Echo state networks are universal

Lyudmila Grigoryeva [a], Juan-Pablo Ortega [b,c,*]

[a] *Department of Mathematics and Statistics, Universität Konstanz, Box 146, D-78457 Konstanz, Germany*
[b] *Universität Sankt Gallen, Faculty of Mathematics and Statistics, Bodanstrasse 6, CH-9000 Sankt Gallen, Switzerland*
[c] *Centre National de la Recherche Scientifique (CNRS), France*

ARTICLE INFO

ABSTRACT

This paper shows that echo state networks are universal uniform approximants in the context of discrete-time fading memory filters with uniformly bounded inputs defined on negative infinite times. This result guarantees that any fading memory input/output system in discrete time can be realized as a simple finite-dimensional neural network-type state-space model with a static linear readout map. This approximation is valid for infinite time intervals. The proof of this statement is based on fundamental results, also presented in this work, about the topological nature of the fading memory property and about reservoir computing systems generated by continuous reservoir maps.

© 2018 Elsevier Ltd. All rights reserved.

1. Introduction

Many recently introduced machine learning techniques in the context of dynamical problems have much in common with system identification procedures developed in the last decades for applications in signal treatment, circuit theory and, in general, systems theory. In these problems, system knowledge is only available in the form of input–output observations and the task consists in finding or *learning* a model that approximates it for mainly forecasting or classification purposes. An important goal in that context is finding families of transformations that are both computationally feasible and versatile enough to reproduce a rich number of patterns just by modifying a limited number of procedural parameters.

The versatility or flexibility of a given machine learning paradigm is usually established by proving its **universality**. We say that a family of transformations is universal when its elements can approximate as accurately as one wants all the elements of a sufficiently rich class containing, for example, all continuous or even all measurable transformations. In the language of learning theory, this is equivalent to the possibility of making approximation errors arbitrarily small (Cucker & Smale, 2002; Cucker & Zhou, 2007; Smale & Zhou, 2003). In more mathematical terms, the universality of a family amounts to its density in a rich class

of the type mentioned above. Well-known universality results are, for example, the uniform approximation properties of feedforward neural networks established in Cybenko (1989), Hornik (1991) and Hornik, Stinchcombe, and White (1989) in the context of static continuous and, more generally, measurable real functions.

A first solution to this problem in the dynamic context was pioneered in the works of Fréchet (1910) and Volterra (1930) one century ago when they proved that finite Volterra series can be used to uniformly approximate continuous functionals defined on compact sets of continuous functions. These results were further extended in the 1950s by the MIT school lead by Brilliant (1958), George (1959) and Wiener (1958) but always under compactness assumptions on the input space and the time interval in which inputs are defined. A major breakthrough was the generalization to infinite time intervals carried out by Boyd and Chua in Boyd and Chua (1985), who formulated a uniform approximation theorem using Volterra series for operators endowed with the so called **fading memory property** on continuous time inputs. An input/output system is said to have fading memory when the outputs associated to inputs that are close in the recent past are close, even when those inputs may be very different in the distant past.

In this paper we address the universality or the uniform approximation problem for transformations or **filters** of discrete time signals of infinite length that have the fading memory property. The approximating set that we use is generated by nonlinear state-space transformations and that is referred to as **reservoir computers (RC)** (Crook, 2007; Jaeger, 2010; Jaeger & Haas, 2004; Lukoševičius & Jaeger, 2009; Maass, 2011; Maass, Natschläger, & Markram, 2002; Verstraeten, Schrauwen, D'Haene, & Stroobandt,

2007) or ***reservoir systems***. These are special types of recurrent neural networks determined by two maps, namely a ***reservoir*** $F : \mathbb{R}^N \times \mathbb{R}^n \longrightarrow \mathbb{R}^N$, $n, N \in \mathbb{N}$, and a ***readout*** map $h : \mathbb{R}^N \to \mathbb{R}^d$ that under certain hypotheses transform (or filter) an infinite discrete-time input $\mathbf{z} = (\ldots, \mathbf{z}_{-1}, \mathbf{z}_0, \mathbf{z}_1, \ldots) \in (\mathbb{R}^n)^{\mathbb{Z}}$ into an output signal $\mathbf{y} \in (\mathbb{R}^d)^{\mathbb{Z}}$ of the same type using the state-space transformation given by:

$$\begin{cases} \mathbf{x}_t = F(\mathbf{x}_{t-1}, \mathbf{z}_t), & \text{(a)} \\ \mathbf{y}_t = h(\mathbf{x}_t), & \text{(b)} \end{cases} \qquad (1.1)$$

where $t \in \mathbb{Z}$ and the dimension $N \in \mathbb{N}$ of the state vectors $\mathbf{x}_t \in \mathbb{R}^N$ is referred to as the number of virtual ***neurons*** of the system. When a RC system has a uniquely determined filter associated to it, we refer to it as the ***RC filter***.

An important advantage of the RC approach is that, under certain hypotheses, intrinsically infinite dimensional problems regarding filters can be translated into analogous questions related to the reservoir and readout maps that generate them and that are defined on much simpler finite dimensional spaces. This strategy has already been used in the literature in relation to the universality question in, for instance, Matthews (1992, 1993), Perryman (1996), Sandberg (1991a, b) and Stubberud and Perryman (1997). The universal approximation properties of feedforward neural networks (Arnold, 1957; Cybenko, 1989; Hornik, 1991, 1993; Hornik et al., 1989; Hornik, Stinchcombe, & White, 1990; Kolmogorov, 1956; Rüschendorf & Thomsen, 1998; Sprecher, 1965, 1996, 1997) were used in those works to find neural networks-based families of filters that are dense in the set of approximately finite memory filters with inputs defined in the positive real half-line. Other works in connection with the universality problem in the dynamic context are Maass, Joshi, and Sontag (2007), Maass et al. (2002), Maass, Natschläger, and Markram (2004) and Maass and Sontag (2000) where RC is referred to as Liquid State Machines. In those references and in the same vein as in Boyd and Chua (1985), universal families of RC systems with inputs defined on infinite continuous time intervals were identified in the fading memory category as a corollary of the Stone–Weierstrass theorem. This approach required invoking the natural hypotheses associated to this result, like the pointwise separation property or the compactness of the input space, that was obtained as a consequence of the fading memory property. Another strand of interesting literature that we will not explore in this work has to do with the Turing computability capabilities of the systems of the type that we just introduced; recent relevant works in this direction are Cabessa and Villa (2016), Kilian and Siegelmann (1996), Siegelmann, Horne, and Giles (1997) and references therein.

The main contribution of this paper is showing that a particularly simple type of RC systems called ***echo state networks (ESNs)*** can be used as *universal approximants in the context of discrete-time fading memory filters with uniformly bounded inputs defined on negative infinite times*. ESNs are RC systems of the form (1.1)(a)–(1.1)(b) given by:

$$\begin{cases} \mathbf{x}_t = \sigma\left(A\mathbf{x}_{t-1} + C\mathbf{z}_t + \boldsymbol{\zeta}\right), & \text{(a)} \\ \mathbf{y}_t = W\mathbf{x}_t. & \text{(b)} \end{cases} \qquad (1.2)$$

In these equations, $C \in \mathbb{M}_{N,n}$ is called the ***input mask***, $\boldsymbol{\zeta} \in \mathbb{R}^N$ is the ***input shift***, and $A \in \mathbb{M}_{N,N}$ is referred to as the ***reservoir matrix***. The map σ in the state-space equation (1.2)(a) is constructed by componentwise application of a sigmoid function (like the hyperbolic tangent or the logistic function) and is called the ***activation function***. Finally, the readout map is linear in this case and implemented via the ***readout matrix*** $W \in \mathbb{M}_{d,N}$. ESNs already appear in Matthews (1992, 1993) under the name of *recurrent networks* but it was only more recently, in the works of Jaeger and Haas (2004), that their outstanding performance in machine learning applications was demonstrated.

The strategy that we follow to prove that statement is a combination of what the literature refers to as ***internal*** and ***external approximation***. External approximation is the construction of a RC filter that approximates a given (not necessarily RC) filter. In the internal approximation problem, one is given a RC filter and builds another RC filter that approximates it by finding reservoir and readout maps that are close to those of the given one. In the external part of our proof we use a previous work (Grigoryeva & Ortega, 2018a) where we constructed a family of RC systems with linear readouts that we called ***non-homogeneous state affine systems (SAS)***. We showed in that paper that the RC filters associated to SAS systems uniformly approximate any discrete-time fading memory filter with uniformly bounded inputs defined on negative infinite times. Regarding the internal approximation, we show that any RC filter, in particular SAS filters, can be approximated by ESN filters using the universal approximation property of neural networks. These two facts put together allow us to conclude that ESN filters are capable of uniformly approximating any discrete-time fading memory filter with uniformly bounded inputs. We emphasize that this result is shown exclusively for deterministic inputs using a uniform approximation criterion; an extension of this statement that accommodates stochastic inputs and L^p approximation criteria can be found in Gonon and Ortega (2018).

The paper is structured in three sections:

- Section 2 introduces the notation that we use all along the paper and, more importantly, specifies the topologies and Banach space structures that we need in order to talk about continuity in the context of discrete-time filters. It is worth mentioningâĂĺ that we characterize the fading memory property as a continuity condition of the filters that have it with respect to the product topology in the input space. In other words, *the fading memory property is not a metric property, as it is usually presented in the literature, but a topological one*. An important conceptual consequence of this fact is that the fading memory property does not contain any information about the rate at which systems that have it "forget" inputs. We formulate several corollaries of this fact that are very instrumental in the developments in the paper. The situation is different for the so called differentiable reservoir filters where the decay rate of the weighting sequence with respect to which they have the fading memory property plays an important role in the analysis (see Grigoryeva and Ortega (2018b)).
- Section 3 contains a collection of general results in relation with the properties of the RC systems generated by continuous reservoir maps. In particular, we provide conditions that guarantee that a unique reservoir filter can be associated to them (the so called ***echo state property***) and we identify situations in which those filters are themselves continuous (they satisfy automatically the fading memory property). We also point out large classes of RC systems for which internal approximation is possible, that is, if the RC systems are close then so are the associated reservoir filters.
- Section 4 shows that *echo state networks are universal uniform approximants in the category of discrete-time fading memory filters with uniformly bounded inputs*.

2. Continuous and fading memory filters

This section introduces the notation of the paper as well as general facts about filters and functionals needed in the developments that follow. The new results are contained in Section 2.3, where we characterize the fading memory property as a continuity condition when the sequence spaces where inputs and outputs are defined are uniformly bounded and are endowed with the product topology. This feature makes this property independent of the weighting sequences that are usually introduced to define it.

2.1. Notation

Vectors and matrices. A column vector is denoted by a bold lower case symbol like $\mathbf{r}$ and $\mathbf{r}^\top$ indicates its transpose. Given a vector $\mathbf{v} \in \mathbb{R}^n$, we denote its entries by v_i, with $i \in \{1, \ldots, n\}$; we also write $\mathbf{v} = (v_i)_{i \in \{1,\ldots,n\}}$. We denote by $\mathbb{M}_{n,m}$ the space of real $n \times m$ matrices with $m, n \in \mathbb{N}$. When $n = m$, we use the symbol $\mathbb{M}_n$ to refer to the space of square matrices of order n. Given a matrix $A \in \mathbb{M}_{n,m}$, we denote its components by A_{ij} and we write $A = (A_{ij})$, with $i \in \{1, \ldots, n\}$, $j \in \{1, \ldots m\}$. Given a vector $\mathbf{v} \in \mathbb{R}^n$, the symbol $\|\mathbf{v}\|$ stands for any norm in $\mathbb{R}^n$ (they are all equivalent) and is not necessarily the Euclidean one, unless it is explicitly mentioned. The open balls with respect to a given norm $\|\cdot\|$, center $\mathbf{v} \in \mathbb{R}^n$, and radius $r > 0$ will be denoted by $B_{\|\cdot\|}(\mathbf{v}, r)$; their closures by $\overline{B_{\|\cdot\|}(\mathbf{v}, r)}$. For any $A \in \mathbb{M}_{n,m}$, $\|A\|_2$ denotes its matrix norm induced by the Euclidean norms in $\mathbb{R}^m$ and $\mathbb{R}^n$, and satisfies (Horn & Johnson, 2013, Example 5.6.6) that $\|A\|_2 = \sigma_{\max}(A)$, with $\sigma_{\max}(A)$ the largest singular value of A. $\|A\|_2$ is sometimes referred to as the spectral norm of A. The symbol $\|\cdot\|$ is reserved for the norms of operators or functionals defined on infinite dimensional spaces.

Sequence spaces. $\mathbb{N}$ denotes the set of natural numbers with the zero element included. $\mathbb{Z}$ (respectively, $\mathbb{Z}_+$ and $\mathbb{Z}_-$) are the integers (respectively, the positive and the negative integers). The symbol $(\mathbb{R}^n)^{\mathbb{Z}}$ denotes the set of infinite real sequences of the form $\mathbf{z} = (\ldots, \mathbf{z}_{-1}, \mathbf{z}_0, \mathbf{z}_1, \ldots)$, $\mathbf{z}_i \in \mathbb{R}^n$, $i \in \mathbb{Z}$; $(\mathbb{R}^n)^{\mathbb{Z}_-}$ and $(\mathbb{R}^n)^{\mathbb{Z}_+}$ are the subspaces consisting of, respectively, left and right infinite sequences: $(\mathbb{R}^n)^{\mathbb{Z}_-} = \{\mathbf{z} = (\ldots, \mathbf{z}_{-2}, \mathbf{z}_{-1}, \mathbf{z}_0) \mid \mathbf{z}_i \in \mathbb{R}^n, i \in \mathbb{Z}_-\}$, $(\mathbb{R}^n)^{\mathbb{Z}_+} = \{\mathbf{z} = (\mathbf{z}_0, \mathbf{z}_1, \mathbf{z}_2, \ldots) \mid \mathbf{z}_i \in \mathbb{R}^n, i \in \mathbb{Z}_+\}$. Analogously, $(D_n)^{\mathbb{Z}}$, $(D_n)^{\mathbb{Z}_-}$, and $(D_n)^{\mathbb{Z}_+}$ stand for (semi-)infinite sequences with elements in the subset $D_n \subset \mathbb{R}^n$. In most cases we endow these infinite product spaces with the Banach space structures associated to one of the following two norms:

- The **supremum norm**: define $\|\mathbf{z}\|_\infty := \sup_{t \in \mathbb{Z}} \{\|\mathbf{z}_t\|\}$. The symbols $\ell^\infty(\mathbb{R}^n)$ and $\ell_\pm^\infty(\mathbb{R}^n)$ are used to denote the Banach spaces formed by the elements in the corresponding infinite product spaces that have a finite supremum norm.
- The **weighted norm**: let $w : \mathbb{N} \longrightarrow (0, 1]$ be a decreasing sequence with zero limit. We define the associated **weighted norm** $\|\cdot\|_w$ on $(\mathbb{R}^n)^{\mathbb{Z}_-}$ associated to the **weighting sequence** w as the map:

$$
\|\cdot\|_w : \quad (\mathbb{R}^n)^{\mathbb{Z}_-} \quad \longrightarrow \quad \overline{\mathbb{R}^+}
$$
$$
\mathbf{z} \quad \longmapsto \quad \|\mathbf{z}\|_w := \sup_{t \in \mathbb{Z}_-} \{\|\mathbf{z}_t w_{-t}\|\}.
$$

Proposition A.2 in Appendix A.11 shows that the space

$$
\ell_-^w(\mathbb{R}^n) := \left\{ \mathbf{z} \in \left(\mathbb{R}^n\right)^{\mathbb{Z}_-} \mid \|\mathbf{z}\|_w < \infty \right\},
$$

endowed with weighted norm $\|\cdot\|_w$ forms also a Banach space.

It is straightforward to show that $\|\mathbf{z}\|_w \leq \|\mathbf{z}\|_\infty$, for all $\mathbf{z} \in (\mathbb{R}^n)^{\mathbb{Z}_-}$. This implies that $\ell_-^\infty(\mathbb{R}^n) \subset \ell_-^w(\mathbb{R}^n)$ and that the inclusion map $(\ell_-^\infty(\mathbb{R}^n), \|\cdot\|_\infty) \hookrightarrow (\ell_-^w(\mathbb{R}^n), \|\cdot\|_w)$ is continuous.

2.2. Filters and systems

Filters. Let $D_n \subset \mathbb{R}^n$ and $D_N \subset \mathbb{R}^N$. We refer to the maps of the type $U : (D_n)^{\mathbb{Z}} \longrightarrow (D_N)^{\mathbb{Z}}$ as **filters** or **operators** and to those like $H : (D_n)^{\mathbb{Z}} \longrightarrow D_N$ (or $H : (D_n)^{\mathbb{Z}_\pm} \longrightarrow D_N$) as $\mathbb{R}^N$-valued **functionals**. These definitions will be sometimes extended to accommodate situations where the domains and the targets of the filters are not necessarily product spaces but just arbitrary subsets of $(\mathbb{R}^n)^{\mathbb{Z}}$ and $(\mathbb{R}^N)^{\mathbb{Z}}$ like, for instance, $\ell^\infty(\mathbb{R}^n)$ and $\ell^\infty(\mathbb{R}^N)$.

A filter $U : (D_n)^{\mathbb{Z}} \longrightarrow (D_N)^{\mathbb{Z}}$ is called **causal** when for any two elements $\mathbf{z}, \mathbf{w} \in (D_n)^{\mathbb{Z}}$ that satisfy that $\mathbf{z}_\tau = \mathbf{w}_\tau$ for any $\tau \leq t$, for a given $t \in \mathbb{Z}$, we have that $U(\mathbf{z})_t = U(\mathbf{w})_t$. Let $T_\tau : (D_n)^{\mathbb{Z}} \longrightarrow (D_n)^{\mathbb{Z}}$ be the **time delay** operator defined by $T_\tau(\mathbf{z})_t := \mathbf{z}_{t-\tau}$. The filter U is called **time-invariant** (TI) when it commutes with the time delay operator, that is, $T_\tau \circ U = U \circ T_\tau$, for any $\tau \in \mathbb{Z}$ (in this expression, the two operators T_τ have to be understood as defined in the appropriate sequence spaces).

We recall (see for instance Boyd & Chua, 1985) that there is a bijection between causal time-invariant filters and functionals on $(D_n)^{\mathbb{Z}_-}$. Indeed, consider the sets $\mathbb{F}_{(D_n)^{\mathbb{Z}_-}}$ and $\mathbb{H}_{(D_n)^{\mathbb{Z}_-}}$ defined by

$$
\mathbb{F}_{(D_n)^{\mathbb{Z}_-}} := \big\{ U : (D_n)^{\mathbb{Z}} \longrightarrow (\mathbb{R}^N)^{\mathbb{Z}} \mid
$$
$$
U \text{ is causal and time-invariant} \big\}, \tag{2.1}
$$

$$
\mathbb{H}_{(D_n)^{\mathbb{Z}_-}} := \big\{ H : (D_n)^{\mathbb{Z}_-} \longrightarrow \mathbb{R}^N \big\}. \tag{2.2}
$$

Then, given a time-invariant filter $U : (D_n)^{\mathbb{Z}} \longrightarrow (\mathbb{R}^N)^{\mathbb{Z}}$, we can associate to it a functional $H_U : (D_n)^{\mathbb{Z}_-} \longrightarrow \mathbb{R}^N$ via the assignment $H_U(\mathbf{z}) := U(\mathbf{z}^e)_0$, where $\mathbf{z}^e \in (\mathbb{R}^n)^{\mathbb{Z}}$ is an arbitrary extension of $\mathbf{z} \in (D_n)^{\mathbb{Z}_-}$ to $(D_n)^{\mathbb{Z}}$. Let $\Psi : \mathbb{F}_{(D_n)^{\mathbb{Z}_-}} \longrightarrow \mathbb{H}_{(D_n)^{\mathbb{Z}_-}}$ be the map such that $\Psi(U) := H_U$. Conversely, for any functional $H : (D_n)^{\mathbb{Z}_-} \longrightarrow \mathbb{R}^N$, we can define a time-invariant causal filter $U_H : (D_n)^{\mathbb{Z}} \longrightarrow (\mathbb{R}^N)^{\mathbb{Z}}$ by $U_H(\mathbf{z})_t := H((\mathbb{P}_{\mathbb{Z}_-} \circ T_{-t})(\mathbf{z}))$, where T_{-t} is the $(-t)$-time delay operator and $\mathbb{P}_{\mathbb{Z}_-} : (\mathbb{R}^n)^{\mathbb{Z}} \longrightarrow (\mathbb{R}^n)^{\mathbb{Z}_-}$ is the natural projection. Let $\Phi : \mathbb{H}_{(D_n)^{\mathbb{Z}_-}} \longrightarrow \mathbb{F}_{(D_n)^{\mathbb{Z}_-}}$ be the map such that $\Phi(H) := U_H$. It is easy to verify that:

$$
\Psi \circ \Phi = \mathbb{I}_{\mathbb{H}_{(D_n)^{\mathbb{Z}_-}}} \quad \text{or, equivalently,} \quad H_{U_H} = H,
$$
$$
\text{for any functional} \quad H : (D_n)^{\mathbb{Z}_-} \to \mathbb{R}^N,
$$

$$
\Phi \circ \Psi = \mathbb{I}_{\mathbb{F}_{(D_n)^{\mathbb{Z}_-}}} \quad \text{or, equivalently,} \quad U_{H_U} = U,
$$
$$
\text{for any causal TI filter} \quad U : (D_n)^{\mathbb{Z}} \to (\mathbb{R}^N)^{\mathbb{Z}},
$$

that is, Ψ and Φ are inverses of each other and hence are both bijections. Additionally, we note that the sets $\mathbb{F}_{(D_n)^{\mathbb{Z}_-}}$ and $\mathbb{H}_{(D_n)^{\mathbb{Z}_-}}$ are vector spaces with naturally defined operations and that Ψ and Φ are linear maps between them, which allows us to conclude that $\mathbb{F}_{(D_n)^{\mathbb{Z}_-}}$ and $\mathbb{H}_{(D_n)^{\mathbb{Z}_-}}$ are linear isomorphic.

When a filter is causal and time-invariant, we work in many situations just with the restriction $U : (D_n)^{\mathbb{Z}_-} \longrightarrow (D_N)^{\mathbb{Z}_-}$ instead of the original filter $U : (D_n)^{\mathbb{Z}} \longrightarrow (D_N)^{\mathbb{Z}}$ without making the distinction, since the former uniquely determines the latter. Indeed, by definition, for any $\mathbf{z} \in (D_n)^{\mathbb{Z}}$ and $t \in \mathbb{Z}$:

$$
U(\mathbf{z})_t = (T_{-t}(U(\mathbf{z})))_0 = (U(T_{-t}(\mathbf{z})))_0, \tag{2.3}
$$

where the second equality holds by the time-invariance of U and the value in the right-hand side depends only on $\mathbb{P}_{\mathbb{Z}_-}(T_{-t}(\mathbf{z})) \in (D_n)^{\mathbb{Z}_-}$, by causality.

Reservoir systems and filters. Consider now the RC system determined by (1.1)(a)–(1.1)(b) with reservoir map defined on subsets $D_N, D_N' \subset \mathbb{R}^N$ and $D_n \subset \mathbb{R}^n$, that is, $F : D_N \times D_n \longrightarrow D_N'$ and $h : D_N' \to \mathbb{R}^d$. There are two properties of reservoir systems that will be crucial in what follows:

- **Existence of solutions** property: this property holds when for each $\mathbf{z} \in (D_n)^{\mathbb{Z}}$ there exists an element $\mathbf{x} \in (D_N)^{\mathbb{Z}}$ that satisfies the relation (1.1)(a) for each $t \in \mathbb{Z}$.
- **Uniqueness of solutions** or **echo state** property **(ESP)**: it holds when the system has the existence of solutions property and, additionally, these solutions are unique.

The echo state property has deserved much attention in the context of echo state networks (Buehner & Young, 2006; Gallicchio & Micheli, 2017; Jaeger, 2010; Jaeger & Haas, 2004; Manjunath & Jaeger, 2013; Wainrib & Galtier, 2016; Yildiz, Jaeger, & Kiebel,

2012; Zhang, Miller, & Wang, 2012). We emphasize that these two properties are genuine conditions that are not automatically satisfied by all RC systems. Later on in the paper, Theorem 3.1 specifies sufficient conditions for them to hold.

The combination of the existence of solutions with the axiom of choice allows us to associate filters $U^F : (D_n)^{\mathbb{Z}} \longrightarrow (D_N)^{\mathbb{Z}}$ to each RC system with that property via the reservoir map and (1.1)(a), that is, $U^F(\mathbf{z})_t := \mathbf{x}_t \in \mathbb{R}^N$, for all $t \in \mathbb{Z}$. We will denote by $U_h^F : (D_n)^{\mathbb{Z}} \longrightarrow (D_d)^{\mathbb{Z}}$ the corresponding filter determined by the entire reservoir system, that is, $U_h^F(\mathbf{z})_t = h\left(U^F(\mathbf{z})_t\right) := \mathbf{y}_t \in \mathbb{R}^d$. U_h^F is said to be a **reservoir filter** or a **response map** associated to the RC system (1.1)(a)–(1.1)(b). The filters U^F and U_h^F are causal by construction. A unique reservoir filter can be associated to a reservoir system when the echo state property holds. We warn the reader that reservoir filters appear in the literature only in the presence of the ESP; that is why we sometimes make the distinction between those that come from reservoir systems that do and do not satisfy the ESP by referring to them as **reservoir filters** and **generalized reservoir filters**, respectively.

In the systems theory literature, the RC equations (1.1)(a)–(1.1)(b) are referred to as the **state-variable** or the **internal representation** point of view and associated filters as the **external representation** of the system.

The next proposition shows that in the presence of the ESP, reservoir filters are not only causal but also time-invariant. In that situation we can hence associate to U_h^F a **reservoir functional** $H_h^F : (D_n)^{\mathbb{Z}-} \longrightarrow \mathbb{R}^d$ determined by $H_h^F := H_{U_h^F}$.

Proposition 2.1. Let $D_N \subset \mathbb{R}^N, D_n \subset \mathbb{R}^n$, and $F : D_N \times D_n \longrightarrow D_N$ be a reservoir map that satisfies the echo state property for all the elements in $(D_n)^{\mathbb{Z}}$. Then, the corresponding filter $U^F : (D_n)^{\mathbb{Z}} \longrightarrow (D_N)^{\mathbb{Z}}$ is causal and time-invariant.

We emphasize that, as it can be seen in the proof in the appendix, it is the autonomous character of the reservoir map that guarantees time-invariance in the previous proposition. An explicit time dependence on time in that map would spoil that conclusion.

Reservoir system morphisms. Let $N_1, N_2, n, d \in \mathbb{N}$ and let $F_1 : D_{N_1} \times D_n \longrightarrow D_{N_1}, h_1 : D_{N_1} \to \mathbb{R}^d$ and $F_2 : D_{N_2} \times D_n \longrightarrow D_{N_2}, h_2 : D_{N_2} \to \mathbb{R}^d$ be two reservoir systems. We say that a map $f : D_{N_1} \longrightarrow D_{N_2}$ is a morphism between the two systems when it satisfies the following two properties:

(i) Reservoir equivariance: $f(F_1(\mathbf{x}_1, \mathbf{z})) = F_2(f(\mathbf{x}_1), \mathbf{z})$, for all $\mathbf{x}_1 \in D_{N_1}$, and $\mathbf{z} \in D_n$.
(ii) Readout invariance: $h_1(\mathbf{x}_1) = h_2(f(\mathbf{x}_1))$, for all $\mathbf{x}_1 \in D_{N_1}$.

When the map f has an inverse and it is also a morphism between the systems determined by the pairs (F_2, h_2) and (F_1, h_1) we say that f is a **system isomorphism** and that the systems (F_1, h_1) and (F_2, h_2) are **isomorphic**. Given a system $F_1 : D_{N_1} \times D_n \longrightarrow D_{N_1}, h_1 : D_{N_1} \to \mathbb{R}^d$ and a bijection $f : D_{N_1} \longrightarrow D_{N_2}$, the map f is a system isomorphism with respect to the system $F_2 : D_{N_2} \times D_n \longrightarrow D_{N_2}, h_2 : D_{N_2} \to \mathbb{R}^d$ defined by

$$F_2(\mathbf{x}_2, \mathbf{z}) := f\left(F_1(f^{-1}(\mathbf{x}_2), \mathbf{z})\right), \quad \text{for all} \quad \mathbf{x}_2 \in D_{N_2}, \mathbf{z} \in D_n, \quad (2.4)$$

$$h_2(\mathbf{x}_2) := h_1(f^{-1}(\mathbf{x}_2)), \quad \text{for all} \quad \mathbf{x}_2 \in D_{N_2}. \quad (2.5)$$

The proof of the following statement is a straightforward consequence of the definitions.

Proposition 2.2. Let $F_1 : D_{N_1} \times D_n \longrightarrow D_{N_1}, h_1 : D_{N_1} \to \mathbb{R}^d$ and $F_2 : D_{N_2} \times D_n \longrightarrow D_{N_2}, h_2 : D_{N_2} \to \mathbb{R}^d$ be two reservoir systems. Let $f : D_{N_1} \longrightarrow D_{N_2}$ be a morphism between them. Then:

(i) If $\mathbf{x}^1 \in \left(D_{N_1}\right)^{\mathbb{Z}}$ is a solution for the reservoir map F_1 associated to the input $\mathbf{z} \in (D_n)^{\mathbb{Z}}$, then the sequence $\mathbf{x}^2 \in \left(D_{N_2}\right)^{\mathbb{Z}}$ defined by $\mathbf{x}_t^2 := f\left(\mathbf{x}_t^1\right), t \in \mathbb{Z}$, is a solution for the reservoir map F_2 associated to the same input.
(ii) If $U_{h_1}^{F_1}$ is a generalized reservoir filter for the system determined by the pair (F_1, h_1) then it is also a reservoir filter for the system (F_2, h_2). Equivalently, given a generalized reservoir filter $U_{h_1}^{F_1}$ determined by (F_1, h_1), there exists a generalized reservoir filter $U_{h_2}^{F_2}$ determined by (F_2, h_2) such that $U_{h_1}^{F_1} = U_{h_2}^{F_2}$.
(iii) If f is a system isomorphism then the implications in the previous two points are reversible.

2.3. Continuity and the fading memory property

In agreement with the notation introduced in the previous section, in the following paragraphs the symbol $U : (D_n)^{\mathbb{Z}-} \longrightarrow (D_N)^{\mathbb{Z}-}$ stands for a causal and time-invariant filter or, strictly speaking, for the restriction of $U : (D_n)^{\mathbb{Z}} \longrightarrow (D_N)^{\mathbb{Z}}$ to $\mathbb{Z}_-$, see (2.3); $H_U : (D_n)^{\mathbb{Z}-} \longrightarrow D_N$ is the associated functional, for some $D_N \subset \mathbb{R}^N$ and $D_n \subset \mathbb{R}^n$. Analogously, U_H is the filter associated to a given functional H.

Definition 2.3 (*Continuous Filters and Functionals*). Let $D_N \subset \mathbb{R}^N$ and $D_n \subset \mathbb{R}^n$ be bounded subsets such that $(D_n)^{\mathbb{Z}-} \subset \ell_-^{\infty}(\mathbb{R}^n)$ and $(D_N)^{\mathbb{Z}-} \subset \ell_-^{\infty}(\mathbb{R}^N)$. A causal and time-invariant filter $U : (D_n)^{\mathbb{Z}-} \longrightarrow (D_N)^{\mathbb{Z}-}$ is called **continuous** when it is a continuous map between the metric spaces $\left((D_n)^{\mathbb{Z}-}, \|\cdot\|_{\infty}\right)$ and $\left((D_N)^{\mathbb{Z}-}, \|\cdot\|_{\infty}\right)$. An analogous prescription can be used to define **continuous functionals** $H : \left((D_n)^{\mathbb{Z}-}, \|\cdot\|_{\infty}\right) \longrightarrow (D_N, \|\cdot\|)$.

The following proposition shows that when filters are causal and time-invariant, their continuity can be read out of their corresponding functionals and vice versa.

Proposition 2.4. Let $D_n \subset \mathbb{R}^n$ and $D_N \subset \mathbb{R}^N$ be such that $(D_n)^{\mathbb{Z}-} \subset \ell_-^{\infty}(\mathbb{R}^n)$ and $(D_N)^{\mathbb{Z}-} \subset \ell_-^{\infty}(\mathbb{R}^N)$. Let $U : (D_n)^{\mathbb{Z}-} \longrightarrow (D_N)^{\mathbb{Z}-}$ be a causal and time-invariant filter, $H : (D_n)^{\mathbb{Z}-} \longrightarrow D_N$ a functional, and let Φ and Ψ be the maps defined in the previous section. Then, if the filter U is continuous then so is the associated functional $\Psi(U) =: H_U$. Conversely, if H is continuous then so is $\Phi(H) =: U_H$. Define now the vector spaces

$$\mathbb{F}_{(D_n)^{\mathbb{Z}-}}^{\infty} := \Big\{ U : (D_n)^{\mathbb{Z}-} \longrightarrow \ell_-^{\infty}(\mathbb{R}^N) \mid$$
$$U \text{ is causal, time-invariant, and continuous}\Big\}, \quad (2.6)$$

$$\mathbb{H}_{(D_n)^{\mathbb{Z}-}}^{\infty} := \left\{ H : (D_n)^{\mathbb{Z}-} \longrightarrow \mathbb{R}^N \mid H \text{ is continuous} \right\}. \quad (2.7)$$

The previous statements guarantee that the maps Ψ and Φ restrict to the maps (that we denote with the same symbol) $\Psi : \mathbb{F}_{(D_n)^{\mathbb{Z}-}}^{\infty} \longrightarrow \mathbb{H}_{(D_n)^{\mathbb{Z}-}}^{\infty}$ and $\Phi : \mathbb{H}_{(D_n)^{\mathbb{Z}-}}^{\infty} \longrightarrow \mathbb{F}_{(D_n)^{\mathbb{Z}-}}^{\infty}$ that are linear isomorphisms and are inverses of each other.

Definition 2.5 (*Fading Memory Filters and Functionals*). Let $w : \mathbb{N} \longrightarrow (0, 1]$ be a weighting sequence and let $D_N \subset \mathbb{R}^N$ and $D_n \subset \mathbb{R}^n$ be such that $(D_n)^{\mathbb{Z}-} \subset \ell_-^w(\mathbb{R}^n)$ and $(D_N)^{\mathbb{Z}-} \subset \ell_-^w(\mathbb{R}^N)$. We say that a causal and time-invariant filter $U : (D_n)^{\mathbb{Z}-} \longrightarrow (D_N)^{\mathbb{Z}-}$ (respectively, a functional $H : (D_n)^{\mathbb{Z}-} \longrightarrow D_N$) satisfies the **fading memory property (FMP)** with respect to the sequence w when it is a continuous map between the metric spaces $\left((D_n)^{\mathbb{Z}-}, \|\cdot\|_w\right)$ and $\left((D_N)^{\mathbb{Z}-}, \|\cdot\|_w\right)$ (respectively, $\left((D_n)^{\mathbb{Z}-}, \|\cdot\|_w\right)$ and $(D_N, \|\cdot\|)$). If the weighting sequence w is such that $w_t = \lambda^t$, for some $\lambda \in (0, 1)$ and all $t \in \mathbb{N}$, then U is said to have the λ-**exponential fading memory**

property. We define the sets

$$\mathbb{F}^w_{(D_n)^{\mathbb{Z}_-},(D_N)^{\mathbb{Z}_-}} := \big\{ U : (D_n)^{\mathbb{Z}_-} \longrightarrow (D_N)^{\mathbb{Z}_-} \mid$$
$$U \text{ causal, time-invariant, and FMP w.r.t. } w \big\}, \tag{2.8}$$

$$\mathbb{H}^w_{(D_n)^{\mathbb{Z}_-},(D_N)^{\mathbb{Z}_-}} := \big\{ H : (D_n)^{\mathbb{Z}_-} \longrightarrow D_N \mid$$
$$H \text{ is FMP w.r.t. } w \big\}. \tag{2.9}$$

These definitions can be extended by replacing the product set $(D_N)^{\mathbb{Z}_-}$ by any subset of $\ell^w_-(\mathbb{R}^N)$ that is not necessarily a product space. In particular, we define the sets

$$\mathbb{F}^w_{(D_n)^{\mathbb{Z}_-}} := \big\{ U : (D_n)^{\mathbb{Z}_-} \longrightarrow \ell^w_-(\mathbb{R}^N) \mid$$
$$U \text{ is causal, time-invariant, and FMP w.r.t. } w \big\}, \tag{2.10}$$

$$\mathbb{H}^w_{(D_n)^{\mathbb{Z}_-}} := \big\{ H : (D_n)^{\mathbb{Z}_-} \longrightarrow \mathbb{R}^N \mid$$
$$H \text{ is FMP w.r.t. } w \big\}. \tag{2.11}$$

Definitions 2.3 and 2.5 can be easily reformulated in terms of more familiar ϵ-δ-type criteria, as they were introduced in Boyd and Chua (1985). For example, the continuity of the functional $H : (D_n)^{\mathbb{Z}_-} \longrightarrow D_N$ is equivalent to stating that for any $\mathbf{z} \in (D_n)^{\mathbb{Z}_-}$ and any $\epsilon > 0$, there exists a $\delta(\epsilon) > 0$ such that for any $\mathbf{s} \in (D_n)^{\mathbb{Z}_-}$ that satisfies that

$$\|\mathbf{z} - \mathbf{s}\|_\infty = \sup_{t \in \mathbb{Z}_-}\{\|\mathbf{z}_t - \mathbf{s}_t\|\} < \delta(\epsilon),$$
$$\text{then} \quad \|H_U(\mathbf{z}) - H_U(\mathbf{s})\| < \epsilon. \tag{2.12}$$

Regarding the fading memory property, it suffices to replace the implication in (2.12) by

$$\|\mathbf{z} - \mathbf{s}\|_w = \sup_{t \in \mathbb{Z}_-}\{\|\mathbf{z}_t - \mathbf{s}_t\| w_{-t}\} < \delta(\epsilon),$$
$$\text{then} \quad \|H_U(\mathbf{z}) - H_U(\mathbf{s})\| < \epsilon. \tag{2.13}$$

A very important part of the results that follow concern ***uniformly bounded*** families of sequences, that is, subsets of $(\mathbb{R}^n)^{\mathbb{Z}_-}$ of the form

$$K_M := \Big\{ \mathbf{z} \in \big(\mathbb{R}^n\big)^{\mathbb{Z}_-} \mid \|\mathbf{z}_t\| \leq M \quad \text{for all} \quad t \in \mathbb{Z}_- \Big\},$$
$$\text{for some } M > 0. \tag{2.14}$$

It is straightforward to show that $K_M \subset \ell^\infty_-(\mathbb{R}^n) \subset \ell^w_-(\mathbb{R}^n)$, for all $M > 0$ and any weighting sequence w. A very useful fact is that the relative topology induced by $(\ell^w_-(\mathbb{R}^n), \|\cdot\|_w)$ in K_M coincides with the one induced by the product topology in $(\mathbb{R}^n)^{\mathbb{Z}_-}$. This is a consequence of the following result that is a slight generalization of Munkres (2014, Theorem 20.5). A proof is provided in Appendix A.3 for the sake of completeness.

Theorem 2.6. *Let* $\|\cdot\| : \mathbb{R}^n \longrightarrow [0, \infty)$ *be a norm in* $\mathbb{R}^n$, $M > 0$, *and let* $w : \mathbb{N} \longrightarrow (0, 1]$ *be a weighting sequence. Let* $\bar{d}_M(\mathbf{a}, \mathbf{b}) := \min\{\|\mathbf{a} - \mathbf{b}\|, M\}$, $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$, *be a bounded metric on* $\mathbb{R}^n$ *and define the* w-**weighted metric** D^M_w *on* $(\mathbb{R}^n)^{\mathbb{Z}_-}$ *as*

$$D^M_w(\mathbf{x}, \mathbf{y}) := \sup_{t \in \mathbb{Z}_-}\big\{\bar{d}_M(\mathbf{x}_t, \mathbf{y}_t) w_{-t}\big\}, \quad \mathbf{x}, \mathbf{y} \in (\mathbb{R}^n)^{\mathbb{Z}_-}. \tag{2.15}$$

Then D^M_w *is a metric that induces the product topology on* $(\mathbb{R}^n)^{\mathbb{Z}_-}$. *The space* $(\mathbb{R}^n)^{\mathbb{Z}_-}$ *is complete relative to this metric.*

An important consequence that can be drawn from this theorem is that all the weighted norms induce the same topology on the subspaces formed by uniformly bounded sequences. An obvious consequence of this fact is that continuity with respect to this topology can be defined without the help of weighting sequences or, equivalently, filters or functionals with uniformly bounded inputs that have the fading memory with respect to a weighting sequence, have the same feature with respect to any

other weighting sequence. We make this more specific in the following statements.

Corollary 2.7. *Let* $M > 0$ *and let* $K_M := \big\{ \mathbf{z} \in (\mathbb{R}^n)^{\mathbb{Z}_-} \mid \|\mathbf{z}_t\| \leq M \text{ for all } t \in \mathbb{Z}_- \big\}$ *be a subset of* $(\mathbb{R}^n)^{\mathbb{Z}_-}$ *formed by uniformly bounded sequences. Let* $w : \mathbb{N} \longrightarrow (0, 1]$ *be an arbitrary weighting sequence. Then, the metric induced by the weighted norm* $\|\cdot\|_w$ *on* K_M *coincides with* D^{2M}_w. *Moreover, since* D^{2M}_w *induces the product topology on* $K_M = \big(\overline{B_{\|\cdot\|}(\mathbf{0}, M)}\big)^{\mathbb{Z}_-}$, *we conclude that all the weighted norms induce the same topology on* K_M. *We recall that* $\overline{B_{\|\cdot\|}(\mathbf{0}, M)}$ *is the closure of the ball with radius* M *centered at the origin, with respect to the norm* $\|\cdot\|$ *in* $\mathbb{R}^n$. *The same conclusion holds when instead of* K_M *we consider the set* $(D_n)^{\mathbb{Z}_-}$, *with* D_n *a compact subset of* $\mathbb{R}^n$.

Theorem 2.6 can also be used to give a quick alternative proof in discrete time to an important compactness result originally formulated in Boyd and Chua (1985, Lemma 1) for continuous time and, later on, in Grigoryeva and Ortega (2018a) for discrete time. The next corollary contains an additional completeness statement.

Corollary 2.8. *Let* K_M *be the set of uniformly bounded sequences, defined as in* (2.14), *and let* $w : \mathbb{N} \longrightarrow (0, 1]$ *be a weighting sequence. Then,* $(K_M, \|\cdot\|_w)$ *is a compact, complete, and convex subset of the Banach space* $(\ell^w_-(\mathbb{R}^n), \|\cdot\|_w)$. *The compactness and the completeness statements also hold when instead of* K_M *we consider the set* $(D_n)^{\mathbb{Z}_-}$, *with* D_n *a compact subset of* $\mathbb{R}^n$; *if* D_n *is additionally convex then the convexity of* $(D_n)^{\mathbb{Z}_-}$ *is also guaranteed.*

It is important to point out that the coincidence between the product topology and the topologies induced by weighted norms that we described in Corollary 2.7 only occurs for uniformly bounded sets of the type introduced in (2.14). As we state in the next result, the norm topology in $\ell^w_-(\mathbb{R}^n)$ is strictly finer than the one induced by the product topology in $(\mathbb{R}^n)^{\mathbb{Z}_-}$.

Proposition 2.9. *Let* $w : \mathbb{N} \longrightarrow (0, 1]$ *be a weighting sequence and let* $(\ell^w_-(\mathbb{R}^n), \|\cdot\|_w)$ *be the Banach space constructed using the corresponding weighted norm on the space of left infinite sequences with elements in* $\mathbb{R}^n$. *The norm topology in* $\ell^w_-(\mathbb{R}^n)$ *is strictly finer than the subspace topology induced by the product topology in* $(\mathbb{R}^n)^{\mathbb{Z}_-}$ *on* $\ell^w_-(\mathbb{R}^n) \subset (\mathbb{R}^n)^{\mathbb{Z}_-}$.

The results that we just proved imply an elementary property of the sets that we defined in (2.8)–(2.9) and (2.10)–(2.11) that we state in the following lemma.

Lemma 2.10. *Let* $M > 0$ *and let* w *be a weighting sequence. Let* $U : K_M \longrightarrow \ell^w_-(\mathbb{R}^N)$ *(respectively,* $H : K_M \longrightarrow \mathbb{R}^N$) *be an element of* $\mathbb{F}^w_{K_M}$ *(respectively,* $\mathbb{H}^w_{K_M}$). *Then there exists* $L > 0$ *such that* $U(K_M) \subset K_L$ *(respectively,* $H(K_M) \subset \overline{B_{\|\cdot\|}(\mathbf{0}, L)}$) *and we can hence conclude that* $U \in \mathbb{F}^w_{K_M, K_L}$ *(respectively,* $H \in \mathbb{H}^w_{K_M, K_L}$). *Conversely, the inclusion* $\mathbb{F}^w_{K_M, K_L} \subset \mathbb{F}^w_{K_M}$ *(respectively,* $\mathbb{H}^w_{K_M, K_L} \subset \mathbb{H}^w_{K_M}$) *holds true for any* $M > 0$. *The sets* $\mathbb{F}^w_{K_M}$ *and* $\mathbb{H}^w_{K_M}$ *are vector spaces.*

The next proposition spells out how the fading memory property is independent of the weighting sequence that is used to define it, which shows its intrinsically topological nature. A conceptual consequence of this fact is that the fading memory property does not contain any information about the rate at which systems that have it "forget" inputs. A similar statement in the continuous time setup has been formulated in Sandberg (2003). Additionally, there is a bijection between FMP filters and functionals.

Proposition 2.11. *Let* $K_M \subset (\mathbb{R}^n)^{\mathbb{Z}_-}$ *and* $K_L \subset \big(\mathbb{R}^N\big)^{\mathbb{Z}_-}$ *be subsets of uniformly bounded sequences defined as in* (2.14) *and let* $w : \mathbb{N} \longrightarrow (0, 1]$ *be a weighting sequence. Let* $U : K_M \longrightarrow K_L$ *be a causal and time-invariant filter and let* $H : K_M \longrightarrow \overline{B_{\|\cdot\|}(\mathbf{0}, L)}$ *be a functional. Then:*

(i) If U (respectively, H) has the fading memory property with respect to the weighting sequence w, then it has the same property with respect to any other weighting sequence. In particular, this implies that

$$\mathbb{F}^w_{K_M,K_L} = \mathbb{F}^{w'}_{K_M,K_L} \quad and \quad \mathbb{H}^w_{K_M,K_L} = \mathbb{H}^{w'}_{K_M,K_L},$$
for any weighting sequence w'.

In what follows we just say that U (respectively, H) has the fading memory property and denote

$$\mathbb{F}^{\mathrm{FMP}}_{K_M,K_L} := \mathbb{F}^w_{K_M,K_L} \quad and \quad \mathbb{H}^{\mathrm{FMP}}_{K_M,K_L} := \mathbb{H}^w_{K_M,K_L},$$
for any weighting sequence w.

The same statement holds true for the vector spaces $\mathbb{F}^w_{K_M}$ and $\mathbb{H}^w_{K_M}$, that will be denoted in the sequel by $\mathbb{F}^{\mathrm{FMP}}_{K_M}$ and $\mathbb{H}^{\mathrm{FMP}}_{K_M}$, respectively.

(ii) Let Φ and Ψ be the maps defined in the previous section. Then, if the filter U has the fading memory property then so does the associated functional $\Psi(U) =: H_U$. Analogously, if H has the fading memory property, then so does $\Phi(H) =: U_H$. This implies that the maps Ψ and Φ restrict to maps (that we denote with the same symbols) $\Psi : \mathbb{F}^{\mathrm{FMP}}_{K_M,K_L} \longrightarrow \mathbb{H}^{\mathrm{FMP}}_{K_M,K_L}$ and $\Phi : \mathbb{H}^{\mathrm{FMP}}_{K_M,K_L} \longrightarrow \mathbb{F}^{\mathrm{FMP}}_{K_M,K_L}$ that are inverses of each other. The same applies to $\Psi : \mathbb{F}^{\mathrm{FMP}}_{K_M} \longrightarrow \mathbb{H}^{\mathrm{FMP}}_{K_M}$ and $\Phi : \mathbb{H}^{\mathrm{FMP}}_{K_M} \longrightarrow \mathbb{F}^{\mathrm{FMP}}_{K_M}$ that, in this case, are linear isomorphisms.

The same statements can be formulated when instead of K_M and K_L we consider the sets $(D_n)^{\mathbb{Z}_-}$ and $(D_N)^{\mathbb{Z}_-}$, with D_n and D_N compact subsets of $\mathbb{R}^n$ and $\mathbb{R}^N$, respectively.

In the conditions of the previous proposition, the vector spaces $\mathbb{F}^{\mathrm{FMP}}_{K_M}$ and $\mathbb{H}^{\mathrm{FMP}}_{K_M}$ can be endowed with a norm. More specifically, let $U : K_M \longrightarrow \ell^w_-(\mathbb{R}^n)$ be a filter and let $H : K_M \longrightarrow \mathbb{R}^N$ be a functional that have the FMP. Define:

$$\|\|U\|\|_\infty := \sup_{\mathbf{z} \in K_M} \{\|U(\mathbf{z})\|_\infty\} = \sup_{\mathbf{z} \in K_M} \left\{ \sup_{t \in \mathbb{Z}_-} \{\|U(\mathbf{z})_t\|\} \right\}, \tag{2.16}$$

$$\|\|H\|\|_\infty := \sup_{\mathbf{z} \in K_M} \{\|H(\mathbf{z})\|\}. \tag{2.17}$$

The compactness of $(K_M, \|\cdot\|_w)$ guaranteed by Corollary 2.8 and the fact that by Lemma 2.10 U and H map into uniformly bounded sequences and a compact subspace of $\mathbb{R}^N$, respectively, ensures that the values in (2.16) and (2.17) are finite, which makes $(\mathbb{F}^{\mathrm{FMP}}_{K_M}, \|\cdot\|_\infty)$ and $(\mathbb{H}^{\mathrm{FMP}}_{K_M}, \|\cdot\|_\infty)$ into normed spaces that, as we will see in the next result, are linearly homeomorphic. For any $L > 0$ these norms restrict to the spaces $\mathbb{F}^{\mathrm{FMP}}_{K_M,K_L}$ and $\mathbb{H}^{\mathrm{FMP}}_{K_M,K_L}$, which are in general not linear but become nevertheless metric spaces.

Proposition 2.12. The linear isomorphism $\Psi : \left(\mathbb{F}^{\mathrm{FMP}}_{K_M}, \|\cdot\|_\infty \right) \longrightarrow \left(\mathbb{H}^{\mathrm{FMP}}_{K_M}, \|\cdot\|_\infty \right)$ and its inverse Φ satisfy that

$$\|\|\Psi(U)\|\|_\infty = \|\|U\|\|_\infty, \quad for \ any \quad U \in \mathbb{F}^{\mathrm{FMP}}_{K_M}. \tag{2.18}$$

$$\|\|\Phi(H)\|\|_\infty = \|\|H\|\|_\infty, \quad for \ any \quad H \in \mathbb{H}^{\mathrm{FMP}}_{K_M}. \tag{2.19}$$

These equalities imply that these two maps are continuous linear bijections and hence the spaces $\left(\mathbb{F}^{\mathrm{FMP}}_{K_M}, \|\cdot\|_\infty \right)$ and $\left(\mathbb{H}^{\mathrm{FMP}}_{K_M}, \|\cdot\|_\infty \right)$ are linearly homeomorphic. Equivalently, the following diagram commutes and all the maps in it are linear and continuous

$$
\begin{array}{ccc}
\left(\mathbb{F}^{\mathrm{FMP}}_{K_M}, \|\cdot\|_\infty \right) & \xrightarrow{\ \Psi\ } & \left(\mathbb{H}^{\mathrm{FMP}}_{K_M}, \|\cdot\|_\infty \right) \\
\mathrm{Id}_{\mathbb{F}^{\mathrm{FMP}}_{K_M}} \uparrow & & \downarrow \mathrm{Id}_{\mathbb{H}^{\mathrm{FMP}}_{K_M}} \\
\left(\mathbb{F}^{\mathrm{FMP}}_{K_M}, \|\cdot\|_\infty \right) & \xleftarrow{\ \Phi\ } & \left(\mathbb{H}^{\mathrm{FMP}}_{K_M}, \|\cdot\|_\infty \right).
\end{array}
$$

For any $L > 0$, the inclusions $\left(\mathbb{F}^{\mathrm{FMP}}_{K_M,K_L}, \|\cdot\|_\infty \right) \hookrightarrow \left(\mathbb{F}^{\mathrm{FMP}}_{K_M}, \|\cdot\|_\infty \right)$ and $\left(\mathbb{H}^{\mathrm{FMP}}_{K_M,K_L}, \|\cdot\|_\infty \right) \hookrightarrow \left(\mathbb{H}^{\mathrm{FMP}}_{K_M}, \|\cdot\|_\infty \right)$ (see Lemma 2.10) are continuous and so are the restricted bijections (that we denote with the same symbols) $\Psi : (\mathbb{F}^{\mathrm{FMP}}_{K_M,K_L}, \|\cdot\|_\infty) \longrightarrow (\mathbb{H}^{\mathrm{FMP}}_{K_M,K_L}, \|\cdot\|_\infty)$ and $\Phi : (\mathbb{H}^{\mathrm{FMP}}_{K_M,K_L}, \|\cdot\|_\infty) \longrightarrow (\mathbb{F}^{\mathrm{FMP}}_{K_M,K_L}, \|\cdot\|_\infty)$ that are inverses of each other. The last statement is a consequence of the following equalities:

$$\|\|\Psi(U_1) - \Psi(U_2)\|\|_\infty = \|\|U_1 - U_2\|\|_\infty,$$
$$for \ any \quad U_1, U_2 \in \mathbb{F}^{\mathrm{FMP}}_{K_M,K_L}, \tag{2.20}$$

$$\|\|\Phi(H_1) - \Phi(H_2)\|\|_\infty = \|\|H_1 - H_2\|\|_\infty,$$
$$for \ any \quad H_1, H_2 \in \mathbb{H}^{\mathrm{FMP}}_{K_M,K_L}. \tag{2.21}$$

The same statements can be formulated when instead of K_M and K_L we consider the sets $(D_n)^{\mathbb{Z}_-}$ and $(D_N)^{\mathbb{Z}_-}$, with D_n and D_N compact subsets of $\mathbb{R}^n$ and $\mathbb{R}^N$, respectively.

It is not difficult to see that as a consequence of Proposition 2.12 a weighted version of the norm (2.16) in the space of filters with respect to a weighting sequence w is identical to $\|\cdot\|_\infty$ for any w.

3. Internal approximation of reservoir filters

This section characterizes situations under which reservoir filters can be uniformly approximated by finding uniform approximants for the corresponding reservoir systems. Such a statement is part of the next theorem that also identifies criteria for the availability of the echo state and the fading memory properties (recall that we used the acronyms ESP and FMP, respectively). As it was already mentioned, a reservoir system has the ESP when it has a unique semi-infinite solution for each semi-infinite input. We also recall that in the presence of uniformly bounded inputs, as it was shown in Section 2.3, the FMP amounts to the continuity of a reservoir filter with respect to the product topologies on the input and output spaces. The completeness and compactness of those spaces established in Corollary 2.8 allows us to use various fixed point theorems to show that solutions for reservoir systems exist under very weak hypotheses and that for contracting and continuous reservoir maps (we define this below) these solutions are unique and depend continuously on the inputs. Said differently, *contracting continuous reservoir maps induce reservoir filters that automatically have the echo state and the fading memory properties.*

Theorem 3.1. Let $K_M \subset (\mathbb{R}^n)^{\mathbb{Z}_-}$ and $K_L \subset (\mathbb{R}^N)^{\mathbb{Z}_-}$ be subsets of uniformly bounded sequences defined as in (2.14) and let $F : \overline{B_{\|\cdot\|}(\mathbf{0}, L)} \times \overline{B_{\|\cdot\|}(\mathbf{0}, M)} \longrightarrow \overline{B_{\|\cdot\|}(\mathbf{0}, L)}$ be a continuous reservoir map.

(i) Existence of solutions: for each $\mathbf{z} \in K_M$ there exists a $\mathbf{x} \in K_L$ (not necessarily unique) that solves the reservoir equation associated to F, that is,

$$\mathbf{x}_t = F(\mathbf{x}_{t-1}, \mathbf{z}_t), \quad for \ all \ t \in \mathbb{Z}_-.$$

(ii) Uniqueness and continuity of solutions (ESP and FMP): suppose that the reservoir map F is a contraction, that is, there exists $0 < r < 1$ such that for all $\mathbf{u}, \mathbf{v} \in B_{\|\cdot\|}(\mathbf{0}, L)$, $\mathbf{z} \in B_{\|\cdot\|}(\mathbf{0}, M)$, one has

$$\|F(\mathbf{u}, \mathbf{z}) - F(\mathbf{v}, \mathbf{z})\| \le r \|\mathbf{u} - \mathbf{v}\|.$$

Then, the reservoir system associated to F has the echo state property. Moreover, this system has a unique associated causal and time-invariant filter $U_F : K_M \longrightarrow K_L$ that has the fading memory property, that is, $U_F \in \mathbb{F}^{\mathrm{FMP}}_{K_M,K_L}$. The set $U_F(K_M)$ of accessible states of the filter U_F is compact.

(iii) Internal approximation property: let $F_1, F_2 : \overline{B_{\|\cdot\|}(\mathbf{0}, L)} \times \overline{B_{\|\cdot\|}(\mathbf{0}, M)} \longrightarrow \overline{B_{\|\cdot\|}(\mathbf{0}, L)}$ be two continuous reservoir maps such that F_1 is a contraction with constant $0 < r < 1$ and F_2 has the existence of solutions property. Let $U_{F_1}, U_{F_2} : K_M \longrightarrow K_L$ be the corresponding filters (if F_2 does not have the ESP, then U_{F_2} is just a generalized filter). Then, for any $\epsilon > 0$, we have that

$$\|F_1 - F_2\|_\infty < \delta(\epsilon)$$
$$:= (1 - r)\epsilon \quad \text{implies that} \quad \left\|\left\|U_{F_1} - U_{F_2}\right\|\right\|_\infty < \epsilon. \quad (3.1)$$

Part **(i)** *also holds true when instead of K_M and K_L we consider the sets $(D_n)^{\mathbb{Z}_-}$ and $(D_N)^{\mathbb{Z}_-}$, with D_n and D_N compact and convex subsets of $\mathbb{R}^n$ and $\mathbb{R}^N$, respectively, that replace the closed balls $\overline{B_{\|\cdot\|}(\mathbf{0}, M)}$ and $\overline{B_{\|\cdot\|}(\mathbf{0}, L)}$. The same applies to parts* **(ii)** *and* **(iii)** *but, this time, the convexity hypothesis is not needed.*

Define the set $\mathbb{K}_{K_M, K_L} := \left\{F : \overline{B_{\|\cdot\|}(\mathbf{0}, L)} \times \overline{B_{\|\cdot\|}(\mathbf{0}, M)} \longrightarrow \overline{B_{\|\cdot\|}(\mathbf{0}, L)} \mid F \text{ is a continuous contraction}\right\}$. Using the notation introduced in the previous section, the statement in (3.1) and part **(ii)** of the theorem automatically imply that the map

$$\Xi : \quad (\mathbb{K}_{K_M, K_L}, \|\cdot\|_\infty) \quad \longrightarrow \quad (\mathbb{F}_{K_M, K_L}^{\text{FMP}}, \|\cdot\|_\infty)$$
$$F \quad \longmapsto \quad U_F$$

is continuous and by Proposition 2.12, the map that associates to each $F \in \mathbb{K}_{K_M, K_L}$ the corresponding functional H_F, that is,

$$\Psi \circ \Xi : \quad (\mathbb{K}_{K_M, K_L}, \|\cdot\|_\infty) \quad \longrightarrow \quad (\mathbb{H}_{K_M, K_L}^{\text{FMP}}, \|\cdot\|_\infty)$$
$$F \quad \longmapsto \quad H_F,$$

is also continuous.

Proof of the theorem. **(i)** We start by defining, for each $\mathbf{z} \in K_M$, the map given by

$$\mathcal{F}_{\mathbf{z}} : \quad K_L \quad \longrightarrow \quad K_L$$
$$\mathbf{x} \quad \longmapsto \quad (\mathcal{F}_{\mathbf{z}}(\mathbf{x}))_t := F(\mathbf{x}_{t-1}, \mathbf{z}_t).$$

We show first that $\mathcal{F}_{\mathbf{z}}$ can be written as a product of continuous functions. Indeed:

$$\mathcal{F}_{\mathbf{z}} = \prod_{t \in \mathbb{Z}_-} F(\cdot, \mathbf{z}_t) \circ p_{t-1}(\mathbf{x}), \quad (3.2)$$

where the projections $p_t : K_L \longrightarrow \overline{B_{\|\cdot\|}(\mathbf{0}, L)}$ are given by $p_t(\mathbf{x}) = \mathbf{x}_t$. These projections are continuous when we consider in K_L the product topology. Additionally, the continuity of the reservoir F implies that $\mathcal{F}_{\mathbf{z}}$ is a product of continuous functions, which ensures that $\mathcal{F}_{\mathbf{z}}$ is itself continuous (Munkres, 2014, Theorem 19.6). Moreover, by Corollary 2.7 and 2.8, the space K_L is a compact and convex subset of the Banach space $\left(\ell_-^w(\mathbb{R}^n), \|\cdot\|_w\right)$ (see Proposition A.2), for any weighting sequence w. Schauder's Fixed Point Theorem (see Shapiro, 2016, Theorem 7.1, page 75) guarantees then that $\mathcal{F}_{\mathbf{z}}$ has at least a fixed point, that is, a point $\mathbf{x} \in K_L$ that satisfies $\mathcal{F}_{\mathbf{z}}(\mathbf{x}) = \mathbf{x}$ or, equivalently,

$$\mathbf{x}_t = F(\mathbf{x}_{t-1}, \mathbf{z}_t), \quad \text{for all } t \in \mathbb{Z}_-,$$

which implies that $\mathbf{x}$ is a solution of F for $\mathbf{z}$, as required.

Proof of part (ii) The main tool in the proof of this part is a parameter dependent version of the Contraction Fixed Point Theorem, that we include here for the sake of completeness and whose proof can be found in Sternberg (2010, Theorem 6.4.1, page 137).

Lemma. *Let (X, d_X) be a complete metric space and let Z be a metric space. Let $K : X \times Z \longrightarrow X$ be a continuous map such that for each $z \in Z$, the map $K_z : X \longrightarrow X$ given by $K_z(x) := K(x, z)$ is a contraction with a constant $0 < r < 1$ (independent of z), that is, $d_X(K(x, z), K(y, z)) \leq rd(x, y)$, for all $x, y \in X$ and all $z \in Z$. Then:*

(i) *For each $z \in Z$, the map K_z has a unique fixed point in X.*

(ii) *The map $U_K : Z \longrightarrow X$ that associates to each point $z \in Z$ the unique fixed point of K_z is continuous.*

Consider now the map

$$\mathcal{F} : \quad K_L \times K_M \quad \longrightarrow \quad K_L$$
$$(\mathbf{x}, \mathbf{z}) \quad \longmapsto \quad (\mathcal{F}(\mathbf{x}, \mathbf{z}))_t := F(\mathbf{x}_{t-1}, \mathbf{z}_t).$$

First, as we did in (3.2), it is easy to show that $\mathcal{F}$ is continuous with respect to the product topologies in K_M and K_L, by writing it down as the product of the composition of continuous functions. Second, we show that the map $\mathcal{F}$ is a contraction. Indeed, since by Corollary 2.7 we can choose an arbitrary weighting sequence to generate the product topologies in K_M and K_L, we select $w : \mathbb{N} \longrightarrow (0, 1]$ given by $w_t := \lambda^t$, with $t \in \mathbb{N}$ and $\lambda > 0$ that satisfies $0 < r < \lambda < 1$. Then, for any $\mathbf{x}, \mathbf{y} \in K_L$ and any $\mathbf{z} \in K_M$, we have

$$\|\mathcal{F}(\mathbf{x}, \mathbf{z}) - \mathcal{F}(\mathbf{y}, \mathbf{z})\|_w = \sup_{t \in \mathbb{Z}_-} \left\{\|F(\mathbf{x}_{t-1}, \mathbf{z}_t) - F(\mathbf{y}_{t-1}, \mathbf{z}_t)\| \lambda^{-t}\right\}$$
$$\leq \sup_{t \in \mathbb{Z}_-} \left\{\|\mathbf{x}_{t-1} - \mathbf{y}_{t-1}\| r\lambda^{-t}\right\},$$

where we used that F is a contraction. Now, since $0 < r < \lambda < 1$ and hence $r/\lambda < 1$, we have

$$\sup_{t \in \mathbb{Z}_-} \left\{\|\mathbf{x}_{t-1} - \mathbf{y}_{t-1}\| r\lambda^{-t}\right\} = \sup_{t \in \mathbb{Z}_-} \left\{\|\mathbf{x}_{t-1} - \mathbf{y}_{t-1}\| \lambda^{-(t-1)}\frac{r}{\lambda}\right\}$$
$$\leq \frac{r}{\lambda} \|\mathbf{x} - \mathbf{y}\|_w.$$

This shows that $\mathcal{F}$ is a family of contractions with constant $r/\lambda < 1$ that is continuously parametrized by the elements in K_M. The lemma above implies the existence of a continuous map $U_F : (K_M, \|\cdot\|_w) \longrightarrow (K_L, \|\cdot\|_w)$ that is uniquely determined by the identity

$$\mathcal{F}(U_F(\mathbf{z}), \mathbf{z}) = U_F(\mathbf{z}), \quad \text{for all } \mathbf{z} \in K_M.$$

Proposition 2.1 implies that U_F is causal and time-invariant. The set $U_F(K_M)$ of accessible states of the filter U_F is compact because it is the image of a compact set (see Corollary 2.8) by a continuous map (see Munkres, 2014, Theorem 26.5, page 166).

Proof of part (iii) Let $\mathbf{z} \in K_M$ and let $U_{F_1}(\mathbf{z})$ be the unique solution for $\mathbf{z}$ of the reservoir systems associated to F_1 available by the part **(ii)** of the theorem that we just proved. Additionally, let $U_{F_2}(\mathbf{z})$ be the value of a generalized filter associated to F_2 that exist by hypothesis. Then, for any $t \in \mathbb{Z}_-$, we have:

$$\|U_{F_1}(\mathbf{z})_t - U_{F_2}(\mathbf{z})_t\| = \left\|F_1(U_{F_1}(\mathbf{z})_{t-1}, \mathbf{z}_t) - F_2(U_{F_2}(\mathbf{z})_{t-1}, \mathbf{z}_t)\right\|$$
$$= \left\|F_1(U_{F_1}(\mathbf{z})_{t-1}, \mathbf{z}_t) - F_1(U_{F_2}(\mathbf{z})_{t-1}, \mathbf{z}_t)\right.$$
$$\left. + F_1(U_{F_2}(\mathbf{z})_{t-1}, \mathbf{z}_t) - F_2(U_{F_2}(\mathbf{z})_{t-1}, \mathbf{z}_t)\right\|$$
$$\leq \left\|F_1(U_{F_1}(\mathbf{z})_{t-1}, \mathbf{z}_t) - F_1(U_{F_2}(\mathbf{z})_{t-1}, \mathbf{z}_t)\right\|$$
$$+ \left\|F_1(U_{F_2}(\mathbf{z})_{t-1}, \mathbf{z}_t) - F_2(U_{F_2}(\mathbf{z})_{t-1}, \mathbf{z}_t)\right\|$$
$$\leq r \left\|U_{F_1}(\mathbf{z})_{t-1} - U_{F_2}(\mathbf{z})_{t-1}\right\|$$
$$+ \left\|F_1(U_{F_2}(\mathbf{z})_{t-1}, \mathbf{z}_t) - F_2(U_{F_2}(\mathbf{z})_{t-1}, \mathbf{z}_t)\right\|.$$

If we now recursively apply n times the same procedure to the first summand of this expression, we obtain that

$$\|U_{F_1}(\mathbf{z})_t - U_{F_2}(\mathbf{z})_t\| \leq r^n \|U_{F_1}(\mathbf{z})_{t-n} - U_{F_2}(\mathbf{z})_{t-n}\|$$
$$+ \left\|F_1(U_{F_2}(\mathbf{z})_{t-1}, \mathbf{z}_t) - F_2(U_{F_2}(\mathbf{z})_{t-1}, \mathbf{z}_t)\right\|$$
$$+ r \left\|F_1(U_{F_2}(\mathbf{z})_{t-2}, \mathbf{z}_{t-1}) - F_2(U_{F_2}(\mathbf{z})_{t-2}, \mathbf{z}_{t-1})\right\|$$
$$+ \cdots$$
$$+ r^{n-1} \left\|F_1(U_{F_2}(\mathbf{z})_{t-n}, \mathbf{z}_{t-(n+1)}) - F_2(U_{F_2}(\mathbf{z})_{t-n}, \mathbf{z}_{t-(n+1)})\right\|. \quad (3.3)$$

If we combine the inequality (3.3) with the hypothesis

$$\|F_1 - F_2\|_\infty = \sup_{\mathbf{x} \in \overline{B_{\|\cdot\|}(\mathbf{0}, L)}, \, \mathbf{z} \in \overline{B_{\|\cdot\|}(\mathbf{0}, M)}} \{\|F_1(\mathbf{x}, \mathbf{z}) - F_2(\mathbf{x}, \mathbf{z})\|\}$$

$$< \delta(\epsilon) := (1 - r)\epsilon,$$

we obtain

$$\|U_{F_1}(\mathbf{z}) - U_{F_2}(\mathbf{z})\|_\infty = \sup_{t \in \mathbb{Z}_-} \left\{ \|U_{F_1}(\mathbf{z})_t - U_{F_2}(\mathbf{z})_t\| \right\}$$

$$\leq 2Lr^n + (1 + \cdots + r^{n-1})\delta(\epsilon) = 2Lr^n + \frac{1 - r^n}{1 - r}\delta(\epsilon). \quad (3.4)$$

Since this inequality is valid for any $n \in \mathbb{N}$, we can take the limit $n \longrightarrow \infty$ and we obtain that

$$\|U_{F_1}(\mathbf{z}) - U_{F_2}(\mathbf{z})\|_\infty \leq \frac{\delta(\epsilon)}{1 - r} = \epsilon.$$

Additionally, as this relation is valid for any $\mathbf{z} \in K_M$, we can conclude that

$$\left\|\|U_{F_1} - U_{F_2}\|\right\|_\infty = \sup_{\mathbf{z} \in K_M} \left\{ \left\|U_{F_1}(\mathbf{z}) - U_{F_2}(\mathbf{z})\right\|_\infty \right\} \leq \epsilon,$$

as required. ∎

As a straightforward corollary of the first part of the previous theorem, it is easy to show that echo state networks always have (generalized) reservoir filters associated as well as to formulate conditions that ensure simultaneously the echo state and the fading memory properties.

We recall that a map $\sigma : \mathbb{R} \longrightarrow [-1, 1]$ is a **squashing function** if it is non-decreasing, $\lim_{x \to -\infty} \sigma(x) = -1$, and $\lim_{x \to \infty} \sigma(x) = 1$.

Corollary 3.2. *Consider echo state network given by*

$$\begin{cases} \mathbf{x}_t = \sigma\left(A\mathbf{x}_{t-1} + C\mathbf{z}_t + \boldsymbol{\zeta}\right), & \text{(a)} \\ \mathbf{y}_t = W\mathbf{x}_t, & \text{(b)} \end{cases} \quad (3.5)$$

where $C \in \mathbb{M}_{N,n}$ for some $N \in \mathbb{N}$, $\boldsymbol{\zeta} \in \mathbb{R}^N$, $A \in \mathbb{M}_{N,N}$, $W \in \mathbb{M}_{d,N}$, and the input signal $\mathbf{z} \in (D_n)^{\mathbb{Z}}$, with $D_n \subset \mathbb{R}^n$ a compact and convex subset. The function $\sigma : \mathbb{R}^N \longrightarrow [-1, 1]^N$ in (3.5)(a) is constructed by componentwise application of a squashing function that we also call σ. Then:

(i) *If the squashing function σ is continuous, then the reservoir equation (3.5)(a) has the existence of solutions property and we can hence associate to the system (3.5)(a)–(3.5)(b) a generalized reservoir filter.*

(ii) *If the squashing function σ is differentiable with Lipschitz constant $L_\sigma := \sup_{x \in \mathbb{R}} \{|\sigma'(x)|\} < \infty$ and the matrix A is such that $\|A\|_2 L_\sigma = \sigma_{\max}(A)L_\sigma < 1$, then the reservoir system (3.5)(a)–(3.5)(b) has the echo state and the fading memory properties and we can hence associate to it a unique time-invariant reservoir filter.*

*The statement in part **(i)** remains valid when $[-1, 1]^N$ is replaced by a compact and convex subset $D_N \subset [-1, 1]^N$ that is left invariant by the reservoir equation (3.5)(a), that is, $\sigma(A\mathbf{x} + C\mathbf{z} + \boldsymbol{\zeta}) \in D_N$ for any $\mathbf{x} \in D_N$ and any $\mathbf{z} \in D_n$. The same applies to part **(ii)** but only the compactness hypothesis is necessary.*

Remark 3.3. The hypothesis $\|A\|_2 L_\sigma < 1$ appears in the literature as a sufficient condition to ensure the echo state property, which has been extensively studied in the ESN literature (Buehner & Young, 2006; Jaeger, 2010; Jaeger & Haas, 2004; Manjunath & Jaeger, 2013; Wainrib & Galtier, 2016; Yildiz et al., 2012; Zhang et al., 2012). Our result shows that this condition implies automatically the fading memory property. Nevertheless, that condition is far from being sharp and has been significantly improved in Buehner and Young (2006) and Yildiz et al. (2012). We point out that the enhanced sufficient conditions for the echo state property contained in those references also imply the fading memory property via part **(ii)** of Theorem 3.1.

4. Echo state networks as universal uniform approximants

The internal approximation property that we introduced in part **(ii)** of Theorem 3.1 tells us that we can approximate any reservoir filter by finding an approximant for the reservoir system that generates it. This reduces the problem of proving a density statement in a space of operators between infinite-dimensional spaces to a space of functions with finite dimensional variables and values. This topic is the subject of many results in approximation theory, some of which we mentioned in the introduction. This strategy allows one to find simple approximating reservoir filters for any reservoir system that has the fading memory property. In the next result we use as approximating family the echo state networks that we presented in the introduction and that, as we see later on, are the natural generalizations of neural networks in a dynamic learning setup, with the important added feature that they are constructed using linear readouts. The combination of this approach with a previously obtained result (Grigoryeva & Ortega, 2018a) on the density of reservoir filters on the fading memory category allows us to prove in the next theorem that echo state networks can approximate any fading memory filter. In other words, *echo state networks are universal.*

All along this section, we use the Euclidean norm for the finite dimensional spaces, that is, for each $\mathbf{x} \in \mathbb{R}^n$, we write $\|\mathbf{x}\| := \left(\sum_{i=1}^n x_i^2\right)^{1/2}$. For any $M > 0$, the symbol $B_{\|\cdot\|}(\mathbf{0}, M)$ (respectively $\overline{B_{\|\cdot\|}(\mathbf{0}, M)}$) denotes here the open (respectively closed) balls with respect to that norm. Additionally, we set $I_n := B_{\|\cdot\|}(\mathbf{0}, 1)$.

Theorem 4.1. *Let $U : I_n^{\mathbb{Z}_-} \longrightarrow \left(\mathbb{R}^d\right)^{\mathbb{Z}_-}$ be a causal and time-invariant filter that has the fading memory property. Then, for any $\epsilon > 0$ and any weighting sequence w, there is an echo state network*

$$\begin{cases} \mathbf{x}_t = \sigma\left(A\mathbf{x}_{t-1} + C\mathbf{z}_t + \boldsymbol{\zeta}\right), & \text{(a)} \\ \mathbf{y}_t = W\mathbf{x}_t. & \text{(b)} \end{cases} \quad (4.1)$$

whose associated generalized filters $U_{\text{ESN}} : I_n^{\mathbb{Z}_-} \longrightarrow \left(\mathbb{R}^d\right)^{\mathbb{Z}_-}$ satisfy that

$$\|U - U_{\text{ESN}}\|_\infty < \epsilon. \quad (4.2)$$

In these expressions $C \in \mathbb{M}_{N,n}$ for some $N \in \mathbb{N}$, $\boldsymbol{\zeta} \in \mathbb{R}^N$, $A \in \mathbb{M}_{N,N}$, and $W \in \mathbb{M}_{d,N}$. The function $\sigma : \mathbb{R}^N \longrightarrow [-1, 1]^N$ in (4.1)(a) is constructed by componentwise application of a continuous squashing function $\sigma : \mathbb{R} \longrightarrow [-1, 1]$ that we denote with the same symbol.

When the approximating echo state network (4.1)(a)–(4.1)(b) satisfies the echo state property, then it has a unique filter U_{ESN} associated which is necessarily time-invariant. The corresponding reservoir functional $H_{\text{ESN}} : I_n^{\mathbb{Z}_-} \longrightarrow \mathbb{R}^d$ satisfies that

$$\|H_U - H_{\text{ESN}}\|_\infty < \epsilon. \quad (4.3)$$

Remark 4.2. Echo state networks are generally used in practice in the following way: the architecture parameters A, C, and $\boldsymbol{\zeta}$ are drawn at random from a given distribution and it is only the readout matrix W that is trained using a teaching signal by solving a linear regression problem. It is important to emphasize that the universality theorem that we just stated does not completely explain the empirically observed robustness of ESNs with respect to the choice of those parameters. In the context of standard feed-forward neural networks this feature has been addressed using, for example, the so called extreme learning machines (Huang, Zhu, & Siew, 2006). In dynamical setups and for ESNs this question remains an open problem that will be addressed in future works.

Proof of the theorem. As we already explained, we proceed by first approximating the filter U by one of the non-homogeneous state-affine system (SAS) reservoir filters introduced in Grigoryeva

and Ortega (2018a), and we later on show that we can approximate that reservoir filter by an echo state network like the one in (4.1)(a)–(4.1)(b).

We start by recalling that a non-homogeneous state-affine system is a reservoir system determined by the state-space transformation:

$$\begin{cases} \mathbf{x}_t = p(\mathbf{z}_t)\mathbf{x}_{t-1} + q(\mathbf{z}_t), & \text{(a)} \\ \mathbf{y}_t = W_1 \mathbf{x}_t, & \text{(b)} \end{cases} \tag{4.4}$$

where the inputs $\mathbf{z_t} \in I_n := B_{\|\cdot\|}(\mathbf{0}, 1)$, the states $\mathbf{x}_t \in \mathbb{R}^{N_1}$, for some $N_1 \in \mathbb{N}$, and $W_1 \in \mathbb{M}_{d,N_1}$. The symbols $p(\mathbf{z_t})$ and $q(\mathbf{z_t})$ stand for polynomials with matrix coefficients and degrees r and s, respectively, of the form:

$$p(\mathbf{z}) = \sum_{\substack{i_1,\ldots,i_n \in \{0,\ldots,r\} \\ i_1+\cdots+i_n \leq r}} z_1^{i_1} \cdots z_n^{i_n} A_{i_1,\ldots,i_n}, \quad A_{i_1,\ldots,i_n} \in \mathbb{M}_{N_1}, \quad \mathbf{z} \in I_n$$

$$q(\mathbf{z}) = \sum_{\substack{i_1,\ldots,i_n \in \{0,\ldots,s\} \\ i_1+\cdots+i_n \leq s}} z_1^{i_1} \cdots z_n^{i_n} B_{i_1,\ldots,i_n}, \quad B_{i_1,\ldots,i_n} \in \mathbb{M}_{N_1,1}, \quad \mathbf{z} \in I_n.$$

Let $L > 0$ and choose a real number K such that

$$0 < K < \frac{L}{L+1} < 1. \tag{4.5}$$

Consider now SAS filters that satisfy that $\max_{\mathbf{z} \in I_n} \sigma_{\max}(p(\mathbf{z})) < K$ and $\max_{\mathbf{z} \in I_n} \sigma_{\max}(q(\mathbf{z})) < K$. It can be shown (Grigoryeva & Ortega, 2018a, Proposition 3.7) that under those hypotheses, the reservoir system (4.4)(a)–(4.4)(b) has the echo state property and defines a unique causal, time-invariant, and fading memory filter $U_{W_1}^{p,q} : I_n^{\mathbb{Z}_-} \longrightarrow (\mathbb{R}^d)^{\mathbb{Z}_-}$. Moreover, Theorem 3.12 in Grigoryeva and Ortega (2018a) shows that for any $\epsilon_1 > 0$, there exists a SAS filter $U_{W_1}^{p,q}$ satisfying the hypotheses that we just discussed, for which

$$\left\| H_U - H_{W_1}^{p,q} \right\|_\infty < \epsilon_1, \tag{4.6}$$

where H_U and $H_{W_1}^{p,q}$ are the reservoir functionals associated to U and $U_{W_1}^{p,q}$, respectively. Proposition 2.12 together with this inequality implies that

$$\left\| U - U_{W_1}^{p,q} \right\|_\infty < \epsilon_1. \tag{4.7}$$

We now show that the SAS filter $U_{W_1}^{p,q}$ can be approximated by the filters generated by an echo state network. Define the map

$$F_{\text{SAS}} : \begin{array}{ccc} \overline{B_{\|\cdot\|}(\mathbf{0}, L)} \times I_n & \longrightarrow & \mathbb{R}^{N_1} \\ (\mathbf{x}, \mathbf{z}) & \longmapsto & p(\mathbf{z})\mathbf{x} + q(\mathbf{z}), \end{array} \tag{4.8}$$

with $B_{\|\cdot\|}(\mathbf{0}, L) \subset \mathbb{R}^{N_1}$ and p and q the polynomials associated to the approximating SAS filter $U_{W_1}^{p,q}$ in (4.7).

The prescription on the choice of the constant K in (4.5) has two main consequences. Firstly, the map F_{SAS} is a contraction. Indeed, for any $(\mathbf{x}, \mathbf{z}), (\mathbf{y}, \mathbf{z}) \in \overline{B_{\|\cdot\|}(\mathbf{0}, L)} \times I_n$:

$$\|F_{\text{SAS}}(\mathbf{x}, \mathbf{z}) - F_{\text{SAS}}(\mathbf{y}, \mathbf{z})\|$$
$$\leq \|p(\mathbf{z})\mathbf{x} - p(\mathbf{z})\mathbf{y}\| \leq \|p(\mathbf{z})\|_2 \|\mathbf{x} - \mathbf{y}\| \leq K \|\mathbf{x} - \mathbf{y}\|. \tag{4.9}$$

The map F_{SAS} is hence a contraction since $K < 1$ by hypothesis. Second, $\|F_{\text{SAS}}\|_\infty < L$ because by (4.5)

$$\|F_{\text{SAS}}\|_\infty = \sup_{(\mathbf{x},\mathbf{z}) \in B_{\|\cdot\|}(\mathbf{0},L) \times I_n} \{\|p(\mathbf{z})\mathbf{x} + q(\mathbf{z})\|\}$$
$$\leq \sup_{(\mathbf{x},\mathbf{z}) \in B_{\|\cdot\|}(\mathbf{0},L) \times I_n} \{\|p(\mathbf{z})\|_2 \|\mathbf{x}\| + \|q(\mathbf{z})\|\} \leq KL + K < L.$$

This implies, in particular, that the map F_{SAS} maps into $\overline{B_{\|\cdot\|}(\mathbf{0}, L)}$ and hence (4.8) can be rewritten as

$$F_{\text{SAS}} : \overline{B_{\|\cdot\|}(\mathbf{0}, L)} \times I_n \longrightarrow \overline{B_{\|\cdot\|}(\mathbf{0}, L)}.$$

Additionally, we set

$$L_1 := \|F_{\text{SAS}}\|_\infty < L. \tag{4.10}$$

The uniform density on compacta of the family of feedforward neural networks with one hidden layer proved in Cybenko (1989) and Hornik et al. (1989) guarantees that for any $\epsilon_2 > 0$, there exist $N \in \mathbb{N}$, $G \in \mathbb{M}_{N,N_1}$, $C \in \mathbb{M}_{N,n}$, $E \in \mathbb{M}_{N_1,N}$, and $\zeta \in \mathbb{R}^N$, such that the map defined by

$$F_{\text{NN}} : \begin{array}{ccc} \overline{B_{\|\cdot\|}(\mathbf{0}, L)} \times I_n & \longrightarrow & \mathbb{R}^{N_1} \\ (\mathbf{x}, \mathbf{z}) & \longmapsto & E\sigma(G\mathbf{x} + C\mathbf{z} + \zeta), \end{array} \tag{4.11}$$

satisfies that

$$\|F_{\text{NN}} - F_{\text{SAS}}\|_\infty = \sup_{\mathbf{x} \in B_{\|\cdot\|}(\mathbf{0},L), \, \mathbf{z} \in I_n} \{\|F_{\text{NN}}(\mathbf{x}, \mathbf{z}) - F_{\text{SAS}}(\mathbf{x}, \mathbf{z})\|\}$$
$$< \epsilon_2. \tag{4.12}$$

The combination of (4.12) with the reverse triangle inequality implies that $\|F_{\text{NN}}\|_\infty - \|F_{\text{SAS}}\|_\infty < \epsilon_2$ or, equivalently,

$$\|F_{\text{NN}}\|_\infty < \|F_{\text{SAS}}\|_\infty + \epsilon_2. \tag{4.13}$$

Given that $\|F_{\text{SAS}}\|_\infty = L_1 < L$, if we choose $\epsilon_2 > 0$ small enough so that $L_1 + \epsilon_2 < L$ or, equivalently,

$$\epsilon_2 < L - L_1, \tag{4.14}$$

then (4.13) guarantees that $\|F_{\text{NN}}\|_\infty < L$, which shows that F_{NN} maps into $B_{\|\cdot\|}(\mathbf{0}, L)$, that is, we can write that

$$F_{\text{NN}} : \overline{B_{\|\cdot\|}(\mathbf{0}, L)} \times I_n \longrightarrow \overline{B_{\|\cdot\|}(\mathbf{0}, L)}. \tag{4.15}$$

The continuity of the map F_{NN} and the first part of Theorem 3.1 imply that the corresponding reservoir equation has the existence of solutions property and that we can hence associate to it a (generalized) filter $U_{F_{\text{NN}}}$. At the same time, as we proved in (4.9), the map F_{SAS} is a contraction with constant $K < 1$. These facts, together with (4.12) and the internal approximation property in Theorem 3.1 allow us to conclude that the (unique) reservoir filter $U_{F_{\text{SAS}}}$ associated to the reservoir map F_{SAS} is such that

$$\left\| U_{F_{\text{NN}}} - U_{F_{\text{SAS}}} \right\|_\infty < \epsilon_2/(1-K). \tag{4.16}$$

Consider now the readout map $h_{W_1} : \mathbb{R}^{N_1} \longrightarrow \mathbb{R}^d$ given by $h_{W_1}(\mathbf{x}) := W_1 \mathbf{x}$ and let $U_{F_{\text{NN}}}^{h_{W_1}} : (I_n)^{\mathbb{Z}_-} \longrightarrow (\mathbb{R}^d)^{\mathbb{Z}_-}$ be the filter given by $U_{F_{\text{NN}}}^{h_{W_1}}(\mathbf{z})_t := W_1 U_{F_{\text{NN}}}(\mathbf{z})_t$, $t \in \mathbb{Z}_-$. Analogously, define $U_{F_{\text{SAS}}}^{h_{W_1}} : (I_n)^{\mathbb{Z}_-} \longrightarrow (\mathbb{R}^d)^{\mathbb{Z}_-}$ and notice that $U_{F_{\text{SAS}}}^{h_{W_1}} = U_{W_1}^{p,q}$. Using these observations and (4.16) we have proved that for any $\epsilon_2 > 0$ we can find a filter of the type $U_{F_{\text{NN}}}^{h_{W_1}}$ that satisfies that

$$\left\| U_{W_1}^{p,q} - U_{F_{\text{NN}}}^{h_{W_1}} \right\|_\infty \leq \|W_1\|_2 \left\| U_{F_{\text{SAS}}} - U_{F_{\text{NN}}} \right\|_\infty$$
$$< \|W_1\|_2 \, \epsilon_2/(1-K). \tag{4.17}$$

Consequently, for any $\epsilon > 0$, if we first set $\epsilon_1 = \epsilon/2$ in (4.6) and we then choose

$$\epsilon_2 := \min \left\{ \frac{\epsilon(1-K)}{2\|W_1\|_2}, \frac{L-L_1}{2} \right\}, \tag{4.18}$$

in view of (4.14) and (4.17), we can guarantee using (4.7) and (4.17) that

$$\left\| U - U_{F_{\text{NN}}}^{h_{W_1}} \right\|_\infty \leq \left\| U - U_{W_1}^{p,q} \right\|_\infty + \left\| U_{W_1}^{p,q} - U_{F_{\text{NN}}}^{h_{W_1}} \right\|_\infty$$
$$\leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon. \tag{4.19}$$

In order to conclude the proof it suffices to show that the filter $U_{F_{\text{NN}}}^{h_{W_1}}$ can be realized as the reservoir filter associated to an echo

state network of the type presented in the statement. We carry that out by using the elements that appeared in the construction of the reservoir F_{NN} in (4.11) to define a new reservoir map F_{ESN} with the architecture of an echo state network. Let $A := GE \in \mathbb{M}_N$ and define

$$F_{ESN} : \quad D_N \times I_n \quad \longrightarrow \quad \mathbb{R}^N \\ (\mathbf{x}, \mathbf{z}) \quad \longmapsto \quad \sigma\,(A\mathbf{x} + C\mathbf{z} + \boldsymbol{\zeta})\,. \tag{4.20}$$

The set D_N in the domain of F_{ESN} is given by

$$D_N := [-1, 1]^N \cap E^{-1}(\overline{B_{\|\cdot\|}(\mathbf{0}, L)}), \tag{4.21}$$

where $E^{-1}(\overline{B_{\|\cdot\|}(\mathbf{0}, L)})$ denotes the preimage of the set $\overline{B_{\|\cdot\|}(\mathbf{0}, L)} \subset \mathbb{R}^{N_1}$ by the linear map $E : \mathbb{R}^N \longrightarrow \mathbb{R}^{N_1}$ associated to the matrix $E \in \mathbb{M}_{N_1, N}$. This set is compact as $E^{-1}(\overline{B_{\|\cdot\|}(\mathbf{0}, L)})$ is closed and $[-1, 1]^N$ is compact and hence D_N is a closed subspace of a compact space which is always compact (Munkres, 2014, Theorem 26.2). Additionally, D_N is also convex because $[-1, 1]^N$ is convex and $E^{-1}(\overline{B_{\|\cdot\|}(\mathbf{0}, L)})$ is also convex because it is the preimage of a convex set by a linear map, which is always convex.

We note now that the image of F_{ESN} is contained in D_N. First, as the squashing function maps into the interval $[-1, 1]$, it is clear that

$$F_{ESN}(D_N, I_n) \subset [-1, 1]^N. \tag{4.22}$$

Second, for any $\mathbf{x} \in D_N$ we have by construction that $\mathbf{x} \in E^{-1}(\overline{B_{\|\cdot\|}(\mathbf{0}, L)})$ and hence $E\mathbf{x} \in \overline{B_{\|\cdot\|}(\mathbf{0}, L)}$. Since by (4.15) F_{NN} maps into $\overline{B_{\|\cdot\|}(\mathbf{0}, L)}$, we can ensure that for any $\mathbf{z} \in I_n$, the image $F_{NN}(E\mathbf{x}, \mathbf{z}) = E\sigma\,(GE\mathbf{x} + C\mathbf{z} + \boldsymbol{\zeta}) = E\sigma\,(A\mathbf{x} + C\mathbf{z} + \boldsymbol{\zeta}) \in \overline{B_{\|\cdot\|}(\mathbf{0}, L)}$ or, equivalently,

$$F_{ESN}(\mathbf{x}, \mathbf{z}) = \sigma\,(A\mathbf{x} + C\mathbf{z} + \boldsymbol{\zeta}) \in E^{-1}(\overline{B_{\|\cdot\|}(\mathbf{0}, L)}). \tag{4.23}$$

The relations (4.22) and (4.23) imply that

$$F_{ESN}(D_N, I_n) \subset D_N, \tag{4.24}$$

and hence, we can rewrite (4.20) as

$$F_{ESN} : D_N \times I_n \longrightarrow D_N.$$

The continuity of the map F_{ESN} and the compactness and convexity of the set $D_N \subset \mathbb{R}^N$ that we established above allow us to use the first part of Theorem 3.1 to conclude that the corresponding reservoir equation has the existence of solutions property and that we can hence associate to it a (generalized) filter $U_{F_{ESN}}$. Let $W := W_1 E \in \mathbb{M}_{d, n}$ and define the readout map $h_{ESN} : D_N \longrightarrow \mathbb{R}^d$ by $h_{ESN}(\mathbf{x}) := W\mathbf{x} = W_1 E\mathbf{x}$. Denote by U_{ESN} any generalized reservoir filter associated to the echo state network system (F_{ESN}, h_{ESN}) that, by construction, satisfies $U_{ESN}(\mathbf{z})_t := h_{ESN}(U_{F_{ESN}}(\mathbf{z})_t) = WU_{F_{ESN}}(\mathbf{z})_t$, for any $\mathbf{z} \in I_n$ and $t \in \mathbb{Z}_-$.

We next show that the map $f : D_N = [-1, 1]^N \cap E^{-1}(\overline{B_{\|\cdot\|}(\mathbf{0}, L)}) \longrightarrow \overline{B_{\|\cdot\|}(\mathbf{0}, L)}$ given by $f(\mathbf{x}) := E\mathbf{x}$ is a morphism between the echo state network system (F_{ESN}, h_{ESN}) and the reservoir system (F_{NN}, h_{W_1}). Indeed, the reservoir equivariance property holds because, for any $(\mathbf{x}, \mathbf{z}) \in D_N \times I_n$, the definitions (4.11) and (4.20) ensure that

$$f(F_{ESN}(\mathbf{x}, \mathbf{z})) = E\sigma\,(A\mathbf{x} + C\mathbf{z} + \boldsymbol{\zeta}) \\ = E\sigma\,(GE\mathbf{x} + C\mathbf{z} + \boldsymbol{\zeta}) = F_{NN}(E\mathbf{x}, \mathbf{z}) = F_{NN}(f(\mathbf{x}), \mathbf{z}).$$

The readout invariance is obvious. This fact and the second part in Proposition 2.2 imply that all the generalized filters U_{ESN} associated to the echo state network are actually filters generated by the system (F_{NN}, h_{W_1}). This means that for each generalized filter U_{ESN} there exists a generalized filter of the type $U_{F_{NN}}^{h_{W_1}}$ such that $U_{ESN} = U_{F_{NN}}^{h_{W_1}}$. The inequality (4.19) proves then (4.2) in the statement of the theorem. The last claim in the theorem is a straightforward consequence of Propositions 2.1 and 2.12. $\blacksquare$

Acknowledgments

Appendix

A.1. Proof of Proposition 2.1

Let $\tau \in \mathbb{N}$ and let $T_\tau^n : (D_n)^{\mathbb{Z}} \longrightarrow (D_n)^{\mathbb{Z}}$ and $T_\tau^N : (D_N)^{\mathbb{Z}} \longrightarrow (D_N)^{\mathbb{Z}}$ be the corresponding time delay operators. For any $\mathbf{z} \in (D_n)^{\mathbb{Z}}$, let $\mathbf{x} \in (D_N)^{\mathbb{Z}}$ be the unique solution of the reservoir system determined by F, that is,

$$\mathbf{x} := U^F(\mathbf{z}). \tag{A.1}$$

Then, for any $t \in \mathbb{Z}$,

$$\left(T_\tau^N \circ U^F(\mathbf{z})\right) = \mathbf{x}_{t-\tau}. \tag{A.2}$$

Analogously, let $\widetilde{\mathbf{x}} \in (D_N)^{\mathbb{Z}}$ be the unique solution of F associated to the input $T_\tau^n(\mathbf{z})$, that is,

$$\widetilde{\mathbf{x}}_t = \left(U^F \circ T_\tau^n(\mathbf{z})\right)_t, \quad \text{for any} \quad t \in \mathbb{Z}. \tag{A.3}$$

By construction, the sequence $\widetilde{\mathbf{x}}$ satisfies that

$$\widetilde{\mathbf{x}}_t = F\left(\widetilde{\mathbf{x}}_{t-1}, T_\tau^n(\mathbf{z})_t\right) = F\left(\widetilde{\mathbf{x}}_{t-1}, \mathbf{z}_{t-\tau}\right), \quad \text{for any} \quad t \in \mathbb{Z}.$$

It we set $s := t - \tau$, this expression can be rewritten as

$$\widetilde{\mathbf{x}}_{s+\tau} = F\left(\widetilde{\mathbf{x}}_{s+\tau-1}, \mathbf{z}_s\right), \quad \text{for any} \quad s \in \mathbb{Z}, \tag{A.4}$$

and if we define $\widehat{\mathbf{x}}_s := \widetilde{\mathbf{x}}_{s+\tau}$, the equality (A.4) becomes

$$\widehat{\mathbf{x}}_s = F\left(\widehat{\mathbf{x}}_{s-1}, \mathbf{z}_s\right), \quad \text{for any} \quad s \in \mathbb{Z},$$

which shows that $\widehat{\mathbf{x}} \in (D_N)^{\mathbb{Z}}$ is a solution of F determined by the input $\mathbf{z} \in (D_n)^{\mathbb{Z}}$. Since the sequence $\mathbf{x} \in (D_N)^{\mathbb{Z}}$ in (A.1) is also a solution of F for the same input, the echo state property hypothesis on the systems determined by F implies that $\mathbf{x} = \widehat{\mathbf{x}}$, necessarily. This implies that $\mathbf{x}_{t-\tau} = \widehat{\mathbf{x}}_{t-\tau}$ for all $t \in \mathbb{Z}$, which is equivalent to $\widetilde{\mathbf{x}}_t = \mathbf{x}_{t-\tau}$. This equality guarantees that (A.2) and (A.3) are equal and since $\widehat{\mathbf{z}} \in (D_n)^{\mathbb{Z}}$ is arbitrary, we have that

$$T_\tau^N \circ U^F = U^F \circ T_\tau^n,$$

as required. $\blacksquare$

A.2. Proof of Proposition 2.4

Suppose first that U is continuous. This implies the existence of a positive function $\delta_U(\epsilon)$ such that if $\mathbf{u}, \mathbf{v} \in (D_n)^{\mathbb{Z}_-}$ are such

that $\|\mathbf{u} - \mathbf{v}\|_\infty < \delta_U(\epsilon)$, then $\|U(\mathbf{u}) - U(\mathbf{v})\|_\infty < \epsilon$. Under that hypothesis, it is clear that:

$$\|H_U(\mathbf{u}) - H_U(\mathbf{v})\| = \|U(\mathbf{u})_0 - U(\mathbf{v})_0\| \leq \sup_{t \in \mathbb{Z}_-} \{\|U(\mathbf{u})_t - U(\mathbf{v})_t\|\}$$
$$= \|U(\mathbf{u}) - U(\mathbf{v})\|_\infty < \epsilon,$$

which shows the continuity of $H_U : \left((D_n)^{\mathbb{Z}_-}, \|\cdot\|_\infty\right) \longrightarrow (D_N, \|\cdot\|)$.

Conversely, suppose that $H : \left((D_n)^{\mathbb{Z}_-}, \|\cdot\|_\infty\right) \longrightarrow (D_N, \|\cdot\|)$ is continuous and let $\delta_H(\epsilon) > 0$ be such that if $\|\mathbf{u} - \mathbf{v}\|_\infty < \delta_H(\epsilon)$ then $\|H(\mathbf{u}) - H(\mathbf{v})\| < \epsilon$. Then, for any $t \in \mathbb{Z}_-$,

$$\|U_H(\mathbf{u})_t - U_H(\mathbf{v})_t\|$$
$$= \left\|H((\mathbb{P}_{\mathbb{Z}_-} \circ T_{-t})(\mathbf{u})) - H((\mathbb{P}_{\mathbb{Z}_-} \circ T_{-t})(\mathbf{v}))\right\| < \epsilon, \quad (A.5)$$

which proves the continuity of U_H. The inequality follows from the fact that for any $\mathbf{u} \in (D_n)^{\mathbb{Z}_-}$, the components of the sequence $(\mathbb{P}_{\mathbb{Z}_-} \circ T_{-t})(\mathbf{u})$ are included in those of $\mathbf{u}$ and hence $\sup_{s \in \mathbb{Z}_-} \left\{\left\|(\mathbb{P}_{\mathbb{Z}_-} \circ T_{-t}(\mathbf{u}))_s\right\|\right\} \leq \sup_{s \in \mathbb{Z}_-} \{\|\mathbf{u}_s\|\}$ or, equivalently, $\|(\mathbb{P}_{\mathbb{Z}_-} \circ T_{-t})(\mathbf{u})\|_\infty \leq \|\mathbf{u}\|_\infty$. This implies that if $\|\mathbf{u} - \mathbf{v}\|_\infty < \delta_H(\epsilon)$ then $\|T_{-t}(\mathbf{u}) - T_{-t}(\mathbf{v})\|_\infty < \delta_H(\epsilon)$ and hence (A.5) holds. ∎

A.3. Proof of Theorem 2.6

We first show that the map $D_w^M : (\mathbb{R}^n)^{\mathbb{Z}_-} \times (\mathbb{R}^n)^{\mathbb{Z}_-} \longrightarrow [0, \infty)$ defined in (2.15) is indeed a metric. It is clear that $D_w^M(\mathbf{x}, \mathbf{y}) \geq 0$ and that $D_w^M(\mathbf{x}, \mathbf{x}) = 0$, for any $\mathbf{x}, \mathbf{y} \in (\mathbb{R}^n)^{\mathbb{Z}_-}$. Conversely, if $D_w^M(\mathbf{x}, \mathbf{y}) = 0$, this implies that $\bar{d}_M(\mathbf{x}_t, \mathbf{y}_t)w_{-t} \leq \sup_{t \in \mathbb{Z}_-} \{\bar{d}_M(\mathbf{x}_t, \mathbf{y}_t)w_{-t}\} = D_w^M(\mathbf{x}, \mathbf{y}) = 0$, which ensures that $\bar{d}_M(\mathbf{x}_t, \mathbf{y}_t) = 0$, for any $t \in \mathbb{Z}_-$, and hence $\mathbf{x} = \mathbf{y}$ necessarily since the map $\bar{d}_M$ is a metric in $\mathbb{R}^n$ (Munkres, 2014, Chapter 2, §20). It is also obvious that $D_w^M(\mathbf{x}, \mathbf{y}) = D_w^M(\mathbf{y}, \mathbf{x})$. Regarding the triangle inequality, notice that for any $\mathbf{x}, \mathbf{y}, \mathbf{z} \in (\mathbb{R}^n)^{\mathbb{Z}_-}$ and $t \in \mathbb{Z}_-$:

$$\bar{d}_M(\mathbf{x}_t, \mathbf{z}_t)w_{-t} \leq \bar{d}_M(\mathbf{x}_t, \mathbf{y}_t)w_{-t} + \bar{d}_M(\mathbf{y}_t, \mathbf{z}_t)w_{-t}$$
$$\leq D_w^M(\mathbf{x}, \mathbf{y}) + D_w^M(\mathbf{y}, \mathbf{z}),$$

which implies that

$$D_w^M(\mathbf{x}, \mathbf{z}) = \sup_{t \in \mathbb{Z}_-} \{\bar{d}_M(\mathbf{x}_t, \mathbf{z}_t)w_{-t}\} \leq D_w^M(\mathbf{x}, \mathbf{y}) + D_w^M(\mathbf{y}, \mathbf{z}).$$

We now show that the metric topology on $(\mathbb{R})^{\mathbb{Z}_-}$ associated to D_w^M coincides with the product topology. Let $\mathbf{x} \in (\mathbb{R}^n)^{\mathbb{Z}_-}$ and let $B_{D_w^M}(\mathbf{x}, \epsilon)$ be an ϵ-ball around it with respect to the metric D_w^M. Let now $N \in \mathbb{N}$ be large enough so that $w_N < \epsilon/M$. We then show that the basis element V for the product topology in $(\mathbb{R}^n)^{\mathbb{Z}_-}$ given by

$$V := \cdots \times \mathbb{R}^n \times \mathbb{R}^n \times B_{\bar{d}_M}(\mathbf{x}_{-N}, \epsilon) \times \cdots B_{\bar{d}_M}(\mathbf{x}_{-1}, \epsilon) \times B_{\bar{d}_M}(\mathbf{x}_0, \epsilon)$$

and that obviously contains the element $\mathbf{x} \in (\mathbb{R}^n)^{\mathbb{Z}_-}$ is such that $V \subset B_{D_w^M}(\mathbf{x}, \epsilon)$. Indeed, since for any $\mathbf{y} \in (\mathbb{R}^n)^{\mathbb{Z}_-}$ and any $t \in \mathbb{Z}_-$ we have that $\bar{d}_M(\mathbf{x}_t, \mathbf{y}_t) \leq M$, we can conclude that

$$\bar{d}_M(\mathbf{x}_t, \mathbf{y}_t)w_{-t} \leq Mw_N, \quad \text{for all} \quad t \leq -N.$$

Therefore, $D_w^M(\mathbf{x}, \mathbf{y}) \leq \max \{Mw_{-N}, \bar{d}_M(\mathbf{x}_{-N}, \mathbf{y}_{-N})w_N, \ldots, \bar{d}_M(\mathbf{x}_{-1}, \mathbf{y}_{-1})w_1, \bar{d}_M(\mathbf{x}_0, \mathbf{y}_0)w_0\}$ and hence if $\mathbf{y} \in V$ this expression is smaller than ϵ which allows us to conclude the desired inclusion $V \subset B_{D_w^M}(\mathbf{x}, \epsilon)$.

Conversely, consider a basis element of the product topology given by $U = \prod_{t \in \mathbb{Z}_-} U_t$ where $U_t = B_{\bar{d}_M}(\mathbf{x}_t, \epsilon_t)$ for a finite set of indices $t \in \{\alpha_1, \ldots, \alpha_r\}$, $\epsilon_t \leq 1$, and $U_t = \mathbb{R}^n$ for the rest. Let $\epsilon := \min_{t \in \{\alpha_1, \ldots, \alpha_r\}} \{\epsilon_t w_{-t}\}$. We now show that $B_{D_w^M}(\mathbf{x}, \epsilon) \subset U$. Indeed, if $\mathbf{y} \in B_{D_w^M}(\mathbf{x}, \epsilon)$ then $\bar{d}_M(\mathbf{x}_t, \mathbf{y}_t)w_{-t} \leq D_w^M(\mathbf{x}, \mathbf{y}) < \epsilon$, for all $t \in \mathbb{Z}_-$. If $t \in \{\alpha_1, \ldots, \alpha_r\}$ then $\epsilon < \epsilon_t w_{-t}$ and hence $\bar{d}_M(\mathbf{x}_t, \mathbf{y}_t)w_{-t} < \epsilon_t w_{-t}$, which ensures that $\bar{d}_M(\mathbf{x}_t, \mathbf{y}_t) < \epsilon_t$ and hence $\mathbf{y} \in U$, as desired.

We conclude by showing that $\left((\mathbb{R}^n)^{\mathbb{Z}_-}, D_w^M\right)$ is a complete metric space. First, notice that since for any $\mathbf{x}, \mathbf{y} \in (\mathbb{R}^n)^{\mathbb{Z}_-}$ and any given $t \in \mathbb{Z}_-$ we have that

$$\bar{d}_M(\mathbf{x}_t, \mathbf{y}_t) \leq \frac{D_w^M(\mathbf{x}, \mathbf{y})}{w_{-t}},$$

we can conclude that if $\{\mathbf{x}(i)\}_{i \in \mathbb{N}}$ is a Cauchy sequence in $(\mathbb{R}^n)^{\mathbb{Z}_-}$, then so are the sequences $\{\mathbf{x}_t(i)\}_{i \in \mathbb{N}}$ in $\mathbb{R}^n$, for any $t \in \mathbb{Z}_-$, with respect to the bounded metric $\bar{d}_M$. Since the completeness with respect to the bounded metric $\bar{d}_M$ and the Euclidean metric are equivalent (Munkres, 2014, Chapter 7, §43) we can ensure that $\{\mathbf{x}_t(i)\}_{i \in \mathbb{N}}$ converges to an element $\mathbf{a}_t \in \mathbb{R}^n$ with respect to the Euclidean metric for any $t \in \mathbb{Z}_-$. We now show that $\{\mathbf{x}(i)\}_{i \in \mathbb{N}}$ converges to $\mathbf{a} := (\mathbf{a}_t)_{t \in \mathbb{Z}_-} \in (\mathbb{R}^n)^{\mathbb{Z}_-}$, with respect to the metric D_w^M, which proves the completeness statement. Indeed, since the metric D_w^M generates the product topology, let $U = \prod_{t \in \mathbb{Z}_-} U_t$ be a basis element such that $\mathbf{a} \in U$ and, as before, $U_t = B_{\bar{d}_M}(\mathbf{a}_t, \epsilon_t)$ for a finite set of indices $t \in \{\alpha_1, \ldots, \alpha_r\}$, $\epsilon_t \leq 1$, and $U_t = \mathbb{R}^n$ for the rest. Let $\epsilon = \min\{\epsilon_{\alpha_1}, \ldots, \epsilon_{\alpha_r}\}$. Since for each $t \in \mathbb{Z}_-$ the sequence $\mathbf{x}_t(i) \overset{i \to \infty}{\longrightarrow} \mathbf{a}_t$, then there exists $N_t \in \mathbb{N}$ such that for any $k > N_t$ we have that $\|\mathbf{x}_t(k) - \mathbf{a}_t\| < \epsilon$. If we take $N_\epsilon = \max\{N_{\alpha_1}, \ldots, N_{\alpha_r}\}$ then it is clear that $\mathbf{x}(i) \in U$, for all $i > N_\epsilon$, as required. ∎

A.4. Proof of Corollary 2.7

Notice first that for any $\mathbf{x}, \mathbf{y} \in K_M$, we have that $\|\mathbf{x}_t - \mathbf{y}_t\| < 2M, t \in \mathbb{Z}_-$, and hence

$$D_w^{2M}(\mathbf{x}, \mathbf{y}) := \sup_{t \in \mathbb{Z}_-} \{\bar{d}_{2M}(\mathbf{x}_t, \mathbf{y}_t)w_{-t}\}$$
$$= \sup_{t \in \mathbb{Z}_-} \{\|\mathbf{x}_t - \mathbf{y}_t\| w_{-t}\} = \|\mathbf{x} - \mathbf{y}\|_w.$$

Hence, the topology induced by the weighted norm $\|\cdot\|_w$ on K_M coincides with the metric topology induced by the restricted metric $D_w^{2M}|_{K_M \times K_M}$ which, by Theorem 2.6, is the subspace topology induced by the product topology on $(\mathbb{R}^n)^{\mathbb{Z}_-}$ on K_M (see Munkres, 2014, Exercise 1, page 133), as well as the product topology on the product $K_M = \left(\overline{B_{\|\cdot\|}(\mathbf{0}, M)}\right)^{\mathbb{Z}_-}$ (see Munkres, 2014, Theorem 19.3, page 116). ∎

A.5. Proof of Corollary 2.8

First, since $K_M = \left(\overline{B_{\|\cdot\|}(\mathbf{0}, M)}\right)^{\mathbb{Z}_-}$, it is clearly the product of compact spaces. By Tychonoff's Theorem (Munkres, 2014, Chapter 5) K_M is compact when endowed with the product topology which, by Corollary 2.7, coincides with the topology associated to the restriction of the norm $\|\cdot\|_w$ to K_M, as well as with the metric topology given by $D_w^{2M}|_{K_M \times K_M}$.

Second, since $(K_M, \|\cdot\|_w)$ is metrizable it is a Hausdorff space. This implies (see Munkres, 2014, Theorem 26.3) that as K_M is a compact subspace of the Banach space $(\ell_-^w(\mathbb{R}^n), \|\cdot\|_w)$ (see Proposition A.2) then it is necessarily closed. This in turn implies (Simmons, 1963, Theorem B, page 72) that $(K_M, \|\cdot\|_w)$ is complete.

Finally, the convexity statement follows from the fact that the product of convex sets is always convex. ∎

A.6. Proof of Proposition 2.9

Let d_w be the metric on $\ell_-^w(\mathbb{R}^n)$ induced by the weighted norm $\|\cdot\|_w$ and let $D_w := D_w^1$ be the w-weighted metric on $(\mathbb{R}^n)^{\mathbb{Z}_-}$ with constant $M = 1$ introduced in Theorem 2.6 and defined using the same underlying norm in $\mathbb{R}^n$ as the one associated to $\|\cdot\|_w$. As we saw in that theorem, the metric D_w induces the product topology on $(\mathbb{R}^n)^{\mathbb{Z}_-}$.

Let now $\mathbf{u} \in \ell_-^w(\mathbb{R}^n)$ and let $\epsilon > 0$. Let now $\mathbf{v} \in \ell_-^w(\mathbb{R}^n)$ be such that $d_w(\mathbf{u}, \mathbf{v}) < \epsilon$. By definition, we have that

$$
\begin{aligned}
D_w(\mathbf{u}, \mathbf{v}) &= \sup_{t \in \mathbb{Z}_-} \left\{ \overline{d}_1(\mathbf{x}_t, \mathbf{y}_t) w_{-t} \right\} \\
&= \sup_{t \in \mathbb{Z}_-} \left\{ (\min\{\|\mathbf{x}_t - \mathbf{y}_t\|, 1\}) w_{-t} \right\} \\
&\leq \sup_{t \in \mathbb{Z}_-} \left\{ \|\mathbf{x}_t - \mathbf{y}_t\| w_{-t} \right\} = d_w(\mathbf{u}, \mathbf{v}) < \epsilon,
\end{aligned}
$$

which shows that $B_{d_w}(\mathbf{u}, \epsilon) \subset B_{D_w}(\mathbf{u}, \epsilon)$ and allows us to conclude that the norm topology in $\ell_-^w(\mathbb{R}^n)$ is finer than the subspace topology induced by the product topology in $(\mathbb{R}^n)^{\mathbb{Z}_-}$.

We now show that this inclusion is strict. Since the weighting sequence w converges to zero, there exists an element $t_0 \in \mathbb{Z}_-$ such that $w_{-t_0} < \epsilon/2$. Let $\lambda > 0$ be arbitrary and define the element $\mathbf{v}^\lambda \in (\mathbb{R}^n)^{\mathbb{Z}_-}$ by setting $\mathbf{v}_{t_0}^\lambda := \lambda \mathbf{u}_{t_0}$ and $\mathbf{v}^{\lambda t} := \lambda \mathbf{u}_t$ when $t \neq t_0$. We now show that $\mathbf{v}^\lambda \in B_{D_w}(\mathbf{u}, \epsilon)$ for any $\lambda > 0$. Indeed,

$$
D_w(\mathbf{u}, \mathbf{v}^\lambda) = \min\{|\lambda - 1| \|\mathbf{u}_{t_0}\|, 1\} w_{-t_0} \leq 1 \cdot w_{-t_0} < \epsilon/2 < \epsilon.
$$

At the same time, by definition,

$$
d_w(\mathbf{u}, \mathbf{v}^\lambda) = |\lambda - 1| \|\mathbf{u}_{t_0}\| w_{-t_0} < \infty,
$$

which shows that $\mathbf{v}^\lambda \in \ell_-^w(\mathbb{R}^n)$. However, since $|\lambda - 1| \|\mathbf{u}_{t_0}\| w_{-t_0}$ can be made as large as desired by choosing λ big enough, we have proved that for any ball $B_{d_w}(\mathbf{u}, \epsilon')$, with $\epsilon' > 0$ arbitrary, the ball $B_{D_w}(\mathbf{u}, \epsilon)$ contains always an element in $\ell_-^w(\mathbb{R}^n)$ that is not included in $B_{d_w}(\mathbf{u}, \epsilon')$. This argument allows us to conclude that the norm topology in $\ell_-^w(\mathbb{R}^n)$ is strictly finer than the subspace topology induced by the product topology. ∎

A.7. Proof of Lemma 2.10

The proof requires the following preparatory lemma that will also be used later on in the proof of Proposition 2.11.

Lemma A.1. Let $M > 0$ and let w be a weighting sequence. Then:

(i) The operator $\mathbb{P}_{\mathbb{Z}_-} \circ T_{-t} : (K_M, \|\cdot\|_w) \longrightarrow (K_M, \|\cdot\|_w)$ is a continuous map, for any $t \in \mathbb{Z}_-$.

(ii) The projections $p_i : (\ell_-^w(\mathbb{R}^n), \|\cdot\|_w) \longrightarrow (\mathbb{R}^n, \|\cdot\|)$, $i \in \mathbb{Z}_-$, given by $p_i(\mathbf{z}) = \mathbf{z}_i$, are continuous.

Proof of the lemma. (i) We show that this statement is true by characterizing $\mathbb{P}_{\mathbb{Z}_-} \circ T_{-t}$ as a Cartesian product of continuous maps between two product spaces endowed with the product topologies and by using Corollary 2.7. Indeed, notice first that the projections $p_i : (K_M, \|\cdot\|_w) \longrightarrow \overline{B_{\|\cdot\|}(\mathbf{0}, M)}$ are continuous since by Corollary 2.7 the topology induced on K_M by the weighted norm $\|\cdot\|_w$ is the product topology. Since $\mathbb{P}_{\mathbb{Z}_-} \circ T_{-t}$ can be written as the infinite Cartesian product of continuous maps $\mathbb{P}_{\mathbb{Z}_-} \circ T_{-t} = \prod_{i=t}^{-\infty} p_i = (\ldots, p_{t-2}, p_{t-1}, p_t)$ it is hence continuous when using the product topology induced by $\|\cdot\|_w$ (see Munkres, 2014, Theorem 19.6).

(ii) Notice first that the projections $p_i : (\ell_-^w(\mathbb{R}^n), \|\cdot\|_w) \longrightarrow (\mathbb{R}^n, \|\cdot\|)$ are obviously continuous when we consider in $\ell_-^w(\mathbb{R}^n)$ the subspace topology induced by the product topology in $(\mathbb{R}^n)^{\mathbb{Z}_-}$. The continuity of $p_i : (\ell_-^w(\mathbb{R}^n), \|\cdot\|_w) \longrightarrow (\mathbb{R}^n, \|\cdot\|)$ then follows directly from Proposition 2.9. ▾

We now proceed with the proof of Lemma 2.10. Let first $H \in \mathbb{H}_{K_M}^w$. The FMP hypothesis implies that the map $H : (K_M, \|\cdot\|_w) \longrightarrow (\mathbb{R}^N, \|\cdot\|)$ is continuous. Given that K_M is compact by Corollary 2.8 then so is $H(K_M) \subset \mathbb{R}^M$. This in turn implies that $H(K_M)$ is closed and bounded (Munkres, 2014, Theorem 27.3) which guarantees the existence of $L > 0$ such that $H(K_M) \subset \overline{B_{\|\cdot\|}(\mathbf{0}, L)}$. The map obtained out of H by restriction of its target to $\overline{B_{\|\cdot\|}(\mathbf{0}, L)}$ (that we denote with the same symbol) is also continuous and hence $H \in \mathbb{H}_{K_M, K_L}^w$.

Let now $U : K_M \longrightarrow \ell_-^w(\mathbb{R}^N)$ in $\mathbb{F}_{K_M}^w$ and consider the composition $p_0 \circ U : K_M \longrightarrow \mathbb{R}^N$. The FMP hypothesis on U and the continuity of $p_0 : (\ell_-^w(\mathbb{R}^n), \|\cdot\|_w) \longrightarrow (\mathbb{R}^n, \|\cdot\|)$ that we established in the second part of Lemma A.1 imply that $p_0 \circ U$ is continuous. This implies, together with the compactness of K_M that we proved in Corollary 2.8, the existence of $L > 0$ such that $p_0 \circ U(K_M) \subset \overline{B_{\|\cdot\|}(\mathbf{0}, L)}$. Equivalently, for any $\mathbf{z} \in K_M$, we have that $U(\mathbf{z})_0 \in \overline{B_{\|\cdot\|}(\mathbf{0}, L)}$. Now, since U is by hypothesis time invariant, we have by (2.3) that

$$
\begin{aligned}
U(\mathbf{z})_t &= (T_{-t}(U(\mathbf{z})))_0 \\
&= U(T_{-t}(\mathbf{z}))_0 \in \overline{B_{\|\cdot\|}(\mathbf{0}, L)}, \ t \in \mathbb{Z}_-, \text{ since } T_{-t}(\mathbf{z}) \in K_M,
\end{aligned}
$$

which proves that $U(K_M) \subset K_L$. The map obtained out of U by restriction of its target to K_L (that we denote with the same symbol) is also continuous since $(K_L, \|\cdot\|_w)$ is a topological subspace of $(\ell_-^w(\mathbb{R}^n), \|\cdot\|_w)$ and hence $U \in \mathbb{F}_{K_M, K_L}^w$, as required.

The inclusion $\mathbb{F}_{K_M, K_L}^w \subset \mathbb{F}_{K_M}^w$ (respectively, $\mathbb{H}_{K_M, K_L}^w \subset \mathbb{H}_{K_M}^w$) is a consequence of the continuity of the inclusion map $(K_L, \|\cdot\|_w) \hookrightarrow (\ell_-^w(\mathbb{R}^n), \|\cdot\|_w)$ (respectively, $(\overline{B_{\|\cdot\|}(\mathbf{0}, L)}, \|\cdot\|) \hookrightarrow (\mathbb{R}^n, \|\cdot\|)$). ∎

A.8. Proof of Proposition 2.11

Proof of part (i) The FMP of U with respect to the sequence w is, by definition, equivalent to the continuity of the map $U : (K_M, \|\cdot\|_w) \longrightarrow (K_L, \|\cdot\|_w)$ (respectively, $H : (K_M, \|\cdot\|_w) \longrightarrow (\overline{B_{\|\cdot\|}(\mathbf{0}, L)}, \|\cdot\|))$. By Corollary 2.8, this is equivalent to the continuity of these maps when K_M and K_L are endowed with the product topology which is, by the same result, generated by any arbitrary weighting sequence.

Consider now $U : (K_M, \|\cdot\|_w) \longrightarrow (\ell_-^w(\mathbb{R}^n), \|\cdot\|_w)$ in $\mathbb{F}_{K_M}^w$ (respectively, $H : (K_M, \|\cdot\|_w) \longrightarrow (\mathbb{R}^n, \|\cdot\|)$ in $\mathbb{H}_{K_M}^w$). By Lemma 2.10 there exists an $L > 0$ such that U (respectively, H) can be considered an element of $\mathbb{F}_{K_M, K_L}^w$ (respectively, $\mathbb{H}_{K_M, K_L}^w$) by restriction of the target. Using the statement that we just proved about the space $\mathbb{F}_{K_M, K_L}^w$ (respectively, $\mathbb{H}_{K_M, K_L}^w$) we can conclude that U (respectively, H) has the FMP with respect to any weighting sequence. Since, again by Lemma 2.10, the inclusion $\mathbb{F}_{K_M, K_L}^w \subset \mathbb{F}_{K_M}^w$ (respectively, $\mathbb{H}_{K_M, K_L}^w \subset \mathbb{H}_{K_M}^w$) holds true for any $M > 0$, and any weighting sequence w, we can conclude that U (respectively, H) is continuous as an element of $\mathbb{F}_{K_M}^w$ (respectively, $\mathbb{H}_{K_M}^w$) for any weighting sequence w, as required.

Proof of part (ii) First, suppose that $H : (K_M, \|\cdot\|_w) \longrightarrow (\overline{B_{\|\cdot\|}(\mathbf{0}, L)}, \|\cdot\|)$ has the FMP and that this map is hence continuous. Given that the associated filter $U_H : (K_M, \|\cdot\|_w) \longrightarrow (K_L, \|\cdot\|_w)$ can be written as $U_H = \prod_{t=0}^{-\infty} H \circ (\mathbb{P}_{\mathbb{Z}_-} \circ T_{-t})$ we can also conclude that it is continuous. Indeed, by part (i) of Lemma A.1, the map $H \circ (\mathbb{P}_{\mathbb{Z}_-} \circ T_{-t})$, $t \in \mathbb{Z}_-$, is a composition of continuous functions and it is hence continuous. Additionally, the product $\prod_{t=0}^{-\infty} H \circ (\mathbb{P}_{\mathbb{Z}_-} \circ T_{-t}) : (K_M, \|\cdot\|_w) \longrightarrow (K_L, \|\cdot\|_w)$ is also continuous because the topology of $(K_L, \|\cdot\|_w)$ coincides with the product topology by Corollary 2.7 and hence the continuity follows from Munkres (2014, Theorem 19.6), which shows that U_H has the FMP. Conversely, if $U : (K_M, \|\cdot\|_w) \longrightarrow (K_L, \|\cdot\|_w)$ has the FMP, so is the case with $H_U = p_0 \circ U : (K_M, \|\cdot\|_w) \longrightarrow (\overline{B_{\|\cdot\|}(\mathbf{0}, L)}, \|\cdot\|)$ as it is the composition of two continuous maps. These arguments show that $\Psi(\mathbb{F}_{K_M, K_L}^{\mathrm{FMP}}) \subset \mathbb{H}_{K_M, K_L}^{\mathrm{FMP}}$ and $\Phi(\mathbb{H}_{K_M, K_L}^{\mathrm{FMP}}) \subset \mathbb{F}_{K_M, K_L}^{\mathrm{FMP}}$, and that the maps $\Psi : \mathbb{F}_{K_M, K_L}^{\mathrm{FMP}} \longrightarrow \mathbb{H}_{K_M, K_L}^{\mathrm{FMP}}$ and $\Phi : \mathbb{H}_{K_M, K_L}^{\mathrm{FMP}} \longrightarrow \mathbb{F}_{K_M, K_L}^{\mathrm{FMP}}$ are hence inverses of each other.

The parallel statement regarding the spaces $\mathbb{F}_{K_M}^{\mathrm{FMP}}$ and $\mathbb{H}_{K_M}^{\mathrm{FMP}}$ can be easily established by mimicking the proof of part (i) using Lemma 2.10. ∎

A.9. Proof of Proposition 2.12

We prove the continuity of Ψ by establishing the equality (2.18). Let $U \in \mathbb{F}_{K_M}^{\text{FMP}}$. By definition

$$\|\Psi(U)\|_\infty = \sup_{\mathbf{z} \in K_M} \{\|\Psi(U)(\mathbf{z})\|\} = \sup_{\mathbf{z} \in K_M} \{\|U(\mathbf{z})_0\|\}. \tag{A.6}$$

Since we have that

$$\sup_{\mathbf{z} \in K_M} \{\|U(\mathbf{z})_0\|\} \leq \sup_{\mathbf{z} \in K_M} \{\sup_{t \in \mathbb{Z}_-} \{\|U(\mathbf{z})_t\|\}\} = \sup_{\mathbf{z} \in K_M} \{\|U(\mathbf{z})\|_\infty\} = \|U\|_\infty,$$

this shows, together with (A.6), that

$$\|\Psi(U)\|_\infty \leq \|U\|_\infty,$$

which implies the continuity of Ψ. Regarding (2.19), let $H \in \mathbb{H}_{K_M}^{\text{FMP}}$. We have:

$$
\begin{aligned}
\|\Phi(H)\|_\infty &= \sup_{\mathbf{z} \in K_M} \left\{ \sup_{t \in \mathbb{Z}_-} \{\|\Phi(H)(\mathbf{z})_t\|\} \right\} \\
&= \sup_{\mathbf{z} \in K_M} \left\{ \sup_{t \in \mathbb{Z}_-} \{\|H((\mathbb{P}_{\mathbb{Z}_-} \circ T_{-t})(\mathbf{z}))\|\} \right\} \\
&\leq \sup_{\mathbf{z} \in K_M} \{\|H(\mathbf{z})\|_\infty\} = \|H\|_\infty,
\end{aligned}
$$

which proves the continuity of Φ. The inequality is a consequence of the fact that the sequence $(\mathbb{P}_{\mathbb{Z}_-} \circ T_{-t})(\mathbf{z}) \in K_M$. The equalities (2.18) and (2.19) are a consequence of the two inequalities that we just proved, together with the fact that $U_{H_U} = U$ and $H_{U_H} = H$. The equalities (2.20) and (2.21) are proved in a similar fashion. ∎

A.10. Proof of Corollary 3.2

(i) Consider the reservoir map $F_{\text{ESN}} : [-1, 1]^N \times D_n \longrightarrow [-1, 1]^N$ given by $F_{\text{ESN}} := \sigma (A\mathbf{x} + C\mathbf{z} + \boldsymbol{\zeta})$. The statement is a direct consequence of the continuity of F_{ESN}, the compactness and convexity of $[-1, 1]^N$ and D_n, and of part **(i)** of Theorem 3.1.

(ii) The result follows from part **(ii)** of Theorem 3.1 since the hypotheses in the statement imply that the reservoir map F_{ESN} is in those circumstances a contraction. Indeed, let $\mathbf{x}, \mathbf{y} \in [-1, 1]^N$ and let $z \in D_n$, then

$$
\begin{aligned}
\|F_{\text{ESN}}(\mathbf{x}, \mathbf{z}) - F_{\text{ESN}}(\mathbf{y}, \mathbf{z})\| &= \|\sigma (A\mathbf{x} + C\mathbf{z} + \boldsymbol{\zeta}) - \sigma (A\mathbf{y} + C\mathbf{z} + \boldsymbol{\zeta})\| \\
&\leq L_\sigma \|A\|_2 \|\mathbf{x} - \mathbf{y}\|.
\end{aligned}
$$

Since by hypothesis $L_\sigma \|A\|_2 < 1$, we can conclude that F_{ESN} is a contraction, as required. The norm $\|\cdot\|$ in the previous expression is the Euclidean norm in $\mathbb{R}^N$. The time invariance of the resulting unique fading memory reservoir filter is a consequence of Proposition 2.1. ∎

A.11. $\left(\ell_-^w(\mathbb{R}^n), \|\cdot\|_w \right)$ is a Banach space

Proposition A.2. *Let* $w : \mathbb{N} \longrightarrow (0, 1]$ *be a weighting sequence and let* $\|\cdot\|_w : (\mathbb{R}^n)^{\mathbb{Z}_-} \longrightarrow \mathbb{R}^+$ *be the corresponding weighted norm. Then, the space* $\left(\ell_-^w(\mathbb{R}^n), \|\cdot\|_w \right)$ *defined by*

$$\ell_-^w(\mathbb{R}^n) := \left\{ \mathbf{z} \in (\mathbb{R}^n)^{\mathbb{Z}_-} \mid \|\mathbf{z}\|_w < \infty \right\},$$

endowed with weighted norm $\|\cdot\|_w$ *is a Banach space.*

Proof. We first show that $\ell_-^w(\mathbb{R}^n)$ is a linear subspace of $(\mathbb{R}^n)^{\mathbb{Z}_-}$. Let $\mathbf{u}, \mathbf{v} \in \ell_-^w(\mathbb{R}^n)$ and let $\lambda \in \mathbb{R}$. Then,

$$
\begin{aligned}
\|\mathbf{u} + \lambda \mathbf{v}\|_w &= \sup_{t \in \mathbb{Z}_-} \{\|\mathbf{u}_t + \lambda \mathbf{v}_t\| w_{-t}\} \\
&\leq \sup_{t \in \mathbb{Z}_-} \{\|\mathbf{u}_t\| w_{-t} + \lambda \|\mathbf{v}_t\| w_{-t}\} \\
&\leq \sup_{t \in \mathbb{Z}_-} \{\|\mathbf{u}_t\| w_{-t}\} + \lambda \sup_{t \in \mathbb{Z}_-} \{\|\mathbf{v}_t\| w_{-t}\} \\
&= \|\mathbf{u}\|_w + \lambda \|\mathbf{v}\|_w.
\end{aligned}
$$

We now show that this space is complete. Let $\{\mathbf{u}(n)\}_{n \in \mathbb{N}} \subset \ell_-^w(\mathbb{R}^n)$ be a Cauchy sequence. This implies that for any $\epsilon > 0$, there exists $N(\epsilon) \in \mathbb{N}$ such that for all $m, n > N(\epsilon)$ we have $\|\mathbf{u}(n) - \mathbf{u}(m)\|_w < \epsilon$. Hence, for any $t \in \mathbb{Z}_-$,

$$
\begin{aligned}
\|\mathbf{u}_t(n) - \mathbf{u}_t(m)\| w_{-t} &\leq \sup_{t \in \mathbb{Z}_-} \{\|\mathbf{u}_t(n) - \mathbf{u}_t(m)\| w_{-t}\} \\
&= \|\mathbf{u}(n) - \mathbf{u}(m)\|_w < \epsilon. \tag{A.7}
\end{aligned}
$$

This implies that taking for each fixed $t \in \mathbb{Z}_-$ the value $N(\epsilon w_{-t})$, the sequences $\{\mathbf{u}_t(n)\}_{n \in \mathbb{N}}$ in $\mathbb{R}^n$ are Cauchy and hence convergent to values $\mathbf{u}_t \in \mathbb{R}^n$. We now show that $\{\mathbf{u}(n)\}_{n \in \mathbb{N}}$ converges to $\mathbf{u} \in (\mathbb{R}^n)^{\mathbb{Z}_-}$. Using (A.7), take $N(\epsilon/2)$ so that for all $m, n > N(\epsilon/2)$ and any $t \in \mathbb{Z}_-$ one has $\|\mathbf{u}_t(n) - \mathbf{u}_t(n)\| w_{-t} \leq \epsilon/2$. If we take the limit $m \to \infty$ in this inequality, we obtain

$$\|\mathbf{u}_t(n) - \mathbf{u}_t\| w_{-t} \leq \epsilon/2, \quad \text{for all } t \in \mathbb{Z}_-.$$

This implies that

$$\|\mathbf{u}(n) - \mathbf{u}\|_w = \sup_{t \in \mathbb{Z}_-} \{\|\mathbf{u}_t(n) - \mathbf{u}_t\| w_{-t}\} \leq \epsilon/2 < \epsilon,$$

which proves that $\{\mathbf{u}(n)\}_{n \in \mathbb{N}}$ converges to $\mathbf{u}$, as required. It remains to be shown that $\mathbf{u} \in \ell_-^w(\mathbb{R}^n)$, that is, that $\|\mathbf{u}\| < \infty$. In order to show that this is indeed the case, let $n \in \mathbb{N}$ be such that $\|\mathbf{u} - \mathbf{u}(n)\| < \epsilon$. This implies that

$$\|\mathbf{u}\|_w - \|\mathbf{u}(n)\|_w < |\|\mathbf{u}\|_w - \|\mathbf{u}(n)\|_w| < \|\mathbf{u} - \mathbf{u}(n)\|_w < \epsilon,$$

and hence $\|\mathbf{u}\|_w < \|\mathbf{u}(n)\|_w + \epsilon < \infty$, as required. ∎

References

Arnold, V. I. (1957). On functions of three variables. *Proceedings of the USSR Academy of Sciences, 114*, 679–681.

Boyd, S., & Chua, L. (1985). Fading memory and the problem of approximating non-linear operators with Volterra series. *IEEE Transactions on Circuits and Systems, 32*(11), 1150–1161.

Brilliant, M. B. (1958). *Theory of the analysis of nonlinear systems. Tech. rep.*. Massachusetts Institute of Technology, Research Laboratory of Electronics.

Buehner, M., & Young, P. (2006). A tighter bound for the echo state property. *IEEE Transactions on Neural Networks, 17*(3), 820–824.

Cabessa, J., & Villa, A. E. (2015). Computational capabilities of recurrent neural networks based on their attractor dynamics. In *2015 international joint conference on neural networks, IJCNN* (pp. 1–8). IEEE.

Cabessa, J., & Villa, A. E. (2016). Expressive power of first-order recurrent neural networks determined by their attractor dynamics. *Journal of Computer and System Sciences, 82*(8), 1232–1250.

Crook, N. (2007). Nonlinear transient computation. *Neurocomputing, 70*, 1167–1176.

Cucker, F., & Smale, S. (2002). On the mathematical foundations of learning. *American Mathematical Society. Bulletin, 39*(1), 1–49.

Cucker, F., & Zhou, D.-X. (2007). *Learning theory : An approximation theory viewpoint*. Cambridge University Press.

Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals, and Systems, 2*(4), 303–314.

Fréchet, M. (1910). Sur les fonctionnelles continues. *Annales Scientifiques de l'Ecole Normale Supérieure. 3ème série., 27*, 193–216.

Gallicchio, C., & Micheli, A. (2017). Echo state property of deep reservoir computing networks. *Cognitive Computation, 9*.

George, D. A. (1959). *Continuous nonlinear systems. Tech. rep.*. Massachusetts Institute of Technology, Research Laboratory of Electronics.

Gonon, L., & Ortega, J.-P. (2018). Reservoir computing universality with stochastic inputs. Preprint.

Grigoryeva, L., & Ortega, J.-P. (2018a). Universal discrete-time reservoir computers with stochastic inputs and linear readouts using non-homogeneous state-affine systems. *Journal of Machine Learning Research (JMLR), 19*, 1–40.

Grigoryeva, L., & Ortega, J.-P. (2018b). Differentiable reservoir computing filters and their Volterra series representations. Preprint.

Horn, R. A., & Johnson, C. R. (2013). *Matrix analysis* (2nd Ed.). Cambridge University Press.

Hornik, K. (1991). Approximation capabilities of muitilayer feedforward networks. *Neural Networks, 4*(1989), 251–257.

Hornik, K. (1993). Some new results on neural network approximation. *Neural Networks, 6*(8), 1069–1072.

Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, *2*(5), 359–366.

Hornik, K., Stinchcombe, M., & White, H. (1990). Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks. *Neural Networks*, *3*(5), 551–560.

Huang, G.-B., Zhu, Q.-Y., & Siew, C.-K. (2006). Extreme learning machine: Theory and applications. *Neurocomputing*, *70*(1–3), 489–501.

Jaeger, H. (2010). *The 'echo state' approach to analysing and training recurrent neural networks with an erratum note. Tech. rep.*. German National Research Center for Information Technology.

Jaeger, H., & Haas, H. (2004). Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication. *Science*, *304*(5667), 78–80.

Kilian, J., & Siegelmann, H. T. (1996). The dynamic universality of sigmoidal neural networks. *Information and Computation*, *128*(1), 48–56.

Kolmogorov, A. N. (1956). On the representation of continuous functions of several variables as superpositions of functions of smaller number of variables. *Soviet Mathematics Doklady*, *108*, 179–182.

Lukoševičius, M., & Jaeger, H. (2009). Reservoir computing approaches to recurrent neural network training. *Computer Science Review*, *3*(3), 127–149.

Maass, W. (2011). Liquid state machines: motivation, theory, and applications. In S. S. Barry Cooper, & A. Sorbi (Eds.), *Computability in context: Computation and logic in the real world* (pp. 275–296). Chap. 8.

Maass, W., Joshi, P., & Sontag, E. D. (2007). Computational aspects of feedback in neural circuits. *PLoS Computational Biology*, *3*(1), e165.

Maass, W., Natschläger, T., & Markram, H. (2002). Real-time computing without stable states: a new framework for neural computation based on perturbations. *Neural Computation*, *14*, 2531–2560.

Maass, W., Natschläger, T., & Markram, H. (2004). Fading memory and kernel properties of generic cortical microcircuit models. *Journal of Physiology Paris*, *98*(4–6, Spec. Iss.), 315–330.

Maass, W., & Sontag, E. D. (2000). Neural systems as nonlinear filters. *Neural Computation*, *12*(8), 1743–1772.

Manjunath, G., & Jaeger, H. (2013). Echo state property linked to an input: exploring a fundamental characteristic of recurrent neural networks. *Neural Computation*, *25*(3), 671–696.

Matthews, M. B. (1992). *On the uniform approximation of nonlinear discrete-time fading-memory systems using neural network models*. (Ph.D. thesis), ETH Zürich.

Matthews, M. B. (1993). Approximating nonlinear fading-memory operators using neural network models. *Circuits, Systems, and Signal Processing*, *12*(2), 279–307.

Munkres, J. (2014). *Topology* (2nd Ed.). Pearson.

Perryman, P. C. (1996). *Approximation theory for deterministic and stochastic nonlinear systems*. (Ph.D. thesis), Irvine: University of California.

Rüschendorf, L., & Thomsen, W. (1998). Closedness of sum spaces and the generalized schrödinger Problem. *Theory of Probability & Its Applications*, *42*(3), 483–494.

Sandberg, I. W. (1991a). Approximation theorems for discrete-time systems. *IEEE Transactions on Circuits and Systems*, *38*(5), 564–566.

Sandberg, I. W. (1991b). Structure theorems for nonlinear systems. *Multidimensional Systems and Signal Processing*, *2*, 267–286.

Sandberg, I. W. (2003). Notes of fading-memory conditions. *Circuits, Systems, and Signal Processing*, *22*(1), 43–55.

Shapiro, J. H. (2016). *A fixed-point farrago*. Switzerland: Springer International Publishing.

Siegelmann, H., Horne, B., & Giles, C. (1997). Computational capabilities of recurrent NARX neural networks. *IEEE Transactions on Systems, Man and Cybernetics, Part B (Cybernetics)*, *27*(2), 208–215.

Simmons, G. F. (1963). *Topology and modern analysis*. McGraw-Hill.

Smale, S., & Zhou, D.-X. (2003). Estimating the approximation error in learning theory. *Analysis and Applications*, *01*(01), 17–41.

Sprecher, D. A. (1965). A representation theorem for continuous functions of several variables. *Proceedings of the Americal Mathematical Society*, *16*(2), 200.

Sprecher, D. A. (1996). A numerical implementation of Kolmogorov's superpositions. *Neural Networks*, *9*(5), 765–772.

Sprecher, D. A. (1997). A numerical implementation of Kolmogorov's superpositions II. *Neural Networks*, *10*(3), 447–457.

Sternberg, S. (2010). *Dynamical Systems*. Dover.

Stubberud, A., & Perryman, P. (1997). Current state of system approximation for deterministic and stochastic systems. In *Conference record of the thirtieth asilomar conference on signals, systems and computers* (pp. 141–145). IEEE Comput. Soc. Press.

Verstraeten, D., Schrauwen, B., D'Haene, M., & Stroobandt, D. (2007). An experimental unification of reservoir computing methods. *Neural Networks*, *20*, 391–403.

Volterra, V. (1930). *Theory of functionals and of integral and integro-differential equations*. Glasgow: Blackie & Son Limited.

Wainrib, G., & Galtier, M. N. (2016). A local echo state property through the largest Lyapunov exponent. *Neural Networks*, *76*, 39–45.

Wiener, N. (1958). *Nonlinear problems in random theory*. The Technology Press of MIT.

Yildiz, I. B., Jaeger, H., & Kiebel, S. J. (2012). Re-visiting the echo state property. *Neural Networks*, *35*, 1–9.

Zhang, Bai, Miller, D. J., & Wang, Yue (2012). Nonlinear system modeling with random matrices: echo state networks revisited. *IEEE Transactions on Neural Networks and Learning Systems*, *23*(1), 175–182.