The Echo Index and multistability in input-driven recurrent neural networks

Andrea Ceni⁵, Peter Ashwin⁶, Lorenzo Livi⁷, Claire Postlethwaite⁸

⁵Department of Computer Science, University of Exeter, Exeter EX4 4QF, UK  
⁶Department of Mathematics, University of Exeter, Exeter EX4 4QF, UK  
⁷Departments of Computer Science and Mathematics, University of Manitoba, Winnipeg, MB R3T 2N2, Canada  
⁸Department of Mathematics, University of Auckland, Auckland 1142, New Zealand

Abstract

A recurrent neural network (RNN) possesses the echo state property (ESP) if, for a given input, it “forgets” any internal states of the driven (nonautonomous) system and asymptotically follows a unique (possibly complex) trajectory. Nonetheless, tasks involving RNNs may require the formation of multiple internal memory states in the absence of inputs, implying multistability of the associated autonomous system. How might these seemingly contradictory properties fit together? We give some insight into this question by introducing a generalisation of the ESP suitable for multistable RNNs subjected to inputs. We introduce an “echo index” to characterise the number of simultaneously stable responses of an RNN. Driven RNNs that display the classical ESP have echo index one. We show that it is possible for this index to change with inputs, highlighting a potential source of computational errors in RNNs.

Keywords: Nonautonomous dynamical systems, Input-driven systems, Recurrent neural networks, Echo state property, Multistability, Machine learning.
1. Introduction

Recurrent neural networks (RNNs) are input-driven (i.e. nonautonomous) dynamical systems \[1\] whose behaviour depends both on training and on inputs to the trained system. In order to describe responses of a trained RNN to the full range of possible inputs, it is necessary to go beyond the theory of autonomous (input-free) dynamical systems and consider more general nonautonomous dynamical systems \[2\], where the equations ruling the dynamics change over time. The theory of nonautonomous dynamical systems is much less developed than that of autonomous systems, and notions such as convergence (and hence attractors) need to be carefully defined \[3\].

Starting with the work of Jaeger \[4\], several authors have proposed that a successfully trained RNN should have the so-called *echo state property* (ESP). The idea of “echo state” gave rise to a training paradigm of RNNs called *reservoir computing* \[5\] and a class of RNNs known as *echo state networks* (ESNs) \[6\]. An ESN is an RNN that is relatively easy to train, since optimisation is restricted to the output layer and recurrent connections are left untouched after a random initialisation. Yildiz and Jaeger propose a necessary condition for successful training of an ESN: the echo state property (ESP) \[7\].

If an RNN possesses the ESP this means that, for a given input sequence, it will asymptotically produce the same sequence of states regardless of initial conditions: it will “forget” any initial internal state and end up following a unique (though possibly very complex) trajectory in response to that input \[8\]. This trajectory represents the solution of the specific problem encoded in the input sequence that is fed to the network and so the system can be seen as acting as a *filter* that transforms the input sequence into a unique sequence of output \[9\]. On the other hand, the simultaneous existence of multiple attractors is common in RNNs without input for solving specific tasks that involve some degree of memory \[10\]. This multistability may well be also a byproduct of conventional training mechanisms \[11\].

In this paper, we analyse the ESP through the lens of nonautonomous dynamical system theory and introduce, in Section 3, a generalisation of the ESP that allows us to relate the ESP and multistability dependent on inputs. We prove in Proposition 3.1 that the ESP corresponds to existence and uniqueness of a global pullback attractor with specific features. The ESP is a special case of an input-driven system that depends not only on the system state but also on the specific input sequence. For a fixed input sequence, different initial conditions for the internal state of an RNN can, in principle, generate different responses. However, for a reliable output of the RNN this might be a *symptom of malfunction* due either to unsuccessful training, or an input that is too far from the training set to be correctly classified. Multiple responses are typically a problem in forecasting tasks, where the only possible correct response consists of closely approximating the input signal some steps ahead.

Nonetheless, input-driven computations dependent on initial conditions are instrumental for some state-dependent classification tasks. For example, consider the addition modulo \(N\), i.e. the addition in a modular...
arithmetic. Given an initial integer value \( n_0 \) and an input sequence of integers \( n_1, n_2, \ldots \), assuming values in \( \{0, 1, \ldots, N-1\} \), the neural network needs to compute online the sum \( N_k := \sum_{i=0}^{k} n_i \mod N \) as time \( k \) runs. Therefore, in this case the system has to respond in \( N \) different ways depending on the initial state.

We account for multistability by defining a simple, yet non-trivial class of attracting solutions for input-driven systems and count how many of these solutions coexist in phase space under the action of inputs. We introduce the \textit{echo index} that quantifies the degree of multistability of the system driven by a specific input. A parallel idea is introduced in [12], where the authors proposed an algorithm to empirically determine the number of simultaneous solutions and hence assess the presence of the ESP.

In this paper, we use the formalism of nonautonomous dynamical systems theory and consider \textit{local point attractors}. In Section 3.2 we show that the existence of a unique local point attractor in phase space coincides with the presence of ESP. The presence of multiple local point attractors can be useful in cases where the RNN has to change behaviour according to control inputs. For example, the task described in [13, Figure 5] shows how a network can learn to switch between two “attracting trajectories” according to control inputs.

The remainder of this paper is structured as follows. Background material is introduced in Section 2. In Section 3 we describe our main contribution: a generalisation of the ESP for multistable, nonautonomous systems driven by inputs. Our main theoretical results are here: in Theorem 3.2 we provide a sufficient condition for existence and uniqueness of a local point attractor, in Theorem 3.4 we prove the existence of multistability in systems perturbed by low amplitude inputs, while in Theorem 3.6 we prove that forcing RNNs with inputs of large enough amplitude will induce the ESP. In Section 4 we report some numerical experiments showing different numbers of local point attractors and highlighting their input-driven nature. In Section 4.1 we show how the echo index depends, for a given RNN, on the particular input driving the dynamics illustrating a bifurcation from the classical ESP to multistable dynamics. Section 4.2 provides another example of RNN dynamics possessing echo index 2. Finally, Section 5 concludes the paper and provides final remarks. Some background definitions, results and proofs are provided in appendices located at the end of the manuscript. In particular, we show in Appendix A that the \textit{natural association} of [8] corresponds to the system’s pullback attractor [2].

2. Nonautonomous dynamics of recurrent neural networks

In this section, we investigate RNN dynamics using tools from nonautonomous dynamical systems theory. In Section 2.1 we highlight the input-driven nature of RNN dynamics before introducing, in Section 2.2, the formalism of nonautonomous dynamical systems [2] in the more general context of input-driven systems.
2.1. Recurrent neural networks

A general discrete-time RNN \[14, 15\] is an evolution of \(x[k] \in X \subset \mathbb{R}^{N_r}\) in discrete time \(k \in \mathbb{Z}\) according to

\[
x[k + 1] = \phi(W_r x[k] + W_i u[k + 1] + W_f z[k]),
\]
\[
z[k + 1] = \psi(x[k + 1]),
\]

where \(\phi(\cdot)\) is a component-wise activation function (e.g. hyperbolic tangent) and \(u[k] \in U\) is the \(N_i\)-dimensional input sequence for \(U \subset \mathbb{R}^{N_i}\). The sets \(X\) and \(U\) denote the state (or phase) and input spaces, respectively. The matrices \(W_r \in \mathbb{R}^{N_r \times N_r}\) and \(W_i \in \mathbb{R}^{N_r \times N_i}\) represent recurrent and input-to-network couplings. The output feedback matrix \(W_f \in \mathbb{R}^{N_r \times N_o}\) injects the last computed output into the state-update equation. The particular implementation of \(\psi(\cdot)\) depends on the task at hand. For instance, in classification \(\psi(\cdot)\) might take the form of a softmax assigning probabilities to predicted classes; in forecasting is usually a linear or nonlinear deterministic function, e.g. \(z[k + 1] = \psi(x[k + 1]) = W_o x[k + 1]\). Sometimes, in place of (1), it is useful to consider RNNs with leaky neurons \(16\), namely:

\[
x[k + 1] = (1 - \alpha)x[k] + \alpha \phi(W_r x[k] + W_i u[k + 1] + W_f z[k]),
\]

where \(\alpha \in (0, 1]\) sets the time-scale of the RNN \(14\).

Training of RNNs is typically implemented by means of stochastic gradient descent (SGD) or variation of thereof \(17\). Learning long-term dependencies with SGD is known to be problematic, as a consequence of the so-called vanishing/exploding gradient problem \(18\). For this purpose, two types of methods have been proposed: (i) gating mechanisms \(19, 20\) and (ii) approaches based on unitary matrices and constant-slope activation functions \(21\). ESNs \(22, 6, 23, 24, 25, 9\), a special class of RNNs, bypass this problem as training targets the output layer weights \(W_o\) only. The recurrent layer, called a reservoir, is randomly instantiated (although more sophisticated methods have been proposed in the literature \(20\)) and the model is modified only offline at the hyper-parameter level. This simple training protocol is not sufficient in many applications requiring to learn memory states. To this end, training mechanisms based on output feedback \(27, 28\) and online training \(13, 29\) have been proposed, with successful applications in physics \(30, 31\), complex systems modeling \(32, 33\), and neuroscience \(34\), just to name a few.

2.2. Input-driven dynamical systems

The state-update equation (1) of a (trained) RNN is a special case of an input-driven dynamical system:

\[
x[k + 1] = G(u[k + 1], x[k]),
\]

for some map \(G : U \times X \rightarrow X\). The action of inputs \(u[k]\) driving the dynamics of \(x[k]\) gives an explicit time-dependence of (1) and means this system is nonautonomous. A given input sequence \(u = \{u[k]\}_{k \in \mathbb{Z}}\)
induces a sequence of maps \( \{f_k\}_{k \in \mathbb{Z}} \), where \( f_k(\cdot) := G(u[k+1], \cdot) : X \to X \) is the map ruling the update of the RNN state at time \( k \), i.e. \( x[k+1] = f_k(x[k]) \). Note that \([4]\) takes into account the output feedback, \( W_f z[k] \). We make the following standing assumption on \([4]\).

Assumption 2.1. We suppose that

(i) \( G \) is continuously differentiable in all arguments, i.e \( G \in C^1(U \times X, X) \);

(ii) for all \( u \in U \), the map \( G(u, \cdot) : X \to X \) is a local diffeomorphism onto its image;

(iii) \( U \subset \mathbb{R}^N_u \) is compact and \( X \subset \mathbb{R}^N_r \) is usually the compact closure of a \( N_r \)-dimensional Cartesian product of real intervals.\(^1\)

Note that for \([2]\)-\([3]\) we have \( G(u, x) = (1 - \alpha)x + \alpha \phi(W_r x + W_j u + W_f \psi(x)) \), and as long as the neuronal activation function \( \phi \) and the readout function \( \psi \) are both regular of class \( C^1 \), then the map \( G \) will be too. Moreover, for all \( u \in U \), the map \( G(u, \cdot) : X \to X \) is a local diffeomorphism whenever the “effective recurrent matrix" \( W_r + W_f D_x \psi(\cdot) \) is invertible, where \( D_x \psi(\cdot) \) denotes the Jacobian matrix of \( \psi \) \([2]\). In addition, if \( \phi \) is bounded with image \((-L, L)\) then the compact state space \( X \) can be assumed to be the hypercube \([-L, L]^N_r \) := \( \{(x_1, \ldots, x_N_r) \in \mathbb{R}^{N_r} \mid x_i \in [-L, L], i = 1, \ldots, N_r\} \), e.g. if \( \phi = \tanh \) then \( L = 1 \), see Proposition \([A.1]\).

Assumption 2.1(ii) implies that the preimage of any zero measure set is also zero measure, hence given any input sequence \( \{u[k]\}_{k \in \mathbb{Z}} \) assuming values in \( U \), for any \( Z \subset X \) with \( \lambda(Z) = 0 \), then \( \lambda(f_k^{-1}(Z)) = 0 \) for all \( k \in \mathbb{Z} \), where \( \lambda \) denotes Lebesgue measure on \( X \subset \mathbb{R}^{N_r} \). This is weaker than assuming that \( G(u, \cdot) \) is invertible for fixed \( u \), but it means that phase space volume cannot “suddenly collapse”.

2.3. The cocycle formalism

There are several ways to describe nonautonomous systems \([2]\): the process and the skew product (also called cocycle) formalism. In this paper, we use the cocycle formalism that is convenient when describing the input evolution as a shift in sequence space.

Let \((X, d_X)\) and \((U, d_U)\) be compact metric spaces. Time evolution will be parametrised through the ring of integers \( \mathbb{Z} \) or a subset of it. We write \( \mathbb{Z}^+ := \{k \in \mathbb{Z} : k \geq 1\} \) and \( \mathbb{Z}_0^- := \{k \in \mathbb{Z} : k \leq 0\} \). Let \( T \) be one of the sets \( \mathbb{Z}, \mathbb{Z}^+, \text{ or } \mathbb{Z}_0^- \), we consider the set \( U^T := \{u = \{u[k]\}_{k \in T} : u[k] \in U, \forall k \in T\} \) of all input sequences assuming values in set \( U \), and we will denote \( U := U^\mathbb{Z}, U^+ = U^{\mathbb{Z}_0^+} \) and \( U^- = U^{\mathbb{Z}_0^-} \). Moreover, given \( u \in U \), we will denote with \( u^+ \) and \( u^- \) the projection of \( u \) to \( U^+ \), and \( U^- \), respectively. The set \( U \) is a compact metric space for the product topology induced by the metric

\[
d_U(u, v) := \sum_{k \in \mathbb{Z}} \frac{d_U(u[k], v[k])}{2^{|k|}}. \quad \tag{5} \label{eq:metric-inj}
\]

\(^1\)We make exceptions for Appendix \([A]\) where \( X \) is only assumed to be a complete metric space.
The dynamics on this space of input sequences is described by means of the *shift operator*, which is an application \( \sigma : \mathcal{U} \to \mathcal{U} \) defined as follows:

\[
\sigma \left( \{ u[k] \}_{k \in \mathbb{Z}} \right) := \{ u[k+1] \}_{k \in \mathbb{Z}}.
\]

The composition \( \sigma^n \) defines a discrete dynamical system on the metric space \((\mathcal{U}, d_{\mathcal{U}})\) \([2]\). Setting \( \sigma^0(u) := u \) we have that \( \{ \sigma^n(u) \}_{n \in \mathbb{Z}} \) represents a group of homeomorphisms on \( \mathcal{U} \), which expresses the sequential forward or backward shift in time of all input sequences. Moreover, defining \( p : \mathcal{U} \to \mathcal{U} \) as the projection mapping

\[
p \left( \{ u[k] \}_{k \in \mathbb{Z}} \right) := u[0],
\]

we get in \( \mathcal{U} \) the current value of input sequence \( u \) as \( p(\sigma^n(u)) = u[n] \).

We describe the dynamics in response to input using a *cocycle map* \([2\text{, Definition 2.1, page 28}]\) for RNNs.

**Definition 2.1.** The nonautonomous dynamical system \((4)\) can be described using a cocycle mapping \( \Phi : \mathbb{Z}_0^+ \times \mathcal{U} \times X \to X \) as follows:

\[
\Phi(0, u, x_0) := x_0, \quad \forall x_0 \in X,
\]

\[
\Phi(n, u, x_0) := G(p(\sigma^n(u)), \Phi(n - 1, u, x_0)), \quad \forall x_0 \in X, \forall u \in \mathcal{U}, \forall n \in \mathbb{Z}_0^+.
\]

The set \( \{ \Phi(n, \cdot, \cdot) \}_{n \in \mathbb{Z}_0^+} \) forms a semigroup of continuous functions from \( \mathcal{U} \times X \) to \( X \). In particular, relations \((8)\)-\((9)\) imply the cocycle property:

\[
\Phi(m + n, u, x_0) = \Phi(n, \sigma^m(u), \Phi(m, u, x_0)),
\]

for any \( m, n \in \mathbb{Z}_0^+ \), \( x_0 \in X \) and \( u \in \mathcal{U} \).

Note that \((4)\) implies that the forward map is always defined, i.e. given a point \( x_0 \in X \) and any input \( u \in \mathcal{U} \), the forward trajectory is uniquely defined by \((4)\). On the other hand, it is possible that a backward trajectory does not exist, or may not be unique if one does exist. For example, although the one-dimensional map \( G(u, x) := \tanh(\mu x + u) \) is invertible for any fixed \( \mu \in (0, 1) \), every backward trajectory constructed from a \( x \neq 0 \) leads outside the compact set \([-1, 1]\) in a finite number of backward steps, thus making impossible to obtain a further preimage of \( \tanh \). This means it may not be possible to extend the cocycle mapping backward in time. Trajectories that are well-defined in the infinite past play an important role in the nonautonomous dynamics, as expressed in the next definition \([8]\).\(^2\)

**Definition 2.2.** An entire solution for the system in \((4)\) with input \( u := \{ u[k] \}_{k \in \mathbb{Z}} \in \mathcal{U} \) is a bi-infinite sequence of states \( \{ x[k] \}_{k \in \mathbb{Z}} \) that satisfies \((4)\) for all \( k \in \mathbb{Z} \). In other words,

\[
\Phi(s, \sigma^m(u), x[m]) = x[m + s]
\]

Note we do not consider invariant sets or entire solutions in terms of the skew product formalism \([2\text{, Definition 2.19}]\) but rather in terms of the process \([2\text{, Definition 2.14}]\) induced by a given input sequence \( v \).
for all \( m \in \mathbb{Z} \) and \( s \in \mathbb{Z}_0^+ \).

Assuming the existence of an entire solution \( \{ x[k] \}_{k \in \mathbb{Z}} \) for input \( u \in \mathcal{U} \), and exploiting the forward definition of the cocycle mapping, we can write the past evolution as follows:

\[
x[-n] = \Phi(m-n, \sigma^{-m}(u), x[-m]), \quad \forall n \in \mathbb{Z}_0^+ \quad \text{and} \quad m \in \mathbb{Z}_0^+ \quad \text{with} \quad m \geq m.
\]

(11) \( \text{eq:cocycle_back} \)

Such a relation expresses the fact that the point \( x[-n] \) is the resulting state of the system if we start from \( x[-m] \) and drive the dynamics with the sequence of input values \( u[-m+1], \ldots, u[-n] \).

2.4. Pullback attractors

In this section we introduce some basic definitions of the theory of nonautonomous dynamical systems, including the definition of pullback attractor: the ESP notion in [4] is formulated using left-infinite sequences and a pullback argument. The following definition extends the notion of entire solution.

**Definition 2.3.** Consider a nonautonomous system defined by a cocycle mapping as in (2.1) with input sequence \( u \in \mathcal{U} \). A family of nonempty compact sets \( A = \{ A_n \}_{n \in \mathbb{Z}} \) is called an invariant nonautonomous set for input \( u \) if

\[
\Phi(s, \sigma^m(u), A_m) = A_{s+m}.
\]

for all \( m \in \mathbb{Z} \) and \( s \in \mathbb{Z}_0^+ \).

Entire solutions are invariant nonautonomous sets where each \( A_n \) is a single point. Invariant nonautonomous sets turn out to be composed by entire solutions [2, Lemma 2.15]. Replacing “=” with “\( \subseteq \)” in Definition 2.3, we obtain the definition of a positively invariant family of sets.

**Definition 2.4.** A family of nonempty compact sets \( B = \{ B_n \}_{n \in \mathbb{Z}} \) is called a positively invariant nonautonomous set for input \( u \) (or simply \( u \)-positively invariant) if

\[
\Phi(s, \sigma^m(u), B_m) \subseteq B_{s+m}.
\]

for all \( m \in \mathbb{Z} \) and \( s \in \mathbb{Z}_0^+ \).

Note that the assumption that \( G \) is well defined as a map to \( X \) implies that \( X \) itself is positively invariant. Invariant sets play an important role for understanding the behaviour of a dynamical system. The most relevant ones are those invariant sets that attract the surrounding trajectories. Since, in nonautonomous systems, the equations ruling the dynamics change with time, the notion of attraction can be formulated in various ways, leading to definitions of forward attraction and pullback attraction. It is interesting to note that pullback attractors share more properties with the autonomous counterparts [3].

We introduce notion of pullback attraction of a point and the (global) attractor [2]. Let \( h \) be Hausdorff semi-distance for the metric space \( (X, d_X) \).
Definition 2.5. [2, Definition 3.3] An invariant nonautonomous set $A = \{A_n\}_{n \in \mathbb{Z}}$ for input $u$ pullback attracts $x_0 \in X$ if
\[
\lim_{k \to \infty} h(\Phi(k, \sigma^{-k+n}(u), x_0), A_n) = 0, \quad \forall n \in \mathbb{Z}.
\] (12) \hspace{1cm} \text{eq:pullback_attraction}

Note that as $X$ is bounded in our context, all bounded subsets of $X$ are pullback attracted to $A_n$ [2, Definition 3.4] if and only if $X$ is: we therefore define pullback attractor as follows:

Definition 2.6. An invariant nonautonomous set $A = \{A_n\}_{n \in \mathbb{Z}}$ for input $u$ consisting of nonempty compact sets is called a (global) pullback attractor of the system (4) driven by $u$ if
\[
\lim_{k \to \infty} h(\Phi(k, \sigma^{-k+n}(u), X), A_n) = 0, \quad \forall n \in \mathbb{Z}.
\] (13) \hspace{1cm} \text{eq:pullback_attractor}

It is easy to show from this that if $A = \{A_n\}$ is a pullback attractor for input $u$ then, for any $k \in \mathbb{Z}$, \{A_{n+m}\} is a pullback attractor for $\sigma^m(u)$.

Manjunath and Jaeger [8] introduce the notion of natural association as follows.

Definition 2.7. [8, Definition 4] Consider an input sequence $v := \{v[k]\}_{k \in \mathbb{Z}} \in U$. The sequence of sets $\{X_n(v)\}_{n \in \mathbb{Z}}$ defined by
\[
X_n(v) := \bigcap_{m \leq n} \Phi(n - m, \sigma^m(v), X), \quad n \in \mathbb{Z},
\] (14) \hspace{1cm} \text{eq:natural_association}
is called the natural association of the process induced by input sequence $v$.

In Theorem [A.4] we show that the natural association in Definition 2.7 is a pullback attractor. As a consequence, we can consider (14) as the pullback attractor whenever it is convenient. We define the notion of local pullback attractor [2, Definition 3.48] as follows.

Definition 2.8. We say a positively invariant nonautonomous set $Q = \{Q_n\}_{n \in \mathbb{Z}}$ for input $u$ made by nonempty compact sets is pullback attracted by an invariant nonautonomous set $A = \{A_n\}_{n \in \mathbb{Z}}$ for input $u$ if
\[
A_n := \bigcap_{m \leq n} \Phi(n - m, \sigma^m(u), Q_m), \quad n \in \mathbb{Z}.
\] (15) \hspace{1cm} \text{eq:local_pullback}

An invariant nonautonomous set $A = \{A_n\}_{n \in \mathbb{Z}}$ for input $u$ is a local pullback attractor for input $u$ if there exists $\eta > 0$ and a positively invariant nonautonomous set $Q = \{Q_n\}_{n \in \mathbb{Z}}$ for input $u$, such that $A$ pullback attracts $Q$ and $Q$ contains a uniform neighbourhood of the invariant nonautonomous set, i.e. there is an $\eta > 0$ such that $B_\eta(A_n) \subset Q_n$, for all $n \in \mathbb{Z}$.

3. An echo index for recurrent neural networks

We consider the Echo State Property for systems with specific inputs as in [S], and generalise this to give a notion of multistability under the action of inputs.
3.1. Entire solutions and the Echo State Property

Recall from Definition 2.2 that an entire solution for input $v = \{v[k]\}_{k \in \mathbb{Z}} \in U$ is a sequence $x = \{x[k]\}_{k \in \mathbb{Z}}$ that satisfies (4), for all $k \in \mathbb{Z}$.

**Definition 3.1.** [Definition 2] The nonautonomous system (4) has the Echo State Property (ESP) for input $v \in U$ if there exists a unique entire solution $x$.

We give a simple proof of the fact that, if the ESP holds for a given $v \in U$, then the entire solution is the pullback attractor as in Definition 2.6 using the notation of the cocycle mapping introduced in Section 2.3.

**Proposition 3.1.** If (4) has the ESP for input $v \in U$, then the (unique) entire solution is the global pullback attractor of the system (4) with input $v$.

**Proof.** First, note that if $\{x^*[k]\}_{k \in \mathbb{Z}}$ is an entire solution for input $v \in U$, then by the definition of $\Phi$, it holds true that $x^*[n] = \Phi(n - m, \sigma^m(v), x^*[m]) \forall m < n, \forall n \in \mathbb{Z}$. Compactness of $X$ guarantees that $x^*[m] \in X$ from which it follows that $x^*[n] \in \bigcap_{m < n} \Phi(n - m, \sigma^m(v), X) \forall n \in \mathbb{Z}$. The ESP hypothesis means $\{x^*[k]\}_{k \in \mathbb{Z}}$ is the only entire solution for input $v \in U$. Therefore, we need to prove that for all $n \in \mathbb{Z}$ the pullback attractor is composed of the singleton $x^*[n] = \bigcap_{m < n} \Phi(n - m, \sigma^m(v), X)$. We prove this by contradiction: suppose there exists a $n_0 \in \mathbb{Z}$, where w.l.o.g. $n_0 = 0$, and a $y_0 \in X$ with $y_0 \neq x^*[0]$ such that

$$\{x^*[0], y_0\} \subset \bigcap_{m < 0} \Phi(-m, \sigma^m(v), X).$$

This means that $y_0 \in \Phi(-m, \sigma^m(v), X) \forall m \leq -1$, i.e. for each $m \leq -1$ there exists at least one point, let us denote it $y[m] \in X$, such that $y_0 = \Phi(-m, \sigma^m(v), y[m])$. In other words, there exists a backward trajectory $y^- := \{\ldots, y[-2], y[-1], y[0] := y_0\}$ for the point $y_0$. Now, by iterating (4) forward in time from $y[0] := y_0$, we get the entire solution $y := \{y[k]\}_{k \in \mathbb{Z}}$. This gives a contradiction. □

Jaeger [4] shows that if the ESP holds for all inputs in a given compact set then the unique entire solution is forward attracting [4, Definition 4]. In Appendix A.1 Proposition A.6, we provide an alternative proof of this fact by applying a result found in the literature to RNNs as special class of nonautonomous dynamical systems. Note that in Proposition A.6 we do consider the feedback of the output.

3.2. Local point attractors and the echo index

If the ESP does not hold, then clearly a wide variety of behaviours are possible. We will consider the special case where there are only a finite number of possible attracting asymptotic responses – in this case we will say there is a “finite echo index”. For this we use the notion of uniform local point attractor; [3, cf Definition 3.11].
Definition 3.2. Consider a fixed input sequence \( u \in \mathcal{U} \), an entire solution \( \{x[k]\}_{k \in \mathbb{Z}} \) and a positively invariant nonautonomous set \( \{B[k]\}_{k \in \mathbb{Z}} \) composed of compact sets.

(i) If

\[
\lim_{k \to \infty} \left( \sup_{j \in \mathbb{Z}} h(\Phi(k, \sigma^j(u), B[j]), x[j + k]) \right) = 0
\]

then we say \( \{B[k]\}_{k \in \mathbb{Z}} \) is uniformly attracted to \( \{x[k]\}_{k \in \mathbb{Z}} \).

(ii) We say \( \{x[k]\}_{k \in \mathbb{Z}} \) is a uniform local point attractor if there is a neighbourhood \( \{B[k]\}_{k \in \mathbb{Z}} \) of \( \{x[k]\}_{k \in \mathbb{Z}} \) that is uniformly attracted to \( \{x[k]\}_{k \in \mathbb{Z}} \).

Note that if \( \{x[k]\}_{k \in \mathbb{Z}} \) is a uniform local point attractor, then it is both a forward attractor and a pullback attractor; [35] discusses this in more detail.

Uniform local point attractors allow us to define the echo index as a function of the specific input driving the dynamics, as expressed by the following definition.

Definition 3.3. We say the system \( (4) \) with input \( u \) admits a decomposition into \( n > 0 \) uniform local point attractors if there are \( n \) uniform local point attractors \( \{x_1[k]\}_{k \in \mathbb{Z}}, \ldots, \{x_n[k]\}_{k \in \mathbb{Z}} \) such that, for all \( \eta > 0 \) and \( i = 1, \ldots, n \), there are neighbourhoods \( \{B_1^\eta[k]\}_{k \in \mathbb{Z}} \) uniformly attracted by \( \{x_i[k]\}_{k \in \mathbb{Z}} \) and

\[
\lambda(X \setminus \bigcup_{i=1}^n B_1^\eta[k]) < \eta, \quad \forall k \in \mathbb{Z}.
\]

We say this is a proper decomposition if in addition

\[
\inf_{k \in \mathbb{Z}} d_X(x_i[k], x_j[k]) > 0, \quad \forall i, j \in \{1, \ldots, n\}, \ i \neq j.
\]

Note that (16) together with (18) imply that \( B_1^\eta[k] \cap B_j^\eta[k] = \emptyset \) for all \( k \in \mathbb{Z} \). Condition (17) means we can exclude a neighbourhood of the repelling dynamics that may sit on basin boundaries, while (18) implies there is a uniform threshold that can be used to separate all attractors.

Definition 3.4. We say the system \( (4) \) driven by input \( u \in \mathcal{U} \) has echo index \( n \), and write

\[ I(u) = n, \]

if it admits a proper decomposition into \( n \) uniform local point attractors. In this case, we say \( (4) \) has the \( n \)-Echo State Property \((n\text{-ESP})\) for input \( u \).

Note that if \( u \) has echo index \( n \) then so does \( \sigma^k(u) \) for any \( k \in \mathbb{Z} \).

Definition 3.4 describes the case where the system has a finite number of possible trajectories that can be “stably” followed forward in time for the same input: which of these is actually followed will depend on initial conditions. Whenever the echo index is 1, there is a single solution that is a global point attractor,
implying the ESP. Our definition of echo index assumes that all local attractors are point attractor. This is a restriction even when the input is null. In fact, there can be invariant curves or chaotic sets that attract some portion of the phase space. Note that the ESP and having echo index 1 are not equivalent [35].

For all those cases where a proper decomposition in \( n \) local point attractors does not exist, we say the system \((4)\) driven by \( v \) has *indefinite* echo index and we denote it as \( I(v) = \infty \). This covers a number of cases that we discuss in detail in [35].

Condition (18) guarantees that the local point attractors do not merge into each other, implying that the number of stable responses to an input in the infinite past coincides with the number of stable responses in the infinite future. Hence, the echo index characterises those input-driven systems whose degree of multistability does not change over time.

It is possible to extend the definition of echo index to sets of inputs as follows.

**Definition 3.5.** Consider a system \((4)\) with a set of possible inputs \( V \subset U \) and define

\[
V_{[n]} = \{ v \in V : I(v) = n \}.
\]

We say that such a system has echo index set \( \mathcal{N}(V) \) for a subset \( V \subset U \) if

\[
\mathcal{N}(V) = \{ n \in \mathbb{N} : V_{[n]} \neq \emptyset \}. \tag{19}
\]

We can split a set of inputs \( V \) into a disjoint union according to the echo index:

\[
V = \left( \bigcup_{n \in \mathcal{N}(V)} V_{[n]} \right) \cup V_{[\infty]}, \tag{20}
\]

where \( V_{[\infty]} \) are those inputs that give an indefinite echo index.
3.3. Existence of local point attractors

In this section, we give some necessary conditions for the existence and uniqueness for local point attractors. To this end, it is important to delimit the region of uniform contraction of the map $G: U \times X \to X$.

**Definition 3.6.** Given a positive real $\mu$ such that $0 < \mu < 1$, we define the set of linear $\mu$-contraction uniform in $U$ of the map $G$ as

$$C(\mu) := \{ x \in X : \sup_{u \in U} ||D_x G(u, x)|| \leq \mu \}, \quad (21)$$

where $\|\cdot\|$ denotes the matrix norm induced by the Euclidean norm on $\mathbb{R}^{N_r}$.

**Remark 3.1.** The set in $(21)$ is the phase space region where the nonautonomous dynamics contract at each time step with a rate of at most $\mu$, i.e. where each autonomous map $G(u, \cdot) : X \to X$, with $u \in U$, contracts with a rate of at most $\mu$.

In the following, we will need the following definition, which describes a particular useful circumstance where a single set $B \subseteq X$ is positively invariant for all input sequences in $V \subseteq U$.

**Definition 3.7.** A nonempty compact subset $B \subseteq X$ is called positively invariant for $V \subseteq U$ (or $V$-positively invariant) if

$$\Phi(k, u, B) \subseteq B, \quad \forall u \in V, \ k \in \mathbb{Z}^+$$

**Remark 3.2.** If the family of input sequences under consideration is $V := \{ \sigma^n(v) \}_{n \in \mathbb{Z}}$, for some $v \in U$, then we simply say that $B$ is a $v$-positively invariant set. This coincides with the case in Definition 2.4 of a constant nonautonomous set $B_n \equiv B$, which is positively invariant for the input $v$.

The following theorem provides some ingredients that are sufficient to prove the existence and uniqueness of a uniform local point attractor, allowing to detect it inside a compact and convex set. The proof of the theorem exploits the contraction of the family of autonomous maps $\{ G(u, \cdot) \}_{u \in U}$ defining the nonautonomous system.

**Theorem 3.2.** Let $\mu$ be a positive real such that $\mu < 1$. Suppose $Q_\mu$ is a $U$-positively invariant nonempty compact set such that it is contained inside $C(\mu)$ of Definition 3.6, and suppose further that $Q_\mu$ is convex. Then, for all $v \in U$ the system (4) driven by the input sequence $v$ admits a unique entire solution inside $Q_\mu$. In particular, if such entire solution is contained inside the interior of $Q_\mu$ then this entire solution is a local point attractor.
Proof. The assumption of convexity of $Q_\mu$ implies that for all $x_0, y_0 \in Q_\mu$ the line segment $l_{[x_0, y_0]} := \{ z = (1-s)x_0 + sy_0 \mid s \in [0,1] \}$ lies inside $Q_\mu$ and the following inequality holds
\[
d_X(G(u, x_0), G(u, y_0)) = \| G(u, x_0) - G(u, y_0) \| = \left\| \int_{s=0}^1 D_x G(u, (1-s)x_0 + sy_0) \cdot y \, ds \right\| 
\leq \sup_{z \in l_{[x_0, y_0]}} \| D_x G(u, z) \| \| x_0 - y_0 \| = \sup_{z \in l_{[x_0, y_0]}} \| D_x G(u, z) \| d_X(x_0, y_0) \leq \mu \, d_X(x_0, y_0).
\] (22) eq:contractivity

Now, let us take an input sequence $v \in U^\mathbb{Z}$ and an arbitrarily chosen initial time step $k_0 \in \mathbb{Z}$. Denote $x_k := \Phi(k, \sigma^{k_0}(v), x_0)$ and $y_k := \Phi(k, \sigma^{k_0}(v), y_0)$, for some $k > 0$. Hence, since $p(\sigma^s(v)) = v[s] \in U$ for any $s \in \mathbb{Z}$, then (22) reads $d_X(x_1, y_1) \leq \mu \, d_X(x_0, y_0)$. By hypothesis $Q_\mu$ is positively invariant for all input sequences $u \in U^\mathbb{Z}$, thus $x_1, y_1 \in Q_\mu$ if $x_0, y_0 \in Q_\mu$. Therefore, we can repeat the same argument on the pair $(x_1, y_1)$ with $v[k_0+1]$ in place of $(x_0, y_0)$ with $v[k_0]$, obtaining that $d_X(x_2, y_2) \leq \mu \, d_X(x_1, y_1) \leq \mu^2 \, d_X(x_0, y_0)$, and so on for all pairs $(x_k, y_k)$ forming the forward trajectories. In other words, by induction we have
\[
d_X(\Phi(k, \sigma^{k_0}(u), x_0), \Phi(k, \sigma^{k_0}(u), y_0)) \leq \mu^k \, d_X(x_0, y_0) \quad \forall u \in U, \forall k_0 \in \mathbb{Z}, \forall x_0, y_0 \in Q_\mu, \forall k > 0.
\] eq:cocycle_contraction

Moreover, for any given input signal $v \in U$, thanks to the fact that $Q_\mu$ is $v$-positively invariant (see Definition 2.8), the following nonautonomous invariant set
\[
A_n := \bigcap_{m \leq n} \Phi(n-m, \sigma^m(v), Q_\mu), \quad n \in \mathbb{Z}
\] eq:pullback_counter

is a local pullback attractor of the system driven by $v$ which is confined inside $Q_\mu$. We will prove that such a local pullback attractor is indeed a uniform local point attractor. Let us first prove that it is an entire solution. By contradiction, let us assume there are $x_s, y_s \in A_s$ distinct points, i.e. $x_s \neq y_s$, for some $s \in \mathbb{Z}$. By the invariance of $\{A_n\}_{n \in \mathbb{Z}}$, it follows that $\Phi(k, \sigma^{-k+s}(v), A_{-k+s}) = A_s$. Therefore, there exist $x', y' \in A_{-k+s}$ such that $x_s = \Phi(k, \sigma^{-k+s}(v), x')$ and $y_s = \Phi(k, \sigma^{-k+s}(v), y')$. Hence, the following inequalities hold:
\[
0 < d_X(x_s, y_s) = d_X(\Phi(k, \sigma^{-k+s}(v), x'), \Phi(k, \sigma^{-k+s}(v), y')) \leq \mu^k \, d_X(x', y') \leq \mu^k \, \text{diam}(Q_\mu).
\]

Therefore, by the compactness of $Q_\mu$, there cannot exist two distinct points in $A_s$, i.e. the local pullback attractor (24) is an entire solution $x = \{ x[n] \}_{n \in \mathbb{Z}}$ for input $v$. This fact implies that inside $Q_\mu$ there can be only one entire solution. Indeed, if there exists another entire solution $y = \{ y[n] \}_{n \in \mathbb{Z}}$ for input $v$, then it means that $\Phi(k, \sigma^{-k+s}(v), y[-k+s]) = y[s]$ holds for all $s \in \mathbb{Z}$ and for all $k > 0$. Therefore, if such entire solution is contained in $Q_\mu$, i.e. $y[n] \in Q_\mu$ for all $n \in \mathbb{Z}$, then $y[s] \in \bigcap_{k \geq 0} \Phi(k, \sigma^{-k+s}(v), Q_\mu) = \bigcap_{m \leq s} \Phi(s - m, \sigma^m(v), Q_\mu) = A_s$, which coincides with $x[s]$.

Now, we prove that such unique entire solution is uniformly attracting according to (16) of Definition 2.8.
Let us take a sequence of integers \( \{ j_k \} \). Note that

\[
h(\Phi(k, \sigma^{j_k}(u), Q_\mu), x[j_k + k]) = \max_{x \in Q_\mu} d_X(\Phi(k, \sigma^{j_k}(u), x), x[j_k + k])
\]

\[
= \max_{x \in Q_\mu} d_X(\Phi(k, \sigma^{j_k}(v), x), \Phi(k, \sigma^{j_k}(v), x[j_k])).
\]

Now, relation (23) implies that \( \max_{x \in Q_\mu} d_X(\Phi(k, \sigma^{j_k}(u), x), x[j_k + k]) \leq \mu^k \text{diam}(Q_\mu) \) and hence

\[
\lim_{k \to \infty} h(\Phi(k, \sigma^{j_k}(u), Q_\mu), x[j_k + k]) = 0,
\]

regardless of the sequence of integers \( \{ j_k \} \). Therefore,

\[
\lim_{k \to \infty} \sup_{n \in \mathbb{Z}} h(\Phi(k, \sigma^{n}(u), Q_\mu), x[n + k]) = 0.
\]

In particular, all forward trajectories starting at the same time step from initial conditions inside \( Q_\mu \) synchronise to a (unique) common solution which is indeed \( x \).

Moreover, as long as \( x[n] \in \text{int}(Q_\mu), \forall n \in \mathbb{Z} \), then we can find a neighbourhood of \( x = \{ x[n] \} \) such that Definition 3.2 is satisfied.

### 3.3.1. n-ESP for systems perturbed by low-amplitude inputs

In this section, we analyse the input-driven dynamics of an autonomous system perturbed by low (in amplitude) input sequences. A special case of (4) is the constant-input case:

\[
x[k + 1] = G(u_0, x[k]),
\]

i.e. with \( u \equiv u_0 \) constant, which gives an autonomous nonlinear system. We will look at how the autonomous system (27), possessing a uniformly stable point \( x^* \), reacts under small external perturbations represented by a forcing input sequence. Therefore, input values are taken in a neighbourhood of \( u_0 \), i.e. \( \mathcal{U} := B_r(u_0)\mathbb{Z} \) with \( r > 0 \) but small. The idea is that stable fixed points will become local point attractors in the nonautonomous setting.

In the following, we will assume that the autonomous map (27) possesses a uniformly attracting stable fixed point according to the following definition.

**Definition 3.8.** Let us be given an autonomous map \( F : X \to X \) and a fixed point \( x^* \in X \) for \( F \), i.e. \( F(x^*) = x^* \). We call \( x^* \) a Uniformly Attracting Stable Point (UASP) if there is an \( 0 < M < 1 \) and \( \delta > 0 \), such that

\[
d_X(F(z), x^*) < Md_X(z, x^*) \quad \forall z \in B_\delta(x^*).
\]

The property (28) imposes some algebraic condition on the linearised map on the fixed point \( x^* \).

**Lemma 3.3.** Let us be given an autonomous map \( F \) with a UASP \( x^* \) characterised by a contraction rate \( 0 < M < 1 \). Let \( F \) be differentiable at \( x^* \). Denote by \( A := D_x F(x^*) \) the Jacobian matrix evaluated onto the fixed point \( x^* \), and its maximum singular value as \( \sigma(A) \). Then \( \sigma(A) \leq M \).

A proof of Lemma 3.3 can be found at the end of Appendix A.
Remark 3.3. It is a known fact that $||A|| = \sigma(A)$, where $|| \cdot ||$ denotes the matrix norm induced by the Euclidean norm on $\mathbb{R}^{N_r}$. Therefore, by the continuity assumption on the function $||D_x F(\cdot)||$ at a UASP $x^*$ characterised by a contraction rate $M$, Proposition 3.3 implies that for all $\varepsilon > 0$ we can always find a small enough $\delta' > 0$ such that $||D_x F(z)|| < M + \varepsilon'$ for all $z \in B_{\delta'}(x^*)$.

Theorem 3.4. If there is a UASP $x^*$ of (27), characterised by a rate of contraction $M$, then for all $\varepsilon > 0$ there exists an $\varepsilon > 0$ such that, for any input sequence $\mathbf{v} \in B_{\varepsilon}(u_0)^2$, the system (4) driven by $\mathbf{v}$ has a unique local point attractor bounded inside $B_{\frac{\varepsilon}{\delta}}(x^*)$.

Proof. Let’s assume $G$ regular of class $C^1$ in a neighbourhood of $(u_0, x^*)$. Then, given two points $(u', y), (u', z)$ in such neighbourhood, the fundamental theorem of calculus allows to write the integral of the gradient

$$D_x G(u', y + s z) = \int_0^1 D_x G(u', y + s z) \cdot z \; ds \leq \max_{x \in \{y + s z; s \in [0, 1]\}} ||D_x G(u', x)|| ||z||.$$ (29)

Now, since $||D_x G||$ is continuous and $||D_x G(u_0, x^*)|| \leq M$ (due to Proposition 3.3), then for all $\varepsilon > 0$ we can find a $r_{\varepsilon} > 0$ and a $\delta_{\varepsilon} > 0$ such that $\max_{x \in B_{r_{\varepsilon}}(x^*)} ||D_x G(u', x)|| \leq M + \varepsilon < 1$, $\forall u' \in B_{r_{\varepsilon}}(u_0)$. Therefore, for all $\varepsilon > 0$ and denoting $\mu := M + \varepsilon < 1$, there exists an $r_{\varepsilon} > 0$ and a $\delta_{\varepsilon} > 0$ such that (29) reads as

$$d_X(G(u', y + z), G(u', y)) \leq \mu ||z|| = \mu d_X(y + z, y).$$ (30)

which holds $\forall y \in B_{\frac{\varepsilon}{\delta}}(x^*), \forall z \in B_{\frac{\varepsilon}{2\delta}}(0), \forall u' \in B_{r_{\varepsilon}}(u_0)$. Therefore, defining $x = y + z$, for all $\varepsilon > 0$, with $\mu = M + \varepsilon < 1$, we can find a $r_{\varepsilon} > 0$ and a $\delta_{\varepsilon} > 0$ such that

$$d_X(G(u', x), G(u', y)) \leq \mu d_X(x, y) \quad \forall x, y \in B_{\frac{\varepsilon}{2\delta}}(x^*), \forall u' \in B_{r_{\varepsilon}}(u_0).$$

Now, we prove that the compact convex set $B_{\frac{\varepsilon}{2\delta}}(x^*)$ is positively invariant for the family of input sequences $\mathcal{U} := B_{r_{\varepsilon}}(u_0)^2$. Assuming $G(\cdot, x) : U \rightarrow X$ is continuous in $u = u_0 \in U$, for each $x \in B_{\delta}(x^*)$, we get that $\forall \varepsilon > 0$, $\exists r > 0 : G(B_{r}(u_0), B_{\delta}(x^*)) \subseteq B_{\varepsilon}(G(u_0, B_{\delta}(x^*)))$.

Hypothesis reads $G(u_0, B_{\delta}(x^*)) \subseteq B_{M\delta}(x^*)$. Therefore, we obtain

$$G(B_{r}(u_0), B_{\delta}(x^*)) \subseteq B_{M\delta + \varepsilon}(x^*).$$

As long as $\varepsilon$ is such that $M\delta + \varepsilon < \delta$, i.e. $\varepsilon < \delta(1 - M)$, we can reuse the same idea with $(\varepsilon, r)$ pair obtaining that, for all $\mathbf{v} \in B_{r}(u_0)^2$,

$$\Phi(n, \mathbf{v}, B_{\delta}(x^*)) \subseteq B_{M\delta + \varepsilon}(x^*) \quad \forall n \in \mathbb{Z}^+.$$ (31)

Furthermore, repeating the above argument in (31) on $M\delta + \varepsilon$ and iteratively so on, we obtain a sequence $\delta_n = M\delta_{n-1} + \varepsilon$, for all $n \geq 1$, starting with $\delta_0 = \delta$, which is strictly decreasing. Unfolding such a recursive
Lemma 3.5. Hence, this proves that for all $\varepsilon > 0$ it is possible to find an $r > 0$ such that the compact convex set $\bar{B}_r(x^\ast)$, with $\frac{\varepsilon}{1 - M} < \rho \leq \min\{\delta, \frac{1}{4}\delta_\varepsilon\}$ is positively invariant for all the input sequences $v \in B_r(u_0)^\mathbb{Z}$. Therefore, Theorem 3.2 implies the existence and uniqueness of a uniform local point attractor inside the ball $B_{\frac{\varepsilon}{1 - M}}(x^\ast)$.

The assumption on $x^\ast$ to be a UASP makes easier to prove the existence of a local point attractor. Nevertheless, the same results can presumably be proved just assuming $x^\ast$ to be a stable hyperbolic fixed point, using exponential dichotomies; see for example [35].

As a corollary of Theorem 3.4, we get that if for a given constant input value $u_0$ the autonomous map $x[k] = G(u_0, x[k-1])$ presents a number $n$ of UASPs then the nonautonomous system $x[k] = G(u[k], x[k-1])$ driven by any deterministic input sequence $u[k]$ assuming values in a neighbourhood of $u_0$ will typically have the $n$-ESP.

3.3.2. ESP for RNNs driven by large-amplitude inputs

It has previously been claimed [7, 8] that the ESP might be trivially gained if the input driving the dynamics is large in amplitude, even considering large recurrent feedback. Here, we rigorously prove this in Theorem 3.6 with minimal assumptions. First we need a technical Lemma. In order to state and prove such Lemma we introduce the following definitions. We will denote with $(W)_{(j)}$ the $j$th row of a matrix $W$.

**Definition 3.9.** We define

$$H_j := \{ u \in \mathbb{R}^{N_i} : (W_{in})_{(j)} \cdot u = 0 \} \quad (32)$$

the hyperplane of input values which vanish the $j$th row of the matrix $W_{in}$. Moreover, given real values $\epsilon > 0$ (meant to be small) and $R > 0$ (meant to be large), we define

$$P_j(\epsilon, R) := \left\{ u \in \mathbb{R}^{N_i} \setminus B_R(0) : \left| \frac{(W_{in})_{(j)} \cdot u}{\|(W_{in})_{(j)}\|} \right| \geq \epsilon \right\} \quad (33)$$

**Remark 3.4.** Note that, denoted with $\theta$ the angle between the vectors pointing to $(W_{in})_{(j)}$ and $u$ then $\cos(\theta) = \frac{(W_{in})_{(j)} \cdot u}{\|(W_{in})_{(j)}\|}$. Therefore, the space $P_j(\epsilon, R)$ is basically the set $\mathbb{R}^{N_i} \setminus B_R(0)$ where we cut out all the lines which form an angle $\theta \in (\arccos(\epsilon), \arccos(-\epsilon))$, i.e. those close to the hyperplane $H_j$.

**Lemma 3.5.** Let $W_r, W_{fb}, W_{in}$ be real matrices of dimensions, respectively, $N_r \times N_r$, $N_r \times N_o$, $N_r \times N_i$, and $\psi \in C^0(\mathbb{R}^{N_r}, \mathbb{R}^{N_o})$. Consider the functions $\xi_j : \mathbb{R}^{N_i} \times [-L, L]^{N_r} \rightarrow \mathbb{R}$, for $j = 1, \ldots, N_r$, defined as $\xi_j(u, x) := (W_{in})_{(j)} \cdot u + f_j(x)$, where $f_j(x) := (W_r)_{(j)} \cdot x + (W_{fb})_{(j)} \cdot \psi(x)$. If $(W_{in})_{(j)}$ is not the null vector the following statements are true:

1. There exists an $R_j > 0$ (which depends on $\max_{x \in [-L, L]^{N_r}} |f_j(x)|$ and $\|(W_{in})_{(j)}\|$) such that for all $R \geq R_j$ we have that inf $\{ |\xi_j(u, x)| : u \in \partial B_R(0) \} = 0$ holds for all $x \in [-L, L]^{N_r}$. 

17
2. In the subspace $P_j(\epsilon, R)$ of input values we can make the function $|\xi_j(u, x)|$ large as much as we want.

Precisely, for all $\epsilon > 0$ and for all $\bar{\xi} > 0$ there exists an $R_{\xi, \epsilon} > 0$ such that for all $R \geq R_{\xi, \epsilon}$ we have that $\inf \{ |\xi_j(u, x)| \mid u \in P_j(\epsilon, R) \} \geq \bar{\xi}$ holds for all $x \in [-L, L]^{N_r}$.

**Proof.** If $(W_{in})_{(j)}$ is the null vector then $\xi_j(u, x) = f_j(x)$, i.e. it does not depend on the input, thus let us assume $(W_{in})_{(j)}$ is not the null vector. Then, a hyperplane $H_j = \{ u \in \mathbb{R}^{N_r} : (W_{in})_{(j)} \cdot u = 0 \}$ is defined in the space of input values $\mathbb{R}^{N_r}$ such that $H_j$ is the orthogonal space of the vector pointing to $(W_{in})_{(j)}$.

1. In general,

$$(W_{in})_{(j)} \cdot u = \| (W_{in})_{(j)} \| R \cos(\theta)$$

where $R$ is the norm of $u$ and $\theta$ is the angle between the vectors pointing to $(W_{in})_{(j)}$ and $u$. Let us fix arbitrarily a $R > 0$ and consider $u$ to vary on the surface of the ball of radius $R$ centred on the origin of $\mathbb{R}^{N_r}$, i.e. $u \in \partial B_R(0)$, thus we have that $\| (W_{in})_{(j)} \| R$ is constant and $\cos(\theta)$ can assume any value in $[-1, 1]$. In particular, $(W_{in})_{(j)} \cdot u$ can assume any value in the interval $[-\| (W_{in})_{(j)} \| R, \| (W_{in})_{(j)} \| R]$. Hence, for all $x \in [-L, L]^{N_r}$ there exists a $u \in \partial B_R(0)$ such that $(W_{in})_{(j)} \cdot u = -f_j(x)$, i.e. such that $\xi_j(u, x) = 0$; this is true as long as $(W_{in})_{(j)}$ is not the null vector and $R \geq \frac{\sigma}{\| (W_{in})_{(j)} \|}$, where we denoted $\sigma := \max_{x \in [-L, L]^{N_r}} |f_j(x)|$. Thus, defined $R_j := \frac{\sigma}{\| (W_{in})_{(j)} \|}$ we have that for all $R \geq R_j$

$$\inf_{u \in \partial B_R(0)} |\xi_j(u, x)| = 0$$

holds for all $x \in [-L, L]^{N_r}$.

2. Let us take an arbitrary $\epsilon > 0$ and consider the subspace $P_j(\epsilon, R)$ for some $R > 0$. Note that for any $u \in P_j(\epsilon, R)$, denoted with $\epsilon := |\cos(\theta)|$, where $\theta$ is the angle between $(W_{in})_{(j)}$ and $u$, and with $\rho$ the norm of $\|u\|$, then $|\epsilon| \geq \epsilon$ and $\rho \geq R$. Therefore, we have that

$$\| (W_{in})_{(j)} \| u \geq \| (W_{in})_{(j)} \| R \epsilon$$

holds for all $u \in P_j(\epsilon, R)$. Now, for all $\bar{\xi} > 0$ we can find an $R_{\xi, \epsilon} > 0$ large enough such that

$$\| (W_{in})_{(j)} \| R_{\xi, \epsilon} \epsilon - \max_{x \in [-L, L]^{N_r}} |f_j(x)| \geq \bar{\xi}_r.$$  \hspace{1cm} (35) \hspace{1cm} \text{eq:xi_over}

Therefore, the reverse triangle inequality leads to

$$|\xi_j(u, x)| \geq |f_j(x)| - |(W_{in})_{(j)} \cdot u| \geq |(W_{in})_{(j)} \cdot u| - |f_j(x)|,$$

where the second inequality holds for all $x \in [-L, L]^{N_r}$ and for all $u \in P_j(\epsilon, R)$ as long as $R \geq \frac{\sigma}{\epsilon \| (W_{in})_{(j)} \|}$, where we denoted $\sigma := \max_{x \in [-L, L]^{N_r}} |f_j(x)|$.

Hence, if $R \geq R_{\xi, \epsilon} := \bar{\xi}_r \frac{\sigma}{\epsilon \| (W_{in})_{(j)} \|}$ then thanks to (34), (35), and (36) we get that the following

$$|\xi_j(u, x)| \geq |(W_{in})_{(j)} \cdot u| - |f_j(x)| \geq \| (W_{in})_{(j)} \| R_{\xi, \epsilon} \epsilon - \max_{x \in [-L, L]^{N_r}} |f_j(x)| \geq \bar{\xi}_r$$
holds for all \( x \in [-L, L]^{N_r} \) and for all \( u \in P_j(\epsilon, R) \). In conclusion, we have that for all \( \epsilon > 0 \) and for all \( \xi > 0 \) there exists an \( R_{\xi, \epsilon} > 0 \) such that for all \( R \geq R_{\xi, \epsilon} \) we have that
\[
\inf_{u \in P_j(\epsilon, R)} |\xi_j(u, x)| \geq \xi
\]
holds for all \( x \in [-L, L]^{N_r} \).

\[\square\]

We now state the main result of this section.

**Theorem 3.6.** Consider the RNN \((2)-(3)\) with a bounded \( \phi \in C^1(\mathbb{R}, (-L, L)) \) that is monotonically increasing and \( \phi' \) has a unique maximum point at \( \xi = 0 \) and \( \psi \in C^1(\mathbb{R}^{N_r}, \mathbb{R}^{N_r}) \). Assume the matrix \( W_{in} \) is not the zero matrix. Then, for all \( \epsilon > 0 \) there exists an \( R_{\epsilon} > 0 \) such that for all compact sets \( U \subset \bigcap_{j=1}^{N_r} P_j(\epsilon, R_{\epsilon}) \) and for all input sequences \( u \in U^Z \) the RNN \((2)-(3)\) driven by \( u \) admits exactly one entire solution, i.e. it has the ESP as in Definition 3.1. Moreover, if the following condition holds,
\[
\phi'(0)\|W_r + W_{fb}D_x\psi(x)\| < 1 \tag{37}
\]
for all \( x \in [-L, L]^{N_r} \), then for all compact sets \( U \subset \mathbb{R}^{N_i} \) the RNN \((2)-(3)\) driven by any input sequence \( u \in U^Z \) admits exactly one entire solution, i.e. it has the ESP.

**Proof.** First, let us consider in \((2)-(3)\) neurons without leakage, i.e. \( \alpha = 1 \). The Jacobian matrix of \( G \) is
\[
D_xG(u, x) = S(u, x)M(x),
\]
where \( M(x) = W_r + W_{fb}D_x\psi(x) \) and \( S(u, x) = \text{diag}[\phi'(\xi_j(u, x))] \) with \( j = 1, \ldots, N_r \),
\[
\xi_j(u, x) := (W_r)_{(j)} \cdot x + (W_{fb})_{(j)} \cdot \psi(x) + (W_{in})_{(j)} \cdot u.
\]
Given a subset of input values \( V \subset \mathbb{R}^{N_i} \) let us denote
\[
l_V(x) := \sup_{u \in V} ||S(u, x)||.
\]
Then, for all \( x \in [-L, L]^{N_r} \) we have that
\[
\sup_{u \in V} ||D_xG(u, x)|| \leq l_V(x) \|M(x)\|.
\]
Now, since \( S(u, x) \) is a diagonal matrix, then
\[
l_V(x) = \sup_{u \in V} ||S(u, x)|| = \sup_{u \in V} \max_{j=1,\ldots,N_r} |\phi'(\xi_j(u, x))| = \sup_{u \in V} \phi'\left(\min_{j=1,\ldots,N_r} |\xi_j(u, x)|\right) = \phi'\left(\inf_{u \in V} \left\{\min_{j=1,\ldots,N_r} |\xi_j(u, x)|\right\}\right),
\]
\[\text{eq:ESP_max_sing_value}\]
where the last two equalities hold because $\phi'$ is a continuous function with the only maximum point in $\xi = 0$.

We want to prove that for all $\epsilon > 0$ there exists an $R_\epsilon > 0$, such that $\sup_{u \in V_\epsilon} ||D_xG(u, x)|| < 1$ strictly, for all $x \in [-L, L]^{N_r}$, where we denoted $V_\epsilon := \bigcap_{j=1}^{N_r} P_j(\epsilon, R_\epsilon)$. Thanks to Lemma 3.5 we know that for all $\tilde{\xi} > 0$ there exists an $R_{\tilde{\xi}, x} > 0$ such that

$$\inf_{u \in V_\epsilon} \left\{ \min_{j=1, \ldots, N_r} \left| \xi_j(u, x) \right| \right\} \geq \tilde{\xi}.$$ 

Now, $\phi \in C^1(\mathbb{R}, (-L, L))$ is monotonic and its image is $(-L, L)$, hence $\lim_{|\xi| \to \infty} \phi'(\xi) = 0$. Hence, $l_{V_\epsilon}(x)$ can be made arbitrarily low as long as $u \in V_\epsilon$ is such that $||u|| > R$ with $R$ large enough. More precisely, fixing an arbitrary $0 < \mu < 1$ and denoting $\bar{\sigma} := \max_{x \in [-L, L]^{N_r}} ||M(x)||$, then for all $\epsilon > 0$ there exists an $R_\epsilon > 0$, such that $l_{V_\epsilon}(x) \leq \frac{\mu}{\bar{\sigma}}$ holds for all $x \in [-L, L]^{N_r}$, and consequently such that

$$\sup_{u \in V_\epsilon} ||D_xG(u, x)|| \leq l_{V_\epsilon}(x) ||M(x)|| \leq \mu,$$

holds for all $x \in [-L, L]^{N_r}$. From this, it follows that the whole space $[-L, L]^{N_r}$ is contained inside $C(\mu)$ of \cite{21} w.r.t any $V_\epsilon = \bigcap_{j=1}^{N_r} P_j(\epsilon, R_\epsilon)$ with a $R_\epsilon > 0$ large enough. Therefore we can make the RNN dynamics arbitrarily contractive as long as we stay inside the space $V_\epsilon$. Thus, taking any compact set $U \subset V_\epsilon$, we conclude applying Theorem 3.2 with $\mathcal{U} := U^Z$ on the whole space $X = [-L, L]^{N_r}$ which is a $U$-positively invariant (thanks to Proposition \ref{prop:positively_invariant}), compact and convex set.

On the other hand, given a subset $V \subseteq \mathbb{R}^{N_r}$, if there exists an $R > 0$ large enough such that $\partial B_R(0) \subseteq V$ then Lemma 3.5 implies that $l_{V}(x) \equiv \phi'(0)$. Therefore, in the general case of $V = \mathbb{R}^{N_r}$ we need to impose the condition

$$\phi'(0)||W_r + W_{fb}D_x\psi(x)|| < 1, \quad \forall x \in [-L, L]^{N_r},$$

in order to apply Theorem 3.2 with any compact set $U \subset \mathbb{R}^{N_r}$.

In the case of a leaky RNN, the Jacobian matrix reads

$$D_xG(u, x) = (1 - \alpha)I_{N_r} + \alpha S(u, x)M(x),$$

where $I_{N_r}$ denotes the identity matrix of dimension $N_r$. Hence,

$$||D_xG(u, x)|| \leq (1 - \alpha)^{N_r} + \alpha^{N_r}||S(u, x)M(x)|| \leq 1 - \alpha(1 - ||S(u, x)M(x)||),$$

where the second inequality holds since the function $f(x) = (1-\alpha)^x + \alpha^x s$ is strictly monotonically decreasing for all $\alpha \in (0, 1)$ and for all $s \geq 0$. Now, note that $||D_xG(u, x)|| < 1 - \alpha$, if and only if $0 \leq ||S(u, x)M(x)|| < 1$ holds, which leads back to the already proved case of no leakage. Therefore, we can apply Theorem 3.2 on the whole space $[-L, L]^{N_r}$ which is contained inside $C(1 - \alpha)$. \hfill \Box

20
Remark 3.5. Remarkably, the ESP holds without any particular assumption on the matrices \(W_r, W_{fb}\), and the readout \(\psi\), as long as the input values are large enough in amplitude and stand a bit far from the space \(H = \bigcup_{j=1}^{N_r} H_j\). The need to exclude the sets \(H_j\) is due to the fact that for all input sequences \(u \in H_f^2\) in some directions the dynamics might be repulsive unless we impose some conditions on the matrices \(W_r, W_{fb}\) and the function \(\psi\), as the condition in (37). Nevertheless, note that the set \(H = \bigcup_{j=1}^{N_r} H_j\) is the union of a number \(m = \text{rank}(W_{in})\) of hyperplanes in \(\mathbb{R}^{N_i}\), which has a zero Lebesgue measure. Roughly speaking, this means that if the input sequence \(u\) is a realisation of a process which generates values inside \(U\) according to a uniform distribution, then for all \(R > 0\) such that \(U \subset \mathbb{R}^{N_i} \setminus B_R(0)\) the probability to observe values of the input outside of the compact space \(U \cap \bigcap_{j=1}^{N_r} P_j(\epsilon, R)\) will be proportional to \(\epsilon\).

Note that, in the presence of feedback of the output with linear readout, i.e. \(\psi(x) = W_o x\), whenever the activation function is such that \(\phi'(0) = 1\), as for example \(\phi = \tanh\), then from (37) we obtain the condition on the maximum singular value \(||M|| < 1\), where \(M = W_r + W_{fb} W_o\) is the “effective recurrent matrix”. In particular, when there is no feedback, i.e. \(W_{fb} = 0\), we recover the well-known condition \(||W_r|| < 1\).

4. Examples of input-driven RNNs with multistable dynamics

In this section, we consider the RNN (1) with no output feedback \((z[k] = 0)\) interpreted as a nonautonomous system in the sense (4). We use a linear layer for the RNN output (2). The results here are illustrative of some of the behaviours we expect in more general cases. In Section 4.1, we discuss a one dimensional example of RNN with maximum singular value greater than 1, where the echo index changes according to the amplitude of the input injected into the system. Notably, we show through a simulation how the bifurcation of the system from 2-ESP to 1-ESP can occur. Then, in Section 4.2, we show a two dimensional example of RNN with echo index 2, highlighting the need for concepts introduced around Def 3.3.

4.1. An example with input-dependent echo index

In this section, we consider a simple example of a one dimensional RNN and we show how the echo index can vary according to the specific input signal driving the dynamics. Let us consider the input-driven system

\[
x[k+1] = G(u[k+1], x[k]),
\]

where \(w\) is a positive real constant. Note that in this example the value 1.01 represents the one dimensional reservoir which hence coincides with its maximum singular value (and spectral radius as well). Therefore, the autonomous system defined by the map \(F(x) := G(0, x)\) is expanding around the (unstable) fixed point \(x = 0\) and there exist two (uniformly attracting) stable points \(x^*_1 \approx -0.17, x^*_2 \approx 0.17\).
Theorems 3.4 and 3.6 suggest that, in order to gain or lose the ESP, scaling the matrix $W_{in}$ might be as relevant as scaling $W_r$; indeed, we show that for different values of the parameter $w$, we obtain different values of echo index. Note that the parameter $w$ scales the inputs provided to the system; accordingly, we force the autonomous map $F$ with inputs assuming values in $[-w,w]$. We can analytically compute the value of $w$ such that a fold bifurcation occurs in the system (41), see Eq. 36 of [10] for the details, and it is approximately $w_c \approx 0.0007$. This implies that perturbing the autonomous dynamics by means of any input sequence with amplitude less than $w_c$ will induce an input-driven system with echo index 2.

We generated an input sequence according to a uniform distribution in $[-w,w]$, then we considered three values of the parameter $w = 0.0006, 0.01, 0.05$. Figure 2 shows the evolution of ten initial conditions of the system for the three values of $w$. In the first case of $w = 0.0006$ (top Figure 2), the input-driven system exhibited echo index 2, in accordance with the fact that the critical value $w_c$ is greater than $w = 0.0006$. The second case of $w = 0.01$ (centre Figure 2) produced an interesting dynamics with echo index 1: the unique local point attractor is characterised by a switching behaviour between the “ghosts” of the two randomly perturbed autonomous stable solutions of $F$. Finally, in the third case of $w = 0.05$ (bottom Figure 2), the nonautonomous dynamics converge toward a unique local point attractor (i.e. echo index 1), which wanders randomly around the origin.
Figure 2: Nonautonomous dynamics of the system (41) driven by an input sequence generated according to a uniform distribution in \([-1,1]\) and then scaled by means of a positive real parameter \(w\), for three values of the parameter \(w = 0.0006, 0.01, 0.05\). Ten initial conditions have been run in all three cases. **Top:** case of \(w = 0.0006\). The input driven system exhibits echo index 2, i.e. there are 2 local point attractors. **Centre:** case of \(w = 0.01\). The input driven system has echo index 1. There exists a unique local point attractor whose behaviour is affected by the vicinity of the fold bifurcation of the underlying autonomous map \(F(x) = G(0, x)\). The resulting dynamics manifest a switching motion. **Bottom:** case of \(w = 0.05\). The input-driven system presents echo index 1 and the corresponding local point attractor does not exhibit any switching behaviour.
4.2. An example of switching system with echo index 2

We now report some numerical experiments on a simple RNN that can be thought of as switching dynamics between two autonomous maps. We will consider the space of inputs defined as $\mathcal{U} := \{u_1, u_2\}^2$, i.e. as made by sequences assuming just two possible values: $u_1$ and $u_2$. For the purpose we used RNN state-update equations with leaky-integrator neurons $[16]$. Let us consider the map $G_{\alpha}(u,x) := (1 - \alpha)x + \alpha \tanh(W_r x + W_{in} u)$ where $W_r = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{3}{2} \end{bmatrix}$, $W_{in} = I_2$ is the identity matrix and $x$ a real vector of dimension two. What follows can be observed for any value of $\alpha \in (0,1]$. We select $\alpha = \frac{1}{4}$ because a small leak rate highlights transient dynamics.

The space of possible input sequences is $\mathcal{U} = \{u_1, u_2\}^2$, where $u_1 := \begin{bmatrix} 1 \\ \frac{1}{2} \end{bmatrix}$ and $u_2 := -u_1$. Therefore, the nonautonomous dynamics driven by input $u \in \mathcal{U}$ consists of a switching pattern between the two component maps defined as $f_1(x) := G(u_1, x)$ and $f_2(x) := G(u_2, x)$.

The autonomous system $x[k+1] = f_1(x[k])$ has two asymptotically stable fixed points with a saddle between them along the vertical line of $x_1 \approx 0.45$, see Figure 3. Analogously, the autonomous system $x[k+1] = f_2(x[k])$ has two asymptotically stable points with a saddle between them along the vertical line $x_1 \approx -0.45$.

We generated input sequences according to a Markov chain defined as follow: $p_1$ is the probability to switch from $f_1$ to $f_2$; with probability $1 - p_1$ we stay in $f_1$; $p_2$ is the probability to switch from $f_2$ to $f_1$; with probability $1 - p_2$ we stay in $f_2$. Exploiting Theorem 3.2 we are able to prove the existence of two local point attractors for every randomly generated input sequences (with probability 1). The Jacobian matrix of $f_1$ reads

$$D_x f_1(x_1, x_2) = \begin{bmatrix} 1 - \alpha/2 [1 + \tanh^2(x_1/2 + u_1)] & 0 \\ 0 & 1 + \alpha/2 [1 - 3 \tanh^2(3x_2/2 + u_2)] \end{bmatrix}.$$ 

Diagonal elements are thus eigenvalues and also singular values, hence $\sigma(D_x f_1(x_1, x_2)) = 1 + \alpha/2 [1 - 3 \tanh^2(3x_2/2 + u_2)]$. Thus $\|D_x f_1(x_1, x_2)\|_2 = 1 + \alpha/2 [1 - 3 \tanh^2(3x_2/2 + u_2)]$. The region of phase space where contraction occurs, i.e. $\|D_x f_1(x_1, x_2)\|_2 < 1$, is made by 2 connected components divided by the strip $-0.54 < x_2 < 0.34$, approximately. Similarly, for $f_2$ we have that $\|D_x f_2(x_1, x_2)\|_2 < 1$ holds true everywhere except for $-0.34 < x_2 < 0.54$, approximately. Now, it is easy to see that the region $R_+ = \{(x_1, x_2) \in [-1,1]^2 : x_2 > 0.54 \}$ is positively invariant for both maps $f_1$ and $f_2$. Analogously, the region $R_- = \{(x_1, x_2) \in [-1,1]^2 : x_2 < -0.54 \}$ is positively invariant for both maps $f_1$ and $f_2$. Hence, both regions $R_+$ and $R_-$ result to be convex $\mathcal{U}$-positively invariant compact sets of the nonautonomous system where contraction occurs. Therefore, Theorem 3.2 ensures that inside the set $R_+$, there exists a local point attractor and, similarly, there exists another local point attractor inside $R_-$. In the top panel of Figure 3 are shown the local point attractors in phase space of this nonautonomous system. In the bottom plot are
shown time series of the observable $\frac{x_1 + x_2}{2}$ of many initial conditions. As can be seen a local point attractor is substantially different from a fixed point due to its input-driven nature.

Figure 3: All pictures refer to the same input sequence $v \in \{u_1, u_2\}^2$ which has been randomly generated by means of a Markovian process with equal probability $p_1 = 0.5 = p_2$ to switch between $u_1$ and $u_2$. **Top:** Fixed points of both (autonomous) component maps $f_1(x), f_2(x)$ are showed as blue (stable nodes) and red (saddle) dots, respectively. Notably, fixed points vertically aligned on the positive side of the $x_1$ variable characterise the $f_1$ map, while the ones on the negative side characterise $f_2$. On the left panel, the initial condition of coordinates $(0.1, -0.1)$ evolves towards the upper local point attractor solution; on the right panel, the initial condition of coordinates $(0.1, -0.2)$ evolves towards the bottom local point attractor solution. Under the same driving input sequence, almost every initial condition converges to one of those two nonautonomous attractors. **Bottom:** Time series of the observable $x_1 + x_2$ of 20 trajectories starting from different initial conditions are shown. The synchronisation of trajectories around local point attractors is due to the contraction property, which is clearly observable in the first 50 time steps.
4.2.1. The separatrix solutions

The basin boundary of the two local point attractors is, at each fixed time step, a horizontal line which lies in-between the stable manifolds of the two saddles of the component maps. In the top left of Figure 4, this line is shown for time step $k = 0$. In the top right plot of Figure 4, we evolved many different initial conditions in a neighbourhood of such a horizontal line. We randomly chose one of these trajectories as reference trajectory and computed the distances, at each time step, between each trajectory with respect to a randomly chosen one. It is numerically observed that the closer we start to such horizontal line, the more time is needed to converge towards one of the local point attractors. Indeed, the uniform convergence (i.e. in the Hausdorff semi-metric sense) towards a local point attractor holds only excluding a neighbourhood of the boundary of its basin of forward attraction. Moreover, from the top right plot of Figure 4 we observe that each initial condition seems to be attracted for a few time steps to a special trajectory (all the distances go to zero), but are eventually pushed away from it in the long run towards one of the two local point attractors. This suggests the existence of a third attractive entire solution whose basin of forward attraction is exactly the (moving in time) separatrix line. As a matter of fact, each point of such separatrix line represents a fibre of an entire solution. Interestingly, all those entire solutions seem to converge towards a unique solution which appear as in bottom pictures of Figure 4. Note that, in particular, such nonautonomous system admits an infinite number of entire solutions but only two of them essentially characterise its dynamics. Although this third attractive entire solution plays a key role in the nonautonomous dynamics of the system, it should not be considered equivalent to the other two local point attractors. In fact, the probability to pick up an initial condition in phase space converging to this particular entire solution is null. This is the reason why we require for a local point attractor to have a neighbourhood of attraction, as formally stated in Definition 3.2, which in particular implies that its basin of attraction has positive Lesbegue measure. In the bottom panels of Figure 4, we tried to detect the third attractive entire solution letting evolve initial conditions extremely close to the separatrix line. Such a special entire solution appears unstable and it traces an orbit that wanders erratically in the region between the stable manifolds of the two saddles of $f_1$ and $f_2$.

\footnote{Note also that the separatrix solution and all the infinitely many entire solutions which it attracts are part of the global pullback attractor of Definition 2.6.}
Figure 4: All pictures refer to the same randomly generated input signal of Figure 3. **Top left:** computation of the basins of attraction of the two local point attractor solutions at time step $k = 0$. **Top right:** 950 initial conditions have been chosen around the position of the separatrix at time step $k = 0$; the $x_2$ coordinate starting from $-0.113248$ and cumulatively increasing of $10^{-6}$ until the value $-0.113198$, the $x_1$ coordinate starting from $-1$ and cumulatively increasing of $10^{-1}$ until the value $1$. The trajectory starting from the initial condition $(0, -0.113218)$ has been chosen as reference. At each time step, it has been computed and plotted the distance between the reference trajectory and all other trajectories. **Bottom left:** initial condition is $(0.49, -0.11322366651)$. The trajectory rambles in phase space following the separatrix attractive entire solution, then after about 230 time steps it starts to converge towards the upper local point attractor. **Bottom right:** here the initial condition is $(0.49, -0.11322366652)$, hence it differs of $10^{11}$ from the one in the left picture. The trajectory practically coincides with the one in the left picture but eventually converges towards the bottom local point attractor.
5. Conclusions

As main contribution of this paper, we highlight how the echo state property, which guarantees the existence of a unique (stable) response to an input, may hold for some inputs but not others. When it does not hold, we show that a recurrent neural network might still reliably produce several stable responses to an input: the echo index introduced here counts such stable responses. Our theoretical developments are framed within the theory of nonautonomous dynamical systems. Here, we introduce a suitable definition of attractor for input-driven recurrent neural networks that we called local point attractor, which models and characterises stable responses of recurrent neural networks to inputs. The presence of more than one stable response indicates the possibility to observe multiple, yet consistent behaviours of a recurrent neural network driven by an input sequence (for example, when classifying an input sequence, the trajectory followed by the network may be different depending on initial conditions, yet the final decision may remain the same). On the other hand, echo index greater than one might also indicate incorrect training of a network and hence signal possible malfunctionings on a task.

It is worth noting that the echo state property appears to be related to the concept of synchronisation, see e.g. [38, Chapter 15], which has been shown to be a key concept in complex networks [39]. In particular, ESP may be interpreted as an induced synchronisation in a network via nonautonomous forcing [40]; see also [41] where synchronisation is defined by means of mutual convergence of trajectories in a probability theory framework (random dynamical systems).

We believe that the notions and results introduced here will prove fundamental for the more general and ambitious goal of providing mechanistic models describing the behaviour of recurrent neural networks in machine learning tasks.

Acknowledgements

LL gratefully acknowledges partial support of the Canada Research Chairs program. CP and PA gratefully acknowledge partial support from the NZ Marsden fund, grant UOA1722. PA acknowledges funding from EPSRC as part of the Centre for Predictive Modelling in Healthcare grant EP/N014391/1.

References

[1] G. Manjunath, P. Tino, H. Jaeger, Theory of input driven dynamical systems, dice. ucl. ac. be, number April (2012) 25–27 (2012).

[2] P. E. Kloeden, M. Rasmussen, Nonautonomous Dynamical Systems, no. 176, American Mathematical Soc., 2011 (2011).
[3] P. E. Kloeden, M. Yang, Forward attraction in nonautonomous difference equations, Journal of Difference Equations and Applications 22 (8) (2016) 1027–1039 (2016). doi:10.1080/10236198.2015.1107550

[4] H. Jaeger, The “echo state” approach to analysing and training recurrent neural networks-with an erratum note, Bonn, Germany: German National Research Center for Information Technology GMD Technical Report 148 (34) (2001) 13 (2001).

[5] M. Lukoševičius, H. Jaeger, Reservoir computing approaches to recurrent neural network training, Computer Science Review 3 (3) (2009) 127–149 (2009). doi:10.1016/j.cosrev.2009.03.005

[6] H. Jaeger, H. Haas, Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication, Science 304 (5667) (2004) 78–80 (2004). doi:10.1126/science.1091277

[7] I. B. Yildiz, H. Jaeger, S. J. Kiebel, Re-visiting the echo state property, Neural Networks 35 (2012) 1–9 (2012). doi:10.1016/j.neunet.2012.07.005

[8] G. Manjunath, H. Jaeger, Echo state property linked to an input: Exploring a fundamental characteristic of recurrent neural networks, Neural Computation 25 (3) (2013) 671–696 (2013). doi:10.1162/NECO_a_00411

[9] L. Grigoryeva, J.-P. Ortega, Echo state networks are universal, Neural Networks 108 (2018) 495–508 (2018). doi:10.1016/j.neunet.2018.08.025

[10] A. Ceni, P. Ashwin, L. Livi, Interpreting recurrent neural networks behaviour via excitable network attractors, Cognitive Computation (2019) 1–27 (Mar. 2019). doi:10.1007/s12559-019-09634-2

[11] D. Haviv, A. Rivkind, O. Barak, Understanding and controlling memory in recurrent neural networks, in: K. Chaudhuri, R. Salakhutdinov (Eds.), Proceedings of the 36th International Conference on Machine Learning, Vol. 97, PMLR, Long Beach, California, USA, 2019, pp. 2663–2671 (Jun 2019).

[12] C. Gallicchio, Chasing the echo state property, arXiv preprint arXiv:1811.10892 (2018).

[13] G. M. Hoerzer, R. Legenstein, W. Maass, Emergence of complex computational structures from chaotic neural networks through reward-modulated Hebbian learning, Cerebral Cortex 24 (3) (2012) 677–690 (2012). doi:10.1093/cercor/bhs348

[14] C. Tallec, Y. Ollivier, Can recurrent neural networks warp time? in: International Conference on Learning Representations, 2018 (2018).

URL https://openreview.net/forum?id=SJcKhk-Ab
F. M. Bianchi, E. Maiorino, M. C. Kampffmeyer, A. Rizzi, R. Jenssen, Recurrent Neural Networks for Short-Term Load Forecasting: An Overview and Comparative Analysis, Springer, 2017 (2017). doi:10.1007/978-3-319-70338-1

H. Jaeger, M. Lukoševičius, D. Popovic, U. Siewert, Optimization and applications of echo state networks with leaky-integrator neurons, Neural Networks 20 (3) (2007) 335–352 (2007). doi:10.1016/j.neunet.2007.04.016

S. Ruder, An overview of gradient descent optimization algorithms, arXiv preprint arXiv:1609.04747 (2016).

R. Pascanu, T. Mikolov, Y. Bengio, On the difficulty of training recurrent neural networks, in: Proceedings of the 30th International Conference on Machine Learning, Vol. 28, Atlanta, Georgia, USA, 2013, pp. 1310–1318 (2013).

S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Computation 9 (8) (1997) 1735–1780 (1997).

J. Chung, C. Gulcehre, K. Cho, Y. Bengio, Empirical evaluation of gated recurrent neural networks on sequence modeling, arXiv preprint arXiv:1412.3555 (2014).

S. Wisdom, T. Powers, J. Hershey, J. Le Roux, L. Atlas, Full-capacity unitary recurrent neural networks, in: D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, R. Garnett (Eds.), Advances in Neural Information Processing Systems, Curran Associates, Inc., Barcelona, Spain, 2016, pp. 4880–4888 (Dec. 2016).

J. Qiao, F. Li, H. Han, W. Li, Growing echo-state network with multiple subreservoirs, IEEE Transactions on Neural Networks and Learning Systems (2016) 1–14 (Jan. 2016). doi:10.1109/TNNLS.2016.2514275

R. Pascanu, H. Jaeger, A neurodynamical model for working memory, Neural Networks 24 (2) (2011) 199–207 (2011). doi:10.1016/j.neunet.2010.10.003

P. Tiño, Asymptotic Fisher memory of randomized linear symmetric echo state networks, Neurocomputing 298 (2018) 4–8 (2018). doi:10.1016/j.neucom.2017.11.076

L. Livi, F. M. Bianchi, C. Alippi, Determination of the edge of criticality in echo state networks through Fisher information maximization, IEEE Transactions on Neural Networks and Learning Systems 29 (3) (2018) 706–717 (Mar. 2018). doi:10.1109/TNNLS.2016.2644268
[26] A. Rodan, P. Tiño, Simple deterministically constructed cycle reservoirs with regular jumps, Neural Computation 24 (7) (2012) 1822–1852 (2012). \doi{10.1162/NECO_a_00297}

[27] F. Mastrogiuseppe, S. Ostojic, A geometrical analysis of global stability in trained feedback networks, Neural Computation 31 (6) (2019) 1139–1182 (June 2019). \doi{10.1162/neco_a_01187}

[28] A. Rivkind, O. Barak, Local dynamics in trained recurrent neural networks, Physical Review Letters 118 (2017) 258101 (Jun. 2017). \doi{10.1103/PhysRevLett.118.258101}

[29] D. Sussillo, L. F. Abbott, Generating coherent patterns of activity from chaotic neural networks, Neuron 63 (4) (2009) 544–557 (2009). \doi{10.1016/j.neuron.2009.07.018}

[30] L. F. Seoane, Evolutionary aspects of reservoir computing, Philosophical Transactions of the Royal Society B 374 (1774) (2019) 20180377 (2019). \doi{10.1098/rstb.2018.0377}

[31] Z. Lu, B. R. Hunt, E. Ott, Attractor reconstruction by machine learning, Chaos: An Interdisciplinary Journal of Nonlinear Science 28 (6) (2018) 061104 (2018). \doi{10.1063/1.5039508}

[32] D. Ibáñez-Soria, J. Garcia-Ojalvo, A. Soria-Frisch, G. Ruffini, Detection of generalized synchronization using echo state networks, Chaos: An Interdisciplinary Journal of Nonlinear Science 28 (3) (2018) 033118 (2018). \doi{10.1063/1.5010285}

[33] T. L. Carroll, Using reservoir computers to distinguish chaotic signals, Physical Review E 98 (2018) 052209 (Nov 2018). \doi{10.1103/PhysRevE.98.052209}

[34] D. V. Buonomano, W. Maass, State-dependent computations: spatiotemporal processing in cortical networks, Nature Reviews Neuroscience 10 (2) (2009) 113–125 (2009). \doi{10.1038/nrn2558}

[35] A. Ceni, P. Ashwin, L. Livi, C. Postlethwaite, Point attractors and the echo index of driven recurrent neural networks, (in prep) XX (X) (2020).

[36] C. Pötzsche, Nonautonomous continuation of bounded solutions, Commun. Pure Appl. Anal 10 (3) (2011) 937–961 (2011).

[37] K. Falconer, Fractal Geometry: Mathematical Foundations and Applications, John Wiley & Sons, New York, NY, 2004 (2004).

[38] A. Pikovsky, M. Rosenblum, J. Kurths, Synchronization: A Universal Concept in Nonlinear Sciences, Vol. 12, Cambridge university press, 2003 (2003).

[39] A. Arenas, A. Díaz-Guilera, J. Kurths, Y. Moreno, C. Zhou, Synchronization in complex networks, Physics Reports 469 (3) (2008) 93–153 (2008). \doi{10.1016/j.physrep.2008.09.002}
[40] M. Lucas, D. Fanelli, A. Stefanovska, Nonaustonomous driving induces stability in network of identical oscillators, Physical Review E 99 (1) (2019) 012309 (2019).

[41] J. Newman, Necessary and sufficient conditions for stable synchronization in random dynamical systems, Ergodic Theory and Dynamical Systems 38 (5) (2018) 1857–1875 (2018).
Appendix A  Pullback attractors of input-driven RNNs

In this section we will apply results found in the literature to input-driven RNNs dynamics, therefore we will deal with a generic leaky RNN state-update map:

\[
G(u, x) = (1 - \alpha)x + \alpha\phi(W_r x + W_i u + W_f \psi(x)).
\] (42)

In the remainder of this section the state space is a complete metric space, denoted as \((Y, d_Y)\) with \(Y \subseteq \mathbb{R}^{N_r}\) closed and \(d_Y\) the induced Euclidean distance. Hence we relax the hypothesis of compactness and deal with a state space not necessarily bounded.

For the purpose of this section it will be useful to introduce the definition of pullback absorbing set. In this regard we wish to recall from Section 2.3 that \(\sigma : U \rightarrow U\) (the shift operator) defines an autonomous dynamical system acting on a compact metric space \((U, d_U)\) (the space of all admissible input sequences), and \(\Phi : \mathbb{Z}_0^+ \times U \times Y \rightarrow Y\) (the cocycle map) describes the nonautonomous dynamics on a complete metric space \((Y, d_Y)\).

**Definition A.1.** [2, Definition 3.17] A nonempty compact subset \(B \subseteq Y\) is called pullback absorbing for a family of inputs \(V \subseteq U\) if

\[
\forall u \in V, \forall \text{ bounded } D \subseteq Y, \exists N = N(u, D) \in \mathbb{Z}_0^+ : \Phi(n, \sigma^{-n}(u), D) \subseteq B \ \forall n \geq N.
\]

Analogously, a nonempty compact subset \(B \subseteq Y\) is called forward absorbing for a family of inputs \(V \subseteq U\) if

\[
\forall u \in V, \forall \text{ bounded } D \subseteq Y, \exists N = N(u, D) \in \mathbb{Z}_0^+ : \Phi(n, u, D) \subseteq B \ \forall n \geq N.
\]

While, a nonempty compact subset \(B \subseteq Y\) is called forward absorbing uniformly for a family of inputs \(V \subseteq U\) if

\[
\forall \text{ bounded } D \subseteq Y, \exists N = N(D) \in \mathbb{Z}_0^+ : \forall u \in V, \Phi(k, u, D) \subseteq B \ \forall k \geq N.
\]

**Remark A.1.** Note that a forward absorbing set uniformly for a family of inputs \(V \subseteq U\) is necessarily a pullback absorbing set uniformly for the same family of inputs \(V\), and vice versa a pullback absorbing set uniformly for the family \(V\) is also a forward absorbing set uniformly for \(V\).

Recall Definition 3.7 of \(U\)-positively invariant set. Below we prove, under very few assumptions, that for a generic RNN with leaky-integrator neurons \([42]\), if the image of the activation function \(\phi\) is \((-L, L)\) then the hypercube \([-L, L]^{N_r}\) of phase space is a pullback and uniformly forward absorbing, positively invariant set for all input sequences \(u \in U\).

**Proposition A.1.** Let us consider a compact subspace \(U \subseteq \mathbb{R}^{N_i}\) as the set of admissible input values and \(U := U^Z\) as the set of admissible input sequences. Let us consider \(Y := \mathbb{R}^{N_r}\) equipped with the Euclidean distance. For all \(u \in U\), the dynamics of \([42]\) is described through the cocycle mapping

\[
x[k] := \Phi(1, \sigma^k(u), x[k - 1]) = G(u[k], x[k - 1])
\] (43)
by means of \((42)\). If \(\phi, \psi\) are upper semi-continuous functions and \(\phi\) is non-decreasing with image \((-L, L)\), then \(L \cdot I^N_r := [-L, L]^N_r\) is a \(\mathcal{U}\)-positively invariant, pullback and uniformly forward absorbing set for inputs \(\mathcal{U}\) of the leaky RNN dynamics ruled by \((42)\).

**Proof.** The proof is divided in two parts. First we prove that \(L \cdot I^N_r\) is a \(\mathcal{U}\)-positively invariant set, then we show that \(L \cdot I^N_r\) is a uniformly forward absorbing set.

(i) **Positively invariant.**

Let be given an initial condition \(x[0]\) such that \(||x[0]\|_\infty \leq L\). Thanks to the triangle inequality applied on \((42)\), we have

\[
||x[1]\|_\infty \leq (1 - \alpha)||x[0]\|_\infty + \alpha||\phi(W_r x[0] + W_{in} u[1] + W_{fb} \psi(x[0]))||_\infty \leq (1 - \alpha)L + \alpha L = L.
\]

Analogously, if at any time step it holds that \(||x[N]\|_\infty \leq L\), then it will be \(||x[k]\|_\infty \leq L, \forall k \geq N\).

(ii) **Uniformly forward absorbing.**

First of all, note that in this framework the universe of possible past input sequences coincides with the universe of possible future sequences. In other words, for all \(u \in \mathcal{U}\) there exists a \(v \in \mathcal{U}\) such that \(v[k] = u[-k - 1]\) for all \(k \geq 1\), thus driving the system in pullback sense with the past sequence \(u^{-}\) is equivalent to drive the system in forward sense with the future sequence \(v^{+}\). Therefore, a pullback absorbing set of Definition A.1 for the family \(\mathcal{U}\) is such if and only if it is a forward absorbing set for the family \(\mathcal{U}\). We will prove that \(L \cdot I^N_r\) is a uniformly (in \(\mathcal{U}\)) forward absorbing set, i.e. we will prove that

\[
\forall \alpha \in (0, 1), \forall x[0] \in Y, \exists N = N(\alpha, x[0]) : \forall u^{+} \in \mathcal{U}^{+}, ||x[k]\|_\infty \leq L, \forall k \geq N.
\]

The case of \(\alpha = 1\) brings trivially to the thesis. Thus let us suppose that \(\alpha \in (0, 1)\). Let be given the initial condition \(x[0]\), where we assume \(||x[0]\|_\infty > L\), otherwise the argument of (i) brings to the thesis. Note that for all \(x[k]\) such that \(||x[k]\|_\infty > L\) it holds that

\[
||x[k + 1]\|_\infty \leq (1 - \alpha)||x[k]\|_\infty + \alpha L < ||x[k]\|_\infty
\]

for all \(k \geq 0\). Now, since \(\phi, \psi\) are upper semi-continuous functions and \(\phi\) is non-decreasing then the function \(\nu : U \times [-R, R]^N_r \rightarrow \mathbb{R}_{\geq 0}\) defined as \(\nu(u, x) := ||\phi(W_r x + W_{in} u + W_{fb} \psi(x))||_\infty\) is upper semi-continuous. Hence, defined \(R := ||x[0]\|_\infty\) and since \(U\) is compact, there exists a maximum value

\[
\eta := \max_{u \in U, x} \max_{||x||_\infty \leq R} ||\phi(W_r x + W_{in} u + W_{fb} \psi(x))||_\infty.
\]
Exploiting recursively the triangle inequality on (42), the following holds
\[ ||x[k]||_{\infty} \leq (1 - \alpha)||x[k-1]||_{\infty} + \alpha \eta \]
\[ \leq (1 - \alpha)^2||x[k-2]||_{\infty} + \alpha \eta (1 - \alpha) + \alpha \eta \]
\[ \leq (1 - \alpha)^k||x[0]||_{\infty} + \eta \alpha \sum_{j=0}^{k-1} (1 - \alpha)^j \]
\[ = (1 - \alpha)^k||x[0]||_{\infty} + \eta \left[ 1 - \alpha \sum_{j=k}^{\infty} (1 - \alpha)^j \right], \]
where the last equality holds true in virtue of the geometric series limit \( \alpha \sum_{j=0}^{\infty} (1 - \alpha)^j = 1 \). Now the following inequalities are equivalent,
\[ (1 - \alpha)^k||x[0]||_{\infty} + \eta \left[ 1 - \alpha \sum_{j=k}^{\infty} (1 - \alpha)^j \right] \leq L \iff \]
\[ (1 - \alpha)^k||x[0]||_{\infty} \leq L - \eta + \eta \alpha \sum_{j=k}^{\infty} (1 - \alpha)^j \iff \]
\[ ||x[0]||_{\infty} \leq \eta \alpha \sum_{j=k}^{\infty} (1 - \alpha)^j + \frac{L - \eta}{1 - \alpha} \iff \]
\[ ||x[0]||_{\infty} \leq \eta \alpha \sum_{j=0}^{\infty} (1 - \alpha)^j + \frac{L - \eta}{1 - \alpha} = 1 \]
\[ ||x[0]||_{\infty} - \eta \leq \frac{L - \eta}{(1 - \alpha)^k}. \]
Note that, by the boundedness hypothesis of \( \phi \), it holds that \( \eta \leq L \). If \( \eta = L \) then we conclude from the last inequality that
\[ ||x[0]||_{\infty} \leq \eta = L, \]
which is in contradiction with the assumption of \( ||x[0]||_{\infty} > L \). Accordingly, \( \eta \leq L \) must hold, leading to the following inequality:
\[ (1 - \alpha)^k \leq \frac{L - \eta}{||x[0]||_{\infty} - \eta} \iff \]
\[ k \geq \frac{\ln(L - \eta) - \ln(||x[0]||_{\infty} - \eta)}{\ln(1 - \alpha)}. \]

Therefore, after a number of time steps given by \( N(\alpha,x[0]) := \frac{\ln(L - \eta) - \ln(||x[0]||_{\infty} - \eta)}{\ln(1 - \alpha)} \), the internal state of a leaky ESN (42) will surely lie inside the hypercube \( L \cdot I^{N_r} \). □

For the sake of clarity, we report here below without proof [2, Theorem 3.20] but using our notation.

**Theorem A.2.** [2, Theorem 3.20] Let \( U \subset \mathbb{R}^{N_i} \) be compact and \( (\mathcal{U},d_{\mathcal{U}}) \) be the compact metric space of admissible input sequences, where \( \mathcal{U} := U^{\mathbb{Z}} \) and \( d_{\mathcal{U}} \) as defined in (5). Let \( \sigma : \mathcal{U} \rightarrow \mathcal{U} \) be the shift operator.
Let \((\sigma, \Phi)\) be the skew product flow on a complete metric space \((Y, d_Y)\), and \(\Phi\) defined as Definition 2.1. If there exists a nonempty compact subset \(B \subset Y\) which is pullback absorbing and positively invariant for \(U\), then there exists a unique pullback attractor \(A = \{A_u\}_{u \in U}\) with fibres in \(B\) uniquely determined by

\[
A_u := \bigcap_{m \geq 0} \bigcup_{s \geq m} \Phi(s, \sigma^{-s}(u), B), \quad \forall u \in U. \tag{44}
\]

In addition, since \((U, d_U)\) is a compact metric space the subset \(A(U) := \bigcup_{u \in U} A_u \subset B\) uniformly (in \(U\)) forward attracts every bounded set of the phase space, that is

\[
\lim_{k \to \infty} \sup_{u \in U} h(\Phi(k, u, D), A(U)) = 0, \quad \forall \text{ bounded } D \subseteq X. \tag{45}
\]

Thanks to Proposition A.1 we can apply Theorem A.2 on a RNN taking values with phase space the whole \(\mathbb{R}^{N_r}\) and use the set \(B = [-L, L]^{N_r}\) in order to construct the pullback attractor. Moreover, since \(U\) is compact the second part of Theorem A.2 implies that the entire nonautonomous dynamics of a RNN is uniformly attracted to a closed subset inside \(B = [-L, L]^{N_r}\). This justifies our assumption of considering the whole space as \(X = [-L, L]^{N_r}\) whenever referring to the RNN nonautonomous dynamics even with leaky neurons.

Therefore, pullback attractor has component sets made by

\[
A_u := \bigcap_{m \geq 0} \bigcup_{s \geq m} \Phi(s, \sigma^{-s}(u), X), \quad \forall u \in U. \tag{46}
\]

Note that under this formalism the resulting set \(A_u\) actually depends only by the left-infinite sequence \(u^- = \{\ldots, u[-2], u[-1], u[0]\}\). Furthermore, since \(U = U^\mathbb{Z}\) is shift-invariant, i.e. \(\sigma^n(U) = U\) for all \(n \in \mathbb{Z}\), we can equivalently write \((46)\) as follows

\[
A_{\sigma^n(u)} := \bigcap_{m \geq -n} \bigcup_{s \geq m} \Phi(s + n, \sigma^{-s}(u), X) = \bigcap_{m' \leq n} \bigcup_{s' \leq m'} \Phi(n - s', \sigma^{s'}(u), X), \quad \forall u \in U, \tag{47}
\]

where the last equality is obtained transforming indices as \(m' = -m\) and \(s' = -s\).

As a consequence, for any fixed input sequence \(v \in U\), from \((47)\) we get the component sets of the pullback attractor with regard to such input sequence:

\[
A(v) = \{A_{\sigma^n(v)}\}_{n \in \mathbb{Z}}. \tag{48}
\]

Theorem A.4 below implies that \((48)\) is exactly the natural association \((14)\) defined in Section 2.3 for a given input sequence \(v \in U\).

**Lemma A.3.** Assume the hypotheses of Theorem A.2 hold. If there exists a \(U\)-positively invariant set \(B \subset X\), then

\[
\Phi(N + 1, v, B) \subseteq \Phi(N, \sigma(v), B) \quad \forall N \in \mathbb{Z}_0^+, \forall v \in U. \tag{49}
\]
Theorem A.4. Let \( U \subset \mathbb{R}^N \) be compact and \((U,d_U)\) be the compact metric space of admissible input sequences, where \( U := U^2 \) and \( d_U \) as defined in (9). Let \( \sigma : U \rightarrow U \) be the shift operator. Let \((\sigma,\Phi)\) be the skew product flow on a complete metric space \((X,d_X), X \subseteq \mathbb{R}^N\) closed and \( d_X \) the Euclidean distance, and \( \Phi \) defined as Definition 2.1. Let us be given an input sequence \( v \in U \), yielding the subset \( V \subseteq U \) of input sequences \( V := \{ \ldots, \sigma^{-2}(v), \sigma^{-1}(v), v, \sigma(v), \sigma^2(v), \ldots \} \). If there exists a nonempty compact subset \( B \subset X \) which is pullback absorbing and positively invariant for \( V \), then there exists the (global) pullback attractor \( A(v) := \{ A_n \}_{n \in \mathbb{Z}} \) of the dynamics driven by \( v \) and it has component sets

\[
A_n := \bigcap_{m \leq n} \Phi(n - m, \sigma^m(v), B), \quad \forall n \in \mathbb{Z}. \tag{50}
\]

In addition, if \( V \) is compact in \((U,d_U)\) then \( A(V) := \bigcup_{n \in \mathbb{Z}} A_n \subset B \) uniformly (in time) forward attracts the whole phase space driven by \( v \), that is

\[
\lim_{k \to \infty} \sup_{n \in \mathbb{Z}} h(\Phi(k, \sigma^n(v), X), A(V)) = 0. \tag{51}
\]

Proof. The proof is a direct application of Theorem A.2. We need to prove that component sets of (44) coincide with component sets of (50). Let be given a \( v \in U \). Fixed a \( n \in \mathbb{Z} \), let us define a sequence of sets \( B_s := \Phi(n + s, \sigma^s(v), B) \), for \( s \geq -n \). \( B_s = \overline{B_s} \) holds as \( B \) is compact and \( \forall N \in \mathbb{Z}_0^+ \), \( \forall v \in U \) the function \( \Phi(N, v, \cdot) : X \rightarrow X \) is a closed map, i.e. it maps closed sets in closed sets\(^4\) hence compact sets in compact sets in our case. It is known that, for a nonincreasing sequence of closed sets \( \{ B_s \}_{s \geq -n} \), the limit set \( L(n) := \bigcap_{m \geq -n} \bigcup_{s \geq m} B_s \) exists and it also holds that \( B_m = \bigcup_{s \geq m} B_s \). Therefore, it is sufficient to prove that \( B_{s+1} \subseteq B_s \) for having that \( L(n) := \bigcap_{m \geq -n} B_m \). Relation \( B_{s+1} \subseteq B_s \) holds thanks to Lemma A.3. Concluding, component sets (44) can be written as \( L(n) := \bigcap_{m \geq -n} B_m \), which reads

\[
A_n := \bigcap_{m \geq -n} \Phi(n + m, \sigma^m(v), B) = \bigcap_{m' \leq n} \Phi(n - m', \sigma^{m'}(v), B),
\]

that is (50) of thesis.

To conclude, (51) follows from (45) by noting that we are interested in the subset of input sequences given by \( V = \{ \ldots, \sigma^{-2}(v), \sigma^{-1}(v), v, \sigma(v), \sigma^2(v), \ldots \} \). Therefore, the supremum \( \sup_{u \in V} \) can be expressed with the supremum \( \sup_{n \in \mathbb{Z}} \) of (51) of thesis.

\[^4\]This is ensured by the closed map lemma, which holds since \( X \) is a compact space (as domain) and a Hausdorff space (as codomain) and \( \Phi(n, u, \cdot) : X \rightarrow X \) is continuous \( \forall n \in \mathbb{Z}, \forall u \in U \).
Remark A.2. Note that, in general, for a given $v \in U$ the set $V = \{\sigma^n(v)\}_{n \in \mathbb{Z}}$ is not compact in $U$. In particular, if $v$ is aperiodic then $V$ will have limit points that are not contained in $V$.

A.1 ESP in RNNs implies uniformly (in time) forward convergence

For the sake of completeness, in Theorem A.5 we show, using our framework, a result found in the literature, which links pullback attraction with forward attraction in the particular case where the global pullback attractor is an entire solution.

Theorem A.5. [2, Theorem 3.44] Let $U \subset \mathbb{R}^{N_i}$ be compact and $(\mathcal{U}, d_{\mathcal{U}})$ be the compact metric space of admissible input sequences, where $\mathcal{U} := U^\mathbb{Z}$ and $d_{\mathcal{U}}$ as defined in (5). Let $\sigma : \mathcal{U} \rightarrow \mathcal{U}$ be the shift operator. Let $(\sigma, \Phi)$ be the skew product flow on the complete metric space $Y = \mathbb{R}^{N_r}$, with the Euclidean distance, and $\Phi$ defined as Definition 2.1. Assume there exists a nonempty compact subset $B \subset Y$ such that:

- $B$ is a $\mathcal{U}$-positively invariant set; and
- $B$ is forward absorbing uniformly in $\mathcal{U}$.

Suppose that for all $u \in \mathcal{U}$ the (global) pullback attractor for input $u$ is an entire solution. Then, for any given input sequence $v \in \mathcal{U}$ such entire solution for input $v$, denoted as $x = \{x[k]\}_{k \in \mathbb{Z}}$, is also attracting in forward sense uniformly in time:

$$\lim_{k \to \infty} \sup_{n \in \mathbb{Z}} h(\Phi(k, \sigma^n(v), D), x[n + k]) = 0, \quad \forall \text{ bounded } D \subseteq Y.$$  

In Proposition A.1 we proved that $B = [-L, L]^{N_r}$ fulfils the hypothesis of Theorem A.5 for a generic leaky RNNs with $\phi, \psi$ upper semi-continuous functions and $\phi$ non-decreasing with bounded image. Therefore, Theorem A.5 for a generic RNN reads as: if the global pullback attractor is an entire solution for all $u \in U^\mathbb{Z}$ then such unique entire solution of the system is also forward attracting uniformly in time. More formally we can state the following result.

Proposition A.6. Let be given a $\alpha \in (0, 1]$ and real matrices $W_r, W_i, W_f$ of dimensions, respectively, $N_r \times N_r$, $N_r \times N_i$, $N_r \times N_o$. Let $\phi : \mathbb{R} \rightarrow (-L, L)$, $\psi : \mathbb{R}^{N_r} \rightarrow \mathbb{R}^{N_o}$ be upper semi-continuous functions and $\phi$ non-decreasing. Consider the following input-driven RNN

$$x[k] = G(u[k], x[k - 1]), \quad x \in \mathbb{R}^{N_r}, \ u \in U \subset \mathbb{R}^{N_i} \text{ compact}, \tag{52}$$

$$G(u,x) = (1 - \alpha)x + \alpha \phi(W_r x + W_i u + W_f \psi(x)). \tag{53}$$

If the ESP as originally introduced in [4, Definition 1] holds for the input-driven RNN (52) w.r.t the compact input space $\mathcal{U} = U^\mathbb{Z}$ then the unique entire solution of such nonautonomous system is uniformly state contracting [4, Definition 4].
Proof of Lemma 3.3

Let us proceed by contradiction assuming that $\sigma(A) > M$. By definition,

$$\sigma(A) = \max_{y \in \mathbb{R}^N \setminus \{0\}} \frac{\|Ay\|}{\|y\|}$$

Therefore, for a small enough $\varepsilon > 0$, it must exist a unit vector $\vec{v}$ such that

$$\|A(c\vec{v})\| \geq (M + \varepsilon)\|c\vec{v}\| \quad \forall c \in \mathbb{R} \setminus \{0\}. \tag{54}$$

Now let us move on the line $z = x^* + c\vec{v}$ and consider the linearisation of the map $F$ around $x^*$, which reads

$$F(z) = x^* + A(z - x^*) + R(z - x^*), \tag{55}$$

where the rest of the expansion $R(z - x^*)$ is such that

$$\lim_{z \to x^*} \frac{\|R(z - x^*)\|}{\|z - x^*\|} = 0, \tag{56}$$

whenever the map $F$ is regular enough in $x^*$. Now, since $x^*$ is a UASP, there exists a $\delta > 0$ such that

$$\|F(z) - x^*\| < M\|z - x^*\|, \quad \forall z \in B_\delta(x^*). \tag{57}$$

By means of the expansion $\text{(55)}$ and applying the reverse triangle inequality we get

$$\left|\|A(z - x^*)\| - \|R(x - x^*)\|\right| \leq \|A(x - x^*) + R(x - x^*)\| = \|F(x) - x^*\| < M\|x - x^*\|. \tag{58}$$

Now, if $z$ is close enough to $x^*$ along the direction pointed by $\vec{v}$, then it holds that $\|A(z - x^*)\| \geq \|R(z - x^*)\|$. Indeed, $\text{(56)}$ implies that for any $\epsilon > 0$ there exists a $c_\epsilon > 0$ such that $\|R(c\vec{v})\| \leq \epsilon|c\vec{v}|$ for all $|c| \leq c_\epsilon$. Moreover, exploiting the fact that $z$ is chosen such that $z - x^* = c\vec{v}$ satisfies $\text{(54)}$, then it holds that

$$(M + \varepsilon)|c\vec{v}| = (M + \varepsilon)|z - x^*| \leq \|A(z - x^*)\|. \tag{59}$$

Hence, for the choice $\varepsilon = (M + \varepsilon)$ there exists a $c_{M+\varepsilon} > 0$ such that $\|R(z - x^*)\| \leq \|A(z - x^*)\|$ for all $z - x^* = c\vec{v}$ having $|c| \leq c_{M+\varepsilon}$. Therefore, we can rewrite $\text{(58)}$ as follows

$$\|R(z - x^*)\| > \|A(z - x^*)\| - M\|z - x^*\|. \tag{59}$$

Finally, applying again $\text{(54)}$ in $\text{(59)}$, we obtain

$$\|R(z - x^*)\| > (M + \varepsilon - M)|z - x^*|, \tag{60}$$

i.e. for all $z - x^* = c\vec{v}$ having $|c| \leq c_{M+\varepsilon}$ holds $\|R(z - x^*)\| > \varepsilon|z - x^*|$, which is in contradiction with relation $\text{(56)}$. \qed