})\neq\mathbf{0}$ unless $(m_1,m_2)\in\Delta$ because $f_{(\phi,\omega,F)}|_{V_P}$ is injective by the Step 1. Now Ψ_n is continuous so there is an open neighbourhood $U_{\mathbf{0}}\subset\mathbb{R}^{\ell(n)}$ of $\mathbf{0}\in\mathbb{R}^{\ell(n)}$ such that for all $\boldsymbol{\epsilon}\in U_{\mathbf{0}}$ we have $\Psi_n(m_1,m_2,\boldsymbol{\epsilon})\neq\mathbf{0}$ for all $(m_1,m_2)\in(M\times M)\setminus W$ unless $(m_1,m_2)\in\Delta$. So all that remains is to find $\boldsymbol{\epsilon}\in U_{\mathbf{0}}\subset\mathbb{R}^{\ell(n)}$ such that $\Psi_n(m_1,m_2,\boldsymbol{\epsilon})\neq 0$ for all $(m_1,m_2)\in W$.

We start by noting that if $\mathbf{0}\in\mathbb{R}^N$ is not in the image of $\Psi_n|_{W\times\{\mathbf{0}\}}$ then we are done so we shall assume the opposite. In that case we proceed by showing that $\Psi_n|_{W\times\{\mathbf{0}\}}$ is a submersion. If that is the case, then for some open set $X\subset(M\times M\times\mathbb{R}^{\ell(n)})$ containing $W\times\{\mathbf{0}\}$ then the restriction $\Psi_n|_X$ is also a submersion and hence by the Submersion Theorem [Abra 88, theorem 3.5.4] the inverse image $\Psi_n|_X^{-1}(\mathbf{0})$ is a closed submanifold of dimension $2q+\ell(n)-N$ of the open submanifold $X\subset M\times M\times\mathbb{R}^{\ell}$. Moreover, if $\pi:M\times M\times\mathbb{R}^{\ell(n)}\to\mathbb{R}^{\ell(n)}$ is the canonical projection defined by $\pi(m_1,m_2,\boldsymbol{\epsilon}):=\boldsymbol{\epsilon}$, in these circumstances the complement $\mathbb{R}^{\ell(n)}\setminus\pi\left(\Psi_n|_X^{-1}(\mathbf{0})\right)$ is a dense subset of $\mathbb{R}^{\ell(n)}$. Indeed, since π is a continuously differentiable map, then so is its restriction $\pi|_{\Psi_n|_X^{-1}(\mathbf{0})\times\mathbb{R}^{\ell}}:\Psi_n|_X^{-1}(\mathbf{0})\longrightarrow\mathbb{R}^{\ell}$ which by [Hirs 76, chapter 3, proposition 1.2] guarantees the density of $\mathbb{R}^{\ell(n)}\setminus\pi\left(\Psi_n|_X^{-1}(\mathbf{0})\right)$. This implies that we can choose $\boldsymbol{\epsilon}\in\left(\mathbb{R}^{\ell(n)}\setminus\pi\left(\Psi_n|_X^{-1}(\mathbf{0})\right)\right)$ as small as we want so that $\boldsymbol{\epsilon}\in U_{\mathbf{0}}$ and $\omega_n\in\mathcal{N}$. We fix this ϵ and see that for any $(m_1,m_2)\in W$ the map $\Psi_n(m_1,m_2,\boldsymbol{\epsilon})\neq 0$, as required. Consequently, all that remains to be done is to find n sufficiently large so that $\Psi_n|_{W\times\{\mathbf{0}\}}$ is a submersion, and then the proof will be complete.

We start by observing that by (4.7)

$$\omega_n \circ \phi^{-t} = \omega \circ \phi^{-t} + \sum_{i=1}^{\ell(n)} \epsilon_i \lambda_i \circ \phi^{-t}$$

hence

$$\frac{\partial(\omega_n \circ \phi^{-t})}{\partial \epsilon_j} = \lambda_j \circ \phi^{-t}.$$

Now we consider an arbitrary $(m_1, m_2) \in W$ assuming once again without loss of generality that $m_1 \in M \setminus V_p$. For each point in the orbit $\{\phi^{-t}(m_1)\}_{t=0,\ldots,T(n)-1}$ there exists a $j(t) \in J_n$ such that $\lambda_{j(t)}(\phi^{-t}(m_1)) \geqslant 1/2$ by the special property that we imposed earlier on the partition of unity $\{\lambda_i \mid i \in J_n\}$. Now the support of $\lambda_{j(t)}$ is a ball $B_{j(t)}$ of radius $2^{-(n+1)}$ which contains $\phi^{-t}(m_1)$. Hence the ball $B_{j(t)} \subset B_{2^{-n}}(\phi^{-t}(m_1))$. Now since the sets in the family $\{B_{2^{-n}}(\phi^{-t}(m_1))\}_{t=0,\ldots,T(n)-1}$ are disjoint then so are $\{B_{j(t)}\}_{t=0,\ldots,T(n)-1}$. Furthermore, since $B_{2^{-n}}(\phi^{-t}(m_1)) \cap B_{2^{-n}}(\phi^{-s}(m_2)) = \emptyset$ for all $t, s \in \{0, \ldots, T(n)-1\}$ hence $\lambda_{j(t)}(\phi^{-t}(m_2)) = 0$ for $t \in \{0, \ldots, T(n)-1\}$. Thus

$$\frac{\partial(\omega_n \circ \phi^{-t})}{\partial \epsilon_{j(t)}}(m_2) = 0.$$

Now,

$$\Psi_n(m_1, m_2, \epsilon) = \sum_{t=0}^{T(n)-1} A^t \mathbf{C}(\omega_n(\phi^{-t}(m_1)) - \omega_n(\phi^{-t}(m_2))) \\ + \sum_{t=T(n)}^{\infty} A^t \mathbf{C}(\omega_n(\phi^{-t}(m_1)) - \omega_n(\phi^{-t}(m_2)))$$

hence for $t = 0, \ldots, T(n)-1$

$$\frac{\partial \Psi_n}{\partial \epsilon_{j(t)}}(m_1, m_2, \epsilon) = A^t \mathbf{C}(\lambda_{j(t)}(\phi^{-t}(m_1))) \\ + \sum_{t=T(n)}^{\infty} A^t \mathbf{C}(\lambda_{j(t)}(\phi^{-t}(m_1)) - \lambda_{j(t)}(\phi^{-t}(m_2))).$$

By assumption $\{A^t\mathbf{C}\}_{t=0,\ldots,N-1}$ are linearly independent, hence the vectors $\{A^t\mathbf{C}(\lambda_{j(t)}(\phi^{-t}(m_1)))\}_{t\in\{0,\ldots,T(n)-1\}}$ necessarily span $\mathbb{R}^N$ because since $n > \nu$ then $T(n) \geqslant N$. Crucially, for any n the property $\lambda_{j(t)}(\phi^{-t}(m_1)) \geqslant 1/2$ holds and therefore, the residual term

$$\sum_{t=T(n)}^{\infty} A^t \mathbf{C}(\lambda_{j(t)}\phi^{-t}(m_1) - \lambda_{j(t)}\phi^{-t}(m_2))$$

may only spoil the spanning property of the vectors $\{A^t\mathbf{C}(\lambda_{j(t)}\phi^{-t}(m_1))\}_{t=0,\ldots,N-1}$ if it is sufficiently large. Since by hypothesis $\rho(A) < 1$, the residual term converges uniformly over $(m_1, m_2) \in W$ to 0 as n grows. We choose consequently n large enough so that for all $(m_1, m_2) \in W$ the residual term is too small to spoil the spanning property of $\{A^t\mathbf{C}(\lambda_{j(t)}\phi^{-t}(m_1))\}_{t\in\{0,\ldots,N-1\}}$. With this choice of n we have that $\Psi_n|_{W\times\{0\}}$ is a submersion and the proof is complete. □

4.1. Linear reservoir embeddings

We conclude the theoretical part of the paper by showing that the embeddings whose existence we proved in theorem 4.1 using generic observation maps $\omega \in C^2(M, \mathbb{R}^N)$ may be almost surely obtained, as it is customary in RC, by randomly drawing the connectivity matrix A and the vector $\mathbf{C}$ of the linear system $F(\mathbf{x}, z) := A\mathbf{x} + \mathbf{C}z$. This result hinges on an important fact in random matrix theory whose proof has been kindly communicated to us by Friedrich Philipp and that is contained in the following statement. We recall that a random variable $X : \Omega \longrightarrow T$ defined on a probability space $(\mathbb{P}, \mathcal{F}, \mathbb{P})$ and with values on a Borel measurable space T is *regular* or *non-singular* whenever $\mathbb{P}(X = a) = 0$ for all $a \in T$.

Proposition 4.1 (Friedrich Philipp). *Let $N \in \mathbb{N}$, $A \in \mathbb{M}_{N,N}$, and $\mathbf{C} \in \mathbb{R}^N$ and assume that the entries of A and $\mathbf{C}$ are drawn using independent regular real-valued distributions. Then the following statements hold:*

(i) *The vectors $\mathbf{C}, A\mathbf{C}, A^2\mathbf{C}, \ldots, A^{N-1}\mathbf{C}$ are linearly independent almost surely*
(ii) *Given m distinct complex numbers $\lambda_1, \ldots, \lambda_m \in \mathbb{C}$, where $m \leqslant N$, the event that $1, \lambda_1, \ldots, \lambda_m \notin \sigma(A)$ ($\sigma(A)$ is the spectrum of A) and that the vectors*

$$(\mathbb{I} - \lambda_j A)^{-1}(\mathbb{I} - A)^{-1}(\mathbb{I} - A^N)\mathbf{C}, \quad j = 1, \ldots, m$$

are linearly independent holds almost surely.

Proof. The vectors $\mathbf{C}, A\mathbf{C}, A^2\mathbf{C}, \ldots, A^{N-1}\mathbf{C}$ are linearly independent if and only if

$$\det\left(\mathbf{C}|A\mathbf{C}|A^2\mathbf{C}|\cdots|A^{N-1}\mathbf{C}\right) = 0$$

which, in the notation of lemma 6.3 in appendix ‘Two lemmas about random matrices’, can be written as

$$\det\left(p_0(A)\mathbf{C}, p_1(A)\mathbf{C}, p_2(A)\mathbf{C}, \ldots, p_{N-1}(A)\mathbf{C}\right) = 0$$

using the linearly independent polynomials $p_j(A) := A^j$, $j \in \{0, \ldots, N-1\}$. Part **(i)** of the statement hence follows directly from lemma 6.3 in appendix ‘Two lemmas about random matrices’. Now we turn our attention to part **(ii)**. First of all, λ_j is an eigenvalue of A if and only if λ_j is a root of the characteristic polynomial of A. This event has probability 0 by lemma 6.2 in appendix ‘Two lemmas about random matrices’ and hence $1, \lambda_1, \ldots, \lambda_m \notin \sigma(A)$ almost surely. On this event, the inverses $(\mathbb{I} - \lambda_j A)^{-1}$ and $(\mathbb{I} - A)^{-1}$ exist. Furthermore, the product

$$\prod_{i=1}^{m}(\mathbb{I} - \lambda_i A)$$

is an invertible matrix. Therefore, the vectors

$$(\mathbb{I} - \lambda_j A)^{-1}(\mathbb{I} - A)^{-1}(\mathbb{I} - A^N)\mathbf{C}, \quad \text{with } j = 1, \ldots, m,$$

are linearly independent if and only if

$$\prod_{i=1}^{m}(\mathbb{I} - \lambda_i A)(\mathbb{I} - \lambda_j A)^{-1}(\mathbb{I} - A)^{-1}(\mathbb{I} - A^N)\mathbf{C}, \quad \text{with } j = 1, \ldots, m, \tag{4.8}$$

are linearly independent. We can now rewrite the vectors in (4.8) as

$$\prod_{i=1}^{m}(\mathbb{I}-\lambda_i A)(\mathbb{I}-\lambda_j A)^{-1}(\mathbb{I}-A)^{-1}(\mathbb{I}-A^N)\mathbf{C} = \prod_{i\neq j}^{m}(\mathbb{I}-\lambda_i A)(\mathbb{I}-A)^{-1}(\mathbb{I}-A^N)\mathbf{C}$$
$$= \prod_{i\neq j}^{m}(\mathbb{I}-\lambda_i A)\sum_{k=0}^{N-1}A^k\mathbf{C},$$

where we used the relation

$$(\mathbb{I}-A^N) = (\mathbb{I}-A)\sum_{k=0}^{N-1}A^k \quad \text{and hence that} \quad (\mathbb{I}-A)^{-1}(\mathbb{I}-A^N) = \sum_{k=0}^{N-1}A^k.$$

Now, if we are able to show that the family

$$p_j(x) = \prod_{i\neq j}^{m}(1-\lambda_i x)\sum_{k=0}^{N-1}x^k, \quad \text{with} \quad j\in\{1,\ldots,m\}$$

is linearly independent, then we can conclude by lemma 6.3 in appendix ‘Two lemmas about random matrices’ that the vectors (4.8) are linearly independent almost surely, which would complete the proof. This is indeed the case because if $\mu_1,\ldots,\mu_n\in\mathbb{R}$ are such that

$$\sum_{j=1}^{n}\mu_j p_j(x) = 0 \quad \text{then} \quad \left(\sum_{k=0}^{N-1}x^k\right)\left(\sum_{j=1}^{n}\mu_j\prod_{i\neq j}^{m}(1-\lambda_i x)\right) = 0.$$

Given that the polynomial $\sum_{k=0}^{N-1}x^k$ is non-zero, the previous equality is equivalent to $\sum_{j=1}^{n}\mu_j\prod_{i\neq j}^{m}(1-\lambda_i x) = 0$ which, evaluated at $x = 1/\lambda_k$, implies that

$$0 = \sum_{j=1}^{n}\mu_j\prod_{i\neq j}^{m}\left(1-\lambda_i\frac{1}{\lambda_k}\right) = \mu_k\prod_{i\neq k}^{m}\left(1-\lambda_i\frac{1}{\lambda_k}\right).$$

Given that, by hypothesis, the values $\lambda_1,\ldots,\lambda_m\in\mathbb{C}$ are all different, we can conclude that $\prod_{i\neq k}^{m}\left(1-\frac{\lambda_i}{\lambda_k}\right)\neq 0$ and hence $\mu_k = 0$, necessarily. Since this procedure can be repeated to obtain that $\mu_1,\ldots,\mu_n = 0$, the result follows. □

This proposition together with theorem 4.1 can be used to prove the following statement which is the main result of the paper.

Theorem 4.2 (Linear reservoir embeddings). *Let $\phi\in \mathrm{Diff}^2(M)$ be a dynamical system on a compact manifold M of dimension q that exhibits finitely many periodic orbits. Suppose that for each periodic orbit m of ϕ with period $n\in\mathbb{N}$, the derivative $T_m\phi^{-n}$ has q distinct eigenvalues $\lambda_1,\lambda_2,\ldots,\lambda_q$. Let now $\ell\in\mathbb{N}$ be the lowest common multiple of all the periods of the finite periodic points of ϕ and let $N\in\mathbb{N}$ such that $N>\max\{2q,\ell\}$.*

Construct now $\overline{A}\in\mathbb{M}_{N,N}$ and $\overline{\mathbf{C}}\in\mathbb{R}^N$ by drawing their entries using independent regular real-valued distributions. Then, there exist rescaled versions A and $\mathbf{C}$ of $\overline{A}$ and $\overline{\mathbf{C}}$, respectively, such that the GS $f_{(\phi,\omega,F)}\in C^2(M,\mathbb{R})$ associated to the state map $F(\mathbf{x},z) := A\mathbf{x}+\mathbf{C}z$ is almost surely an embedding for generic $\omega\in C^2(M,\mathbb{R}^N)$.

Proof. Proposition 4.1 guarantees that the randomly drawn elements $\overline{A}$ and $\overline{\mathbf{C}}$ satisfy almost surely the hypotheses in parts **(i)** and **(ii)** of the statement of theorem 3.1. However, in order to be able to invoke theorem 4.1, we need to use a linear state map F whose connectivity matrix A is such that $\rho(A) < 1$ and, for any observation map $\omega \in C^2(M, \mathbb{R})$, the corresponding GS $f_{(\phi,\omega,F)} \in C^2(M, \mathbb{R}^N)$ and the map $\Theta_{(\phi,F)} : C^2(M, \mathbb{R}) \longrightarrow C^2(M, \mathbb{R}^N)$ introduced in (2.8) are continuous. It is obvious from parts **(i)** and **(ii)** in proposition 2.3 that this can be achieved by choosing appropriate rescaling the matrix $\overline{A}$ and hence the statement follows from theorem 4.1. □

We emphasize that appropriate rescalings of $\overline{A}$ and $\overline{\mathbf{C}}$ that satisfy parts **(i)** and **(ii)** in proposition 2.3 do exist. From a practical point of view, the first condition can be easily satisfied by normalizing $\overline{A}$ with respect to its spectral radius and choosing any scaling multiplier in $(0, 1)$. The second condition, however, requires knowledge about the data-generating dynamical system (M, ϕ), to which, in principle, we have no access. This means that, unfortunately, we cannot formulate a practical recipe for choosing the rescaling for this case. Nevertheless, in section 5, we show that for a number of standard examples, choosing the entries of $\overline{\mathbf{C}}$ from a uniform distribution over $[-0.5, 0.5]$ and rescaling of $\overline{A}$ such that $\|A\| = 1$ yields synchronization maps that are amenable to learnability from data.

5. Numerical illustrations of attractor reconstruction, filtering, and forecasting

In this section we illustrate how the embeddings proposed in theorem 4.2 are able to reconstruct the attractor of various dynamical systems out of one-dimensional observations and, additionally, we show that randomly generated linear reservoir systems are efficient in the filtering and prediction of dynamical systems observations in the presence of additive noise.

Following the prescription proposed in the statement of theorem 4.2, we shall randomly generate linear systems of the form $F(\mathbf{x}, z) = A\mathbf{x} + \mathbf{C}z$ to which we shall feed in the input variable z finite-length one-dimensional ω observations of three different dynamical systems ϕ, namely, the Rossler system, the Van der Pol oscillator, and the Lorenz system. For each of these systems we shall create reservoir states $\mathbf{x}_t$ according to the recursion

$$\mathbf{x}_{t+1} = A\mathbf{x}_t + C\omega(\phi^t(m)). \tag{5.1}$$

Due to the results in the paper, we expect that the states $\mathbf{x}_t$ shall approximate $f_{(\phi,\omega,F)}(\phi^t(m))$ as $t \to \infty$ where $f_{(\phi,\omega,F)}$ is the corresponding embedding GS introduced in section 2. This will be done in practice by keeping only the states $\mathbf{x}_t$ for all $t > T$, for some $T \in \mathbb{N}$ where $[0, T]$ is called the washout period. The embedding properties of $f_{(\phi,\omega,F)}$ will become apparent in plots that will show that the dynamics of the original system and the state dynamics induced by its observations are topologically conjugate.

In order to illustrate the embedding properties of $f_{(\phi,\omega,F)}$ for the three dynamical systems we set up a reservoir system by following the steps:

1. Randomly generate a 7 by 7 matrix A' with IID uniform entries in the interval $[-0.5, 0.5]$.
2. Define the reservoir matrix $A := A'/\|A'\|$.
3. Randomly generate a vector $\mathbf{C} \in \mathbb{R}^7$ with IID uniform entries in the interval $[-0.5, 0.5]$.

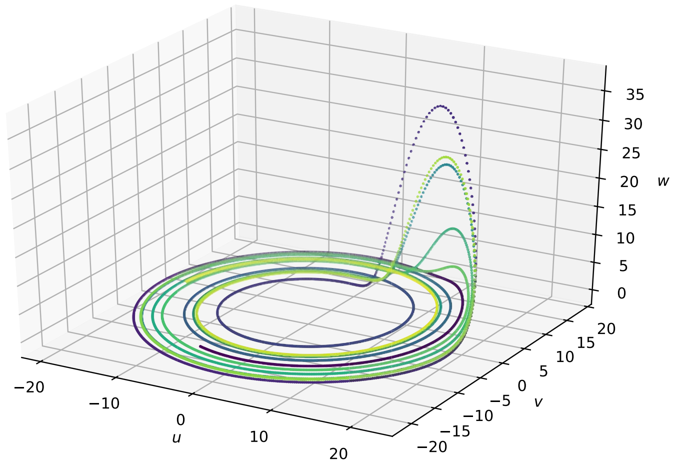

Figure 1. A trajectory of the Rössler system plotted for times in the interval $(60, 120)$. Points at the start of the trajectory are purple, and points later on are yellow.

5.1. The Rössler system

The Rössler system under a popular choice of parameters is described by the differential equations:

$$\begin{aligned}\dot{u} &= -v - w,\\ \dot{v} &= u + v/10,\\ \dot{w} &= 1/10 + w(u - 14).\end{aligned}$$

Using Python 3.7 and `scipy.integrate.odeint` we simulate a trajectory of the Rössler system from the initial condition $(u_0, v_0, w_0) = (2, 1, 5)$ for $T = 120$ time units, with time step $h = 0.01$. The result is plotted in figure 1 after using the interval $[0, 60]$ as washout period.

We then generate reservoir states using the recursion (5.1) and taking as observation map $\omega(u, v, w) := u$, that is, the first component of the Rössler system. A depiction of the projection of the corresponding states x_t onto the first three principal components is shown in figure 2.

5.2. The Van der Pol oscillator

We now repeat the same embedding procedure for the limit cycle of the Van der Pol oscillator with damping parameter μ is described by the differential equation in two dimensions

$$\begin{aligned}\dot{u} &= v,\\ \dot{v} &= \mu(1 - u^2)v - u.\end{aligned}$$

Using the same discretization scheme as before we simulated trajectories of the Van der Pol oscillator for 40 time units with time step $h = 0.01$ for five different damping parameter values $\mu = 0.5, 1, 1.5, 2, 2.5$ but using always the same initial condition $(u_0, v_0) = (-4, 5)$. The result is plotted in figure 3. Regarding the reservoir embedding we use a random linear system of dimension five using the same distributions for the entries as in previous paragraphs and we use as input the first component of the Van der Pol oscillator. A depiction of the projection of the corresponding states x_t onto the first two principal components is shown in figure 4 for each of the five different damping parameter values under consideration.

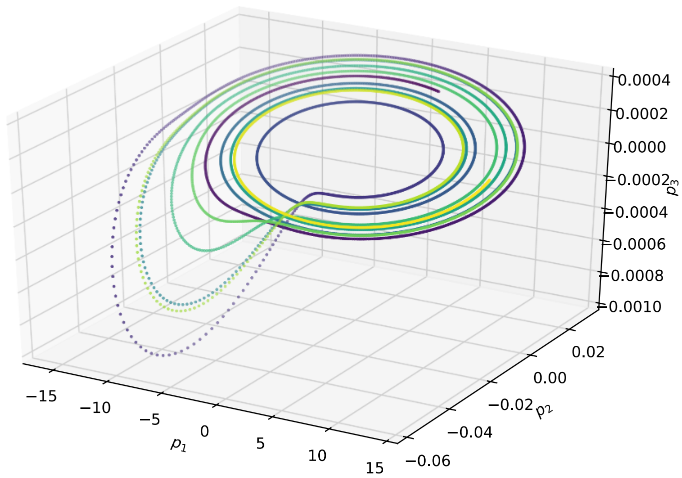

Figure 2. Projection onto the first three principal components of the reservoir states in the interval $(60,120)$ for a system driven by the u-components of the Rössler system. Points at the start of the trajectory are purple, and points later on are yellow. It is worth pointing out the impressive resemblance with the original dynamical system even though this picture has been generated only using one of its components.

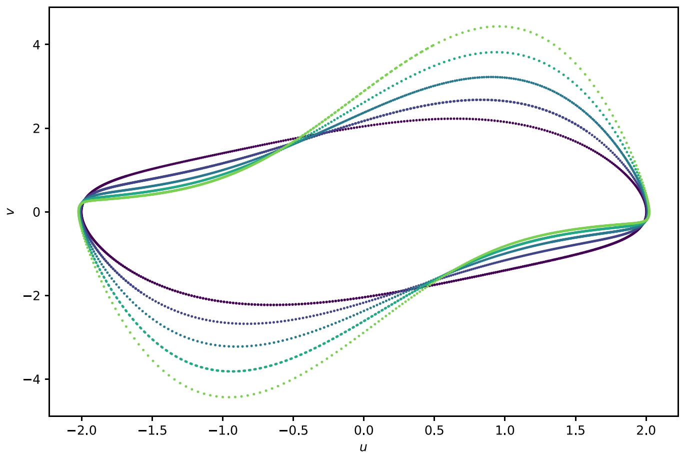

Figure 3. Limit cycles of the Van der Pol oscillator using damping parameters $\mu = 0.5, 1, 1.5, 2, 2.5$ plotted for times in the interval $(30,40)$. Darker colours denote lower values of μ, and brighter colours denote higher values of μ.

5.3. The Lorenz system. Attractor reconstruction

The Lorenz system with the parameter values given in the original paper [Lore 63] is determined by the differential equation

$$\begin{aligned} \dot{u} &= 10(u-v), \\ \dot{v} &= u(28-w)-v, \\ \dot{w} &= uv-8w/3. \end{aligned}$$

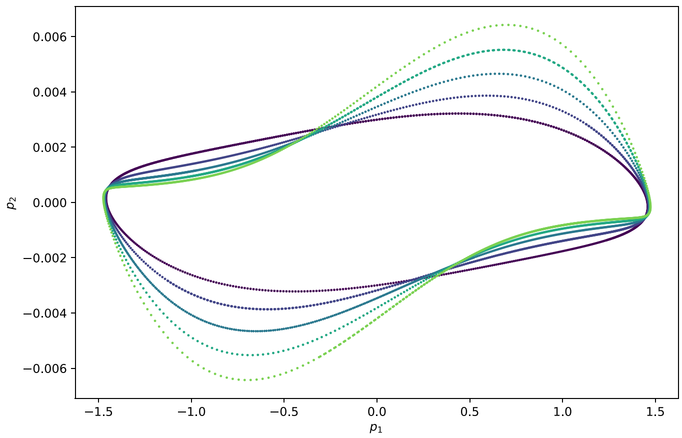

Figure 4. Projection onto the first two principal components of the reservoir states in the interval $(30, 40)$ for a system driven by the u-components of the Van der Pol oscillator with damping parameters $\mu = 0.5, 1, 1.5, 2, 2.5$. Darker colours denote lower values of μ, and brighter colours denote higher values of μ.

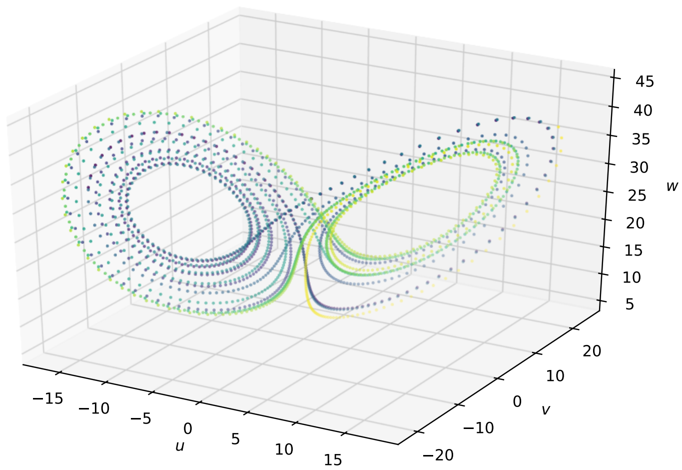

Figure 5. A trajectory of the Lorenz system plotted for times in the interval $(20, 40)$. Points at the start of the trajectory are purple, and points later on are yellow.

The discretization of this differential equation with time step h yields a time evolution operator given by

$$\phi(u_0, v_0, w_0) = (u_0, v_0, w_0) + \int_0^h (\dot{u}(t), \dot{v}(t), \dot{w}(t))\, \mathrm{d}t$$

where the curve $(u(t), v(t), w(t))$ solves the Lorenz equations with initial condition (u_0, v_0, w_0). In this paragraph, we follow the same modelling prescriptions that we used for the Rössler and Van der Pol systems. We take the initial condition as $(u_0, v_0, w_0) = (0, 1, 1.05)$ and time step $h = 0.01$ and the result for $T = 40$ is plotted in figure 5. Regarding the reservoir embedding we use the same dimensionality that we took for Rössler and we also use the u component as the input for state generation. A depiction of the projection of the corresponding states x_t onto the

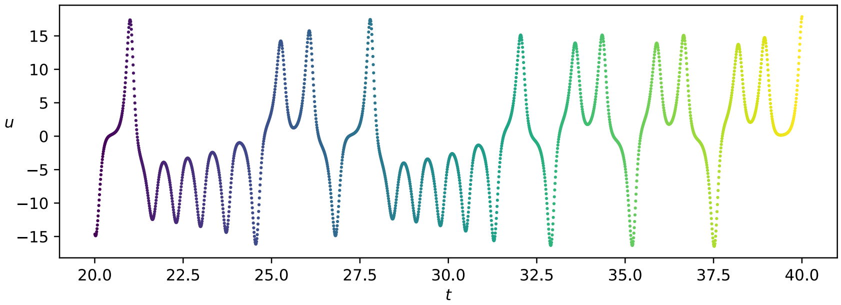

Figure 6. A trajectory of the u-component of the Lorenz system plotted for times in the interval $(20, 40)$. Points at the start of the trajectory are purple, and points later on are yellow.

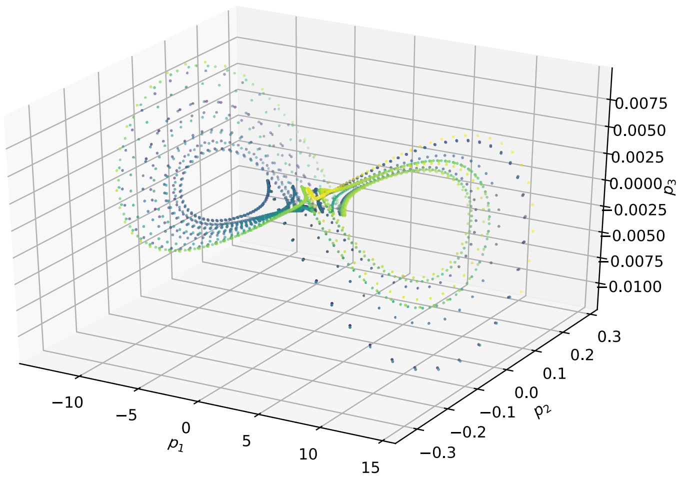

Figure 7. Projection onto the first three principal components of the reservoir states in the interval $(20, 40)$ for a reservoir system driven by the first component of the Lorenz system. Points at the start of the trajectory are purple, and points later on are yellow.

first three principal components is shown in figure 6. The corresponding states $\mathbf{x}_t$ in the interval $(20, 40)$, that is, after a washout period of $(0, 20)$ are plotted in figure 7 after a projection onto the first three principal components.

5.4. The Lorenz system. Forecasting in the presence of noise

In this paragraph we follow closely the experiment design that we previously used for attractor reconstruction but with a few modifications that we now list below. More precisely, we use the same parameters, initial point, and time step, but we consider $T = 11\ 000$ time units with a 1000 time units long washout period. The result for this particular choice of the time interval is plotted in figure 8. Consider now a reservoir system constructed according to the following prescription:

1. Randomly generate a random orthogonal 20 by 20 matrix A' drawn from the unique invariant Haar distribution in the Lie group $\mathrm{O}(20)$.
2. Define the reservoir matrix $A := 0.9 \cdot A'/\|A'\|$.

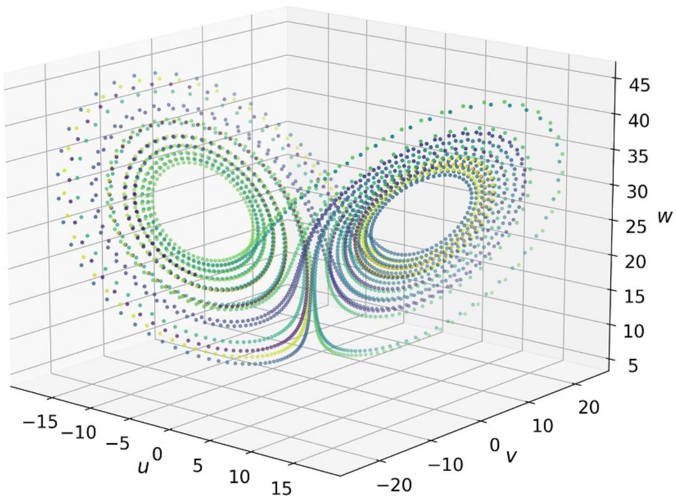

Figure 8. A trajectory of the Lorenz system plotted for times in the interval $(1020, 1050)$. Points at the start of the trajectory are purple, and points later on are yellow.

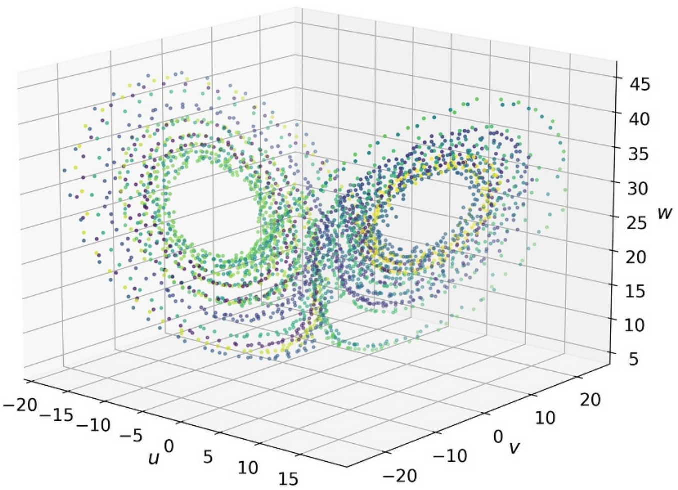

Figure 9. A trajectory of the Lorenz system with the u-component contaminated with additive Gaussian noise and the original v and w components plotted for times in the interval $(1020, 1050)$.

3. Randomly generate a vector $\mathbf{C}' \in \mathbb{R}^{20}$ with IID uniform entries in the interval $[-1, 1]$.
4. Define the reservoir input vector $\mathbf{C} := \mathbf{C}'/\|\mathbf{C}'\|$.

We now choose a readout $h : \mathbb{R}^M \longrightarrow \mathbb{R}^m$ for the state map using a deep neural network with 10 hidden layers of 20 neurons each and taking a scaled logistic map as activation function of the form $\sigma(s) = (z_{\max} - z_{\min})/(1 + e^{-s}) + z_{\min}$. This readout is trained using states that are obtained with inputs of the form $u_t' := u_t + \epsilon_t$, where u_t is the u-component of the Lorenz system and ϵ_t is a Gaussian distributed random variable with mean zero and variance 0.25 for all $t \in (1000, 11\,000)$. We illustrate a Lorenz trajectory for (u', v, w) in figure 9.

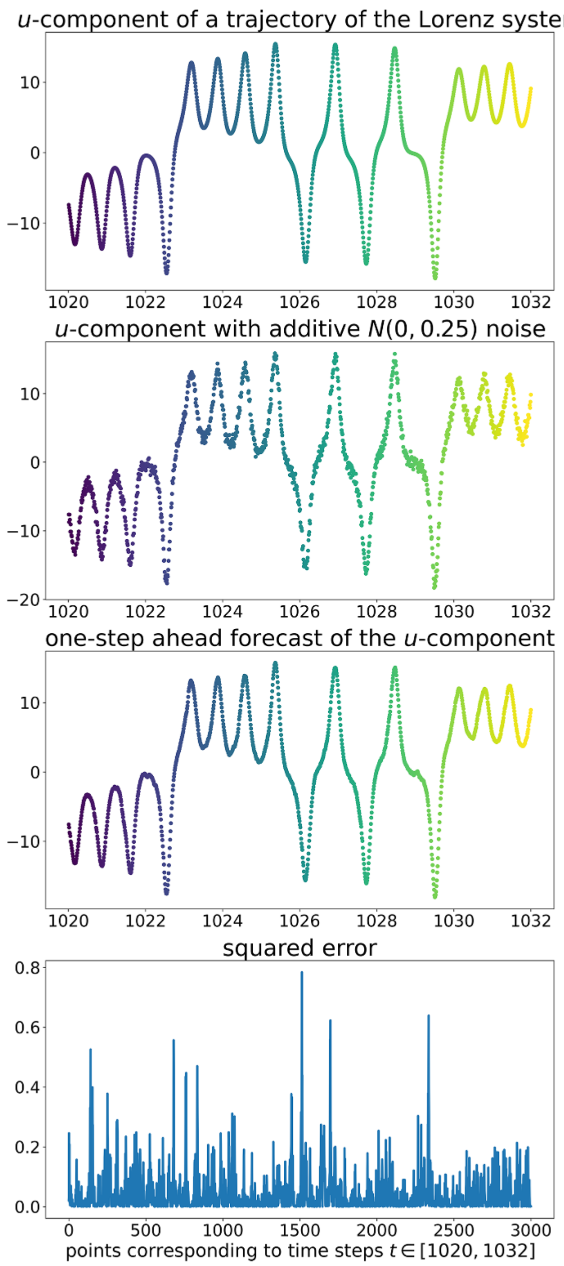

Figure 10. Filtered and forecasted u-component of the Lorenz system. The reservoir system is presented with a noisy version of the u-component time series as input and equipped with a trained deep neural network designed to filter and one-step-ahead forecast the time series.

The corresponding states $\mathbf{x}_t$ are then subsequently used as the input of the deep neural network. The weights of the neural network are obtained by solving the empirical risk minimization problem where the one-step ahead shifted time series of the original u-component

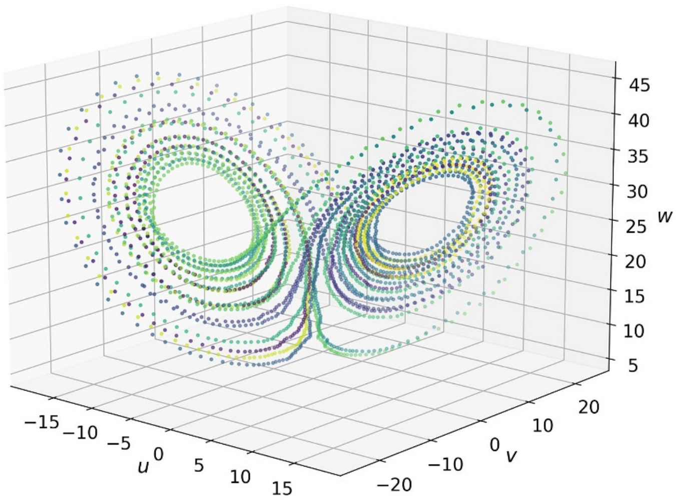

Figure 11. A trajectory of the Lorenz system with the filtered and forecasted u-component plotted for times in the interval $(1020, 1050)$. The reservoir system is presented with a noisy version of the u-component time series as input and equipped with a trained deep neural network designed to filter and one-step-ahead forecast the time series.

of the Lorenz system is taken as the target and the mean squared error is taken as the empirical risk. The learning task hence consists in the filtering of the noisy input and in the one-step ahead forecasting of the u-component of the dynamical Lorenz system. We implement the training of the deep neural network with the help of the Adam Optimizer in Keras in eight iterations with early stopping parametrized by the patience parameter of 500 epochs. Each iteration consists of 7000 epochs of batch size 10 000 and the learning rate is taken in a decreasing manner for each subsequent iteration out of the set of values $\{5 \times 10^{-3}, 3 \times 10^{-3}, 1 \times 10^{-3}, 9 \times 10^{-4}, 7 \times 10^{-4}, 5 \times 10^{-4}, 5 \times 10^{-5}, 3 \times 10^{-5}\}$.

The results on a testing sample are demonstrated in figure 10. We complement the illustration with figure 11 which supports the pertinence of the methodology that we propose for denoising and forecasting since the reconstructed attractor is difficult to distinguish from the original one given in figure 8.

Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: https://github.com/Learning-of-Dynamic-Processes/learnattractor.

Acknowledgments

We thank Friedrich Philipp for kindly communicating to us the proof of proposition 4.1. We also thank Henrik Brautmeier for his assistance with some of the numerical illustrations in the paper. AH is supported by a scholarship from the EPSRC Centre for Doctoral Training in Statistical Applied Mathematics at Bath (SAMBa), Project EP/L015684/1. J P O acknowledges partial financial support coming from the Swiss National Science Foundation (Grant No. 200021_175801/1).

Appendix

Elementary fact in linear algebra

Lemma 6.1. *Let A and B two square matrices of the same size such that* $\det(A) = 0$ *and* $\det(B) \neq 0$. *Then, there exists* $\varepsilon > 0$ *such that*

$$\det(A - \varepsilon B) \neq 0.$$

Proof. Consider the singular matrix $B^{-1}A$ and let λ_0 be its non-zero eigenvalue that has the smallest absolute value. Then, for any $0 < \varepsilon < |\lambda_0|$ we necessarily have that $\det\left(B^{-1}A - \varepsilon\mathbb{I}\right) \neq 0$ because otherwise ε would be an eigenvalue of $B^{-1}A$ which is impossible by the minimality of λ_0. This implies that $C := B^{-1}A - \varepsilon\mathbb{I}$ is invertible and hence so is $BC = A - \varepsilon B$, as required. □

As a corollary of this lemma we can conclude that if $\mathcal{V} := \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ and $\mathcal{W} := \{\mathbf{w}_1, \ldots, \mathbf{w}_n\}$ are two n-sets of vectors in $\mathbb{R}^n$, then there exists $\varepsilon > 0$ such that the set $\{\mathbf{v}_1 + \varepsilon\mathbf{w}_1, \ldots, \mathbf{v}_n + \varepsilon\mathbf{w}_n\}$ is made of linearly independent vectors. This fact is used at the end of the proof of Step 1 of theorem 3.1.

Two lemmas about random matrices

Lemma 6.2. *Let* $X_1, \ldots, X_n$ *be independent real-valued non-singular random variables and let p be a non-trivial polynomial in n complex variables. Then*

$$\mathbb{P}\left(p(X_1, \ldots, X_n) = 0\right) = 0.$$

Proof. Define $\mu_j(\cdot) := \mathbb{P}(X_j \in \cdot)$, $j = 1, \ldots, n$, and let $Z = \{\mathbf{x} \in \mathbb{C}^n \mid p(\mathbf{x}) = 0\}$ be the set of complex roots of the polynomial p. Then, since $X_1, \ldots, X_n$ are independent we have that

$$\mathbb{P}\left(p(X_1, \ldots, X_n) = 0\right) = \mathbb{P}\left((X_1, \ldots, X_n) \in Z\right) = (\mu_1 \otimes \ldots \otimes \mu_n)(Z).$$

We now proceed by induction over n. For $n = 1$ we have that $\mathbb{P}\left(p(X_1) = 0\right) = \mu_1(Z) = 0$ since Z is finite and X_1 is non singular. Let the claim be true for $n - 1$. For fixed $x_1 \in \mathbb{C}$ set $p_{x_1}(x_2, \ldots, x_n) := p(x_1, \ldots, x_n)$ and let

$$Z_{x_1} := \{(x_2, \ldots, x_n) \in \mathbb{C}^{n-1} \mid p_{x_1}(x_2, \ldots, x_n) = 0\}.$$

The set $F := \{x_1 \in \mathbb{R} \mid p_{x_1} \equiv 0\}$ is a finite set, so

$$\begin{aligned}\mathbb{P}\left(p(X_1, \ldots, X_n) = 0\right) &= \int_{\mathbb{C}} (\mu_2 \otimes \ldots \otimes \mu_n)(Z_{x_1}) \mathrm{d}\mu_1(x_1) \\ &= \int_{\mathbb{C}-F} (\mu_2 \otimes \ldots \otimes \mu_n)(Z_{x_1}) \mathrm{d}\mu_1(x_1) = 0,\end{aligned}$$

since we assumed that X_1 is non-singular and $(\mu_2 \otimes, \ldots, \otimes \mu_n)(Z_{x_1}) = 0$ for $x_1 \notin F$ by the induction hypothesis. □

Lemma 6.3. *Let* $N \in \mathbb{R}^N$, *let A be a real* $N \times N$ *matrix, and let* $\mathbf{C}$ *be a random vector in* $\mathbb{R}^N$. *Assume the entries of A and* $\mathbf{C}$ *have been drawn using independent non-singular real-valued random variables. Moreover, let* $p_1, \ldots, p_n \in \mathbb{C}[x]$ *be linearly independent polynomials in one variable of degree at most* $n - 1$. *Then*

$$\mathbb{P}(\det\left(p_1(A)\mathbf{C} | p_2(A)\mathbf{C} | \cdots | p_n(A)\mathbf{C}\right) = 0) = 0.$$

Equivalently, the vectors $p_1(A)\mathbf{C}, p_2(A)\mathbf{C}, \ldots, p_n(A)\mathbf{C}$ *are linearly independent almost surely.*

Proof. The expression $\det(p_1(A)\mathbf{C}|\cdots|p_n(A)\mathbf{C})$ is a polynomial p in the n^2+n variables a_{ij} and b_j, $i,j\in\{1,\ldots,n\}$ that constitute the entries of A and $\mathbf{C}$, respectively, and which in turn are by hypothesis non-singular random variables. As long as the polynomial p is not identically zero, the result follows directly from the lemma 6.2. So all that remains is to show that p is not identically zero, that is, that there exist particular choices of A and $\mathbf{C}$ such that $\det(p_1(A)\mathbf{C}|\cdots|p_n(A)\mathbf{C})$ is non-zero. So, we choose $\mathbf{C}=(1,\ldots,1)^\top$ and $A=\mathrm{diag}(a_1,\ldots,a_n)$ with distinct real numbers $a_1,\ldots,a_n$. We expand the polynomials p_j in terms of their coefficients $\gamma_{j1},\ldots,\gamma_{jn}$ so

$$p_j(x)=\sum_{k=0}^{n-1}\gamma_{jk}x^k.$$

The vectors $\gamma_j=(\gamma_{j1},\ldots,\gamma_{jn})^\top$, $j\in\{1,\ldots,n\}$, are by hypothesis linearly independent. We now show that with these choices, the vectors $p_1(A)\mathbf{C},p_2(A)\mathbf{C},\ldots,p_n(A)\mathbf{C}$ are linearly independent and hence $\det(p_1(A)\mathbf{C}|p_2(A)\mathbf{C}|\cdots|p_n(A)\mathbf{C})\neq 0$, as required. Indeed, let $c_1,\ldots,c_n\in\mathbb{R}$ and suppose that $\sum_{j=1}^n c_jp_j(A)\mathbf{C}=\mathbf{0}$. Additionally, we can write

$$\sum_{j=1}^n c_jp_j(A)\mathbf{C}=\sum_{j=1}^n c_j\sum_{k=0}^{n-1}\gamma_{kj}A^k\mathbf{C}=\sum_{k=0}^{n-1}\left(\sum_{j=1}^n c_j\gamma_{jk}\right)\begin{bmatrix}a_1^k\\ a_2^k\\ \vdots\\ a_n^k\end{bmatrix}=V\mathbf{x}$$

where

$$V=\begin{pmatrix}1 & a_1 & a_1^2 & \ldots & a_1^{n-1}\\ 1 & a_2 & a_2^2 & \ldots & a_2^{n-1}\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ 1 & a_n & a_n^2 & \ldots & a_n^{n-1}\end{pmatrix}\quad\text{and}\quad \mathbf{x}=\sum_{j=1}^n c_j\gamma_j.$$

Since the diagonal entries of A are all different and the determinant of the Vandermonde matrix V is given by

$$\det(V)=\prod_{1\leqslant i<j\leqslant n}(a_i-a_j)$$

we can conclude that V is invertible and hence the identity $\sum_{j=1}^n c_jp_j(A)\mathbf{C}=V\mathbf{x}=\mathbf{0}$ implies that $\mathbf{x}=\mathbf{0}$. By the linear independence of the vectors $\gamma_1,\ldots,\gamma_n$ we have that $c_1,\ldots,c_n=0$ necessarily. It follows that the vectors $p_1(A)\mathbf{C},p_2(A)\mathbf{C},\ldots,p_n(A)\mathbf{C}$ are linearly independent, as required. □

References

[Abra 67] Abraham R and Robbin J 1967 *Transversal Mappings and Flows* (W. A. Benjamin, Inc.)

[Abra 88] Abraham R, Marsden J E and Ratiu T S 1988 *Manifolds, Tensor Analysis and Applications* (*Applied Mathematical Sciences* vol 75) (Springer)

[Apos 74] Apostol T 1974 *Mathematical Analysis* 2nd edn (Addison Wesley)

[Bocc 02] Boccaletti S, Kurths J, Osipov G, Valladares D L and Zhou C S 2002 The synchronization of chaotic systems *Phys. Rep.* **366** 1–101

[Boot 03] Boothby W M 2003 *An Introduction to Differentiable Manifolds and Riemannian Geometry* 2nd edn (Academic)

[Carm 92] do Carmo M P 1992 *Riemannian Geometry* (Birkhäuser)

[Carr 18] Carroll T L 2018 Using reservoir computers to distinguish chaotic signals *Phys. Rev.* E **98** 52209
[Erog 17] Eroglu D, Lamb J S W and Pereira T 2017 Synchronisation of chaos and its applications *Contemp. Phys.* **58** 207–43
[Gaut 21] Gauthier D J, Bollt E, Griffith A and Barbosa W A S 2021 Next generation reservoir computing (arXiv:2106.07688)
[Gono 20a] Gonon L, Grigoryeva L and Ortega J-P 2020 Memory and forecasting capacities of nonlinear recurrent networks *Physica* D **414** 1–13
[Gono 20b] Gonon L and Ortega J-P 2020 Reservoir computing universality with stochastic inputs *IEEE Trans. Neural Netw. Learn. Syst.* **31** 100–12
[Gono 21] Gonon L and Ortega J-P 2021 Fading memory echo state networks are universal *Neural Netw.* **138** 10–13
[Grig 18] Grigoryeva L and Ortega J-P 2018 Echo state networks are universal *Neural Netw.* **108** 495–508
[Grig 21a] Grigoryeva L, Hart A G and Ortega J-P 2021 Chaos on compact manifolds: differentiable synchronizations beyond the Takens theorem *Phys. Rev.* E **103** 062204
[Grig 21b] Grigoryeva L and Ortega J-P 2021 Dimension reduction in recurrent networks by canonicalization *J. Geom. Mech.* **13** 647–77
[Hart 20] Hart A G, Hook J L and Dawes J H P 2020 Embedding and approximation theorems for echo state networks *Neural Netw.* **128** 234–47
[Hart 21] Hart A G, Hook J L and Dawes J H P 2021 Echo state networks trained by Tikhonov least squares are $L^2(\mu)$ approximators of ergodic dynamical systems *Physica* D **421** 132882
[Hirs 76] Hirsch M W 1976 *Differential Topology* (Springer)
[Huke 06] Huke J P 2006 Embedding nonlinear dynamical systems: a guide to Takens' theorem *Technical Report* (Manchester Institute for Mathematical Sciences, The University of Manchester)
[Jaeg 04] Jaeger H and Haas H 2004 Harnessing nonlinearity: predicting chaotic systems and saving energy in wireless communication *Science* **304** 78–80
[Jaeg 10] Jaeger H 2010 The 'echo state' approach to analysing and training recurrent neural networks with an erratum note *Technical Report* (German National Research Center for Information Technology)
[Kalm 10] Kalman R 2010 Lectures on controllability and observability *Controllability and Observability* (Springer) pp 1–149
[Kant 03] Kantz H and Schreiber T 2003 *Nonlinear Time Series Analysis* 2nd edn (Cambridge University Press)
[Kryl 31] Krylov A N 1931 On the numerical solution of equation by which are determined in technical problems the frequencies of small vibrations of material systems *News Acad. Sci. USSR* **7** 491–539
[Kupk 63] Kupka I 1963 Contributiona la théorie des champs génériques *Contrib. Differ. Equ.* **2** 457–84
[Lax 02] Lax P 2002 *Functional Analysis* (Wiley-Interscience)
[Lore 63] Lorenz E N 1963 Deterministic nonperiodic flow *J. Atmos. Sci.* **20** 130–41
[Lu 18] Lu Z, Hunt B R and Ott E 2018 Attractor reconstruction by machine learning *Chaos* **28** 061104
[Lu 20] Lu Z and Bassett D S 2020 Invertible generalized synchronization: a putative mechanism for implicit learning in neural systems *Chaos* **30** 063133
[Luko 09] Lukoševičius M and Jaeger H 2009 Reservoir computing approaches to recurrent neural network training *Comput. Sci. Rev.* **3** 127–49
[Maas 00] Maass W and Sontag E D 2000 Neural systems as nonlinear filters *Neural Comput.* **12** 1743–72
[Maas 02] Maass W, Natschläger T and Markram H 2002 Real-time computing without stable states: a new framework for neural computation based on perturbations *Neural Comput.* **14** 2531–60
[Maas 04] Maass W, Natschläger T and Markram H 2004 Fading memory and kernel properties of generic cortical microcircuit models *J. Physiol. Paris* **98** 315–30
[Maas 07] Maass W, Joshi P and Sontag E D 2007 Computational aspects of feedback in neural circuits *PLoS Comput. Biol.* **3** e165
[Manj 13] Manjunath G and Jaeger H 2013 Echo state property linked to an input: exploring a fundamental characteristic of recurrent neural networks *Neural Comput.* **25** 671–96

[Manj 20] Manjunath G 2020 Stability and memory-loss go hand-in-hand: three results in dynamics and computation *Proc. R. Soc.* A **476** 1–25
[Matt 92] Matthews M B 1992 On the uniform approximation of nonlinear discrete-time fading-memory systems using neural network models *PhD Thesis* ETH Zürich
[Matt 93] Matthews M B 1993 Approximating nonlinear fading-memory operators using neural network models *Circuits Syst. Signal Process.* **12** 279–307
[Munk 14] Munkres J 2014 *Topology* 2nd edn (Pearson)
[Nats 02] Natschläger T, Maass W and Markram H 2002 The "liquid computer": a novel strategy for real-time computing on time series *Telematik* **8** 39–43
[Ott 02] Ott E 2002 *Chaos in Dynamical Systems* 2nd edn (Cambridge University Press)
[Path 17] Pathak J, Lu Z, Hunt B R, Girvan M and Ott E 2017 Using machine learning to replicate chaotic attractors and calculate Lyapunov exponents from data *Chaos* **27** 121102
[Path 18] Pathak J, Hunt B, Girvan M, Lu Z and Ott E 2018 Model-free prediction of large spatiotemporally chaotic systems from data: a reservoir computing approach *Phys. Rev. Lett.* **120** 024102
[Peco 97] Pecora L M, Carroll T L, Johnson G A, Mar D J and Heagy J F 1997 Fundamentals of synchronization in chaotic systems, concepts and applications *Chaos* **7** 520–43
[Rulk 95] Rulkov N F, Sushchik M M, Tsimring L S and Abarbanel H D I 1995 Generalized synchronization of chaos in directionally coupled chaotic systems *Phys. Rev.* E **51** 980
[Saue 91] Sauer T, Yorke J A and Casdagli M 1991 Embedology *J. Stat. Phys.* **65** 579–616
[Smal 63] Smale S 1963 Stable manifolds for differential equations and diffeomorphisms *Ann. Scuola Norm. Super. Pisa-Classe Sci.* **17** 97–116
[Sont 98] Sontag E 1998 *Mathematical Control Theory: Deterministic Finite Dimensional Systems* (Springer)
[Take 81] Takens F 1981 *Detecting Strange Attractors in Turbulence* (Springer) pp 366–81
[Tana 19] Tanaka G, Yamane T, Héroux J B, Nakane R, Kanazawa N, Takeda S, Numata H, Nakano D and Hirose A 2019 Recent advances in physical reservoir computing: a review *Neural Netw.* **115** 100–23
[Verz 21] Verzelli P, Alippi C and Livi L 2021 Learn to synchronize, synchronize to learn *Chaos* **31** 083119