Graduate School of Decision Sciences

Universität Konstanz

GSDS Working Paper No. 2019-01

Differentiable reservoir computing

Lyudmila Grigoryeva
Juan-Pablo Ortega

February 2019

Graduate School of Decision Sciences

All processes within our society are based on decisions – whether they are individual or collective decisions. Understanding how these decisions are made will provide the tools with which we can address the root causes of social science issues.

The GSDS offers an open and communicative academic environment for doctoral researchers who deal with issues of decision making and their application to important social science problems. It combines the perspectives of the various social science disciplines for a comprehensive understanding of human decision behavior and its economic and political consequences.

The GSDS primarily focuses on economics, political science and psychology, but also encompasses the complementary disciplines computer science, sociology and statistics. The GSDS is structured around four interdisciplinary research areas: (A) Behavioural Decision Making, (B) Intertemporal Choice and Markets, (C) Political Decisions and Institutions and (D) Information Processing and Statistical Analysis.

GSDS – Graduate School of Decision Sciences
University of Konstanz
Box 146
78457 Konstanz

Phone: +49 (0)7531 88 3633
Fax: +49 (0)7531 88 5193
E-mail: gsds.office@uni-konstanz.de

-gsds.uni-konstanz.de

ISSN: 2365-4120
February 2019

Differentiable reservoir computing

Lyudmila Grigoryeva[1] and Juan-Pablo Ortega[2,3]

Abstract

Much effort has been devoted in the last two decades to characterize the situations in which a reservoir computing system exhibits the so called echo state and fading memory properties. These important features amount, in mathematical terms, to the existence and continuity of global reservoir system solutions. That research is complemented in this paper where the differentiability of reservoir filters is fully characterized for very general classes of discrete-time deterministic inputs. The local nature of the differential allows the formulation of conditions that ensure both the local and global existence of differentiable and, in passing, fading memory solutions, which links to existing research on the input-dependent nature of the echo state property. A Volterra-type series representation for reservoir filters with semi-infinite discrete-time inputs is constructed in the analytic case using Taylor's theorem and corresponding approximation bounds are provided. Finally, it is shown as a corollary of these results that any fading memory filter can be uniformly approximated by a finite Volterra series with finite memory.

Key Words: reservoir computing, fading memory property, finite memory, echo state property, differentiable reservoir filter, Volterra series representation, state-space systems, system identification, machine learning.

1 Introduction

This paper studies the continuity and differentiability properties of transformations or ***filters*** of discrete time signals of infinite length induced by nonlinear state-space transformations of the type

$$\begin{cases} \mathbf{x}_t = F(\mathbf{x}_{t-1}, \mathbf{z}_t), & (1.1) \\ \mathbf{y}_t = h(\mathbf{x}_t), & (1.2) \end{cases}$$

that, under certain hypotheses, process infinite discrete-time inputs $\mathbf{z} = (\ldots, \mathbf{z}_{-1}, \mathbf{z}_0, \mathbf{z}_1, \ldots) \in (\mathbb{R}^n)^{\mathbb{Z}}$ into output signals $\mathbf{y} \in (\mathbb{R}^d)^{\mathbb{Z}}$ of the same type. In the context of supervised machine learning or system identification, these transformations are special types of recurrent neural networks and are sometimes referred to as ***reservoir computers (RC)*** [Jaeg 10, Jaeg 04, Maas 02, Maas 11, Croo 07, Vers 07, Luko 09] or ***reservoir systems***. In that setup, the map $F : \mathbb{R}^N \times \mathbb{R}^n \longrightarrow \mathbb{R}^N$, $n, N \in \mathbb{N}^+$, is called the ***reservoir***, it is usually randomly generated and $h : \mathbb{R}^N \rightarrow \mathbb{R}^d$ is the ***readout***, which is estimated via a supervised learning procedure.

[1]Department of Mathematics and Statistics. Universität Konstanz. Box 146. D-78457 Konstanz. Germany. `Lyudmila.Grigoryeva@uni-konstanz.de`

[2]Universität Sankt Gallen. Faculty of Mathematics and Statistics. Bodanstrasse 6. CH-9000 Sankt Gallen. Switzerland. `Juan-Pablo.Ortega@unisg.ch`

[3]Centre National de la Recherche Scientifique (CNRS). France.

Very general situations have been characterized in which the continuity of the reservoir and readout maps in (1.1)-(1.2) translates into the continuity of the associated filter $U_h^F : (\mathbb{R}^n)^{\mathbb{Z}} \longrightarrow (\mathbb{R}^d)^{\mathbb{Z}}$ when this one exists. For example, suppose that we restrict ourselves to inputs that are uniformly bounded by a constant $M > 0$, that is, consider the space K_M of semi-infinite sequences given by

$$K_M := \left\{ \mathbf{z} \in (\mathbb{R}^n)^{\mathbb{Z}_-} \mid \|\mathbf{z}_t\| \leq M \quad \text{for all} \quad t \in \mathbb{Z}_- \right\}, \quad M > 0, \tag{1.3}$$

and assume that the reservoir map F is continuous and a contraction on the first entry that maps $F : \overline{B_{\|\cdot\|}(\mathbf{0}, L)} \times \overline{B_{\|\cdot\|}(\mathbf{0}, M)} \longrightarrow \overline{B_{\|\cdot\|}(\mathbf{0}, L)}$, with $L > 0$ (the symbol $\overline{B_{\|\cdot\|}(\mathbf{v}, r)}$ denotes the closure of the open ball $B_{\|\cdot\|}(\mathbf{v}, r)$ with respect to a given norm $\|\cdot\|$, center $\mathbf{v}$, and radius $r > 0$). In that case, it can be shown (see [Grig 18a, Theorem 3.1]) that for any $\mathbf{z} \in K_M$ there exists a unique $\mathbf{x} \in K_L := \{\mathbf{x} \in \left(\mathbb{R}^N\right)^{\mathbb{Z}_-} \mid \|\mathbf{x}_t\| \leq L \quad \text{for all} \quad t \in \mathbb{Z}_-\}$ that satisfies (1.1). This existence and uniqueness feature is called the ***echo state property (ESP)*** and allows one to associate unique filters $U^F : K_M \longrightarrow K_L$ and $U_h^F : K_M \longrightarrow (\mathbb{R}^d)^{\mathbb{Z}_-}$ to the reservoir map F and the reservoir system (1.1)-(1.2), respectively. The continuity of F and h implies that both U^F and U_h^F are continuous when we consider either the uniform or the product topologies in the domain and target spaces. The continuity with respect to the product topology is called in this context the ***fading memory property (FMP)*** since, as we recall below, it can be characterized using weighted norms in the spaces of input and output sequences, which shows that recent inputs are more represented in the outputs of FMP filters than older ones. Equivalently, the outputs produced by FMP filters associated to inputs that are close in the recent past are close, even when those inputs may be very different in the distant past.

The restriction to uniformly bounded inputs of the type (1.3) simplifies enormously the characterization of the fading memory property. Indeed, it has been shown in [Sand 03, Grig 18a] that in that case the fading memory property is not a metric but an exclusively topological property that does not depend on the weighted norm used to define it. Therefore, the FMP does not contain in that situation any information about the rate at which the dependence on the past inputs in the system output declines. This is not the case anymore when we consider unbounded input sets where, as we show in Theorem 4.1, *reservoir systems have the FMP only with respect to weighting sequences that converge to zero faster than the divergence rate of their outputs.*

There are several reasons to study reservoir computing systems with unbounded inputs. First, even though we only deal in this paper with the deterministic setup, any random component in the data generating process of the inputs, like a Gaussian perturbation, would most surely imply unboundedness. Second, when dealing with reservoir systems associated to physical systems, it is certainly reasonable to assume boundedness in the input due to the saturation effects that most of those systems present. Nevertheless, the value of the bounding constant is in general unknown beforehand, which makes uniform boundedness hypotheses unrealistic.

Finally, an important goal of this paper is extending the continuity statements in the RC literature and to study the differentiability properties of reservoir computers. In particular, we aim at characterizing the situations in which one can obtain the differentiability of reservoir filters out of the differentiability properties of the maps that define the corresponding reservoir system. Differentiability (of Fréchet type) is only defined on open subsets of normed spaces. It is easy to see that any open set in the Banach space of inputs with a weighted norm contains unbounded sequences, which forces us to deal with that situation.

There are important consequences that can be drawn from the differentiability of reservoir filters. First, the local nature of the differential allows the formulation of conditions that ensure both the local and global existence of differentiable and, in passing, fading memory solutions, which links to existing research [Manj 13] on the input-dependence of the echo state property. Second, when filters are analytic they obviously admit a Taylor series expansion which coincides with the so called discrete-time Volterra series representation [Volt 30, Sche 80, Rugh 81, Prie 88] and, moreover, different Taylor remainders can be used to provide bounds on the approximation errors that are committed when those series are truncated. This path has been explicitly explored in [Sand 98a, Sand 99] for analytic filters with respect to the supremum norm and with inputs with a finite past. We extend this work and we characterize the inputs for which an analytic fading memory reservoir filter with respect to a weighted norm admits a Volterra series representation with semi-infinite inputs. Additionally, we use the causality and time-invariance hypotheses to show that the corresponding Volterra series representations have time-independent coefficients (this feature is not available in the case studied in [Sand 99]) that automatically satisfy the convergence conditions spelled out in [Sand 98b, Sand 98c].

We conclude the paper by showing that these statements can be combined with the results in [Grig 18b] to provide an alternative proof of the following Volterra series universality theorem that was stated for the first time in [Boyd 85, Theorems 3 and 4]: *any time-invariant and causal fading memory filter can be uniformly approximated by a finite Volterra series with finite memory.*

The paper is organized as follows:

- Section 2 introduces the Banach sequence spaces where the semi-infinite inputs and outputs of the reservoir systems that we study are defined. Various elementary facts about weighted and supremum norm topologies are stated, and the notions of fading memory, continuity, and differentiability of maps between sequence spaces are carefully introduced.

- In Section 3 it is studied in detail the differentiability of causal and time-invariant filters defined on the sequence spaces introduced in Section 2. Those result results are put to work in Section 3.1 to easily show well-known results that link the continuity of a filter with input and output spaces endowed with weighted norms with its asymptotic independence on the remote past input. A particular attention is paid in Section 3.2 to the relation between the FMP and the differentiability of causal and time-invariant filters with that of their associated functionals.

- Starting from Section 4 the paper focuses on reservoir filters. The main result in this section is Theorem 4.1 that provides a sufficient (but not necessary) condition for the ESP and FMP to hold in the presence of inputs that are not necessarily bounded. This is a significant generalization with respect to the "standard compactness conditions" imposed in [Jaeg 10] or the uniform boundedness in the inputs that was required in similar results in, for instance, [Grig 18a]. An important observation in Theorem 4.1 is that for general inputs, the FMP depends on the weighting sequence that is used to define it and establishes that, roughly speaking, *reservoir systems have the FMP only with respect to weighting sequences that converge to zero faster than the divergence rate of their outputs.* This newly introduced FMP condition is spelled out for several widely used families of reservoir systems. The above mentioned results involving uniform boundedness hypotheses can be obtained as a corollary (see Corollary 4.5) of the results in this section. Another statement that we prove (see Theorem 4.8) is that when the target of the reservoir map is a compact set then the echo state property is in that situation guaranteed *for no matter what input in* even though the FMP may obviously not hold in that case.

- Section 5 is the core of the paper and studies the differentiability properties of reservoir filters determined by differentiable reservoir maps. The main results are contained in Theorems 5.1 and 5.7. The first theorem provides an explicit and easy-to-verify sufficient condition for the ESP and the FMP to hold around a given input for which we know that the reservoir system associated to a differentiable reservoir map has a solution. Theorem 5.7 is a global extension of the previous result that, unlike Theorems 4.1 and 5.1, fully characterizes the ESP and the differentiability (and hence the FMP) of the reservoir filter associated to a differentiable reservoir map. In Section 5.2 we show that the global conditions in Theorem 5.7 are much stronger than the local ones in Theorem 5.1 by introducing an example that shows how *the ESP and the FMP are structural features of a reservoir* *system when considered globally but are mostly input dependent when considered only locally.* This important observation has already been noticed in [Manj 13] where, using tools coming from the theory of non-autonomous dynamical systems, sufficient conditions have been formulated (see, for instance, [Manj 13, Theorem 2]) that ensure the ESP in connection to a given specific input. The differentiability conditions that we impose to our reservoir systems allow us to draw similar conclusions and, additionally, to automatically establish the FMP of the resulting locally defined reservoir filters. In Section 5.3 we show how for globally differentiable reservoir filters we can formulate a non-uniform version of the well-known input forgetting property for FMP filters that we recovered in Section 3.1 for inputs that are not necessarily bounded. Moreover, a novel uniform differential version of that result is provided in Theorem 5.15.

- Section 6 contains two main results. First, Theorem 6.1 shows the availability of discrete-time Volterra series representations for analytic, causal, time-invariant, and FMP filters. This result extends a similar statement formulated in [Sand 98a, Sand 99] to inputs with a semi-infinite past that are not necessarily bounded. Second, in Theorem 6.3, we combine the previous result with a universality statement in [Grig 18b] to provide an alternative proof of the Volterra series universality theorem stated for the first time in [Boyd 85, Theorems 3 and 4].

2 The input and output spaces for reservoir systems

This paper studies input/output systems that are causal, that is, the output depends only on the past history of the input and that, in general, have infinite memory. This makes us consider the spaces of left infinite sequences with values in $\mathbb{R}^n$, that is, $(\mathbb{R}^n)^{\mathbb{Z}_-} = \{\mathbf{z} = (\ldots, \mathbf{z}_{-2}, \mathbf{z}_{-1}, \mathbf{z}_0) \mid \mathbf{z}_i \in \mathbb{R}^n, i \in \mathbb{Z}_-\}$. Analogously, $(D_n)^{\mathbb{Z}_-}$ stands for the space of semi-infinite sequences with elements in the subset $D_n \subset \mathbb{R}^n$. The space $\mathbb{R}^n$ will be considered as a normed space with a norm denoted by $\|\cdot\|$ which is not necessarily the Euclidean one (even though they are all equivalent), unless it is explicitly mentioned.

We endow these infinite product spaces with the Banach space structures associated to one of the following two norms. First, the ***supremum norm*** $\|\mathbf{z}\|_\infty := \sup_{t\in\mathbb{Z}_-}\{\|\mathbf{z}_t\|\}$. The symbol $\ell_-^\infty(\mathbb{R}^n)$ is used to denote the Banach space formed by the elements that have a finite supremum norm. Second, given a strictly decreasing sequence with zero limit $w : \mathbb{N} \longrightarrow (0, 1]$ and that $w_0 = 1$, we define the ***weighted norm*** $\|\cdot\|_w$ on $(\mathbb{R}^n)^{\mathbb{Z}_-}$ associated to w by $\|\mathbf{z}\|_w := \sup_{t\in\mathbb{Z}_-}\{\|\mathbf{z}_t w_{-t}\|\}$. It can be shown (see [Grig 18a]) that the set $\ell_-^w(\mathbb{R}^n)$ formed by the elements that have a finite w-weighted norm is a Banach space. Moreover, it is easy 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.

The Banach spaces $(\ell_-^\infty(\mathbb{R}^n), \|\cdot\|_\infty)$ and $(\ell_-^w(\mathbb{R}^n), \|\cdot\|_w)$ are particular cases of weighted Banach sequence spaces $(\ell_-^{p,w}(\mathbb{R}^n), \|\cdot\|_{p,w})$ where

$$\|\mathbf{z}\|_{p,w} := \left(\sum_{t\in\mathbb{Z}_-} \|\mathbf{z}_t\|^p\, w_{-t}\right)^{\frac{1}{p}}, \quad \text{with } 1 \leq p < +\infty,\ \mathbf{z} \in (\mathbb{R}^n)^{\mathbb{Z}_-}, \text{ and } w \text{ a sequence.} \tag{2.1}$$

When $p = +\infty$ we set $\|\cdot\|_{p,w} := \|\cdot\|_w$. We then define

$$\ell_-^{p,w}(\mathbb{R}^n) := \left\{\mathbf{z} \in \mathbb{R}^n \mid \|\mathbf{z}\|_{p,w} < +\infty\right\}. \tag{2.2}$$

These spaces are defined in the literature (see, for instance, [Reki 15, Guna 15]) without the requirement that w is a weighting sequence in the sense of the definition above. Indeed, the standard Banach spaces $(\ell_-^p(\mathbb{R}^n), \|\cdot\|_p)$, with $1 \leq p \leq +\infty$, are particular cases of $(\ell_-^{p,w}(\mathbb{R}^n), \|\cdot\|_{p,w})$ that are obtained by taking as sequence w the constant sequence w^{ι} given by $w^{\iota}_t := 1$, for all $t \in \mathbb{N}$. This observation is used in the paper to obtain many results for the spaces $\ell^{\infty}_{-}(\mathbb{R}^n)$ as a particular case of those proved for $\ell^{w}_{-}(\mathbb{R}^n)$.

We emphasize that w^{ι} is not a weighting sequence and that the spaces $(\ell^{w}_{-}(\mathbb{R}^n), \|\cdot\|_w)$ considered in this paper are all based on sequences w of weighting type. It can be proved (see [Reki 15, Theorems 3.3 and 4.1]) that, in that case:

$$\ell^{p,w}_{-}(\mathbb{R}^n) \subset \ell^{w}_{-}(\mathbb{R}^n), \quad \text{for any } 1 \leq p < +\infty, \tag{2.3}$$

and that,

$$\ell^{p}_{-}(\mathbb{R}^n) \subset \ell^{p,w}_{-}(\mathbb{R}^n), \quad \text{for any } 1 \leq p \leq +\infty. \tag{2.4}$$

All the results in this paper are formulated for the weighted spaces $(\ell^{w}_{-}(\mathbb{R}^n), \|\cdot\|_w)$ even though many of the statements that we provide are also valid for $(\ell^{\infty}_{-}(\mathbb{R}^n), \|\cdot\|_\infty)$ and $(\ell^{p,w}_{-}(\mathbb{R}^n), \|\cdot\|_{p,w})$. That will be explicitly pointed out in the statements or in remarks when it is the case.

2.1 The topologies induced by weighted and supremum norms

An important feature of the topology generated by weighted norms is that they coincide with the product topology on subsets made of uniformly bounded sequences like the space K_M in (1.3). This fact holds true for any weighting sequence w and has important consequences (see [Grig 18a] for the details). First, the fading memory property that we brought up in the introduction and that we spell out in detail later on is independent of the weighting sequence used to define it. Second, the subsets $K_M \subset \ell^{w}_{-}(\mathbb{R}^n)$ are compact in the topology induced by the weighted norms $\|\cdot\|_w$. We emphasize that these statements are valid exclusively in the context of uniformly bounded subsets which, as we see in the next result, are never open in the weighted topology.

We adopt in the sequel the following notation for product sets and functions: for any family $\{A_t\}_{t\in\mathbb{Z}_-}$, of subsets $A_t \subset \mathbb{R}^n$ the symbol

$$\prod_{t\in\mathbb{Z}_-} A_t := \left\{ \mathbf{z} \in (\mathbb{R}^n)^{\mathbb{Z}_-} \mid \mathbf{z}_t \in A_t, \quad \text{for all } t \in \mathbb{Z}_- \right\},$$

denotes the Cartesian product of the sets in the family. When all the elements in the family are identical to a given subset A, we will exchangeably use the symbols $\prod_{t\in\mathbb{Z}_-} A$ and $(A)^{\mathbb{Z}_-}$. A similar notation is adopted for the Cartesian product of maps: let V be a set and let $f_t : V \longrightarrow A_t$ be a map, $t \in \mathbb{Z}_-$. The symbol $\prod_{t\in\mathbb{Z}_-} f_t$ denotes the map

$$\begin{array}{rccc} \prod_{t\in\mathbb{Z}_-} f_t : & V & \longrightarrow & \prod_{t\in\mathbb{Z}_-} A_t \\ & v & \longmapsto & (\ldots, f_{-2}(v), f_{-1}(v), f_0(v)) . \end{array} \tag{2.5}$$

Lemma 2.1 *Let w be a weighting sequence and $n \in \mathbb{N}^+$. Then:*

(i) *For any $\mathbf{z} \in \ell^{w}_{-}(\mathbb{R}^n)$ and $r > 0$,*

$$B_{\|\cdot\|_w}(\mathbf{z}, r) = \bigcup_{\delta<r} \left(\prod_{t\in\mathbb{Z}_-} B_{\|\cdot\|}\left(\mathbf{z}_t, \frac{\delta}{w_{-t}}\right) \right). \tag{2.6}$$

In particular, this implies that

$$B_{\|\cdot\|_w}(\mathbf{z}, r) \subset \prod_{t\in\mathbb{Z}_-} B_{\|\cdot\|}\left(\mathbf{z}_t, \frac{r}{w_{-t}}\right) \subset \overline{B_{\|\cdot\|_w}(\mathbf{z}, r)}. \tag{2.7}$$

The identity (2.6) implies that any open ball $B_{\|\cdot\|_w}(\mathbf{z}, r)$ in $\ell^{w}_{-}(\mathbb{R}^n)$ contains unbounded sequences.

(ii) *Let* $\{A_t\}_{t\in\mathbb{Z}_-}$ *be a family of subsets* $A_t \subset \mathbb{R}^n$ *such that there exists a sequence* $\{c_t\}_{t\in\mathbb{Z}_-}$ *that satisfies*

$$\sup_{\mathbf{z}_t\in A_t}\{\|\mathbf{z}_t\|\, w_{-t}\} < c_t, \text{ for each } t\in\mathbb{Z}_- \text{ and } \sup_{t\in\mathbb{Z}_-}\{c_t\} < +\infty, \tag{2.8}$$

then the product set

$$\prod_{t\in\mathbb{Z}_-} A_t \subset \ell_-^w(\mathbb{R}^n).$$

(iii) *For every family* $\{A_t\}_{t\in\mathbb{Z}_-}$ *of subsets* $A_t \subset \mathbb{R}^n$ *such that the product set satisfies* $\prod_{t\in\mathbb{Z}_-} A_t \subset \ell_-^w(\mathbb{R}^n)$*, we have*

$$\overline{\prod_{t\in\mathbb{Z}_-} A_t} = \prod_{t\in\mathbb{Z}_-} \overline{A_t}. \tag{2.9}$$

These statements, except for the last sentence in part **(i)***, are also valid for the space* $\ell_-^\infty(\mathbb{R}^n)$ *and are obtained by taking as sequence* w *the constant sequence* w^ι *given by* $w_t^\iota := 1$*, for all* $t\in\mathbb{N}$.

Proof. **(i)** We prove (2.6) by double inclusion. First, let $\mathbf{x}\in B_{\|\cdot\|_w}(\mathbf{z},r)$. By definition $\|\mathbf{x}-\mathbf{z}\|_w = \sup_{t\in\mathbb{Z}_-}\{\|\mathbf{x}_t-\mathbf{z}_t\|\, w_{-t}\} < r$ and hence for any $\delta_x>0$ such that $\|\mathbf{x}-\mathbf{z}\|_w<\delta_x<r$ we have that $\|\mathbf{x}_t-\mathbf{z}_t\|<\delta_x/w_{-t}$, for all $t\in\mathbb{Z}_-$. This implies that

$$\mathbf{x}\in\prod_{t\in\mathbb{Z}_-} B_{\|\cdot\|}\left(\mathbf{z}_t,\frac{\delta_x}{w_{-t}}\right)\subset\bigcup_{\delta<r}\left(\prod_{t\in\mathbb{Z}_-} B_{\|\cdot\|}\left(\mathbf{z}_t,\frac{\delta}{w_{-t}}\right)\right).$$

Conversely, given an element $\mathbf{x}\in\ell_-^w(\mathbb{R}^n)$ in the right hand side of (2.6), there exists $\delta_x<r$ such that $\mathbf{x}\in\prod_{t\in\mathbb{Z}_-} B_{\|\cdot\|}(\mathbf{z}_t,\delta_x/w_{-t})$. This implies that $\|\mathbf{x}_t-\mathbf{z}_t\|\, w_{-t}<\delta_x$, for all $t\in\mathbb{Z}_-$, and hence $\sup_{t\in\mathbb{Z}_-}\{\|\mathbf{x}_t-\mathbf{z}_t\|\, w_{-t}\}=\|\mathbf{x}-\mathbf{z}\|_w\le\delta_x<r$, which proves the inclusion.

As to (2.7), the first inclusion is a straightforward consequence of (2.6). Let now $\mathbf{x}\in\prod_{t\in\mathbb{Z}_-} B_{\|\cdot\|}\left(\mathbf{z}_t,\frac{r}{w_{-t}}\right)$. By definition this implies that $\|\mathbf{x}_t-\mathbf{z}_t\|\, w_{-t}<r$, for all $t\in\mathbb{Z}_-$, and consequently $\sup_{t\in\mathbb{Z}_-}\{\|\mathbf{x}_t-\mathbf{z}_t\|\, w_{-t}\}\le r$ or, equivalently, $\|\mathbf{x}-\mathbf{z}\|_w\le r$. This implies that $\mathbf{x}\in\overline{B_{\|\cdot\|_w}(\mathbf{z},r)}$ and proves the second inclusion.

(ii) Let $\mathbf{x}\in\prod_{t\in\mathbb{Z}_-} A_t$. Then,

$$\|\mathbf{x}\|_w=\sup_{t\in\mathbb{Z}_-}\{\|\mathbf{x}_t\|\, w_{-t}\}\le\sup_{t\in\mathbb{Z}_-}\{c_t\}<+\infty,$$

as required.

(iii) We first prove that $\prod_{t\in\mathbb{Z}_-}\overline{A_t}\subset\overline{\prod_{t\in\mathbb{Z}_-} A_t}$. If $\mathbf{z}\in\prod_{t\in\mathbb{Z}_-}\overline{A_t}$, then for any $\epsilon>0$ and each $t\in\mathbb{Z}_-$ there exists an element $\mathbf{x}_t\in A_t\cap B_{\|\cdot\|}\left(\mathbf{z}_t,\frac{\epsilon}{2w_{-t}}\right)$. Let $\mathbf{x}:=(\mathbf{x}_t)_{t\in\mathbb{Z}_-}$. By construction:

$$\|\mathbf{x}-\mathbf{z}\|_w=\sup_{t\in\mathbb{Z}_-}\{\|\mathbf{x}_t-\mathbf{z}_t\|\, w_{-t}\}\le\frac{\epsilon}{2}<\epsilon,$$

which implies that $\mathbf{x}\in B_{\|\cdot\|_w}(\mathbf{z},\epsilon)\cap\prod_{t\in\mathbb{Z}_-} A_t$ and, as $\mathbf{z}\in\prod_{t\in\mathbb{Z}_-}\overline{A_t}$ is arbitrary, it guarantees that $\mathbf{z}\in\overline{\prod_{t\in\mathbb{Z}_-} A_t}$.

In order to show the reverse inclusion first note that, as it is proved later on in Lemma 3.1, the projections $p_t:\ell_-^w(\mathbb{R}^n)\longrightarrow\mathbb{R}^n$, $t\in\mathbb{Z}_-$, defined by $p_t(\mathbf{z}):=\mathbf{z}_t$, are continuous. Let $\mathbf{z}\in\overline{\prod_{t\in\mathbb{Z}_-} A_t}$ arbitrary, let $t\in\mathbb{Z}_-$ be arbitrary but fixed, and let V_t be an open set in $\mathbb{R}^n$ that contains $\mathbf{z}_t$. The continuity of p_t implies that $p_t^{-1}(V_t)$ is an open set in $\ell_-^w(\mathbb{R}^n)$ that contains $\mathbf{z}$ and therefore there exists $\mathbf{x}\in\left(\prod_{t\in\mathbb{Z}_-} A_t\right)\cap p_t^{-1}(V_t)$. We consequently have that $\mathbf{x}_t\in A_t$, which guarantees that $\mathbf{z}_t\in\overline{A_t}$, as required. ■

Corollary 2.2 *Let D_n be a subset of $\mathbb{R}^n$ and let w be a weighting sequence. Then:*

(i) *If $(D_n)^{\mathbb{Z}_-} \cap \ell^w_-(\mathbb{R}^n)$ is an open subset of $\ell^w_-(\mathbb{R}^n)$ then $D_n = \mathbb{R}^n$, necessarily.*

(ii) *If $(D_n)^{\mathbb{Z}_-} \subset \ell^w_-(\mathbb{R}^n)$ is a closed subset of $\ell^w_-(\mathbb{R}^n)$ then D_n is necessarily closed in $\mathbb{R}^n$, that is, $D_n = \overline{D_n}$.*

(iii) *The following inclusion always holds*

$$\overline{(D_n)^{\mathbb{Z}_-} \cap \ell^w_-(\mathbb{R}^n)} \subset \left(\overline{D_n}\right)^{\mathbb{Z}_-} \cap \ell^w_-(\mathbb{R}^n). \tag{2.10}$$

In particular, if D_n is closed in $\mathbb{R}^n$ then so is $(D_n)^{\mathbb{Z}_-} \cap \ell^w_-(\mathbb{R}^n)$ in $\ell^w_-(\mathbb{R}^n)$.

These statements in parts **(ii)** *and* **(iii)** *are also valid when the space $\ell^w_-(\mathbb{R}^n)$ is replaced by $\ell^\infty_-(\mathbb{R}^n)$.*

Proof. **(i)** We proceed by contradiction. Suppose that $D_n \neq \mathbb{R}^n$. Let $\mathbf{x}_0 \in \mathbb{R}^n \setminus D_n$ and let $\mathbf{z}_0 \in D_n$. Define the constant sequences $\mathbf{x} := (\mathbf{x}_0)_{t\in\mathbb{Z}_-} \in \ell^w_-(\mathbb{R}^n) \setminus \left((D_n)^{\mathbb{Z}_-} \cap \ell^w_-(\mathbb{R}^n)\right)$ and $\mathbf{z} := (\mathbf{z}_0)_{t\in\mathbb{Z}_-} \in (D_n)^{\mathbb{Z}_-} \cap \ell^w_-(\mathbb{R}^n)$. Since by hypothesis $(D_n)^{\mathbb{Z}_-} \cap \ell^w_-(\mathbb{R}^n)$ is an open subset of $\ell^w_-(\mathbb{R}^n)$ there exists $\epsilon > 0$ such that $B_{\|\cdot\|_w}(\mathbf{z}, 2\epsilon) \subset (D_n)^{\mathbb{Z}_-} \cap \ell^w_-(\mathbb{R}^n)$. By the relation (2.7) in Lemma 2.1 we also have

$$B_{\|\cdot\|_w}(\mathbf{z}, \epsilon) \subset \prod_{t\in\mathbb{Z}_-} B_{\|\cdot\|}\left(\mathbf{z}_0, \frac{\epsilon}{w_{-t}}\right) \subset \overline{B_{\|\cdot\|_w}(\mathbf{z}, \epsilon)} \subset B_{\|\cdot\|_w}(\mathbf{z}, 2\epsilon) \subset (D_n)^{\mathbb{Z}_-} \cap \ell^w_-(\mathbb{R}^n),$$

and, in particular,

$$\prod_{t\in\mathbb{Z}_-} B_{\|\cdot\|}\left(\mathbf{z}_0, \frac{\epsilon}{w_{-t}}\right) \subset (D_n)^{\mathbb{Z}_-} \cap \ell^w_-(\mathbb{R}^n), \text{ which implies } B_{\|\cdot\|}\left(\mathbf{z}_0, \tfrac{\epsilon}{w_{-t}}\right) \subset D_n, \text{ for all } t \in \mathbb{Z}_-. \tag{2.11}$$

Let $r_0 := \|\mathbf{x}_0 - \mathbf{z}_0\|$ and let $t_0 \in \mathbb{Z}_-$ be such that for all $t < t_0$ we have that $\epsilon/w_{-t_0} > r_0$. By (2.11) we have that $\mathbf{x}_0 \in B_{\|\cdot\|}\left(\mathbf{z}_0, \frac{\epsilon}{w_{-t}}\right) \subset D_n$, which contradicts the assumption on the choice of $\mathbf{x}_0$.

(ii) By Lemma 2.1 **(iii)** we have that

$$\overline{(D_n)^{\mathbb{Z}_-}} = \left(\overline{D_n}\right)^{\mathbb{Z}_-}. \tag{2.12}$$

Since by hypothesis $(D_n)^{\mathbb{Z}_-}$ is closed and hence it holds true that

$$\overline{(D_n)^{\mathbb{Z}_-}} = (D_n)^{\mathbb{Z}_-}. \tag{2.13}$$

Consequently, by (2.12) and (2.13) we have that $\left(\overline{D_n}\right)^{\mathbb{Z}_-} = (D_n)^{\mathbb{Z}_-}$ which implies that $\overline{D_n} = D_n$ as required.

(iii) Let $\mathbf{x} \in \overline{(D_n)^{\mathbb{Z}_-} \cap \ell^w_-(\mathbb{R}^n)} \subset \ell^w_-(\mathbb{R}^n)$ and consider a sequence $\{\mathbf{x}^m\}_{m\in\mathbb{N}^+} \subset (D_n)^{\mathbb{Z}_-} \cap \ell^w_-(\mathbb{R}^n)$ with $\lim_{m\to\infty} \mathbf{x}^m = \mathbf{x}$, that is for each $\epsilon > 0$ there exists such $N(\epsilon) \in \mathbb{N}^+$ such that for all $m > N(\epsilon)$ it holds that $\|\mathbf{x}^m - \mathbf{x}\|_w < \epsilon$. Hence for all $s \in \mathbb{Z}_-$ one has that

$$w_{-s}\|\mathbf{x}^m_s - \mathbf{x}_s\| \leq \sup_{t\in\mathbb{Z}_-}\left\{\|\mathbf{x}^m_t - \mathbf{x}_t\|\, w_{-t}\right\} = \|\mathbf{x}^m - \mathbf{x}\|_w \leq \epsilon,$$

which immediately implies that

$$\|\mathbf{x}^m_s - \mathbf{x}_s\| < \frac{\epsilon}{w_{-s}}$$

and hence one gets that $\mathbf{x}_s \in \overline{D_n}$ and therefore (2.10) holds as required.
The last claim in part **(iii)** follows from (2.10). Indeed, if $D_n = \overline{D_n}$ then by (2.10) we have that

$$\overline{(D_n)^{\mathbb{Z}_-} \cap \ell_-^w(\mathbb{R}^n)} \subset \left(\overline{D_n}\right)^{\mathbb{Z}_-} \cap \ell_-^w(\mathbb{R}^n) = (D_n)^{\mathbb{Z}_-} \cap \ell_-^w(\mathbb{R}^n).$$

Since the reverse inclusion obviously always holds, we finally have that

$$\overline{(D_n)^{\mathbb{Z}_-} \cap \ell_-^w(\mathbb{R}^n)} = (D_n)^{\mathbb{Z}_-} \cap \ell_-^w(\mathbb{R}^n). \quad \blacksquare$$

We also recall (see [Grig 18a, Proposition 2.9]) that 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 $\ell_-^w(\mathbb{R}^n) \subset (\mathbb{R}^n)^{\mathbb{Z}_-}$. We complement this fact by comparing the norm topology on $(\ell_-^\infty(\mathbb{R}^n), \|\cdot\|_\infty)$ with the relative topology induced by $(\ell_-^w(\mathbb{R}^n), \|\cdot\|_w)$ on it.

Corollary 2.3 *The relative topology $\tau_{w,\infty}$ induced by the norm topology τ_w of $(\ell_-^w(\mathbb{R}^n), \|\cdot\|_w)$ on $\ell_-^\infty(\mathbb{R}^n)$ is strictly coarser than the norm topology τ_∞ on $(\ell_-^\infty(\mathbb{R}^n), \|\cdot\|_\infty)$, that is, $\tau_{w,\infty} \subsetneq \tau_\infty$.*

Proof. Since, as we already saw, $\|\mathbf{z}\|_w \leq \|\mathbf{z}\|_\infty$, for all $\mathbf{z} \in (\mathbb{R}^n)^{\mathbb{Z}_-}$, we have that $\ell_-^\infty(\mathbb{R}^n) \subset \ell_-^w(\mathbb{R}^n)$ (see (2.4)) and the inclusion $\iota : \ell_-^\infty(\mathbb{R}^n) \hookrightarrow \ell_-^w(\mathbb{R}^n)$ is continuous. Consequently, for any open $U \in \tau_w$ the set $\iota^{-1}(U) = U \cap \ell_-^\infty(\mathbb{R}^n) \in \tau_{w,\infty}$ is also open in τ_∞. This immediately implies that

$$\tau_{w,\infty} \subset \tau_\infty.$$

In order to establish that this inclusion is strict, one needs to notice that, given an arbitrary open ball $B_{\|\cdot\|_\infty}(\mathbf{z}, r)$, $r > 0$, around $\mathbf{z} \in \ell_-^\infty(\mathbb{R}^n)$, all the open balls $B_{\|\cdot\|_w}(\mathbf{z}, \epsilon)$ for all $\epsilon > 0$ contain elements that are not included in $B_{\|\cdot\|_\infty}(\mathbf{z}, r)$ by Lemma 2.1 **(i)**. $\blacksquare$

Lemma 2.4 *Let w be a weighting sequence and $n \in \mathbb{N}^+$. We denote by w^a, $a \in \mathbb{R}$, the sequence with terms w_t^a, $t \in \mathbb{N}$. Then, the following inclusions are continuous:*

$$\left(\ell_-^\infty(\mathbb{R}^n), \|\cdot\|_\infty\right) \hookrightarrow \cdots \hookrightarrow \left(\ell_-^{w^{\frac{1}{k+1}}}(\mathbb{R}^n), \|\cdot\|_{w^{\frac{1}{k+1}}}\right) \hookrightarrow \left(\ell_-^{w^{\frac{1}{k}}}(\mathbb{R}^n), \|\cdot\|_{w^{\frac{1}{k}}}\right) \hookrightarrow \cdots \hookrightarrow \left(\ell_-^w(\mathbb{R}^n), \|\cdot\|_w\right), \tag{2.14}$$

$$\left(\ell_-^w(\mathbb{R}^n), \|\cdot\|_w\right) \hookrightarrow \cdots \hookrightarrow \left(\ell_-^{w^k}(\mathbb{R}^n), \|\cdot\|_{w^k}\right) \hookrightarrow \left(\ell_-^{w^{k+1}}(\mathbb{R}^n), \|\cdot\|_{w^{k+1}}\right) \hookrightarrow \cdots \hookrightarrow (\mathbb{R}^n)^{\mathbb{Z}_-}, \tag{2.15}$$

where $k \in \mathbb{N}^+$ and in $(\mathbb{R}^n)^{\mathbb{Z}_-}$ we consider the trivial topology. Define

$$S_w := \bigcap_{k \in \mathbb{N}^+} \ell_-^{w^{\frac{1}{k}}}(\mathbb{R}^n) \quad \textit{and} \quad S^w := \bigcup_{k \in \mathbb{N}^+} \ell_-^{w^k}(\mathbb{R}^n). \tag{2.16}$$

Then, in general,

$$\ell_-^\infty(\mathbb{R}^n) \subsetneq S_w \quad \textit{and} \quad S^w \subsetneq (\mathbb{R}^n)^{\mathbb{Z}_-}. \tag{2.17}$$

Proof. The continuity of the inclusions (2.14) and (2.15) is a consequence of the fact that:

$$\|\mathbf{z}\|_{w^{\frac{1}{k}}} \leq \|\mathbf{z}\|_{w^{\frac{1}{k+1}}}, \quad \text{for all } k \in \mathbb{N}^+ \text{ and } \mathbf{z} \in \ell_-^{w^{\frac{1}{k}}}(\mathbb{R}^n), \tag{2.18}$$

$$\|\mathbf{z}\|_{w^{k+1}} \leq \|\mathbf{z}\|_{w^k}, \quad \text{for all } k \in \mathbb{N}^+ \text{ and } \mathbf{z} \in \ell_-^{w^k}(\mathbb{R}^n). \tag{2.19}$$

Regarding (2.17), the first inclusion follows from the fact that $\ell_-^\infty(\mathbb{R}^n) \subset \ell_-^w(\mathbb{R}^n)$ for any weighting sequence. In order to show that this inclusion is in general not an equality it suffices to consider the following example: let $\mathbf{z} \in (\mathbb{R})^{\mathbb{Z}_-}$ given by $z_t := -t$, $t \in \mathbb{Z}_-$, and let w be the weighting sequence defined by $w_t := \lambda^t$, with $t \in \mathbb{N}$ and $0 < \lambda < 1$. A simple application of the L'Hôpital rule shows that, for any $k \in \mathbb{N}^+$,

$$\lim_{t \to -\infty} z_t w_{-t}^{1/k} = 0,$$

which proves, in particular, that $\|\mathbf{z}\|_{w^{1/k}} < \infty$ and hence that $\mathbf{z} \in \ell_-^{w^{1/k}}(\mathbb{R})$, for any $k \in \mathbb{N}^+$. This implies that $\mathbf{z} \in S_w$. However, $\mathbf{z}$ is an unbounded sequence and hence it does not belong to $\ell_-^\infty(\mathbb{R})$. In order to show that the second inclusion in (2.17) is also strict, take $\mathbf{z} \in (\mathbb{R})^{\mathbb{Z}_-}$ given by $z_t := \lambda^{-t}$ with $\lambda > 1$ and $t \in \mathbb{Z}_-$ and let w be the weighting sequence defined by $w_0 := 1$ and $w_t := \frac{1}{t}$, for any $t \in \mathbb{N}^+$. The L'Hôpital rule shows that, for any $k \in \mathbb{N}^+$,

$$\lim_{t \to -\infty} |z_t w_{-t}^k| = +\infty,$$

and consequently $\mathbf{z}$ does not belong to any of the spaces $\ell_-^{w^k}(\mathbb{R})$ and hence $\mathbf{z} \notin S^w$. ■

2.2 Continuity and differentiability of maps on infinite sequence spaces

Much of this paper is related to the continuity and the differentiability of maps of the type $f : U \subset \ell_-^{w^1}(\mathbb{R}^n) \longrightarrow V \subset \ell_-^{w^2}(\mathbb{R}^N)$, with w^1, w^2 weighting sequences and U and V subsets of $\ell_-^{w^1}(\mathbb{R}^n)$ and $\ell_-^{w^2}(\mathbb{R}^N)$, respectively, that in the case of differentiable maps are necessarily open. Maps that are continuous with respect to topologies generated by weighted norms will be generically referred to as ***fading memory maps*** (or we say that they have the ***fading memory property (FMP)***) while when the topology considered is generated by the supremum norm, we just say that the map is ***continuous***. Most of the definitions that we provide in what follows for the weighted norms case can be adapted to the supremum norm case by replacing the weighting sequences by the constant sequence w^ι given by $w_t^\iota := 1$, for all $t \in \mathbb{N}$.

Suppose now that U and V are open subsets. The map $f : U \subset \ell_-^{w^1}(\mathbb{R}^n) \longrightarrow V \subset \ell_-^{w^2}(\mathbb{R}^N)$ is ***(Fréchet) differentiable*** at $\mathbf{u}_0 \in U$ when there exists a bounded linear map $Df(\mathbf{u}_0) : \ell_-^{w^1}(\mathbb{R}^n) \longrightarrow \ell_-^{w^2}(\mathbb{R}^N)$ that satisfies

$$\lim_{\mathbf{u} \to \mathbf{u}_0} \frac{f(\mathbf{u}) - f(\mathbf{u}_0) - Df(\mathbf{u}_0) \cdot (\mathbf{u} - \mathbf{u}_0)}{\|\mathbf{u} - \mathbf{u}_0\|_{w^1}} = \mathbf{0}. \tag{2.20}$$

We say that $f : U \subset \ell_-^{w^1}(\mathbb{R}^n) \longrightarrow V \subset \ell_-^{w^2}(\mathbb{R}^N)$ is of class $C^1(U)$ when it is differentiable at any point in U and the induced map $Df : U \longrightarrow L\left(\ell_-^{w^1}(\mathbb{R}^n), \ell_-^{w^2}(\mathbb{R}^N)\right)$ is continuous, where the space of linear maps $L\left(\ell_-^{w^1}(\mathbb{R}^n), \ell_-^{w^2}(\mathbb{R}^N)\right)$ is endowed with the operator norm $|||\cdot|||_{w^1,w^2}$ defined by

$$|||A|||_{w^1,w^2} := \sup_{\mathbf{u} \in \ell_-^{w^1}(\mathbb{R}^n)} \left\{ \frac{\|A(\mathbf{u})\|_{w^2}}{\|\mathbf{u}\|_{w^1}} \,\middle|\, \mathbf{u} \neq \mathbf{0} \right\}, \quad A \in L\left(\ell_-^{w^1}(\mathbb{R}^n), \ell_-^{w^2}(\mathbb{R}^N)\right). \tag{2.21}$$

When in the domain and the range we use the same weighting sequence w, we will write $|||A|||_w$ instead of $|||A|||_{w^1,w^2}$. The higher order derivatives

$$D^r f(\mathbf{u}_0) : \underbrace{\ell_-^{w^1}(\mathbb{R}^n) \times \cdots \times \ell_-^{w^1}(\mathbb{R}^n)}_{r \text{ times}} \longrightarrow \ell_-^{w^2}(\mathbb{R}^N), \quad r \in \mathbb{N}^+,$$

are inductively defined and the map f is said to be of class $C^r(U)$ when it is r-times differentiable at any point in U and the induced map $D^r f : U \longrightarrow L^r\left(\ell_-^{w^1}(\mathbb{R}^n), \ell_-^{w^2}(\mathbb{R}^N)\right)$ into the normed space of

r-multilinear maps is continuous. We recall that the operator norm $|||\cdot|||_{w^1,w^2}$ in $L^r\left(\ell_-^{w^1}(\mathbb{R}^n),\ell_-^{w^2}(\mathbb{R}^N)\right)$ is given by

$$|||A|||_{w^1,w^2}:=\sup_{\mathbf{u}_1,\ldots,\mathbf{u}_r\in\ell_-^{w^1}(\mathbb{R}^n)}\left\{\frac{\|A(\mathbf{u}_1,\ldots,\mathbf{u}_r)\|_{w^2}}{\|\mathbf{u}_1\|_{w^1}\cdots\|\mathbf{u}_r\|_{w^1}}\middle|\,\mathbf{u}_1,\ldots,\mathbf{u}_r\neq\mathbf{0}\right\},\quad A\in L^r\left(\ell_-^{w^1}(\mathbb{R}^n),\ell_-^{w^2}(\mathbb{R}^N)\right). \tag{2.22}$$

We recall that differentiable functions are automatically continuous and we denote the class of continuous functions by $C^0(U)$. When f is of class $C^r(U)$ in U for any $r\in\mathbb{N}^+$, we say that f is ***smooth*** in U and we denote this class by $C^\infty(U)$. When f is smooth in U we can construct for it a Taylor power series expansion. We say that f is ***analytic*** in U when the convergence domain of that power series includes U. The analytic class is denoted by $C^\omega(U)$.

We emphasize that, as we pointed out in Lemma 2.1, for any weighting sequence w, any open set in $\left(\ell_-^w(\mathbb{R}^n),\|\cdot\|_w\right)$ contains unbounded sequences. For instance, let $B_{\|\cdot\|_w}(\mathbf{0},\epsilon)$ be the ball of radius $\epsilon>0$ around the zero sequence and let $\mathbf{v}\in\mathbb{R}^n$ be a vector such that $\|\mathbf{v}\|=1$. The divergent sequence $\mathbf{z}$ defined by $\mathbf{z}_t:=\epsilon\mathbf{v}/2w_{-t}$ is such that $\|\mathbf{z}\|_w=\epsilon/2$ and hence $\mathbf{z}\in B_{\|\cdot\|_w}(\mathbf{0},\epsilon)\subset\ell_-^w(\mathbb{R}^n)$.

The following lemma spells out conditions under which infinite Cartesian products of continuous and differentiable functions are continuous and differentiable when we use weighted and supremum norms.

Lemma 2.5 *Let $W\subset V$ with $(V,\|\cdot\|)$ a normed space and let $D_N\subset\mathbb{R}^N$ be a subset of $\mathbb{R}^N$. Let $H_t:W\longrightarrow D_N$, $t\in\mathbb{Z}_-$, be a family of maps. Consider the corresponding product map $\mathcal{H}:W\longrightarrow(D_N)^{\mathbb{Z}_-}$, defined as in (2.5):*

$$\mathcal{H}:=\prod_{t\in\mathbb{Z}_-}H_t:=(\ldots,H_{-2},H_{-1},H_0)\,,\ \text{or equivalently,}\ (\mathcal{H}(\mathbf{z}))_t:=H_t(\mathbf{z}),\ \mathbf{z}\in W,\ t\in\mathbb{Z}_-. \tag{2.23}$$

(i) *Endow $W\subset V$ with the subspace topology. If D_N is a compact subset of $\mathbb{R}^N$ then $(D_N)^{\mathbb{Z}_-}\subset\ell_-^w(\mathbb{R}^N)$ for any weighting sequence w. If each of the functions H_t is continuous then $\mathcal{H}:W\longrightarrow(D_N)^{\mathbb{Z}_-}\subset\ell_-^w(\mathbb{R}^N)$ is also continuous.*

(ii) *Let w be a weighting sequence and suppose that W contains a point $\mathbf{z}^0$ such that $\mathcal{H}(\mathbf{z}^0)\in\ell_-^w(\mathbb{R}^N)$. If each of the functions H_t is Lipschitz continuous with Lipschitz constant c_t^0 and the sequence $c^0:=(c_t^0)_{t\in\mathbb{Z}_-}$ formed by these Lipschitz constants satisfies that $c^0\in\ell_-^w(\mathbb{R})$, then $\mathcal{H}:W\longrightarrow(D_N)^{\mathbb{Z}_-}\cap\ell_-^w(\mathbb{R}^N)$ is Lipschitz continuous with Lipschitz constant $c_{\mathcal{H}}^0\leq\left\|c^0\right\|_w$.*

(iii) *Suppose that W is an open convex subset of $(V,\|\cdot\|)$ and that it contains a point $\mathbf{z}^0$ such that $\mathcal{H}(\mathbf{z}^0)\in\ell_-^w(\mathbb{R}^N)$. Suppose also that the maps H_t are of class $C^r(W)$, $r\geq1$, and let c_t^r be finite constants such that $\sup_{\mathbf{z}\in W}\{|||D^rH_t(\mathbf{z})|||\}\leq c_t^r<+\infty$. If $c^r:=(c_t^r)_{t\in\mathbb{Z}_-}\in\ell_-^w(\mathbb{R})$ then $\mathcal{H}$ is differentiable of order r when considered as a map $\mathcal{H}:W\subset(V,\|\cdot\|)\longrightarrow\left(\ell_-^w(\mathbb{R}^N),\|\cdot\|_w\right)$ and*

$$|||D^r\mathcal{H}(\mathbf{z})|||\leq\|c^r\|_w,\ \text{for any}\ \mathbf{z}\in W. \tag{2.24}$$

Additionally, if $c^j\in\ell_-^w(\mathbb{R})$ for all $j\in\{1,\ldots,r\}$, then $\mathcal{H}$ is of class $C^{r-1}(W)$ and the map $D^{r-1}\mathcal{H}:(W,\|\cdot\|)\longrightarrow\left(L^{r-1}(V,\ell_-^w(\mathbb{R}^N)),|||\cdot|||\right)$ is Lipschitz continuous with Lipschitz constant $c_{\mathcal{H}}^r\leq\|c^r\|_w$.

(iv) *Suppose that W is an open convex subset of $(V,\|\cdot\|)$ and that it contains a point $\mathbf{z}^0$ such that $\mathcal{H}(\mathbf{z}^0)\in\ell_-^w(\mathbb{R}^N)$. If the maps H_t are smooth and $\|c^r\|_w<+\infty$, for each $r\in\mathbb{N}^+$, then so is $\mathcal{H}:W\subset(V,\|\cdot\|)\longrightarrow(\ell_-^w(\mathbb{R}^N),\|\cdot\|_w)$. Suppose, additionally, that the maps H_t are analytic and that $\rho_t>0$ is the radius of convergence of the series expansion of H_t. If $\rho:=\inf_{t\in\mathbb{Z}_-}\{\rho_t\}>0$ then $\mathcal{H}$ is analytic when considered as a map $\mathcal{H}:W\subset(V,\|\cdot\|)\longrightarrow(\ell_-^w(\mathbb{R}^N),\|\cdot\|_w)$ and the radius of convergence $\rho_{\mathcal{H}}$ of its series expansion satisfies that $\rho_{\mathcal{H}}\geq\rho>0$.*

Parts **(ii)**, **(iii)**, *and* **(iv)** *also hold true when the Banach space* $(\ell_-^w(\mathbb{R}^N), \|\cdot\|_w)$ *is replaced by* $(\ell_-^\infty(\mathbb{R}^N), \|\cdot\|_\infty)$. *Part* **(i)** *is in general false in that situation.*

Proof. **(i)** The compactness of D_N guarantees [Munk 14, Theorem 27.3] that there exists $L > 0$ such that $D_N \subset \overline{B_{\|\cdot\|}(\mathbf{0}, L)}$ and hence $(D_N)^{\mathbb{Z}_-} \subset \ell_-^w(\mathbb{R}^N)$ necessarily. It can also be shown (see [Grig 18a, Corollary 2.7]) that when D_N is compact, the relative topology $(D_N)^{\mathbb{Z}_-}$ induced by the weighted norm $\|\cdot\|_w$ in $\ell_-^w(\mathbb{R}^N)$ coincides with the product topology. This implies (see [Munk 14, Theorem 19.6]) that if the functions H_t are continuous then so is $\mathcal{H}$.

(ii) Let $\mathbf{z}^1, \mathbf{z}^2 \in W$. Then,

$$\left\|\mathcal{H}(\mathbf{z}^1) - \mathcal{H}(\mathbf{z}^2)\right\|_w = \sup_{t \in \mathbb{Z}_-} \left\{\left\|H_t(\mathbf{z}^1) - H_t(\mathbf{z}^2)\right\| w_{-t}\right\} \leq \sup_{t \in \mathbb{Z}_-} \left\{c_t^0 \left\|\mathbf{z}^1 - \mathbf{z}^2\right\| w_{-t}\right\} \leq \|c^0\|_w \left\|\mathbf{z}^1 - \mathbf{z}^2\right\|, \tag{2.25}$$

which proves simultaneously that $\mathcal{H}$ is Lipschitz continuous and that it maps into $\ell_-^w(\mathbb{R}^N)$. Regarding the last point, recall that by hypothesis there exists a point $\mathbf{z}^0$ such that $\mathcal{H}(\mathbf{z}^0) \in \ell_-^w(\mathbb{R}^N)$ and hence by (2.25) we have, for any $\mathbf{z} \in W$,

$$\|\mathcal{H}(\mathbf{z})\|_w \leq \|c^0\|_w \left\|\mathbf{z} - \mathbf{z}^0\right\| + \left\|\mathcal{H}(\mathbf{z}^0)\right\|_w < +\infty. \tag{2.26}$$

(iii) First, it is easy to prove recursively that for any $\mathbf{z} \in W$, the map $D^r\mathcal{H}(\mathbf{z}) := \prod_{t \in \mathbb{Z}_-} D^r H_t(\mathbf{z})$ satisfies the condition (2.20). In order to prove the first statement of the lemma, it suffices to show that the multilinear map

$$D^r\mathcal{H}(\mathbf{z}) : \underbrace{(V, \|\cdot\|) \times \cdots \times (V, \|\cdot\|)}_{r \text{ times}} \longrightarrow (\ell_-^w(\mathbb{R}^n), \|\cdot\|_w),$$

is bounded for any $\mathbf{z} \in W$. Let $\left(\mathbf{v}^1, \ldots, \mathbf{v}^r\right) \in V^r$. Using the r-order differentiability of H_t we can write

$$\begin{aligned}\left\|D^r\mathcal{H}(\mathbf{z}) \cdot \left(\mathbf{v}^1, \ldots, \mathbf{v}^r\right)\right\|_w &= \left\|\prod_{t \in \mathbb{Z}_-} D^r H_t(\mathbf{z}) \cdot \left(\mathbf{v}^1, \ldots, \mathbf{v}^r\right)\right\|_w = \sup_{t \in \mathbb{Z}_-} \left\{\left\|D^r H_t(\mathbf{z}) \cdot \left(\mathbf{v}^1, \ldots, \mathbf{v}^r\right)\right\| w_{-t}\right\} \\ &\leq \sup_{t \in \mathbb{Z}_-} \left\{\|\!|D^r H_t(\mathbf{z})|\!\| \left\|\mathbf{v}^1\right\| \cdots \left\|\mathbf{v}^r\right\| w_{-t}\right\} \\ &\leq \left\|\mathbf{v}^1\right\| \cdots \left\|\mathbf{v}^r\right\| \sup_{t \in \mathbb{Z}_-} \left\{c_t^r w_{-t}\right\} \leq \|c^r\|_w \left\|\mathbf{v}_1\right\| \cdots \left\|\mathbf{v}_r\right\|, \end{aligned} \tag{2.27}$$

which proves the boundedness of $D^r\mathcal{H}(\mathbf{z})$ and the inequality in (2.24).

We now assume that $c^j \in \ell_-^w(\mathbb{R})$ for all $j \in \{1, \ldots, r\}$ and show that $\mathcal{H}$ maps into $\ell_-^w(\mathbb{R}^N)$ and that it is of class $C^{r-1}(W)$. Notice, first of all, that for any $t \in \mathbb{Z}_-$ and any $\mathbf{z}^1, \mathbf{z}^2 \in V_n$, we have by the convexity of W, the mean value theorem [Abra 88], and the hypothesis $H_t \in C^r(W)$, that for all $j \in \{1, \ldots, r\}$,

$$\left|\!\left|\!\left|D^{j-1}H_t(\mathbf{z}^1) - D^{j-1}H_t(\mathbf{z}^2)\right|\!\right|\!\right|_w \leq \sup_{\mathbf{z} \in W} \left\{\left|\!\left|\!\left|D^j H_t(\mathbf{z})\right|\!\right|\!\right|\right\} \left\|\mathbf{z}^1 - \mathbf{z}^2\right\| = c_t^j \left\|\mathbf{z}^1 - \mathbf{z}^2\right\|. \tag{2.28}$$

Taking $j = 1$ in the previous inequality, we see that the functions H_t are Lipschitz continuous with constants c_t^1 that form a sequence that by hypothesis belongs to $\ell_-^w(\mathbb{R})$. This guarantees by part **(ii)** that $\mathcal{H}$ maps into $\ell^w_-(\mathbb{R}^N)$ necessarily. Now, using the inequality (2.28), we have that for any $\mathbf{z}^1, \mathbf{z}^2 \in W$,

$$
\begin{aligned}
&\left|\!\left|\!\left| D^{r-1}\mathcal{H}(\mathbf{z}^1) - D^{r-1}\mathcal{H}(\mathbf{z}^2) \right|\!\right|\!\right|_w \\
&\quad = \sup_{\substack{\mathbf{v}^1,\ldots,\mathbf{v}^{r-1}\in V \\ \mathbf{v}^1,\ldots,\mathbf{v}^{r-1}\neq \mathbf{0}}} \left\{ \frac{\left\| \left(D^{r-1}\mathcal{H}(\mathbf{z}^1) - D^{r-1}\mathcal{H}(\mathbf{z}^2) \right) \cdot \left(\mathbf{v}^1, \ldots, \mathbf{v}^{r-1} \right) \right\|_w}{\|\mathbf{v}^1\| \cdots \|\mathbf{v}^{r-1}\|} \right\} \\
&\quad = \sup_{\substack{\mathbf{v}^1,\ldots,\mathbf{v}^{r-1}\in V \\ \mathbf{v}^1,\ldots,\mathbf{v}^{r-1}\neq \mathbf{0}}} \left\{ \frac{\sup_{t\in\mathbb{Z}_-} \left\{ \left\| \left(D^{r-1}H_t(\mathbf{z}^1) - D^{r-1}H_t(\mathbf{z}^2) \right) \cdot \left(\mathbf{v}^1, \ldots, \mathbf{v}^{r-1} \right) \right\| w_{-t} \right\}}{\|\mathbf{v}^1\| \cdots \|\mathbf{v}^{r-1}\|} \right\} \\
&\quad \leq \sup_{\substack{\mathbf{v}^1,\ldots,\mathbf{v}^{r-1}\in V \\ \mathbf{v}^1,\ldots,\mathbf{v}^{r-1}\neq \mathbf{0}}} \left\{ \frac{\sup_{t\in\mathbb{Z}_-} \left\{ c^r_t w_{-t} \left\| \mathbf{z}^1 - \mathbf{z}^2 \right\| \cdot \|\mathbf{v}^1\| \cdots \|\mathbf{v}^{r-1}\| \right\}}{\|\mathbf{v}^1\| \cdots \|\mathbf{v}^{r-1}\|} \right\} = \|c^r\|_w \left\| \mathbf{z}^1 - \mathbf{z}^2 \right\|, \quad (2.29)
\end{aligned}
$$

which shows that the map $D^{r-1}\mathcal{H} : (W, \|\cdot\|_w) \longrightarrow \left(L^{r-1}(\ell^w_-(\mathbb{R}^n), \ell^w_-(\mathbb{R}^N)), |\!|\!|\cdot|\!|\!|_w \right)$ is Lipschitz continuous with Lipschitz constant $c^r_{\mathcal{H}} \leq \|c^r\|_w$.

(iv) The previous part of the lemma together with the hypothesis $c := \sup_{r\in\mathbb{N}^+} \{\|c^r\|_w\} < +\infty$ guarantees that the differentiability of any order in the functions H_t gets translated into the differentiability of any order of the map $\mathcal{H} : W \subset (V, \|\cdot\|) \longrightarrow (\ell^w_-(\mathbb{R}^N), \|\cdot\|_w)$. Moreover, let $\mathbf{u} \in \ell^w_-(\mathbb{R}^n)$ and let $\mathbf{u}^r := (\mathbf{u}, \ldots, \mathbf{u}) \in \left(\ell^w_-(\mathbb{R}^n) \right)^r$, $r \in \mathbb{N}^+$. The Taylor series expansion of $\mathcal{H}$ around $\mathbf{0} \in$ is

$$
\mathcal{H}(\mathbf{0}) + \sum_{r=1}^{\infty} \frac{1}{r!} D^r \mathcal{H}(\mathbf{0}) \cdot \mathbf{u}^r = \prod_{t\in\mathbb{Z}_-} \left(H_t(\mathbf{0}) + \sum_{r=1}^{\infty} \frac{1}{r!} D^r H_t(\mathbf{0}) \cdot \mathbf{u}^r \right). \quad (2.30)
$$

The expansion in the left hand side of this equality is convergent if and only if each of the series in the product in the right hand side is convergent. This is the case when $\|\mathbf{u}\|_w < \rho_t$, for all $t \in \mathbb{Z}_-$, which guarantees the convergence of the Taylor series expansion in (2.30) for all the elements $\mathbf{u} \in \ell^w_-(\mathbb{R}^n)$ that satisfy $\|\mathbf{u}\|_w < \inf_{t\in\mathbb{Z}_-} \{\rho_t\} = \rho$. Since by hypothesis $\rho > 0$, we have proved that $\mathcal{H}$ is analytic with radius of convergence $\rho_{\mathcal{H}} \geq \rho$.

The proof of the statements in **(ii)**, **(iii)**, and **(iv)** for the space $\ell^\infty_-(\mathbb{R}^N)$ is obtained by mimicking the proofs that we just provided, replacing the weighting sequence w by the constant sequence w^ι that is equal to 1 for each $t \in \mathbb{Z}_-$. In order to show that part **(i)** is in general false in that situation take $W = (-1, 1)$, $D_N = [-1, 1]$, and define $H_t(z) := \tanh(-tz)$, with $t \in \mathbb{Z}$ and $z \in (-1, 1)$. Given that $H_t^{-1}\left(-\frac{1}{2}, \frac{1}{2}\right) = \left(-\frac{1}{t}\tanh^{-1}(-\frac{1}{2}), -\frac{1}{t}\tanh^{-1}(\frac{1}{2})\right)$ it is clear that

$$
\mathcal{H}^{-1}\left(B_{\|\cdot\|_\infty}\left(\mathbf{0}, \frac{1}{2} \right) \right) = \bigcap_{t\in\mathbb{Z}_-} \left(-\frac{1}{t}\tanh^{-1}\left(-\frac{1}{2}\right), -\frac{1}{t}\tanh^{-1}\left(\frac{1}{2}\right) \right) = \{0\}.
$$

This equality shows that the preimage by the product map $\mathcal{H}$ of an open set is not open and hence $\mathcal{H}$ is not continuous. ■

3 Differentiable time-invariant filters and functionals

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 can be easily extended to accommodate situations where the domains and the targets of the filters are not necessarily product spaces but just arbitrary subsets V_n and V_N of $(\mathbb{R}^n)^{\mathbb{Z}}$ and $\left(\mathbb{R}^N\right)^{\mathbb{Z}}$ like, for instance, $\ell^\infty(\mathbb{R}^n)$ and $\ell^\infty(\mathbb{R}^N)$, or $\ell^w_-(\mathbb{R}^n)$ and $\ell^w_-(\mathbb{R}^N)$, for some weighting sequence w.

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 $T_\tau^{\mathbb{Z}} : (\mathbb{R}^n)^{\mathbb{Z}} \longrightarrow (\mathbb{R}^n)^{\mathbb{Z}}$ be the ***time delay*** operator defined by $T_\tau^{\mathbb{Z}}(\mathbf{z})_t := \mathbf{z}_{t-\tau}$, $\tau \in \mathbb{Z}$. A subset $V_n \subset (\mathbb{R}^n)^{\mathbb{Z}}$ is called time-invariant when $T_\tau^{\mathbb{Z}}(V_n) = V_n$, for all $\tau \in \mathbb{Z}$. The filter U is called ***time-invariant*** when it is defined on a time-invariant set and commutes with the time delay operator, that is, $T_\tau^{\mathbb{Z}} \circ U = U \circ T_\tau^{\mathbb{Z}}$, for any $\tau \in \mathbb{Z}$ (in this expression, the two operators $T_\tau^{\mathbb{Z}}$ have to be understood as defined in the appropriate sequence spaces).

We recall that there is a bijection between causal time-invariant filters and functionals on $(D_n)^{\mathbb{Z}_-}$. Indeed, 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 $(D_n)^{\mathbb{Z}}$. 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 $U_H(\mathbf{z})_t := H((\mathbb{P}_{\mathbb{Z}_-} \circ T_{-t}^{\mathbb{Z}})(\mathbf{z}))$, where $T_{-t}^{\mathbb{Z}}$ 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. Moreover, when considering causal and time-invariant filters $U : (D_n)^{\mathbb{Z}} \longrightarrow (D_N)^{\mathbb{Z}}$ it suffices to work just with the restriction $U : (D_n)^{\mathbb{Z}_-} \longrightarrow (D_N)^{\mathbb{Z}_-}$, that we denote with the same symbol, since the latter uniquely determines the former. Indeed, by definition, for any $\mathbf{z} \in (D_n)^{\mathbb{Z}}$ and $t \in \mathbb{N}^+$:

$$U(\mathbf{z})_t = \left(T_{-t}^{\mathbb{Z}}\left(U(\mathbf{z})\right)\right)_0 = U\left(T_{-t}^{\mathbb{Z}}(\mathbf{z})\right)_0, \tag{3.1}$$

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

In view of this observation, we restrict our study to filters with domain and target in the spaces of left semi-infinite sequences. In particular, we say that a causal and time-invariant filter U has the fading memory property or that it is continuous when the corresponding restricted filter defined on left semi-infinite inputs has those properties, as we defined them in Section 2.2.

Additionally, from now on we consider most of the time time delay operators with domain and target in $(\mathbb{R}^n)^{\mathbb{Z}_-}$ and that we simply denote as $T_{-\tau} : (\mathbb{R}^n)^{\mathbb{Z}_-} \longrightarrow (\mathbb{R}^n)^{\mathbb{Z}_-}$. The definition of these restricted time delay operators $T_{-\tau}$ requires considering two cases:

- $T_{-\tau} : (\mathbb{R}^n)^{\mathbb{Z}_-} \longrightarrow (\mathbb{R}^n)^{\mathbb{Z}_-}$ with τ negative: as before, $T_{-\tau}(\mathbf{z})_t := \mathbf{z}_{t+\tau}$, for any $\mathbf{z} \in (\mathbb{R}^n)^{\mathbb{Z}_-}$ and $t \in \mathbb{Z}_-$. This implies that, in this case,

$$T_{-\tau}(\mathbf{z}) = \mathbb{P}_{\mathbb{Z}_-} \circ T_{-\tau}^{\mathbb{Z}}(\mathbf{z}^e), \quad \mathbf{z} \in (\mathbb{R}^n)^{\mathbb{Z}_-}, \quad \tau < 0,$$

  where $\mathbf{z}^e \in (\mathbb{R}^n)^{\mathbb{Z}}$ is an arbitrary extension of $\mathbf{z} \in (\mathbb{R}^n)^{\mathbb{Z}_-}$ to $(\mathbb{R}^n)^{\mathbb{Z}}$. The map $T_{-\tau}$, $\tau \in \mathbb{Z}_-$, is surjective, that is, $T_\tau((\mathbb{R}^n)^{\mathbb{Z}_-}) = (\mathbb{R}^n)^{\mathbb{Z}_-}$, but it is not injective. The same applies to the restriction of $T_{-\tau}$ to any time-invariant set $V_n \subset (\mathbb{R}^n)^{\mathbb{Z}_-}$ which satisfies $T_{-\tau}(V_n) = V_n$.

- $T_{-\tau} : (\mathbb{R}^n)^{\mathbb{Z}_-} \longrightarrow (\mathbb{R}^n)^{\mathbb{Z}_-}$ with τ positive: there is in principle not a unique way to define the restricted operators $T_{-\tau}$ since that involves the choice of vectors $\mathbf{v}_\tau \in (\mathbb{R}^n)^\tau$ such that $T_{-\tau}(\mathbf{z}) := (\mathbf{z}, \mathbf{v}_\tau)$, for any $\mathbf{z} \in (\mathbb{R}^n)^{\mathbb{Z}_-}$. The choice $\mathbf{v}_\tau = \mathbf{0}$ for all $\tau > 0$ is canonical since it is the only one that makes the resulting maps linear and additionally satisfy

$$T_{-\tau} = \underbrace{T_{-1} \circ \cdots \circ T_{-1}}_{\tau \text{ times}}.$$

  We hence adopt the definition

$$T_{-\tau}(\mathbf{z}) := (\mathbf{z}, \underbrace{\mathbf{0}, \ldots, \mathbf{0}}_{\tau \text{ times}}), \quad \mathbf{z} \in (\mathbb{R}^n)^{\mathbb{Z}_-}, \quad \tau > 0,$$

  for the rest of the paper. In this case $T_{-\tau}$ it is injective but not surjective.

The following lemma gathers some differentiability properties of projections and time delay operators when restricted to normed sequence spaces and that will be used later on. A key element in this result is what we call, for each weighting sequence w, their ***decay ratio*** D_w and ***inverse decay ratio*** L_w, that are defined as:

$$D_w := \sup_{t\in\mathbb{N}}\left\{\frac{w_{t+1}}{w_t}\right\} \quad \text{and} \quad L_w := \sup_{t\in\mathbb{N}}\left\{\frac{w_t}{w_{t+1}}\right\}. \tag{3.2}$$

As w is by definition strictly decreasing we necessarily have that $0 < w_{t+1}/w_t < 1$, for all $t \in \mathbb{N}$, and $1 < w_0/w_1 \le \sup_{t\in\mathbb{N}}\{w_t/w_{t+1}\} = L_w$. Consequently:

$$0 < D_w \le 1 \quad \text{and} \quad 1 < L_w \le +\infty.$$

The decay ratios provide a geometric bound for the convergence speed of w and the divergence rate of w^{-1}. Indeed, it is easy to see that

$$w_t \le D_w^t \quad \text{and} \quad 1/w_t \le L_w^t, \qquad \text{for any } t \in \mathbb{N}. \tag{3.3}$$

Additionally, the fact that for all $t \in \mathbb{N}$ we have that $1 < w_t/w_{t+1}$ and that $0 < w_{t+1}/w_t < 1$ implies that

$$1/\sup_{t\in\mathbb{N}}\left\{\frac{w_t}{w_{t+1}}\right\} = \inf_{t\in\mathbb{N}}\left\{\frac{w_{t+1}}{w_t}\right\} \le \sup_{t\in\mathbb{N}}\left\{\frac{w_{t+1}}{w_t}\right\} \quad \text{and} \quad 1/\sup_{t\in\mathbb{N}}\left\{\frac{w_{t+1}}{w_t}\right\} = \inf_{t\in\mathbb{N}}\left\{\frac{w_t}{w_{t+1}}\right\} \le \sup_{t\in\mathbb{N}}\left\{\frac{w_t}{w_{t+1}}\right\},$$

which, in both cases, implies that

$$L_w D_w \ge 1. \tag{3.4}$$

More generally, in relation with the power weighting sequences that we discussed in Lemma 2.4, we have that:

$$0 < D_{w^n} \le D_w \le D_{w^{1/m}} \le 1 \quad \text{and} \quad 1 < L_{w^{1/m}} \le L_w \le L_{w^n} \le +\infty, \quad \text{for any } m, n \in \mathbb{N}^+. \tag{3.5}$$

Lemma 3.1 *Let w be a weighting sequence and $n \in \mathbb{N}^+$. Then:*

(i) *The projections $p_t : (\ell_-^w(\mathbb{R}^n), \|\cdot\|_w) \longrightarrow (\mathbb{R}^n, \|\cdot\|)$, $t \in \mathbb{Z}_-$, given by $p_t(\mathbf{z}) := \mathbf{z}_t$, $\mathbf{z} \in \ell_-^w(\mathbb{R}^n)$, are linear, smooth, and hence continuous. Moreover, $|||p_t|||_w = 1/w_{-t}$.*

(ii) *Consider the restriction of the time delay operator T_{-t} to $\ell_-^w(\mathbb{R}^n)$ for any $t \in \mathbb{Z}$. We consider two cases. First, if $t < 0$ and the inverse decay ratio L_w of w is finite, then T_{-t} maps into $\ell_-^w(\mathbb{R}^n)$, that is, $\ell_-^w(\mathbb{R}^n)$ is T_{-t}-invariant and $T_{-t} : (\ell_-^w(\mathbb{R}^n), \|\cdot\|_w) \longrightarrow (\ell_-^w(\mathbb{R}^n), \|\cdot\|_w)$ is surjective, open, and a submersion, that is, $\ker T_{-t}$ is a split subspace of $\ell_-^w(\mathbb{R}^n)$. If $t > 0$, then $\ell_-^w(\mathbb{R}^n)$ is always T_{-t}-invariant. $T_{-t} : (\ell_-^w(\mathbb{R}^n), \|\cdot\|_w) \longrightarrow (\ell_-^w(\mathbb{R}^n), \|\cdot\|_w)$ is in that case an immersion, that is, it is injective and its image $\operatorname{Im} T_{-t}$ is split. Moreover, for any $t > 0$, $T_t \circ T_{-t} = \mathbb{I}_{\ell_-^w(\mathbb{R}^n)}$, and in both cases the maps T_{-t} are linear, smooth, and hence continuous. Additionally,*

$$|||T_1|||_w = L_w, \quad |||T_{-1}|||_w = D_w, \quad |||T_{-t}|||_w \le L_w^{-t}, \quad \textit{and} \quad |||T_t|||_w \le D_w^{-t}, \quad \textit{for all } t \in \mathbb{Z}_-. \tag{3.6}$$

(iii) *For any $t_1, t_2 \in \mathbb{Z}_-$ we have*

$$p_{t_1+t_2} = p_{t_1} \circ T_{-t_2} = p_{t_2} \circ T_{-t_1}. \tag{3.7}$$

These statements also hold true when $(\ell_-^w(\mathbb{R}^n), \|\cdot\|_w)$ is replaced by $(\ell_-^\infty(\mathbb{R}^n), \|\cdot\|_\infty)$. In that case one has to take as sequence w the constant sequence w^ι given by $w_t^\iota := 1$, for all $t \in \mathbb{N}$, and L_w is replaced by the constant 1.

Remark 3.2 The decay ratios are easy to compute for many families of weighting sequences. Two cases that we frequently encounter are:

(i) Geometric sequence: $w_t := \lambda^t$, $t \in \mathbb{N}$, with $0 < \lambda < 1$. In this case:

$$L_w := \sup_{t\in\mathbb{N}} \left\{ \frac{\lambda^t}{\lambda^{t+1}} \right\} = \frac{1}{\lambda} > 1 \quad \text{and} \quad D_w := \sup_{t\in\mathbb{N}} \left\{ \frac{\lambda^{t+1}}{\lambda^t} \right\} = \lambda < 1.$$

(ii) Harmonic sequence: $w_t := 1/(1+td)$, $t \in \mathbb{N}$, with $d > 0$. In this case $D_w = 1$ and $L_w = 1 + d$.

We emphasize that the finiteness of the inverse decay ratio is not guaranteed for all weighting sequences. An example that illustrates this fact is the sequence $w_t := \exp(-t^2)$. It is easy to verify that in that case $L_w = +\infty$ and $D_w = 1/e$.

Remark 3.3 The inequalities (3.6) can be combined with Gelfand's formula [Lax 02, page 195] to provide bounds for the spectral radii $\rho(T_{-t})$ and $\rho(T_t)$ for all $t \in \mathbb{Z}_-$. Indeed,

$$\rho(T_{-t}) = \lim_{n\to\infty} |||T_{-t}^n|||_w^{1/n} \leq \lim_{n\to\infty} (L_w^{-tn})^{1/n} = L_w^{-t}, \quad \text{with } t \in \mathbb{Z}_-.$$

Analogously, one shows that $\rho(T_t) \leq D_w^{-t}$.

Proof of Lemma 3.1. **(i)** The linearity of p_t is obvious. Let $\mathbf{u} \in \ell_-^w(\mathbb{R}^n)$ arbitrary. Since $\|p_t(\mathbf{u})\| = \|\mathbf{u}_t\| = \|\mathbf{u}_t\| \, w_{-t}/w_{-t} \leq \sup_{j\in\mathbb{Z}_-} \{\|\mathbf{u}_j\| \, w_{-j}\} / w_{-t} = \|\mathbf{u}\|_w / w_{-t}$, we can conclude that $|||p_t|||_w \leq 1/w_{-t}$. Let now $\mathbf{v} \in \mathbb{R}^n$ such that $\|\mathbf{v}\| = 1$ and define the element $\mathbf{z} \in \ell_-^w(\mathbb{R}^n)$ by $\mathbf{z}_t := \mathbf{v}/w_{-t}$, for all $t \in \mathbb{Z}_-$. It is clear that $\|\mathbf{z}\|_w = 1$ and that $\|p_t(\mathbf{z})\| / \|\mathbf{z}\|_w = 1/w_{-t}$, which shows that $|||p_t|||_w = 1/w_{-t}$, as required.

(ii) We first prove the statements in this part in the case $t < 0$. Suppose that the inverse decay ratio L_w is finite and let $\mathbf{u} \in \ell_-^w(\mathbb{R}^n)$ arbitrary. Then

$$\begin{aligned} \|T_1(\mathbf{u})\|_w &= \sup_{t\in\mathbb{Z}_-} \{\|\mathbf{u}_{t-1}\| \, w_{-t}\} = \sup_{t\in\mathbb{Z}_-} \left\{ \|\mathbf{u}_{t-1}\| \, w_{-(t-1)} \frac{w_{-t}}{w_{-(t-1)}} \right\} \\ &\leq \sup_{t\in\mathbb{Z}_-} \{\|\mathbf{u}_{t-1}\| \, w_{-(t-1)}\} \sup_{t\in\mathbb{Z}_-} \left\{ \frac{w_{-t}}{w_{-(t-1)}} \right\} \leq \|\mathbf{u}\|_w \, L_w. \end{aligned} \tag{3.8}$$

This inequality shows that T_1 maps $\ell_-^w(\mathbb{R}^n)$ into $\ell_-^w(\mathbb{R}^n)$ and that $|||T_1|||_w \leq L_w$. Given that for any $t \in \mathbb{Z}_-$ we can write

$$T_{-t} = \underbrace{T_1 \circ \cdots \circ T_1}_{-t \text{ times}},$$

the previous conclusion also proves that T_{-t} maps $\ell_-^w(\mathbb{R}^n)$ into $\ell_-^w(\mathbb{R}^n)$ and that $|||T_{-t}|||_w = |||T_1 \circ \cdots \circ T_1|||_w \leq |||T_1|||_w \cdots |||T_1|||_w \leq L_w^{-t}$. It remains to be shown that $|||T_1|||_w = L_w$. In order to do so, take an element $\mathbf{v} \in \mathbb{R}^n$ such that $\|\mathbf{v}\| = 1$ and define the element $\mathbf{u} \in \ell_-^w(\mathbb{R}^n)$ by $\mathbf{u}_t := \mathbf{v}/w_{-t}$, for all $t \in \mathbb{Z}_-$. Notice that by construction $\|\mathbf{u}\|_w = 1$ and, moreover,

$$\frac{\|T_1(\mathbf{u})\|_w}{\|\mathbf{u}\|_w} = \sup_{t\in\mathbb{Z}_-} \{\|\mathbf{u}_{t-1}\| \, w_{-t}\} = \sup_{t\in\mathbb{Z}_-} \left\{ \frac{\|\mathbf{v}\|}{w_{-(t-1)}} w_{-t} \right\} = L_w,$$

which proves the required identity.

We now show that $T_{-t} : (\ell_-^w(\mathbb{R}^n), \|\cdot\|_w) \longrightarrow (\ell_-^w(\mathbb{R}^n), \|\cdot\|_w)$ is surjective. Indeed, it is clear that for any $\mathbf{u} \in \ell_-^w(\mathbb{R}^n)$, the element

$$\widetilde{\mathbf{u}} := (\mathbf{u}, \underbrace{\mathbf{0}, \ldots, \mathbf{0}}_{-t \text{ times}}) \quad \text{is such that} \quad T_{-t}(\widetilde{\mathbf{u}}) = \mathbf{u}.$$

We hence just need to show that $\widetilde{\mathbf{u}} \in \ell_-^w(\mathbb{R}^n)$. This is the case because

$$\begin{aligned}\|\widetilde{\mathbf{u}}\|_w &= \sup_{s\in\mathbb{Z}_-}\{\|\widetilde{\mathbf{u}_s}\|\, w_{-s}\} = \sup_{s\in\mathbb{Z}_-}\{\|\mathbf{u}_{s+t}\|\, w_{-s}\} = \sup_{s\in\mathbb{Z}_-}\left\{\|\mathbf{u}_{s+t}\|\, w_{-(s+t)}\frac{w_{-s}}{w_{-(s+t)}}\right\}\\ &= \sup_{s\in\mathbb{Z}_-}\left\{\|\mathbf{u}_{s+t}\|\, w_{-(s+t)}\frac{w_{-s}}{w_{-(s+1)}}\frac{w_{-(s+1)}}{w_{-(s+2)}}\cdots\frac{w_{-(s+t-1)}}{w_{-(s+t)}}\right\} \leq \|\mathbf{u}\|_w\, L_w^{-t} < +\infty, \end{aligned} \tag{3.9}$$

because $\mathbf{u} \in \ell_-^w(\mathbb{R}^n)$ and by hypothesis $L_w < +\infty$. Now, since we already showed that $T_{-t} : (\ell_-^w(\mathbb{R}^n), \|\cdot\|_w) \longrightarrow (\ell_-^w(\mathbb{R}^n), \|\cdot\|_w)$ is continuous, then the Banach-Schauder Open Mapping Theorem [Abra 88, Theorem 2.2.15] implies that T_{-t} is necessarily an open map.

It remains to be shown that $T_{-t} : (\ell_-^w(\mathbb{R}^n), \|\cdot\|_w) \longrightarrow (\ell_-^w(\mathbb{R}^n), \|\cdot\|_w)$ is a submersion (see [Abra 88, Section 3.5] for context and definitions). First, it is obvious that

$$\ker T_{-t} = \left\{(\ldots, \mathbf{0}, \mathbf{0}, \mathbf{v}) \mid \mathbf{v} \in (\mathbb{R}^n)^{-t}\right\}.$$

Since T_{-t} is linear and bounded, in order to show that it is a submersion it suffices to show that $\ker T_{-t}$ is split, that is, it has a closed complement in $(\ell_-^w(\mathbb{R}^n), \|\cdot\|_w)$. We now prove that such a complement is given by the subspace

$$C_{-t} := \left\{(\mathbf{u}, \underbrace{\mathbf{0}, \ldots, \mathbf{0}}_{-t \text{ times}}) \mid \mathbf{u} \in \ell_-^w(\mathbb{R}^n)\right\}. \tag{3.10}$$

The inequality (3.9) implies that

$$C_{-t} \subset \ell_-^w(\mathbb{R}^n). \tag{3.11}$$

Additionally, C_{-t} is clearly closed in $\ell_-^w(\mathbb{R}^n)$. We conclude by showing by double inclusion that

$$\ell_-^w(\mathbb{R}^n) = \ker T_{-t} \oplus C_{-t}. \tag{3.12}$$

Let first $\mathbf{u} \in \ell_-^w(\mathbb{R}^n)$ and define

$$\mathbf{u}_1 := (\ldots, \mathbf{u}_{t-2}, \mathbf{u}_{t-1}, \mathbf{u}_t, \underbrace{\mathbf{0}, \ldots, \mathbf{0}}_{-t \text{ times}}) \quad \text{and} \quad \mathbf{u}_2 := (\ldots, \mathbf{0}, \mathbf{0}, \mathbf{u}_{t+1}, \mathbf{u}_{t+2}, \ldots, \mathbf{u}_0).$$

It is clear that $\mathbf{u} = \mathbf{u}_1 + \mathbf{u}_2$. Additionally, the sequence $\mathbf{u}_2$ is obviously in $\ker T_{-t}$ and using an argument similar to the one in (3.9) it is easy to show that $\mathbf{u}_1 \in C_{-t}$, which proves the inclusion $\ell_-^w(\mathbb{R}^n) \subseteq \ker T_{-t} \oplus C_{-t}$.

Conversely, let $\mathbf{u}_1 \in C_{-t}$ and $\mathbf{u}_2 \in \ker T_{-t}$. By (3.11) we have that $\|\mathbf{u}_1\|_w < +\infty$ and it is also clear that $\|\mathbf{u}_2\|_w < +\infty$. Therefore $\|\mathbf{u}_1 + \mathbf{u}_2\|_w \leq \|\mathbf{u}_1\|_w + \|\mathbf{u}_2\|_w < +\infty$ and hence $\mathbf{u}_1 + \mathbf{u}_2 \in \ell_-^w(\mathbb{R}^n)$, which shows that T_{-t} is a submersion.

Finally, the statements in the case $t > 0$ are proved in a similar fashion. In particular, it is easy to see that

$$T_{-t} \circ T_t = \mathbb{P}_{C_t}, \quad \text{for any } t > 0,$$

where $\mathbb{P}_{C_t}$ is the projection onto the subspace C_t defined in (3.10) according to the splitting (3.12). Moreover, it is easy to see that T_{-t} is injective and that its image $\operatorname{Im} T_{-t}$ is split because $\operatorname{Im} T_{-t} = C_t$ and by (3.12)

$$\ell_-^w(\mathbb{R}^n) = \ker T_t \oplus \operatorname{Im} T_{-t}, \quad t > 0,$$

which proves that T_{-t} is an immersion.

(iii) Straightforward consequence of the definitions.

The proofs for the space $(\ell_-^\infty(\mathbb{R}^n), \|\cdot\|_\infty)$ can be obtained by replacing in the previous arguments the weighting sequence w by the constant sequence w^ι. ■

Remark 3.4 Lemma 3.1 remains valid when instead of the spaces $\ell_-^w(\mathbb{R}^n)$ we use the spaces $\ell_-^{p,w}(\mathbb{R}^n)$ that we introduced in Section 2, for any $1 \leq p < +\infty$. In that case, and for any $t \in \mathbb{Z}_-$,

$$|||p_t|||_{p,w} = \frac{1}{w_{-t}^{1/p}}, \tag{3.13}$$

$$|||T_1|||_{p,w} = L_{w,p}, \ |||T_{-1}|||_{p,w} = D_{w,p}, \ |||T_{-t}|||_{p,w} \leq L_{w,p}^{-t}, \text{ and } |||T_t|||_{p,w} \leq D_{w,p}^{-t}, \text{ for all } t \in \mathbb{Z}_-. \tag{3.14}$$

where $D_{w,p}$ and $L_{w,p}$ are adapted versions to these norms of the decay ratio D_w and L_w, respectively, given by

$$D_{w,p} := \sup_{t \in \mathbb{N}} \left\{ \left(\frac{w_{t+1}}{w_t} \right)^{1/p} \right\} \quad \text{and} \quad L_{w,p} := \sup_{t \in \mathbb{N}} \left\{ \left(\frac{w_t}{w_{t+1}} \right)^{1/p} \right\}. \tag{3.15}$$

Indeed, (3.13) can be obtained by noting that for any $\mathbf{u} \in \ell_-^{p,w}(\mathbb{R}^n)$,

$$\|p_t(\mathbf{u})\|^p = \|\mathbf{u}_t\|^p = \frac{\|\mathbf{u}_t\|^p w_{-t}}{w_{-t}} \leq \frac{1}{w_{-t}} \sum_{s \in \mathbb{Z}_-} \|\mathbf{u}_s\|^p w_{-s} = \frac{\|\mathbf{u}\|_{p,w}^p}{w_{-t}},$$

which shows that $|||p_t|||_{p,w} \leq 1/w_{-t}^{1/p}$. Consider now the vector $\mathbf{v} \in \ell_-^{p,w}(\mathbb{R}^n)$ given by $\mathbf{v}_s := \delta_{s,t}\widetilde{\mathbf{v}}/w_{-t}^{1/p}$, where $s \in \mathbb{Z}_-$, the symbol $\delta_{s,t}$ stands for Kronecker's delta, and $\widetilde{\mathbf{v}} \in \mathbb{R}^n$ is such that $\|\widetilde{\mathbf{v}}\| = 1$. Notice that $\|\mathbf{v}\|_{p,w}^p = \sum_{s \in \mathbb{Z}_-} \|\mathbf{v}_s\|^p w_{-s} = \|\widetilde{\mathbf{v}}\| = 1$ and given that $\|p_t(\mathbf{v})\| = \|\widetilde{\mathbf{v}}\| / w_{-t}^{1/p} = 1/w_{-t}^{1/p}$, this implies that

$$|||p_t|||_{p,w} = \sup_{\|\mathbf{u}\|_{p,w}=1} \{\|p_t(\mathbf{u})\|\} = \frac{1}{w_{-t}^{1/p}},$$

as required. Regarding (3.14), we only sketch the proof for positive time shifts. As in the proof of Lemma 3.1, it suffices to show that $|||T_1|||_{p,w} = L_{w,p}$. In order to prove this equality, notice first that for any $t \in \mathbb{Z}_-$, the following straightforward inequality holds

$$\frac{w_{-t}}{w_{-(t-1)}} = \left(\left(\frac{w_{-t}}{w_{-(t-1)}} \right)^{1/p} \right)^p \leq \left(\sup_{s \in \mathbb{Z}_-} \left\{ \left(\frac{w_{-s}}{w_{-(s-1)}} \right)^{1/p} \right\} \right)^p = L_{w,p}^p.$$

Now, for any $\mathbf{u} \in \ell_-^{p,w}(\mathbb{R}^n)$,

$$\|T_1(\mathbf{u})\|_{p,w} = \left(\sum_{t \in \mathbb{Z}_-} \|\mathbf{u}_{t-1}\|^p w_{-t} \right)^{1/p} = \left(\sum_{t \in \mathbb{Z}_-} \|\mathbf{u}_{t-1}\|^p w_{-(t-1)} \frac{w_{-t}}{w_{-(t-1)}} \right)^{1/p}$$
$$\leq \left(\sum_{t \in \mathbb{Z}_-} \|\mathbf{u}_{t-1}\|^p w_{-(t-1)} L_{w,p}^p \right)^{1/p} \leq \|\mathbf{u}\|_{p,w} L_{w,p},$$

which proves that $|||T_1|||_{p,w} \leq L_{w,p}$. In order to establish the equality we prove the reverse inequality by considering the family of vectors $\mathbf{v}^t \in \ell_-^{p,w}(\mathbb{R}^n)$, $t \in \mathbb{Z}_-$ defined by $\mathbf{v}_s^t := \delta_{s,t}\widetilde{\mathbf{v}}^t/w_{-(t-1)}^{1/p}$, where $\widetilde{\mathbf{v}}^t \in \mathbb{R}^n$ is such that $\|\widetilde{\mathbf{v}}^t\| = \left(w_{-(t-1)}/w_{-t}\right)^{1/p}$. Notice that for all $t \in \mathbb{Z}_-$,

$$\|\mathbf{v}^t\|_{p,w}^p = \sum_{s \in \mathbb{Z}_-} \|\mathbf{v}_s^t\|^p w_{-s} = \frac{\|\widetilde{\mathbf{v}}^t\|^p}{w_{-(t-1)}} w_{-t} = \frac{w_{-(t-1)}}{w_{-t}} \frac{w_{-t}}{w_{-(t-1)}} = 1 \quad \text{and} \quad \left\|T_1(\mathbf{v}^t)\right\|_{p,w}^p = \frac{w_{-(t+1)}}{w_{-t}} \geq 1,$$

which implies that

$$|||T_1|||_{p,w} = \sup_{\|\mathbf{u}\|_{p,w}=1} \left\{ \|T_1(\mathbf{u})\|_{p,w} \right\} \geq \sup_{t\,\in\,\mathbb{Z}_-} \left\{ \left\|T_1(\mathbf{v}^t)\right\|_{p,w} \right\} = \sup_{t\in\mathbb{Z}_-} \left\{ \left(\frac{w_{-(t+1)}}{w_{-t}} \right)^{1/p} \right\}, = L_{w,p}$$

which proves the required inequality.

Remark 3.5 Some of the properties of time delays operators that we just studied have interesting interpretations in a Hilbert space context. See [Lind 15] for a detailed study.

3.1 The fading memory property and remote past input independence

The properties of time delay operators that we enunciated in Lemma 3.1 allow us to show how the fading memory property, defined as the continuity of a filter linking input and output spaces endowed with weighted norms, (see Section 2.2) can be interpreted as its asymptotic independence on the remote past input [Wien 58, page 89]. Analogously, we can see that the FMP amounts to the attribute that, in the words of Volterra [Volt 30, page 188], the influence of the input a long time before the given moment fades out. This property has also been characterized as a ***unique steady-state property*** in [Boyd 85] and referred to as the ***input forgetting property*** in [Jaeg 10]. All these characterizations were proved under various compactness and/or uniformly boundedness hypotheses on the inputs. The next result shows that property as a straightforward corollary of Lemma 3.1 that, later on in Section 5.3, will be generalized to situations where the inputs are eventually unbounded.

In the following statement we will be using the following notation: given the sequences $\mathbf{u} \in (\mathbb{R}^n)^{\mathbb{Z}_-}$ and $\mathbf{v} \in (\mathbb{R}^n)^t$, $t \in \mathbb{N}$, the symbol $(\mathbf{u}, \mathbf{v}) \in (\mathbb{R}^n)^{\mathbb{Z}_-} \times (\mathbb{R}^n)^t$ denotes the ***concatenation*** of $\mathbf{u}$ and $\mathbf{v}$.

Theorem 3.6 (FMP and the uniform input forgetting property) *Let $M, L > 0$, $n, N \in \mathbb{N}^+$ and let $K_M \subset (\mathbb{R}^n)^{\mathbb{Z}_-}$, $K_L \subset (\mathbb{R}^M)^{\mathbb{Z}_-}$ (respectively, $K_M^+ \subset (\mathbb{R}^n)^{\mathbb{N}^+}$, $K_L^+ \subset (\mathbb{R}^M)^{\mathbb{N}^+}$) be the sets of uniformly bounded left (respectively, right) semi-infinite sequences defined in (1.3). Let $U : K_M \longrightarrow U_L$ be a causal and time-invariant fading memory filter. Then, for any $\mathbf{u}, \mathbf{v} \in K_M$ and $\mathbf{z} \in K_M^+$ we have that*

$$\lim_{t\to+\infty} \|U(\mathbf{u}, \mathbf{z})_t - U(\mathbf{v}, \mathbf{z})_t\| = 0, \tag{3.16}$$

where in this expression the filter U is defined by time-invariance on positive times using (3.1). The convergence in (3.16) is uniform on $\mathbf{u}, \mathbf{v}$, and $\mathbf{z}$ in the sense that there exists a monotonously decreasing sequence w^U with zero limit such that for all $\mathbf{u}, \mathbf{v} \in K_M$, $\mathbf{z} \in K_M^+$, and $t \in \mathbb{N}$,

$$\|U(\mathbf{u}, \mathbf{z})_t - U(\mathbf{v}, \mathbf{z})_t\| \leq w_t^U. \tag{3.17}$$

*Filters that satisfy condition (3.16) for any $\mathbf{u}, \mathbf{v} \in K_M$ and $\mathbf{z} \in K_M^+$ are said to have the **input forgetting property** and we refer to (3.17) as the **uniform input forgetting property**.*

Proof. We start by recalling that in the presence of uniformly bounded inputs, the FMP can be characterized as the continuity of the map $U : K_M \longrightarrow K_L$ with the sets K_M and K_L endowed with the relative topology induced either by the product topology on $(\mathbb{R}^n)^{\mathbb{Z}_-}$ and $(\mathbb{R}^N)^{\mathbb{Z}_-}$, respectively, or by the weighted norms in the spaces $\ell_-^w(\mathbb{R}^n)$ and $\ell_-^w(\mathbb{R}^N)$, with w any weighting sequence (see [Grig 18a, Corollary 2.7 and Proposition 2.11]). Moreover, the sets K_M and K_L are compact in this topology [Grig 18a, Corollay 2.8] and hence the FMP filter $U : K_M \longrightarrow K_L$ is not only continuous but also uniformly continuous. Consequently, once we have fixed a weighting sequence w, an increasing modulus of continuity $\omega_U : \mathbb{R}^+ \longrightarrow \mathbb{R}^+$ can be associated to the map $U : (K_M, \|\cdot\|_w) \longrightarrow (K_L, \|\cdot\|_w)$. We emphasize that ω_U depends on w since it is a metric and not a purely topological notion. Now, using (3.1) and an arbitrary weighting sequence w that we choose with $D_w < 1$, we can write for any $t \in \mathbb{N}$

$$\begin{aligned}\|U(\mathbf{u},\mathbf{z})_t - U(\mathbf{v},\mathbf{z})_t\| &= \left\|U\left(\mathbb{P}_{\mathbb{Z}_-}\left(T^{\mathbb{Z}}_{-t}(\mathbf{u},\mathbf{z})\right)\right)_0 - U\left(\mathbb{P}_{\mathbb{Z}_-}\left(T^{\mathbb{Z}}_{-t}(\mathbf{v},\mathbf{z})\right)\right)_0\right\| \\ &= \left\|p_0 \circ U\left(\mathbb{P}_{\mathbb{Z}_-}\left(T^{\mathbb{Z}}_{-t}(\mathbf{u},\mathbf{z})\right)\right) - p_0 \circ U\left(\mathbb{P}_{\mathbb{Z}_-}\left(T^{\mathbb{Z}}_{-t}(\mathbf{v},\mathbf{z})\right)\right)\right\| \\ &\leq \left\|U\left(\mathbb{P}_{\mathbb{Z}_-}\left(T^{\mathbb{Z}}_{-t}(\mathbf{u},\mathbf{z})\right)\right) - U\left(\mathbb{P}_{\mathbb{Z}_-}\left(T^{\mathbb{Z}}_{-t}(\mathbf{v},\mathbf{z})\right)\right)\right\|_w, \end{aligned} \tag{3.18}$$

where we used that $|||p_0|||_w = 1$ by the first part of Lemma 3.1. We now notice that

$$\mathbb{P}_{\mathbb{Z}_-}\left(T^{\mathbb{Z}}_{-t}(\mathbf{u},\mathbf{z})\right) = T_{-t}(\mathbf{u}) + (\ldots,\mathbf{0},\mathbf{z}_1,\ldots,\mathbf{z}_t), \quad \text{and} \quad \mathbb{P}_{\mathbb{Z}_-}\left(T^{\mathbb{Z}}_{-t}(\mathbf{v},\mathbf{z})\right) = T_{-t}(\mathbf{v}) + (\ldots,\mathbf{0},\mathbf{z}_1,\ldots,\mathbf{z}_t),$$

which substituted in (3.18) and using the second part of Lemma 3.1 yields

$$\begin{aligned}\|U(\mathbf{u},\mathbf{z})_t - U(\mathbf{v},\mathbf{z})_t\| &\leq \omega_U\left(\|T_{-t}(\mathbf{u}-\mathbf{v})\|_w\right) \\ &\leq \omega_U\left(|||T_{-t}|||_w \|\mathbf{u}-\mathbf{v}\|_w\right) \leq \omega_U\left(D_w^t \|\mathbf{u}-\mathbf{v}\|_w\right) \leq \omega_U\left(2MD_w^t\right). \end{aligned} \tag{3.19}$$

Now, as w has been chosen so that $D_w < 1$ and $\lim_{t\to 0} \omega_U(t) = 0$, we set $w_t^U := \omega_U\left(2MD_w^t\right)$, and we have that

$$\lim_{t\to+\infty} w_t^U = \lim_{t\to+\infty} \omega_U\left(2MD_w^t\right) = 0, \tag{3.20}$$

which using the inequality (3.19) proves the claim. ■

3.2 Equivalence of FMP and differentiability in filters and functionals

The facts established in Lemma 3.1 can be used to show the equivalence between the continuity and the differentiability of causal and time-invariant filters and that of their associated functionals. The following result focuses on continuity and the fading memory property and generalizes to the context of eventually unbounded inputs the equivalence between fading memory filters and functionals established in [Grig 18a, Propositions 2.11 and 2.12] for uniformly bounded inputs. In the results that follow we work in a setup slightly more general than the one that is customary in the literature as we will allow for the weighting sequences considered in the domain and the target of the filters to be different. This degree of generality is needed later on in the text.

Proposition 3.7 *Let $V_n \subset (\mathbb{R}^n)^{\mathbb{Z}_-}$ and $V_N \subset \left(\mathbb{R}^N\right)^{\mathbb{Z}_-}$ be time-invariant subsets and let $D_N \subset \mathbb{R}^N$. Let w^1, w^2 be weighting sequences with inverse decay ratios L_{w^1} and L_{w^2}, respectively.*

(i) *Let $U : V_n \subset \ell_-^{w^1}(\mathbb{R}^n) \longrightarrow V_N \subset \ell_-^{w^2}(\mathbb{R}^N)$ be a causal and time-invariant filter. If U has the fading memory property then so does its associated functional $H_U : V_n \longrightarrow p_0(V_N)$. The same conclusion holds for continuous filters $U : V_n \subset \ell_-^{\infty}(\mathbb{R}^n) \longrightarrow V_N \subset \ell_-^{\infty}(\mathbb{R}^N)$.*

(ii) *Let $H : V_n \subset \ell_-^{w^1}(\mathbb{R}^n) \longrightarrow D_N$ be a fading memory functional. If L_{w^1} is finite and D_N is compact then the associated causal and time-invariant filter $U_H : V_n \subset \ell_-^{w^1}(\mathbb{R}^n) \longrightarrow (D_N)^{\mathbb{Z}_-} \subset \ell_-^{w^2}(\mathbb{R}^N)$ has also the fading memory property.*

(iii) *Let $H : V_n \subset \ell_-^{w^1}(\mathbb{R}^n) \longrightarrow D_N$ be a fading memory functional and suppose that V_n contains a point $\mathbf{z}^0$ such that $U_H(\mathbf{z}^0) \in \ell_-^{w^2}(\mathbb{R})$, where U_H is the causal and time-invariant filter associated to H. If H is Lipschitz, c_H is a Lipschitz constant, and the weighting sequences satisfy one of the following two conditions*

$$\textit{either } R_{w^1,w^2} := \sup_{s,t\in\mathbb{N}} \left\{\frac{w_t^1 w_s^2}{w_{t+s}^1}\right\} < +\infty \textit{ or the sequence } \mathcal{L}_{w^1} := \left(L_{w^1}^{-t}\right)_{t\in\mathbb{Z}_-} \in \ell_-^{w^2}(\mathbb{R}), \tag{3.21}$$

then $U_H : V_n \subset \ell_-^{w^1}(\mathbb{R}^n) \longrightarrow (D_N)^{\mathbb{Z}_-} \cap \ell_-^{w^2}(\mathbb{R}^N)$ *has also the fading memory property, it is Lipschitz, and* $R_{w^1,w^2} c_H$ *or* $\|\mathcal{L}_{w^1}\|_{w^2} c_H$*, respectively, is a Lipschitz constant of* U_H*. The same conclusion holds for continuous functionals* $H : V_n \subset \ell_-^{\infty}(\mathbb{R}^n) \longrightarrow D_N$ *where the condition (3.21) is not needed.*

Remark 3.8 When in part **(iii)** we consider the same weighting sequence w for the domain and the target, it is easy to see that

$$R_w := \sup_{s,t\in\mathbb{N}} \left\{ \frac{w_t w_s}{w_{t+s}} \right\}$$

satisfies that $R_w \leq \|\mathcal{L}_w\|_w$ and therefore the second condition in (3.21) implies the first one. Indeed,

$$R_w = \sup_{s,t\in\mathbb{N}} \left\{ \frac{w_t w_s}{w_{t+s}} \right\} = \sup_{s,t\in\mathbb{N}} \left\{ \frac{w_t}{w_{t+1}} \frac{w_{t+1}}{w_{t+2}} \cdots \frac{w_{t+s-1}}{w_{t+s}} w_s \right\} \leq \sup_{s\in\mathbb{N}} \{ L_w^s w_s \} = \|\mathcal{L}_w\|_w , \quad \text{as required.}$$

In this setup, the condition (3.21) is satisfied by many families of commonly used weighting sequences. In the two examples considered in Remark 3.2 we have that $R_w = \|\mathcal{L}_w\|_w = 1$ for the geometric sequence; for the harmonic sequence $\|\mathcal{L}_w\|_w = +\infty$ but $R_w = 1$ and hence (3.21) is still satisfied.

We emphasize that condition (3.21) is not automatically satisfied by all weighting sequences. For example, as we saw in Remark 3.2, the sequence $w_t := \exp(-t^2)$ is such that $L_w = +\infty$ and, additionally, it is easy to see that $R_w := \sup_{s,t\in\mathbb{N}} \{\exp(2st)\} = +\infty$.

Proof. **(i)** As H_U is given by $H_U = p_0 \circ U$, the FMP (respectively, continuity) of U and the first part of Lemma 3.1 prove the statement.

(ii) Notice first that as

$$U_H = \prod_{t\in\mathbb{Z}_-} H \circ T_{-t} \tag{3.22}$$

then, as L_{w^1} is finite, U_H is by the second part of Lemma 3.1 the Cartesian product of continuous functions $H_t := H \circ T_{-t} : V_n \subset \ell_-^{w^1}(\mathbb{R}^n) \longrightarrow D_N$. Since D_N is by hypothesis compact, the result follows from the first part of Lemma 2.5.

(iii) Let $\mathbf{z}^1, \mathbf{z}^2 \in V_n$ arbitrary. Then by (3.22) and the Lipschitz hypothesis on H, we have that

$$\begin{aligned} \left\|U_H(\mathbf{z}^1) - U_H(\mathbf{z}^2)\right\|_{w^2} &= \sup_{t\in\mathbb{Z}_-} \left\{ \left\|H(T_{-t}(\mathbf{z}^1)) - H(T_{-t}(\mathbf{z}^2))\right\| w_{-t}^2 \right\} \\ &\leq c_H \sup_{t\in\mathbb{Z}_-} \left\{ \left\|T_{-t}(\mathbf{z}^1) - T_{-t}(\mathbf{z}^2)\right\|_{w^1} w_{-t}^2 \right\}. \end{aligned} \tag{3.23}$$

If the first condition in (3.21) is satisfied, this expression is bounded above by

$$\begin{aligned} c_H \sup_{t,s\in\mathbb{Z}_-} \left\{ \left\|T_{-t}(\mathbf{z}^1)_s - T_{-t}(\mathbf{z}^2)_s\right\| w_{-t}^2 w_{-s}^1 \right\} &= c_H \sup_{t,s\in\mathbb{Z}_-} \left\{ \left\|\mathbf{z}_{t+s}^1 - \mathbf{z}_{t+s}^2\right\| w_{-(t+s)}^1 \frac{w_{-t}^2 w_{-s}^1}{w_{-(t+s)}^1} \right\} \\ &\leq R_{w^1,w^2} c_H \left\|\mathbf{z}^1 - \mathbf{z}^2\right\|_{w^1} \end{aligned}$$

which proves that in that case U_H has the fading memory property, it is Lipschitz, and $R_{w^1,w^2} c_H$ is a Lipschitz constant. If the second condition in (3.21) is satisfied then the inverse decay ratio L_{w^1} is necessarily finite and hence (3.23) can be bounded using the second part of Lemma 3.1 as

$$c_H \sup_{t\in\mathbb{Z}_-} \left\{ \left\|T_{-t}(\mathbf{z}^1) - T_{-t}(\mathbf{z}^2)\right\|_{w^1} w_{-t}^2 \right\} \leq c_H \left\|\mathbf{z}^1 - \mathbf{z}^2\right\|_{w^1} \sup_{t\in\mathbb{Z}_-} \left\{ L_{w^1}^{-t} w_{-t}^2 \right\} = \|\mathcal{L}_{w^1}\|_{w^2} c_H \left\|\mathbf{z}^1 - \mathbf{z}^2\right\|_{w^1} ,$$

which proves that in that case $\|\mathcal{L}_{w^1}\|_{w^2} c_H$ is a Lipschitz constant of U_H. The Lipschitz continuity of U_H together with the hypothesis on the existence of a point $\mathbf{z}^0$ such that $U_H(\mathbf{z}^0) \in \ell_-^{w^2}(\mathbb{R})$ guarantee that U_H maps into $\ell_-^{w^2}(\mathbb{R}^N)$ using a strategy similar to the one followed in (2.26).

The proof for the spaces $\ell_-^{\infty}(\mathbb{R}^n)$ and $\ell_-^{\infty}(\mathbb{R}^N)$ is obtained by taking as weighting sequences the constant sequence w^ι given by $w_t^\iota := 1$, for all $t \in \mathbb{N}$, that automatically satisfies any of the two conditions in (3.21). ■

Proposition 3.9 *Let w^1 and w^2 be two weighting sequences with inverse decay ratios L_{w^1} and L_{w^1}, respectively. Let $V_n \subset \ell_-^{w^1}(\mathbb{R}^n)$ and $V_N \subset \ell_-^{w^2}(\mathbb{R}^N)$ be time-invariant open subsets, and let D_N be an open subset of $\mathbb{R}^N$.*

(i) *Let $U : V_n \subset \ell_-^{w^1}(\mathbb{R}^n) \longrightarrow V_N \subset \ell_-^{w^2}(\mathbb{R}^N)$ be a causal and time-invariant filter. If U is of class $C^r(V_n)$ (respectively, smooth or analytic) when considered as a map $U : V_n \subset \left(\ell_-^{w^1}(\mathbb{R}^n), \|\cdot\|_w\right) \longrightarrow V_N \subset \left(\ell_-^{w^2}(\mathbb{R}^N), \|\cdot\|_w\right)$, then so is the associated functional $H_U : V_n \subset \left(\ell_-^{w^1}(\mathbb{R}^n), \|\cdot\|_w\right) \longrightarrow p_0(V_N) \subset \mathbb{R}^N$. Moreover,*

$$\left\|\!\left| D^r H_U(\mathbf{z}) \right|\!\right\|_{w^1} \leq \left\|\!\left| D^r U(\mathbf{z}) \right|\!\right\|_{w^1,w^2}, \quad \text{for any } \mathbf{z} \in V_n. \tag{3.24}$$

The same conclusion holds when the weighted sequence spaces are replaced by $\left(\ell_-^{\infty}(\mathbb{R}^n), \|\cdot\|_\infty\right)$ and $\left(\ell_-^{\infty}(\mathbb{R}^N), \|\cdot\|_\infty\right)$.

(ii) *Let $H : V_n \subset \ell_-^{w^1}(\mathbb{R}^n) \longrightarrow D_N$ be a functional and suppose that V_n is convex and contains a point $\mathbf{z}^0$ such that $U_H(\mathbf{z}^0) \in \ell_-^{w^2}(\mathbb{R})$, where U_H is the causal and time-invariant filter associated to H. If the functional H is of class $C^r(V_n)$ and for any $j \in \{1, \ldots, r\}$ we have that $c^j := \sup_{\mathbf{z} \in V_n} \left\{ \left\|\!\left| D^j H(\mathbf{z}) \right|\!\right\|_{w^1} \right\} < +\infty$ and the weighting sequences satisfy that*

$$\mathcal{L}_{w^1,j} := (L_{w^1}^{-jt})_{t \in \mathbb{Z}_-} \in \ell_-^{w^2}(\mathbb{R}), \tag{3.25}$$

then the associated causal and time-invariant filter U_H is differentiable of order r when considered as a map $U_H : V_n \subset \left(\ell_-^{w^1}(\mathbb{R}^n), \|\cdot\|_{w^1}\right) \longrightarrow (D_N)^{\mathbb{Z}_-} \cap \left(\ell_-^{w^2}(\mathbb{R}^N), \|\cdot\|_{w^2}\right)$. Moreover, for any $\mathbf{z} \in V_n$,

$$\left\|\!\left| D^r U_H(\mathbf{z}) \right|\!\right\|_{w^1,w^2} \leq c^r \|\mathcal{L}_{w^1,r}\|_{w^2}. \tag{3.26}$$

Additionally, U_H is of class $C^{r-1}(V_n)$ and the map

$$D^{r-1} U_H : (V_n, \|\cdot\|_{w^1}) \longrightarrow \left(L^{r-1}\left(\ell_-^{w^1}(\mathbb{R}^n), \ell_-^{w^2}(\mathbb{R}^N)\right), \left\|\!\left|\cdot\right|\!\right\|_{w^1,w^2} \right)$$

is Lipschitz continuous with Lipschitz constant $c^r \|\mathcal{L}_{w^1,r}\|_{w^2}$. The same conclusion holds when the weighted sequence spaces are replaced by $\left(\ell_-^{\infty}(\mathbb{R}^n), \|\cdot\|_\infty\right)$ and $\left(\ell_-^{\infty}(\mathbb{R}^N), \|\cdot\|_\infty\right)$. In that case the inequality (3.26) holds with $\|\mathcal{L}_{w^1,r}\|_{w^2} = 1$.

(iii) *Let $H : V_n \subset \ell_-^{w^1}(\mathbb{R}^n) \longrightarrow D_N$ be a functional and suppose that V_n is convex and contains a point $\mathbf{z}^0$ such that $U_H(\mathbf{z}^0) \in \ell_-^{w^2}(\mathbb{R})$, where U_H is the causal and time-invariant filter associated to H. If the functional H is smooth and $c^r < +\infty$ for all $r \in \mathbb{N}^+$, then so is the associated causal and time-invariant filter $U_H : V_n \subset \left(\ell_-^{w^1}(\mathbb{R}^n), \|\cdot\|_{w^1}\right) \longrightarrow (D_N)^{\mathbb{Z}_-} \cap \left(\ell_-^{w^2}(\mathbb{R}^N), \|\cdot\|_{w^2}\right)$. The same conclusion holds when the weighted spaces are replaced by $\left(\ell_-^{\infty}(\mathbb{R}^n), \|\cdot\|_\infty\right)$ and $\left(\ell_-^{\infty}(\mathbb{R}^N), \|\cdot\|_\infty\right)$. In that case, if H is analytic then so is U_H and the radius of convergence of the series expansion of U_H is bigger or equal than that of H.*

Remark 3.10 An important consequence of part **(ii)** in this proposition and, in particular, of the condition (3.25) is that, in general, one cannot obtain (higher order) differentiable filters out of differentiable functionals using the same weighted norm in the domain and the target of the filter. The weighted norm in the target needs to be chosen so that it satisfies the nonautomatic condition (3.25) that, additionally, depends on the differentiability degree that we want to preserve. Weighted norms that satisfy that property are relatively easy to find in most cases. For example, if we take as w^1 the geometric sequence in Remark 3.2, then $\mathcal{L}_{w^1,j} = \left(\lambda^{-jt}\right)_{t\in\mathbb{N}}$ and hence condition (3.25) is satisfied if we take as w^2 any sequence of the type $\left(w^1\right)^r$ (using the notation in Lemma 2.4) with $r \geq j$.

Proof. **(i)** Recall first that H_U can be written as $H_U = p_0 \circ U$. The chain rule and the linearity of the projection p_0 imply that $D^r H_U(\mathbf{z}) = p_0 \circ D^r U(\mathbf{z})$ for any $\mathbf{z} \in V_n$. The first part of Lemma 3.1 guarantees then that H_U is of class $C^r(V_n)$ and that

$$|||D^r H_U(\mathbf{z})|||_{w^1} = |||p_0 \circ D^r U(\mathbf{z})|||_{w^1} \leq |||p_0|||_{w^2} \cdot |||D^r U(\mathbf{z})|||_{w^1,w^2} = |||D^r U(\mathbf{z})|||_{w^1,w^2}, \text{ for any } \mathbf{z} \in V_n,$$

as required. The proof for the spaces $\ell_-^\infty(\mathbb{R}^n)$ and $\ell_-^\infty(\mathbb{R}^N)$ is obtained by taking as sequence w the constant sequence w^ι given by $w_t^\iota := 1$, for all $t \in \mathbb{N}$.

(ii) First of all, notice that the hypothesis on c^1 and the convexity of V_n imply via the mean value theorem [Abra 88] that H is Lipschitz. Moreover, the hypothesis on $\mathcal{L}_{w,1}$ in the statement implies that condition (3.21) is satisfied and hence the third part in Proposition 3.7 guarantees that U_H maps into $\ell_-^{w^2}(\mathbb{R}^N)$.

Now, the expression (3.22) implies that for any $\mathbf{z} \in V_n$,

$$D^r U_H(\mathbf{z}) = \prod_{t\in\mathbb{Z}_-} D^r H(T_{-t}(\mathbf{z})) \circ \underbrace{(T_{-t}, \ldots, T_{-t})}_{r \text{ times}}, \quad r \geq 1. \tag{3.27}$$

In order to prove (3.24) consider $\mathbf{u}^1, \ldots, \mathbf{u}^r \in \ell_-^{w^1}(\mathbb{R}^n)$ arbitrary and notice that by the second part of Lemma 3.1 we have

$$\begin{aligned}
\left\|D^r U_H(\mathbf{z})\left(\mathbf{u}^1, \ldots, \mathbf{u}^r\right)\right\|_{w^2} &= \sup_{t\in\mathbb{Z}_-} \left\{\left\|D^r H(T_{-t}(\mathbf{z})) \cdot \left(T_{-t}(\mathbf{u}^1), \ldots, T_{-t}(\mathbf{u}^r)\right)\right\| w^2_{-t}\right\} \\
&\leq c^r \sup_{t\in\mathbb{Z}_-} \left\{\left\|T_{-t}(\mathbf{u}^1)\right\|_{w^1} \cdots \left\|T_{-t}(\mathbf{u}^r)\right\|_{w^1} w^2_{-t}\right\} \\
&\leq c^r \sup_{t\in\mathbb{Z}_-} \left\{\left\|\mathbf{u}^1\right\|_{w^1} \cdots \left\|\mathbf{u}^r\right\|_{w^1} L_{w^1}^{-rt} w^2_{-t}\right\} \leq c^r \left\|\mathcal{L}_{w^1,r}\right\|_{w^2} \left\|\mathbf{u}^1\right\|_{w^1} \cdots \left\|\mathbf{u}^r\right\|_{w^1},
\end{aligned}$$

as required. We now show that U_H is of class $C^{r-1}(V_n)$. Let $\mathbf{z}^1, \mathbf{z}^2 \in V_n$ arbitrary. Then, using a strategy similar to that one in the last inequality in the previous expression, we have

$$\begin{aligned}
&\left|\!\left|\!\left| D^{r-1}U_H(\mathbf{z}^1) - D^{r-1}U_H(\mathbf{z}^2) \right|\!\right|\!\right|_{w^1,w^2} \\
&\qquad = \sup_{\substack{\mathbf{u}^1,\ldots,\mathbf{u}^{r-1}\in \ell_-^{w^1}(\mathbb{R}^n)\\ \mathbf{u}^1,\ldots,\mathbf{u}^{r-1}\neq \mathbf{0}}} \left\{ \frac{\left\| \left(D^{r-1}U_H(\mathbf{z}^1) - D^{r-1}U_H(\mathbf{z}^2)\right)\cdot\left(\mathbf{u}^1,\ldots,\mathbf{u}^{r-1}\right)\right\|_{w^2}}{\left\|\mathbf{u}^1\right\|_{w^1}\cdots\left\|\mathbf{u}^{r-1}\right\|_{w^1}} \right\} \\
&= \sup_{\substack{\mathbf{u}^1,\ldots,\mathbf{u}^{r-1}\in \ell_-^{w^1}(\mathbb{R}^n)\\ \mathbf{u}^1,\ldots,\mathbf{u}^{r-1}\neq \mathbf{0}}} \left\{ \frac{\sup_{t\in\mathbb{Z}_-}\left\{\left\|\left(D^{r-1}H(T_{-t}(\mathbf{z}^1)) - D^{r-1}H(T_{-t}(\mathbf{z}^2))\right)\cdot\left(T_{-t}(\mathbf{u}^1),\ldots,T_{-t}(\mathbf{u}^{r-1})\right)\right\| w^2_{-t}\right\}}{\left\|\mathbf{u}^1\right\|_{w^1}\cdots\left\|\mathbf{u}^{r-1}\right\|_{w^1}} \right\} \\
&\leq \sup_{\substack{\mathbf{u}^1,\ldots,\mathbf{u}^{r-1}\in \ell_-^{w^1}(\mathbb{R}^n)\\ \mathbf{u}^1,\ldots,\mathbf{u}^{r-1}\neq \mathbf{0}}} \left\{ \frac{\sup_{t\in\mathbb{Z}_-}\left\{c^r\left\|T_{-t}(\mathbf{z}^1) - T_{-t}(\mathbf{z}^2)\right\|_{w^1}\cdot\left\|T_{-t}(\mathbf{u}^1)\right\|_{w^1}\cdots\left\|T_{-t}(\mathbf{u}^{r-1})\right\|_{w^1}\cdot w^2_{-t}\right\}}{\left\|\mathbf{u}^1\right\|_{w^1}\cdots\left\|\mathbf{u}^{r-1}\right\|_{w^1}} \right\} \\
&\qquad = \sup_{\substack{\mathbf{u}^1,\ldots,\mathbf{u}^{r-1}\in \ell_-^{w^1}(\mathbb{R}^n)\\ \mathbf{u}^1,\ldots,\mathbf{u}^{r-1}\neq \mathbf{0}}} \left\{ \frac{\sup_{t\in\mathbb{Z}_-}\left\{c^r\left\|\mathbf{z}^1 - \mathbf{z}^2\right\|_{w^1}\cdot\left\|\mathbf{u}^1\right\|_{w^1}\cdots\left\|\mathbf{u}^{r-1}\right\|_{w^1} L_{w^1}^{-rt} w^2_{-t}\right\}}{\left\|\mathbf{u}^1\right\|_{w^1}\cdots\left\|\mathbf{u}^{r-1}\right\|_{w^1}} \right\} \\
&\qquad\qquad \leq c^r\left\|\mathcal{L}_{w^1,r}\right\|_{w^2}\left\|\mathbf{z}^1-\mathbf{z}^2\right\|_{w^1},
\end{aligned}$$

which shows that the map $D^{r-1}U_H : (V_n, \|\cdot\|_{w^1}) \longrightarrow \left(L^{r-1}(\ell_-^{w^1}(\mathbb{R}^n), \ell_-^{w^2}(\mathbb{R}^N)), |\!|\!|\cdot|\!|\!|_{w^1,w^2}\right)$ is Lipschitz continuous with Lipschitz constant $c^r\left\|\mathcal{L}_{w^1,r}\right\|_{w^2}$.

(iii) First, the condition $c^r < +\infty$ for all $r \in \mathbb{N}^+$ implies by part **(ii)** that U_H is smooth if H is. Suppose now that we work with the supremum norm. The expression (3.27) shows that the point $\mathbf{z} \in V_n$ belongs to the domain of convergence of the series expansion of U_H if and only if all the points $T_{-t}(\mathbf{z})$ belong to the domain of convergence of the series expansion of H. Finally, suppose that $\mathbf{z} \in V_n$ belongs to the domain of convergence of the series expansion of H. Since $|\!|\!|T_{-t}|\!|\!|_\infty \leq 1$ for all $t \in \mathbb{Z}_-$ by Lemma 3.1, we have that $\|T_{-t}(\mathbf{z})\|_\infty \leq \|\mathbf{z}\|_\infty$, which guarantees that all the points $T_{-t}(\mathbf{z})$ belong to the domain of convergence of the series expansion of H and hence, by the argument above, $\mathbf{z} \in V_n$ belongs to the domain of convergence of the series expansion of U_H, which proves the statement. ■

4 The fading memory property in reservoir filters with unbounded inputs

Starting in this section we focus on filters defined by reservoir systems of the type introduced in (1.1)–(1.2), but this time we consider reservoir maps $F : D_N \times D_n \longrightarrow D_N$ where the input variable takes values on a set $D_n \subset \mathbb{R}^N$ that is not necessarily bounded. All along this section, the reservoir map F will be assumed to be continuous and a contraction on the first entry with constant $0 < c < 1$, that is,

$$\left\|F(\mathbf{x}^1, \mathbf{z}) - F(\mathbf{x}^2, \mathbf{z})\right\| \leq c\left\|\mathbf{x}^1 - \mathbf{x}^2\right\|, \quad \text{for all } \mathbf{x}^1, \mathbf{x}^2 \in D_N \text{ and } \mathbf{z} \in D_n.$$

When the inputs are assumed to be uniformly bounded by a constant $M > 0$ and F maps into a ball $\overline{B_{\|\cdot\|}(\mathbf{0}, L)} \subset \mathbb{R}^N$, $L > 0$, it has been proved (see [Grig 18a, Proposition 2.1 and Theorem 3.1]) that we can associate to this system unique filters $U^F : K_M \longrightarrow K_L$ and $U_h^F : K_M \longrightarrow (\mathbb{R}^d)^{\mathbb{Z}_-}$ (the sets K_M and K_L are introduced in (1.3)) that are causal, time-invariant, continuous and, moreover, satisfy the fading memory property with respect to any weighting sequence w. We recall that U^F is the filter associated to the solutions of the reservoir equation (1.1) and assigns to any input sequence $\mathbf{z} \in K_M$

the output $U^F(\mathbf{z})$ that satisfies

$$U^F(\mathbf{z})_t = F(U^F(\mathbf{z})_{t-1}, \mathbf{z}_t), \quad \text{for any} \quad t \in \mathbb{Z}_-. \tag{4.1}$$

Recall also that $U_h^F : K_M \longrightarrow (\mathbb{R}^d)^{\mathbb{Z}_-}$ is the filter associated to the full system (1.1)–(1.2) and is given by $U_h^F := h \circ U^F$. We denote by $H^F : K_M \longrightarrow \overline{B_{\|\cdot\|}(\mathbf{0}, L)}$ and $H_h^F : K_M \longrightarrow \mathbb{R}^d$ the corresponding reservoir functionals. The reservoir functionals are related to the corresponding reservoir filters via the identities:

$$H^F(\mathbf{z}) = U^F(\mathbf{z})_0 = F(U^F(\mathbf{z})_{-1}, \mathbf{z}_0) \quad \text{and} \quad H_h^F(\mathbf{z}) = h\left(U^F(\mathbf{z})\right), \tag{4.2}$$

for all $\mathbf{z} \in K_M$.

The next theorem is the most important result in this section and shows that the results that we just recalled about the ESP and the FMP for reservoir filters with uniformly bounded inputs remain valid in the presence of unbounded inputs. However, in that case, the fading memory property depends on the weighting sequence that is used to define it. The sufficient condition for the FMP spelled out in the next theorem asserts, roughly speaking, that reservoir systems have the FMP only with respect to weighting sequences that converge to zero faster than the divergence rate of their outputs.

Theorem 4.1 (ESP and FMP with continuous reservoir maps) *Let $F : D_N \times D_n \longrightarrow D_N$ be a continuous reservoir map where $D_n \subset \mathbb{R}^n$, $D_N \subset \mathbb{R}^N$, $n, N \in \mathbb{N}^+$. Assume, additionally, that it is a contraction on the first entry with constant $0 < c < 1$. Let w be a weighting sequence with finite inverse decay ratio L_w and let $V_n \subset (D_n)^{\mathbb{Z}_-} \cap \ell_-^w(\mathbb{R}^n)$ be a time-invariant set. We consider two situations regarding the target D_N of the reservoir map:*

(i) *D_N is a compact subset of $\mathbb{R}^N$.*

(ii) *$(D_N)^{\mathbb{Z}_-} \cap \ell_-^w(\mathbb{R}^N)$ is a complete subset of the Banach space $\left(\ell_-^w(\mathbb{R}^N), \|\cdot\|_w\right)$, F is Lipschitz continuous, and the reservoir system (1.1) associated to F has a solution $(\mathbf{x}^0, \mathbf{z}^0) \in (D_N)^{\mathbb{Z}_-} \cap \ell_-^w(\mathbb{R}^N) \times V_n$, that is, $\mathbf{x}_t^0 = F(\mathbf{x}_{t-1}^0, \mathbf{z}_t^0)$, for all $t \in \mathbb{Z}_-$*

In both cases, if

$$cL_w < 1 \tag{4.3}$$

then the reservoir system associated to F with inputs in V_n has the echo state property and hence determines a unique continuous, causal, and time-invariant reservoir filter $U^F : (V_n, \|\cdot\|_w) \longrightarrow ((D_N)^{\mathbb{Z}_-} \cap \ell_-^w(\mathbb{R}^N), \|\cdot\|_w)$ that has the fading memory property with respect to w. Moreover, if F is Lipschitz on the second component (which is always the case under the hypotheses in **(ii)***) with constant L_z, that is,*

$$\left\|F(\mathbf{x}, \mathbf{z}^1) - F(\mathbf{x}, \mathbf{z}^2)\right\| \leq L_z \left\|\mathbf{z}^1 - \mathbf{z}^2\right\|, \quad \text{for any} \quad \mathbf{x} \in D_N, \ \mathbf{z}^1, \mathbf{z}^2 \in D_n,$$

then U^F is also Lipschitz with constant

$$L_{U^F} := \frac{L_z}{1 - cL_w}. \tag{4.4}$$

This statement also holds true under the hypotheses in part **(ii)** *when $(\ell_-^w(\mathbb{R}^n), \|\cdot\|_w)$ is replaced by $(\ell_-^\infty(\mathbb{R}^n), \|\cdot\|_\infty)$. In that case L_w is replaced by the constant 1 and hence condition (4.3) is automatically satisfied. The resulting reservoir filter $U^F : (V_n, \|\cdot\|_\infty) \longrightarrow ((D_N)^{\mathbb{Z}_-} \cap \ell_-^\infty(\mathbb{R}^N), \|\cdot\|_w)$ is continuous.*

Remark 4.2 A very common situation that provides the solution $(\mathbf{x}^0, \mathbf{z}^0) \in (D_N)^{\mathbb{Z}_-} \cap \ell_-^w(\mathbb{R}^N) \times V_n$ for the reservoir system needed in part **(ii)**, is the existence of a fixed point $(\overline{\mathbf{x}}^0, \overline{\mathbf{z}}^0) \in D_N \times D_n$ of F that satisfies $F(\overline{\mathbf{x}}^0, \overline{\mathbf{z}}^0) = \overline{\mathbf{x}}^0$. In that case the required solution is given by the constant sequences $\mathbf{x}_t^0 = \overline{\mathbf{x}}^0$, $\mathbf{z}_t^0 = \overline{\mathbf{z}}^0$, for all $t \in \mathbb{Z}_-$.

Remark 4.3 If the target D_N of the reservoir map is a closed subset of $\mathbb{R}^N$, that is $\overline{D_N} = D_N$, then by part **(iii)** of the Corollary 2.2, the set $(D_N)^{\mathbb{Z}_-} \cap \ell_-^w(\mathbb{R}^N)$ is a closed subset of $\left(\ell_-^w(\mathbb{R}^N), \|\cdot\|_w\right)$ and it is hence necessarily complete.

Proof of the theorem. Consider the map

$$\begin{array}{rccc} \mathcal{F}: & (D_N)^{\mathbb{Z}_-} \cap \ell_-^w(\mathbb{R}^N) \times V_n & \longrightarrow & (D_N)^{\mathbb{Z}_-} \\ & (\mathbf{x}, \mathbf{z}) & \longmapsto & (\mathcal{F}(\mathbf{x}, \mathbf{z}))_t := F(\mathbf{x}_{t-1}, \mathbf{z}_t). \end{array} \tag{4.5}$$

We now show that first, under the two sets of hypotheses in the statement, $\mathcal{F}$ actually maps into $(D_N)^{\mathbb{Z}_-} \cap \ell_-^w(\mathbb{R}^N)$ and second, that $\mathcal{F}$ is continuous. Suppose first that we are in the hypotheses in **(i)**. Since D_N is compact then $(D_N)^{\mathbb{Z}_-} \subset \ell_-^w(\mathbb{R}^N)$ and hence $\mathcal{F}$ obviously maps into $(D_N)^{\mathbb{Z}_-} \cap \ell_-^w(\mathbb{R}^N)$. Regarding the continuity, notice that $\mathcal{F}$ can be written as

$$\mathcal{F} = \prod_{t \in \mathbb{Z}_-} F_t \quad \text{with} \quad F_t := F \circ p_t \circ (T_1 \times \mathrm{id}_{V_n}) : (D_N)^{\mathbb{Z}_-} \cap \ell_-^w(\mathbb{R}^N) \times V_n \longrightarrow D_N. \tag{4.6}$$

The continuity of F, the fact that L_w is by hypothesis finite, and Lemma 3.1 imply that all the functions $F_t : (D_N)^{\mathbb{Z}_-} \cap \ell_-^w(\mathbb{R}^N) \times V_n \subset \ell_-^w(\mathbb{R}^N) \oplus \ell_-^w(\mathbb{R}^n) \longrightarrow D_N \subset \mathbb{R}^N$ are continuous and moreover, they map into a compact subset of $\mathbb{R}^N$. An argument mimicking the proof of the first part of Lemma 2.5 allows us to conclude that $\mathcal{F} : (D_N)^{\mathbb{Z}_-} \cap \ell_-^w(\mathbb{R}^N) \times V_n \longrightarrow (D_N)^{\mathbb{Z}_-} \cap \ell_-^w(\mathbb{R}^N)$ is a continuous map.

Suppose now that we are in the hypotheses in part **(ii)**. We now show that since F is Lipschitz then so are all the functions $F_t := F \circ p_t \circ (T_1 \times \mathrm{id}_{V_n})$, $t \in \mathbb{Z}_-$, by Lemma 3.1, where we consider the direct sum of weighted spaces $\ell_-^w(\mathbb{R}^N) \oplus \ell_-^w(\mathbb{R}^n)$ as a Banach space with the sum norm $\|\cdot\|_{w \oplus w}$ defined by $\|(\mathbf{x}, \mathbf{z})\|_{w \oplus w} := \|\mathbf{x}\|_w + \|\mathbf{z}\|_w$, for any $(\mathbf{x}, \mathbf{z}) \in \ell_-^w(\mathbb{R}^N) \oplus \ell_-^w(\mathbb{R}^n)$. Indeed, let c_F be the Lipschitz constant of F and let $(\mathbf{x}^1, \mathbf{z}^1), (\mathbf{x}^2, \mathbf{z}^2) \in (D_N)^{\mathbb{Z}_-} \cap \ell_-^w(\mathbb{R}^N) \times V_n$, then:

$$\begin{aligned} \left\| F \circ p_t \circ (T_1 \times \mathrm{id}_{V_n}) (\mathbf{x}^1, \mathbf{z}^1) - F \circ p_t \circ (T_1 \times \mathrm{id}_{V_n}) (\mathbf{x}^2, \mathbf{z}^2) \right\| &\leq c_F \left\| p_t \circ (T_1 \times \mathrm{id}_{V_n}) (\mathbf{x}^1 - \mathbf{x}^2, \mathbf{z}^1 - \mathbf{z}^2) \right\| \\ \leq \frac{c_F}{w_{-t}} \left\| (T_1 \times \mathrm{id}_{V_n}) (\mathbf{x}^1 - \mathbf{x}^2, \mathbf{z}^1 - \mathbf{z}^2) \right\|_w &\leq \frac{c_F}{w_{-t}} (L_w \left\| \mathbf{x}^1 - \mathbf{x}^2 \right\|_w + \left\| \mathbf{z}^1 - \mathbf{z}^2 \right\|_w) \\ \leq \frac{c_F}{w_{-t}} L_w (\left\| \mathbf{x}^1 - \mathbf{x}^2 \right\|_w + \left\| \mathbf{z}^1 - \mathbf{z}^2 \right\|_w) &= \frac{c_F}{w_{-t}} L_w \left\| (\mathbf{x}^1, \mathbf{z}^1) - (\mathbf{x}^2, \mathbf{z}^2) \right\|_{w \oplus w}. \end{aligned} \tag{4.7}$$

This chain of inequalities show that F_t is a Lipschitz continuous function and that $c_F L_w / w_{-t}$ is a Lipschitz constant. Given that the sequence $c_{\mathcal{F}} := (c_F L_w / w_{-t})_{t \in \mathbb{Z}_-}$ is such that $\|c_{\mathcal{F}}\|_w = c_F L_w < +\infty$, the part **(ii)** of Lemma 2.5 guarantees that $\mathcal{F}$ is Lipschitz continuous and that $c_F L_w$ is a Lipschitz constant, that is,

$$\left\| \mathcal{F}(\mathbf{x}^1, \mathbf{z}^1) - \mathcal{F}(\mathbf{x}^2, \mathbf{z}^2) \right\|_w \leq \ c_F L_w \left\| (\mathbf{x}^1, \mathbf{z}^1) - (\mathbf{x}^2, \mathbf{z}^2) \right\|_{w \oplus w}. \tag{4.8}$$

Moreover, let $\mathbf{u}^0 := (\mathbf{x}^0, \mathbf{z}^0) \in (D_N)^{\mathbb{Z}_-} \cap \ell_-^w(\mathbb{R}^N) \times V_n$. The fact that $\mathbf{u}^0$ is a solution of the reservoir system implies that $\mathcal{F}(\mathbf{u}^0) = \mathbf{x}^0 \in (D_N)^{\mathbb{Z}_-} \cap \ell_-^w(\mathbb{R}^N)$. An argument mimicking (2.26) in the proof of part **(ii)** in Lemma 2.5 proves that in those conditions $\mathcal{F}$ maps into $(D_N)^{\mathbb{Z}_-} \cap \ell_-^w(\mathbb{R}^N)$.

We now show that in the presence of hypothesis (4.3) $\mathcal{F}$ is a contraction on the first entry with constant $cL_w < 1$. Indeed, for any $\mathbf{x}^1, \mathbf{x}^2 \in (D_N)^{\mathbb{Z}_-} \cap \ell_-^w(\mathbb{R}^N)$ and any $\mathbf{z} \in V_n$, we have

$$\left\| \mathcal{F}(\mathbf{x}^1, \mathbf{z}) - \mathcal{F}(\mathbf{x}^2, \mathbf{z}) \right\|_w = \sup_{t \in \mathbb{Z}_-} \left\{ \left\| F(\mathbf{x}^1_{t-1}, \mathbf{z}_t) - F(\mathbf{x}^2_{t-1}, \mathbf{z}_t) \right\| w_{-t} \right\} \leq \sup_{t \in \mathbb{Z}_-} \left\{ \left\| \mathbf{x}^1_{t-1} - \mathbf{x}^2_{t-1} \right\| c w_{-t} \right\}, \tag{4.9}$$

where we used that F is a contraction on the first entry. Now,

$$\sup_{t \in \mathbb{Z}_-} \left\{ \left\| \mathbf{x}^1_{t-1} - \mathbf{x}^2_{t-1} \right\| c w_{-t} \right\} = c \sup_{t \in \mathbb{Z}_-} \left\{ \left\| \mathbf{x}^1_{t-1} - \mathbf{x}^2_{t-1} \right\| w_{-(t-1)} \frac{w_{-t}}{w_{-(t-1)}} \right\} \leq c L_w \left\| \mathbf{x}^1 - \mathbf{x}^2 \right\|_w. \tag{4.10}$$

This shows that $\mathcal{F}$ is a family of contractions with constant $cL_w < 1$ that is continuously parametrized by the elements in V_n. Since by hypothesis, the domain $(D_N)^{\mathbb{Z}_-} \cap \ell_-^w(\mathbb{R}^N)$ is complete, Theorem 6.4.1 in [Ster 10] implies the existence of a continuous map $U^F : (V_n, \|\cdot\|_w) \longrightarrow \left((D_N)^{\mathbb{Z}_-} \cap \ell_-^w(\mathbb{R}^N), \|\cdot\|_w\right)$ that is uniquely determined by the identity

$$\mathcal{F}\left(U^F(\mathbf{z}), \mathbf{z}\right) = U^F(\mathbf{z}), \quad \text{for all } \mathbf{z} \in V_n. \tag{4.11}$$

The causality and the time-invariance of U^F are a consequence of the time invariance of V_n and of Proposition 2.1 in [Grig 18b].

We now assume that F is Lipschitz on the second component and prove (4.4). The relation (4.11) that defines U^F is equivalent to

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

Consequently, for any $\mathbf{z}^1, \mathbf{z}^2 \in V_n$, we have,

$$\begin{aligned}\left\|U^F(\mathbf{z}^1) - U^F(\mathbf{z}^2)\right\|_w &= \sup_{t\in\mathbb{Z}_-}\left\{\left\|U^F(\mathbf{z}^1)_t - U^F(\mathbf{z}^2)_t\right\| w_{-t}\right\}\\ &= \sup_{t\in\mathbb{Z}_-}\left\{\left\|F(U^F(\mathbf{z}^1)_{t-1}, \mathbf{z}^1_t) - F(U^F(\mathbf{z}^2)_{t-1}, \mathbf{z}^2_t)\right\| w_{-t}\right\}\\ &\leq \sup_{t\in\mathbb{Z}_-}\left\{\left(\left\|F(U^F(\mathbf{z}^1)_{t-1}, \mathbf{z}^1_t) - F(U^F(\mathbf{z}^1)_{t-1}, \mathbf{z}^2_t)\right\| + \left\|F(U^F(\mathbf{z}^1)_{t-1}, \mathbf{z}^2_t) - F(U^F(\mathbf{z}^2)_{t-1}, \mathbf{z}^2_t)\right\|\right) w_{-t}\right\}\\ &\leq \sup_{t\in\mathbb{Z}_-}\left\{L_z\left\|\mathbf{z}^1_t - \mathbf{z}^2_t\right\| w_{-t} + c\left\|U^F(\mathbf{z}^1)_{t-1} - U^F(\mathbf{z}^2)_{t-1}\right\| w_{-t}\right\}.\end{aligned}$$

If we repeat this procedure i times, it is easy to see that

$$\begin{aligned}&\left\|U^F(\mathbf{z}^1) - U^F(\mathbf{z}^2)\right\|_w\\ &\leq L_z \sup_{t\in\mathbb{Z}_-}\left\{\sum_{j=0}^{i} c^j\left\|\mathbf{z}^1_{t-j} - \mathbf{z}^2_{t-j}\right\| w_{-t}\right\} + c^{i+1}\sup_{t\in\mathbb{Z}_-}\left\{\left\|U^F(\mathbf{z}^1)_{t-(i+1)} - U^F(\mathbf{z}^2)_{t-(i+1)}\right\| w_{-t}\right\}.\end{aligned} \tag{4.12}$$

We now study separately the two summands in the right hand side of the previous inequality. First, by Lemma 3.1,

$$\begin{aligned}L_z \sup_{t\in\mathbb{Z}_-}\left\{\sum_{j=0}^{i} c^j\left\|\mathbf{z}^1_{t-j} - \mathbf{z}^2_{t-j}\right\| w_{-t}\right\} &= L_z \sup_{t\in\mathbb{Z}_-}\left\{\sum_{j=0}^{i} c^j\left\|\left(T_j(\mathbf{z}^1)\right)_t - \left(T_j(\mathbf{z}^2)\right)_t\right\| w_{-t}\right\}\\ &= L_z \sup_{t\in\mathbb{Z}_-}\left\{\sum_{j=0}^{i} c^j\left\|T_j(\mathbf{z}^1 - \mathbf{z}^2)_t\right\| w_{-t}\right\} \leq L_z \sum_{j=0}^{i} c^j \sup_{t\in\mathbb{Z}_-}\left\{\left\|T_j(\mathbf{z}^1 - \mathbf{z}^2)_t\right\| w_{-t}\right\}\\ &= L_z \sum_{j=0}^{i} c^j\left\|T_j(\mathbf{z}^1 - \mathbf{z}^2)\right\|_w \leq L_z\left\|\mathbf{z}^1 - \mathbf{z}^2\right\|_w \sum_{j=0}^{i} c^j |||T_j|||_w\\ &\leq L_z\left\|\mathbf{z}^1 - \mathbf{z}^2\right\|_w \sum_{j=0}^{i}(cL_w)^j = L_z\left\|\mathbf{z}^1 - \mathbf{z}^2\right\|_w \frac{1-(cL_w)^{i+1}}{1-cL_w},\end{aligned} \tag{4.13}$$

while the second summand can be bounded as follows

$$
\begin{aligned}
c^{i+1} \sup_{t\in\mathbb{Z}_-} \left\{ \left\| U^F(\mathbf{z}^1)_{t-(i+1)} - U^F(\mathbf{z}^2)_{t-(i+1)} \right\| w_{-t} \right\} &= c^{i+1} \sup_{t\in\mathbb{Z}_-} \left\{ \left\| T_{i+1}\left(U^F(\mathbf{z}^1)\right)_t - T_{i+1}\left(U^F(\mathbf{z}^2)\right)_t \right\| w_{-t} \right\} \\
&= c^{i+1} \left\| T_{i+1}(U^F(\mathbf{z}^1) - (U^F(\mathbf{z}^2)) \right\|_w \leq c^{i+1} \vert\vert\vert T_{i+1} \vert\vert\vert_w \left\| U^F(\mathbf{z}^1) - U^F(\mathbf{z}^2) \right\|_w \\
&\leq (cL_w)^{i+1} \left\| U^F(\mathbf{z}^1) - U^F(\mathbf{z}^2) \right\|_w . \quad (4.14)
\end{aligned}
$$

If we now chain the inequalities (4.13) and (4.14) with (4.12) we can conclude that

$$
\left(1 - (cL_w)^{i+1}\right) \left\| U^F(\mathbf{z}^1) - U^F(\mathbf{z}^2) \right\|_w \leq L_z \left\| \mathbf{z}^1 - \mathbf{z}^2 \right\|_w \frac{1 - (cL_w)^{i+1}}{1 - cL_w}, \tag{4.15}
$$

which after simplification using the condition (4.3) results in (4.4). ■

Remark 4.4 A slight modification of this proof can be used to extend the statement of Theorem 4.1 **(ii)** to reservoir systems with inputs and outputs in $\ell_-^{p,w}(\mathbb{R}^n)$ and $\ell_-^{p,w}(\mathbb{R}^N)$, respectively. Indeed, assume that we are under the hypotheses of Theorem 4.1 **(ii)** with those spaces instead of $\ell_-^{w}(\mathbb{R}^n)$ and $\ell_-^{w}(\mathbb{R}^N)$. Suppose, additionally, that

$$
cL_{w,p} < 1 \tag{4.16}
$$

where $L_{w,p}$ was defined in (3.15). Then, there exists a unique causal and time-invariant continuous reservoir filter $U^F : (V_n, \left\|\cdot\right\|_{p,w}) \longrightarrow ((D_N)^{\mathbb{Z}_-} \cap \ell_-^{p,w}(\mathbb{R}^N), \left\|\cdot\right\|_{p,w})$. Additionally, U^F is also Lipschitz with constant

$$
L_{U^F} := \frac{L_z}{1 - cL_{w,p}}.
$$

The proof of this fact is carried out by showing that the map $\mathcal{F}$ in (4.6) is Lipschitz continuous when $\ell_-^{p,w}(\mathbb{R}^n)$ and $\ell_-^{p,w}(\mathbb{R}^N)$ spaces are considered in its domain and target, respectively, with Lipschitz constant $c_F L_{w,p}$ and hence (4.8) holds in that situation. Indeed, for any $(\mathbf{x}^1, \mathbf{z}^1), (\mathbf{x}^2, \mathbf{z}^2) \in (D_N)^{\mathbb{Z}_-} \cap \ell_-^{p,w}(\mathbb{R}^N) \times V_n$ we can show using the statements in Remark 3.4 that

$$
\begin{aligned}
&\left\| \mathcal{F}(\mathbf{x}^1, \mathbf{z}^1) - \mathcal{F}(\mathbf{x}^2, \mathbf{z}^2) \right\|_{p,w}^p = \sum_{t\in\mathbb{Z}_-} \left\| F_t(\mathbf{x}^1, \mathbf{z}^1) - F_t(\mathbf{x}^2, \mathbf{z}^2) \right\|^p w_{-t} \\
&= \sum_{t\in\mathbb{Z}_-} \left\| F(\mathbf{x}^1_{t-1}, \mathbf{z}^1_t) - F_t(\mathbf{x}^2_{t-1}, \mathbf{z}^2_t) \right\|^p w_{-t} \leq c_F^p \sum_{t\in\mathbb{Z}_-} \left\| \mathbf{x}^1_{t-1} - \mathbf{x}^2_{t-1} \right\|^p w_{-t} + c_F^p \sum_{t\in\mathbb{Z}_-} \left\| \mathbf{z}^1_t - \mathbf{z}^2_t \right\|^p w_{-t} \\
&\leq c_F^p \left\| T_1(\mathbf{x}^1 - \mathbf{x}^2) \right\|_{p,w}^p + c_F^p \left\| \mathbf{z}^1 - \mathbf{z}^2 \right\|_{p,w}^p \leq c_F^p L_{w,p}^p \left\| \mathbf{x}^1 - \mathbf{x}^2 \right\|_{p,w}^p + c_F^p \left\| \mathbf{z}^1 - \mathbf{z}^2 \right\|_{p,w}^p \\
&\leq c_F^p L_{w,p}^p \left\| (\mathbf{x}^1, \mathbf{z}^1) - (\mathbf{x}^2, \mathbf{z}^2) \right\|_{p,w\oplus w}^p ,
\end{aligned}
$$

where in the last inequality we used that $L_{w,p} > 1$. We now show that $\mathcal{F}$ is a contraction on the first entry whenever condition (4.16) is satisfied. Indeed,

$$
\begin{aligned}
\left\| \mathcal{F}(\mathbf{x}^1, \mathbf{z}^1) - \mathcal{F}(\mathbf{x}^2, \mathbf{z}^2) \right\|_{p,w}^p &= \sum_{t\in\mathbb{Z}_-} \left\| F(\mathbf{x}^1_{t-1}, \mathbf{z}^1_t) - F_t(\mathbf{x}^2_{t-1}, \mathbf{z}^2_t) \right\|^p w_{-t} \\
&\leq c^p \sum_{t\in\mathbb{Z}_-} \left\| \mathbf{x}^1_{t-1} - \mathbf{x}^2_{t-1} \right\|^p w_{-t} = c^p \left\| T_1(\mathbf{x}^1 - \mathbf{x}^2) \right\|_{p,w}^p \leq c^p L_{w,p}^p \left\| \mathbf{x}^1 - \mathbf{x}^2 \right\|_{p,w}^p .
\end{aligned}
$$

The rest of the proof can be obtained by mimicking the developments after (4.10).

As a corollary of Theorem 4.1 it can be shown that reservoir systems that have by construction uniformly bounded inputs and outputs always have the ESP and FMP properties and that for any weighting sequence w. This result was already shown in [Grig 18a, Theorem 3.1].

Corollary 4.5 *Let $M, L > 0$, let $K_M \subset (\mathbb{R}^n)^{\mathbb{Z}_-}$ and $K_L \subset \left(\mathbb{R}^N\right)^{\mathbb{Z}_-}$ be subsets of uniformly bounded sequences defined as in (1.3), 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. Assume, additionally, that F is a contraction on the first entry with constant $0 < c < 1$. 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 with respect to any weighting sequence w.*

Proof. Given that $\overline{B_{\|\cdot\|}(\mathbf{0}, L)}$ is a compact subset of $\mathbb{R}^N$, the hypothesis in part **(i)** of Theorem 4.1 and condition (4.3) guarantee that there exists a reservoir filter $U^F : K_M \longrightarrow K_L$ associated to F that has the fading memory property with respect to any weighting sequence that satisfies (4.3). Such a sequence always exists as it suffices to take any geometric sequence $w_t := \lambda^t$, $t \in \mathbb{N}$, with $c < \lambda < 1$. However, as it has been shown in [Grig 18a, Corollary 2.7], all the weighted norms induce in the sets K_M and K_L the same topology, namely, the product topology and hence if U^F is continuous with respect to the topology induced by the weighted norm $\|\cdot\|_w$ then so it is with respect to the norm associated to any other weighting sequence. ■

Remark 4.6 This corollary shows that, in general, the condition (4.3) is sufficient but not necessary. Indeed, if the hypotheses in the corollary are satisfied, the resulting filter U^F has the fading memory property with respect to any geometric sequence $w_t := \lambda^t$, with $0 < \lambda < 1$, $t \in \mathbb{N}$ for which (see Remark 3.2) $L_w = 1/\lambda$. In particular, this holds true when λ is chosen so that $0 < \lambda < c$ and hence when (4.3) is not satisfied since in that case $cL_w > 1$. Additional concrete examples that show that the condition (4.3) is sufficient but not necessary are provided in Section 4.1.

Remark 4.7 The FMP condition (4.3) is sufficient but not necessary even in the absence of boundedness conditions like in Corollary 4.5

Another important statement that can be proved when the target of the reservoir map is a compact subset of $\mathbb{R}^N$ is that the echo state property is in that situation guaranteed *for no matter what input*[1] *in* $(\mathbb{R}^n)^{\mathbb{Z}_-}$ even though the FMP may obviously not hold in that case.

Theorem 4.8 (ESP for reservoir maps with compact target) *Let $F : D_N \times D_n \longrightarrow D_N$ be a continuous reservoir map, $D_n \subset \mathbb{R}^n$, $D_N \subset \mathbb{R}^N$, $n, N \in \mathbb{N}^+$, such that D_N is a compact subset of $\mathbb{R}^N$ and F is a contraction on the first entry with constant $0 < c < 1$. Then, the reservoir system associated to F has the echo state property for any input in $(D_n)^{\mathbb{Z}_-}$. Let $U^F : (D_n)^{\mathbb{Z}_-} \longrightarrow (D_N)^{\mathbb{Z}_-}$ be the associated reservoir filter. For any weighting sequence w such that $cL_w < 1$ the map $U^F : (D_n)^{\mathbb{Z}_-} \longrightarrow ((D_N)^{\mathbb{Z}_-}, \|\cdot\|_w)$ is continuous when in $(D_n)^{\mathbb{Z}_-}$ we consider the relative topology induced by the product topology in $(\mathbb{R}^N)^{\mathbb{Z}_-}$. Moreover, if $(D_n)^{\mathbb{Z}_-} \subset \ell_-^w(\mathbb{R}^n)$ then U^F has the fading memory property.*

Proof. Consider the map $\mathcal{F} : (D_N)^{\mathbb{Z}_-} \times (D_n)^{\mathbb{Z}_-} \longrightarrow (D_N)^{\mathbb{Z}_-}$ defined in (4.5) and endow $(D_n)^{\mathbb{Z}_-}$ and $(D_N)^{\mathbb{Z}_-}$ with the relative topologies induced by the product topologies in $(\mathbb{R}^n)^{\mathbb{Z}_-}$ and $(\mathbb{R}^N)^{\mathbb{Z}_-}$, respectively. It is easy to see that the maps p_t and T_1 are continuous with respect to those product topologies and hence $\mathcal{F}$ can be written using (4.6) as a Cartesian product of continuous functions, which is always continuous in the product topology.

Consider now any weighting sequence w such that $cL_w < 1$. Using an argument similar to the proof of Lemma 2.5 **(i)**, we can conclude that $(D_N)^{\mathbb{Z}_-} \subset \ell_-^w(\mathbb{R}^N)$ and that the product topology on $(D_N)^{\mathbb{Z}_-}$ coincides with the norm topology induced by $\|\cdot\|_w$. Now, following the expressions (4.9) and (4.10) it can be shown that $\mathcal{F}$ is a contraction on the first entry and with respect to $\|\cdot\|_w$. In view of these facts and given that the product topology in $(D_n)^{\mathbb{Z}_-} \subset (\mathbb{R}^n)^{\mathbb{Z}_-}$ is metrizable (see [Munk 14, Theorem 20.5]) and that $(D_N)^{\mathbb{Z}_-} \subset (\mathbb{R}^N)^{\mathbb{Z}_-}$ is compact by Tychonoff's Theorem (see [Munk 14, Theorem 37.3]) in

[1]We thank Lukas Gonon for pointing this out.

the product topology and hence complete, Theorem 6.4.1 in [Ster 10] implies the existence of a unique fixed point of $\mathcal{F}$ for each $\mathbf{z} \in (D_n)^{\mathbb{Z}_-}$, which establishes the ESP. Moreover, that result also shows the continuity of the associated filter $U^F : (D_n)^{\mathbb{Z}_-} \longrightarrow ((D_N)^{\mathbb{Z}_-}, \|\cdot\|_w)$.

Finally, if $(D_n)^{\mathbb{Z}_-} \subset \ell_-^w(\mathbb{R}^n)$, we know from [Grig 18a, Proposition 2.9] that the inclusion $\ell_-^w(\mathbb{R}^n) \hookrightarrow (\mathbb{R}^n)^{\mathbb{Z}_-}$ is continuous and hence so is U^F when in $(D_n)^{\mathbb{Z}_-}$ we consider the topology generated by the norm $\|\cdot\|_w$, which establishes the FMP in that situation. ■

The following result shows how the FMP of the filter associated to a reservoir map established in Theorem 4.1 propagates to the FMP of the filter of the full reservoir system in the readout map is continuous.

Corollary 4.9 *In the conditions of Theorem 4.1, let $h : D_N \longrightarrow \mathbb{R}^d$ be a continuous readout map. Consider the following two cases that correspond to the two sets of hypotheses studied in Theorem 4.1:*

(i) *If D_N is a compact subset of $\mathbb{R}^N$ then there is a constant $R > 0$ such that the filter U_h^F defined by $U_h^F(\mathbf{z})_t := h\left(U^F(\mathbf{z})_t\right)$, $t \in \mathbb{Z}_-$, $\mathbf{z} \in V_n$ maps $U_h^F : (V_n, \|\cdot\|_w) \longrightarrow (K_R, \|\cdot\|_w)$ and has the fading memory property.*

(ii) *If $(D_N)^{\mathbb{Z}_-} \cap \ell_-^w(\mathbb{R}^N)$ is a complete subset of $\left(\ell_-^w(\mathbb{R}^N, \|\cdot\|_w\right)$ and h is Lipschitz continuous on D_N such that $U_h^F(\mathbf{z}^0) \in \ell_-^w(\mathbb{R}^d)$, then the reservoir filter $U_h^F : (V_n, \|\cdot\|_w) \longrightarrow (\ell_-^w(\mathbb{R}^d), \|\cdot\|_w)$ has the fading memory property.*

This statement also holds true under the hypotheses in part **(ii)** *when $(\ell_-^w(\mathbb{R}^n), \|\cdot\|_w)$ is replaced by $(\ell_-^\infty(\mathbb{R}^n), \|\cdot\|_\infty)$. The resulting reservoir filter $U_h^F : (V_n, \|\cdot\|_\infty) \longrightarrow (\ell_-^\infty(\mathbb{R}^d), \|\cdot\|_\infty)$ is continuous.*

Proof. Under the hypothesis in part **(i)**, the continuity of h implies that $h(D_N)$ is compact and hence there exists a constant $R > 0$ such that $h(D_N) \subset \overline{B_{\|\cdot\|}(\mathbf{0}, R)}$. The first part of Lemma 2.5 guarantees that the map $\mathcal{H} := \prod_{t \in \mathbb{Z}_-} h : ((D_N)^{\mathbb{Z}_-}, \|\cdot\|_w) \longrightarrow (K_R, \|\cdot\|_w)$ is continuous and as $U_h^F = \mathcal{H} \circ U^F$ and we proved that under the hypotheses **(i)** in the theorem that $U^F : (V_n, \|\cdot\|_w) \longrightarrow (K_L, \|\cdot\|_w)$ is continuous, the claim follows.

We now prove the statement under the hypotheses in part **(ii)**. First, we show that if h is Lipschitz continuous in D^N with constant c_h then so is the map $\mathcal{H}$ in $(D_N)^{\mathbb{Z}_-} \cap \ell_-^w(\mathbb{R}^N)$. Indeed, let $\mathbf{x}^1, \mathbf{x}^2 \in (D_N)^{\mathbb{Z}_-} \cap \ell_-^w(\mathbb{R}^N)$, then

$$\left\|\mathcal{H}(\mathbf{x}^1) - \mathcal{H}(\mathbf{x}^2)\right\|_w = \sup_{t \in \mathbb{Z}_-} \left\{\left\|h(\mathbf{x}_t^1) - h(\mathbf{x}_t^2)\right\| w_{-t}\right\} \leq c_h \left\|\mathbf{x}^1 - \mathbf{x}^2\right\|_w .$$

The hypothesis $U_h^F(\mathbf{z}^0) \in \ell_-^w(\mathbb{R}^d)$ amounts to the fact that the point $U^F(\mathbf{z}^0) \in (D_N)^{\mathbb{Z}_-} \cap \ell_-^w(\mathbb{R}^N)$ is such that $\mathcal{H}(U^F(\mathbf{z}^0)) \in \ell_-^w(\mathbb{R}^d)$. An argument mimicking (2.26) in the proof of part **(ii)** in Lemma 2.5 proves that in those conditions $\mathcal{H}$ maps into $\ell_-^w(\mathbb{R}^d)$. ■

4.1 Examples

In the following paragraphs we show how the sufficient condition (4.3) explicitly looks like for reservoir systems that are widely used and that have been shown to have universality properties in the fading memory category both with deterministic and stochastic inputs [Grig 18b, Grig 18a, Gono 18].

Linear reservoir maps. Consider the reservoir map $F : \mathbb{R}^N \times \mathbb{R}^n \longrightarrow \mathbb{R}^N$ given by

$$F(\mathbf{x}, \mathbf{z}) = A\mathbf{x} + \mathbf{c}\mathbf{z}, \quad \text{with} \quad A \in \mathbb{M}_N, \mathbf{c} \in \mathbb{M}_{N,n}. \tag{4.17}$$

It is easy to see that F is a contraction on the first entry whenever the matrix A satisfies that $|||A||| < 1$. In that case, using the notation in Theorem 4.1, $c = |||A|||$. Indeed, for any $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^N$, $\mathbf{z} \in \mathbb{R}^n$:

$$\|F(\mathbf{x}_1, \mathbf{z}) - F(\mathbf{x}_2, \mathbf{z})\| = \|A(\mathbf{x}_1 - \mathbf{x}_2)\| \leq |||A||| \, \|\mathbf{x}_1 - \mathbf{x}_2\| \, .$$

We now assume that $|||A||| < 1$ and prove the following two statements:

(i) The reservoir system associated to (4.17) has the echo state property and defines a unique reservoir filter $U^F : \ell_-^w(\mathbb{R}^n) \longrightarrow \ell_-^w(\mathbb{R}^N)$ that has the fading memory property with respect to any weighting sequence w that satisfies the condition

$$\sum_{j=0}^{\infty} \frac{|||A^j|||}{w_j} < +\infty. \tag{4.18}$$

The FMP condition (4.3) reads in this case as

$$|||A||| L_w < 1, \tag{4.19}$$

and implies (4.18) but not vice versa.

(ii) If the inputs presented to the reservoir system associated to (4.17) are uniformly bounded then it has the fading memory property with respect to any weighting sequence. This result was already known as it can be easily obtained by combining [Grig 18b, Corollary 11] with [Grig 18a, Corollary 2.7]. We obtain it here directly out of Corollary 4.5 by noting that for any $M > 0$,

$$F(\overline{B_{\|\cdot\|}(\mathbf{0}, L)}, \overline{B_{\|\cdot\|}(\mathbf{0}, M)} \subset \overline{B_{\|\cdot\|}(\mathbf{0}, L)}, \quad \text{with} \quad L := \frac{|||\mathbf{c}|||M}{1 - |||A|||}. \tag{4.20}$$

Proof of statement (i) One can show by mimicking the proof of [Grig 18b, Corollary 11] that whenever condition (4.18) is satisfied for a given weighting sequence w, the reservoir system determined by (4.17) has a unique reservoir filter $U^F : \ell_-^w(\mathbb{R}^n) \longrightarrow \ell_-^w(\mathbb{R}^N)$ associated that is determined by the linear functional $H^F : \ell_-^w(\mathbb{R}^n) \longrightarrow \mathbb{R}^N$ given by

$$H^F(\mathbf{z}) := \sum_{j=0}^{\infty} A^j \mathbf{c} \mathbf{z}_{-j}.$$

This linear functional is bounded because for any $\mathbf{z} \in \ell_-^w(\mathbb{R}^n)$, the hypothesis (4.18) implies that:

$$\left\|H^F(\mathbf{z})\right\| \leq \sum_{j=0}^{\infty} |||A^j||| |||\mathbf{c}||| \, \|\mathbf{z}_{-j}\| = |||\mathbf{c}||| \sum_{j=0}^{\infty} |||A^j||| \, \|\mathbf{z}_{-j}\| \, \frac{w_j}{w_j} \leq |||\mathbf{c}||| \, \|\mathbf{z}\|_w \sum_{j=0}^{\infty} \frac{|||A^j|||}{w_j} < +\infty$$

We now show that for any weighting sequence w that satisfies $|||A||| L_w < 1$, the condition (4.18) always holds. Indeed, using (3.3) we obtain

$$\sum_{j=0}^{\infty} \frac{|||A^j|||}{w_j} \leq \sum_{j=0}^{\infty} \frac{|||A|||^j}{w_j} \leq \sum_{j=0}^{\infty} |||A|||^j L_w^j = \frac{1}{1 - |||A||| L_w} < +\infty, \quad \text{as required.}$$

We finally show that there exist sequences w that satisfy (4.18) but not $|||A||| L_w < 1$, which is one more example of the fact, that we already indicated in Remark 4.6, that the FMP condition (4.3) is sufficient but not necessary. Let w be a harmonic weighting sequence as in Remark 3.2 given by $w_j := 1/(1 + jd)$,

$j \in \mathbb{N}$, with $d > 0$. In this case $L_w = 1 + d$ so we can choose a value d such that $|||A|||(1 + d) > 1$. However, at the same time, the condition (4.18) holds in this case because

$$\sum_{j=0}^{\infty} \frac{|||A^j|||}{w_j} \leq \sum_{j=0}^{\infty} |||A|||^j (1 + jd) = \sum_{j=0}^{\infty} |||A|||^j + d(j+1)|||A|||^j - d|||A|||^j$$
$$= \sum_{j=0}^{\infty} (1-d)|||A|||^j + d(j+1)|||A|||^j = \frac{1-d}{1-|||A|||} + \frac{d}{(1-|||A|||)^2} = \frac{1 + |||A|||(d-1)}{(1-|||A|||)^2} < +\infty.$$

Another example in this direction can be obtained by using a nilpotent matrices. If A is nilpotent then (4.18) is always satisfied for any weighting sequence w. At the same time, there are nilpotent matrices with arbitrarily large norm $|||A|||$ which, once more, shows that (4.18) can hold, and hence the FMP, without (4.19) being necessarily true. We notice too that reservoir systems determined by nilpotent matrices always satisfy the echo state property even though they are not necessarily contractions.

Proof of statement (ii) We first prove the statement (4.20). For any $\mathbf{x} \in \overline{B_{\|\cdot\|}(\mathbf{0}, L)}$ and $\mathbf{z} \in \overline{B_{\|\cdot\|}(\mathbf{0}, M)}$,

$$\|F(\mathbf{x}, \mathbf{z})\| = \|A\mathbf{x} + \mathbf{c}\mathbf{z}\| \leq |||A|||L + |||\mathbf{c}|||M = L, \quad \text{as required.}$$

This implies that the reservoir map F in (4.17) restricts to a map $F_{L,M} : \overline{B_{\|\cdot\|}(\mathbf{0}, L)} \times \overline{B_{\|\cdot\|}(\mathbf{0}, M)} \longrightarrow \overline{B_{\|\cdot\|}(\mathbf{0}, L)}$ that is a contraction on the first entry with constant $|||A||| < 1$ and hence satisfies the hypotheses of Corollary 4.5. This guarantees the existence of a unique associated causal and time-invariant filter $U^F : K_M \longrightarrow K_L$ that has the fading memory property with respect to any weighting sequence w.

Echo state networks (ESN). Let $\sigma : \mathbb{R} \longrightarrow [-1, 1]$ be a squashing function, that is, σ is non-decreasing, $\lim_{x \to -\infty} \sigma(x) = -1$, and $\lim_{x \to \infty} \sigma(x) = 1$. Moreover, assume that $L_\sigma := \sup_{x \in \mathbb{R}} \{|\sigma'(x)|\} < +\infty$. Let $\boldsymbol{\sigma} : \mathbb{R}^N \longrightarrow [-1, 1]^N$ be the map obtained by componentwise application of the the squashing function σ. An echo state network is a reservoir system with linear readout and reservoir map given by

$$F(\mathbf{x}, \mathbf{z}) = \boldsymbol{\sigma}(A\mathbf{x} + \mathbf{c}\mathbf{z} + \boldsymbol{\zeta}), \quad \text{with} \quad A \in \mathbb{M}_N, \mathbf{c} \in \mathbb{M}_{N,n}, \boldsymbol{\zeta} \in \mathbb{R}^N. \tag{4.21}$$

We notice first that if $|||A|||L_\sigma < 1$ then F is a contraction on the first component with constant $|||A|||L_\sigma$ (see the second part in [Grig 18a, Corollary 3.2]). By construction, F maps into the compact space $[-1, 1]^N \subset \mathbb{R}^N$ and hence satisfies the hypotheses in the first part of Theorem 4.1. Consequently, for any weighting sequence w that satisfies

$$|||A|||L_\sigma L_w < 1 \tag{4.22}$$

there exists a unique reservoir filter $U^F : \ell_-^w(\mathbb{R}^n) \longrightarrow \ell_-^w(\mathbb{R}^N)$ associated to F that has the fading memory property with respect to w. By Corollary 4.5 this statement holds true for any w when one considers uniformly bounded inputs.

Non-homogeneous state-affine systems (SAS). These systems are determined by reservoir maps $F : \mathbb{R}^N \times \mathbb{R}^n \longrightarrow \mathbb{R}^N$ of the form

$$F(\mathbf{x}, \mathbf{z}) := p(\mathbf{z})\mathbf{x} + q(\mathbf{z}), \tag{4.23}$$

where p and q are polynomials with matrix and vector coefficients, respectively, that depending on their nature determine the following two families of SAS systems:

(i) Regular SAS. p and q are polynomials of degree r and s of the form:

$$
\begin{aligned}
p(\mathbf{z}) &= \sum_{\substack{i_1,\ldots,i_n\in\{0,\ldots,r\}\\ i_1+\cdots+i_n\leq r}} z_1^{i_1}\cdots z_n^{i_n} A_{i_1,\ldots,i_n}, \quad A_{i_1,\ldots,i_n}\in\mathbb{M}_N, \quad \mathbf{z}\in D_n\subset\mathbb{R}^n,\\
q(\mathbf{z}) &= \sum_{\substack{i_1,\ldots,i_n\in\{0,\ldots,s\}\\ i_1+\cdots+i_n\leq s}} z_1^{i_1}\cdots z_n^{i_n} B_{i_1,\ldots,i_n}, \quad B_{i_1,\ldots,i_n}\in\mathbb{M}_{N,1}, \quad \mathbf{z}\in D_n\subset\mathbb{R}^n.
\end{aligned}
$$

(ii) Trigonometric SAS. We use trigonometric polynomials instead:

$$
\begin{aligned}
p(\mathbf{z}) &= \sum_{k=1}^{r} A_k^p\cos(\mathbf{u}_k^p\cdot\mathbf{z}) + B_k^p\sin(\mathbf{v}_k^p\cdot\mathbf{z}), \quad A_k^p, B_k^p\in\mathbb{M}_N, \quad \mathbf{u}_k^p,\mathbf{v}_k^p\in\mathbb{R}^N, \quad \mathbf{z}\in D_n\subset\mathbb{R}^n,\\
q(\mathbf{z}) &= \sum_{k=1}^{s} A_k^q\cos(\mathbf{u}_k^q\cdot\mathbf{z}) + B_k^q\sin(\mathbf{v}_k^q\cdot\mathbf{z}), \quad A_k^q, B_k^q\in\mathbb{M}_{N,1}, \quad \mathbf{u}_k^q,\mathbf{v}_k^q\in\mathbb{R}^N, \quad \mathbf{z}\in D_n\subset\mathbb{R}^n.
\end{aligned}
$$

In both cases, define

$$
M_p := \sup_{\mathbf{z}\in D_n}\{|||p(\mathbf{z})|||\} \quad \text{and} \quad M_q := \sup_{\mathbf{z}\in D_n}\{|||q(\mathbf{z})|||\}.
$$

Note that for regular SAS defined by nontrivial polynomials, the set D_n needs to be bounded in order for M_p and M_q to be finite. Additionally, it is easy to see that F is a contraction on the first entry with constant M_p whenever $M_p < 1$, which is a condition that we will assume holds true in the rest of this example. Additionally, we assume that $M_q < +\infty$. Regular SAS are a generalization of the linear case that we considered in the first part of this section and hence two statements can be proved that are analogous to the ones in that part, namely:
(i) The reservoir system associated to (4.23) has the echo state property and defines a unique reservoir filter $U^F : \ell_-^w(\mathbb{R}^n) \longrightarrow \ell_-^w(\mathbb{R}^N)$ that has the fading memory property with respect to any weighting sequence w that satisfies the condition

$$
\sum_{j=0}^{\infty}\frac{M_p^j}{w_j} < +\infty. \tag{4.24}
$$

The FMP condition (4.3) that in this case reads as $M_pL_w < 1$ implies (4.24) but not vice versa.

(ii) If the inputs presented to the reservoir system associated to (4.23) are uniformly bounded then it has the fading memory property with respect to any weighting sequence. We obtain this result out of Corollary 4.5 by noting that for any $M > 0$,

$$
F(\overline{B_{\|\cdot\|}(\mathbf{0},L)}, \overline{B_{\|\cdot\|}(\mathbf{0},M)} \subset \overline{B_{\|\cdot\|}(\mathbf{0},L)}, \quad \text{with} \quad L := \frac{M_qM}{1-M_p}.
$$

We emphasize that in the case of regular SAS, this is the only situation for which one can have $M_p < 1$ and $M_q < +\infty$.

We prove only the statement **(i)** since statement **(ii)** can be easily obtained by mimicking the similar statement for the linear case. Indeed, a straightforward generalization of [Grig 18b, Proposition 14] shows that whenever $M_p < 1$ and $M_q < +\infty$, the reservoir system determined by (4.23) has a unique reservoir filter $U^F : \ell_-^w(\mathbb{R}^n) \longrightarrow \ell_-^w(\mathbb{R}^N)$ associated that is determined by the linear functional $H^F : \ell_-^w(\mathbb{R}^n) \longrightarrow \mathbb{R}^N$ given by

$$
H^F(\mathbf{z}) := \sum_{j=0}^{\infty}\left(\prod_{k=0}^{j-1} p(\mathbf{z}_{-k})\right) q(\mathbf{z}_{-j}).
$$

Mimicking the proof of [Grig 18b, Proposition 16] it can be shown that there exists a constant $C_{p,q} > 0$ that depends exclusively of p and q such that for any $\mathbf{z}, \mathbf{s} \in \ell^w_-(\mathbb{R}^n)$

$$\left\| H^F(\mathbf{z}) - H^F(\mathbf{s}) \right\| \leq C_{p,q} \sum_{j=0}^{\infty} M_p^j \left\| \mathbf{z}_{-j} - \mathbf{s}_{-j} \right\| = C_{p,q} \sum_{j=0}^{\infty} M_p^j \left\| \mathbf{z}_{-j} - \mathbf{s}_{-j} \right\| \frac{w_j}{w_j} \leq C_{p,q} \left\| \mathbf{z} - \mathbf{s} \right\|_w \sum_{j=0}^{\infty} \frac{M_p^j}{w_j},$$

which shows that $H^F : \ell^w_-(\mathbb{R}^n) \longrightarrow \mathbb{R}^N$ is Lipschitz continuous whenever the condition (4.24) holds. The last claim regarding the relation between (4.24) and the FMP condition (4.3) is proved by mimicking the similar statement for the linear case.

5 Differentiability in reservoir filters with unbounded inputs

We now extend the results in the previous section from continuity to differentiability. More specifically, we characterize the situations in which one can prove the existence and obtain the differentiability of reservoir filters out of the differentiability properties of the maps that define the reservoir system. This approach gives us in passing new techniques to establish the echo state and the fading memory properties of reservoir systems. In particular, differentiability being a local property, we show how systems that do not globally have any of these properties may still have them in a neighborhood of certain types of inputs. A phenomenon of this type has also been explored in [Manj 13].

It is worth emphasizing that the study of the differentiability properties of fading memory reservoir filters calls naturally for the handling of unbounded inputs since the definition of the Fréchet derivative requires them to be defined on open subsets of the Banach space $\ell^w_-(\mathbb{R}^n)$ that always contain unbounded sequences, as we saw in the first part of Lemma 2.1.

5.1 Differentiable reservoir filters associated to differentiable reservoir maps

The first result in this section shows that under certain conditions, the echo state and the fading memory properties associated to differentiable reservoir systems locally persist, that is, if a reservoir system has a unique filter associated to a specific input and it is continuous and differentiable at it, then the same property holds for neighboring inputs.

Theorem 5.1 (Local persistence of the ESP and FMP properties) *Let $F : \mathbb{R}^N \times \mathbb{R}^n \longrightarrow \mathbb{R}^N$ be a reservoir map and let w be a weighting sequence with finite inverse decay ratio L_w. Suppose that F is of class $C^1(\mathbb{R}^N \times \mathbb{R}^n)$ and that the corresponding reservoir system (1.1) has a solution $(\mathbf{x}^0, \mathbf{z}^0) \in \ell^w_-(\mathbb{R}^N) \times \ell^w_-(\mathbb{R}^n)$, that is, $\mathbf{x}^0_t = F(\mathbf{x}^0_{t-1}, \mathbf{z}^0_t)$, for all $t \in \mathbb{Z}_-$. Suppose, additionally, that*

$$L_F := \sup_{(\mathbf{x},\mathbf{z}) \in \mathbb{R}^N \times \mathbb{R}^n} \{ |||DF(\mathbf{x}, \mathbf{z})||| \} < +\infty. \tag{5.1}$$

Define

$$L_{F_x}(\mathbf{x}^0, \mathbf{z}^0) := \sup_{t \in \mathbb{Z}_-} \left\{ \left|\left|\left| D_x F(\mathbf{x}^0_{t-1}, \mathbf{z}^0_t) \right|\right|\right| \right\}$$

and suppose that

$$L_{F_x}(\mathbf{x}^0, \mathbf{z}^0) L_w < 1. \tag{5.2}$$

Then there exist open time-invariant neighborhoods $V_{\mathbf{x}^0}$ and $V_{\mathbf{z}^0}$ of $\mathbf{x}^0$ and $\mathbf{z}^0$ in $\ell^w_-(\mathbb{R}^N)$ and $\ell^w_-(\mathbb{R}^n)$, respectively, such that the reservoir system associated to F with inputs in $V_{\mathbf{z}^0}$ has the echo state property and hence determines a unique causal and time-invariant reservoir filter $U^F : (V_{\mathbf{z}^0}, \|\cdot\|_w) \longrightarrow (V_{\mathbf{x}^0}, \|\cdot\|_w)$. Moreover, U^F is differentiable at all the points of the form $T_{-t}(\mathbf{z}^0)$, $t \in \mathbb{Z}_-$, it is locally Lipschitz continuous on $V_{\mathbf{z}^0}$, and it hence has the fading memory property.

Remark 5.2 We refer to (5.2) as the ***persistence condition***. We emphasize that this inequality puts into relation the solution $(\mathbf{x}^0, \mathbf{z}^0)$ whose persistence we are studying with the weighting sequence w. In particular, that relation tells us that solutions are more likely to persist with respect to weighting sequences that decay more slowly (that is, L_w is smaller).

Remark 5.3 There is a situation where the persistence condition is particularly easy to verify, namely, when the solution of the reservoir system is constructed as a constant sequence coming from a fixed point of the reservoir map, that is, $(\mathbf{x}^0, \mathbf{z}^0) \in \mathbb{R}^N \times \mathbb{R}^n$ such that $F(\mathbf{x}^0, \mathbf{z}^0) = \mathbf{x}^0$. In that case $L_{F_x}(\mathbf{x}^0, \mathbf{z}^0) := \left|\left|\left| D_x F(\mathbf{x}^0, \mathbf{z}^0) \right|\right|\right|$.

Remark 5.4 The persistence condition (5.2) can be interpreted as a stability condition for the reservoir system determined by F at the solution $(\mathbf{x}^0, \mathbf{z}^0)$ with respect to perturbations in $\ell^w_-(\mathbb{R}^n)$. The persistence of solutions under stability conditions of that type has been thoroughly studied for many types of dynamical systems (see, for instance, [Mont 97a, Mont 97b, Orte 97, Chos 03]).

Remark 5.5 The derivative $DU^F(\mathbf{z}^0)$ at $\mathbf{z}^0$ of the locally defined reservoir filter U^F is determined by the differentiation of the relation (4.1). Indeed, for any $\mathbf{u} \in \ell^w_-(\mathbb{R}^n)$, and $t \in \mathbb{Z}_-$, the directional derivative $DU^F(\mathbf{z}^0) \cdot \mathbf{u}$ is determined by the recursions

$$
\begin{aligned}
(DU^F(\mathbf{z}^0) \cdot \mathbf{u})_t &= DF\left(U^F(\mathbf{z}^0)_{t-1}, \mathbf{z}^0_t\right) \cdot \left(\left(DU^F(\mathbf{z}^0) \cdot \mathbf{u}\right)_{t-1}, \mathbf{u}_t\right) && (5.3)\\
&= D_x F\left(U^F(\mathbf{z}^0)_{t-1}, \mathbf{z}^0_t\right) \cdot \left(DU^F(\mathbf{z}^0) \cdot \mathbf{u}\right)_{t-1} + D_z F(U(\mathbf{z}^0)_{t-1}, \mathbf{z}^0_t) \cdot \mathbf{u}_t. && (5.4)
\end{aligned}
$$

This relation implies, in particular, that $DU^F(\mathbf{z}^0) : \ell^w_-(\mathbb{R}^n) \longrightarrow \ell^w_-(\mathbb{R}^N)$ is a bounded linear operator and that

$$
\left|\left|\left| DU^F(\mathbf{z}^0) \right|\right|\right|_w \leq \frac{L_{F_z}(\mathbf{x}^0, \mathbf{z}^0)}{1 - L_{F_x}(\mathbf{x}^0, \mathbf{z}^0) L_w}, \tag{5.5}
$$

where

$$
L_{F_z}(\mathbf{x}^0, \mathbf{z}^0) := \sup_{t \in \mathbb{Z}_-} \left\{ \left|\left|\left| D_z F(\mathbf{x}^0_{t-1}, \mathbf{z}^0_t) \right|\right|\right| \right\}.
$$

Indeed, notice first that for any $t \in \mathbb{Z}_-$,

$$
\left|\left|\left| D_x F(\mathbf{x}^0_{t-1}, \mathbf{z}^0_t) \right|\right|\right| \leq \left|\left|\left| DF(\mathbf{x}^0_{t-1}, \mathbf{z}^0_t) \right|\right|\right| \quad \text{and} \quad \left|\left|\left| D_z F(\mathbf{x}^0_{t-1}, \mathbf{z}^0_t) \right|\right|\right| \leq \left|\left|\left| DF(\mathbf{x}^0_{t-1}, \mathbf{z}^0_t) \right|\right|\right|, \tag{5.6}
$$

which, using hypothesis (5.1) implies that

$$
L_{F_x}(\mathbf{x}^0, \mathbf{z}^0) \leq L_F < +\infty \quad \text{and} \quad L_{F_z}(\mathbf{x}^0, \mathbf{z}^0) \leq L_F < +\infty. \tag{5.7}
$$

Now, for any $\mathbf{u} \in \ell^w_-(\mathbb{R}^n)$, and $t \in \mathbb{Z}_-$, the relation (5.3) and the inequalities (5.7) imply that

$$
\begin{aligned}
\left\| DU^F(\mathbf{z}^0) \cdot \mathbf{u} \right\|_w &= \sup_{t \in \mathbb{Z}_-} \left\{ \left\| \left(DU^F(\mathbf{z}^0) \cdot \mathbf{u} \right)_t \right\| w_{-t} \right\} \\
&= \sup_{t \in \mathbb{Z}_-} \left\{ \left\| DF\left(U^F(\mathbf{z}^0)_{t-1}, \mathbf{z}^0_t\right) \cdot \left(\left(DU^F(\mathbf{z}^0) \cdot \mathbf{u}\right)_{t-1}, \mathbf{u}_t\right) \right\| w_{-t} \right\} \\
&= \sup_{t \in \mathbb{Z}_-} \left\{ \left\| D_x F\left(U^F(\mathbf{z}^0)_{t-1}, \mathbf{z}^0_t\right) \cdot \left(DU^F(\mathbf{z}^0) \cdot \mathbf{u}\right)_{t-1} + D_z F(U(\mathbf{z}^0)_{t-1}, \mathbf{z}^0_t) \cdot \mathbf{u}_t \right\| w_{-t} \right\} \\
&\leq L_{F_x}(\mathbf{x}^0, \mathbf{z}^0) \sup_{t \in \mathbb{Z}_-} \left\{ \left\| \left(DU^F(\mathbf{z}^0) \cdot \mathbf{u}\right)_{t-1} \right\| w_{-t} \right\} + L_{F_z}(\mathbf{x}^0, \mathbf{z}^0) \sup_{t \in \mathbb{Z}_-} \left\{ \left\| \mathbf{u}_t \right\| w_{-t} \right\} \\
&\leq L_{F_x}(\mathbf{x}^0, \mathbf{z}^0) \sup_{t \in \mathbb{Z}_-} \left\{ \left\| \left(DU^F(\mathbf{z}^0) \cdot \mathbf{u}\right)_{t-1} \right\| w_{-(t-1)} \frac{w_{-t}}{w_{-(t-1)}} \right\} \\
&\qquad + L_{F_z}(\mathbf{x}^0, \mathbf{z}^0) \sup_{t \in \mathbb{Z}_-} \left\{ \left\| \mathbf{u}_t \right\| w_{-t} \right\} \\
&\leq L_{F_x}(\mathbf{x}^0, \mathbf{z}^0) L_w \left\| DU^F(\mathbf{z}^0) \cdot \mathbf{u} \right\|_w + L_{F_z}(\mathbf{x}^0, \mathbf{z}^0) \left\| \mathbf{u} \right\|_w,
\end{aligned}
$$

which implies (5.5).

Proof of the Theorem. We start with a preliminary result whose proof mimics that of Lemma 2.5 and is also a consequence of Lemma 3.1. As we already did in the proof of Theorem 4.1, in the statement we consider the direct sum of weighted spaces $\ell^w_-(\mathbb{R}^N)\oplus\ell^w_-(\mathbb{R}^n)$ as a Banach space with the sum norm $\|\cdot\|_{w\oplus w}$ defined by $\|(\mathbf{u},\mathbf{v})\|_{w\oplus w}:=\|\mathbf{u}\|_w+\|\mathbf{v}\|_w$, for any $(\mathbf{u},\mathbf{v})\in\ell^w_-(\mathbb{R}^N)\oplus\ell^w_-(\mathbb{R}^n)$. Additionally, in all that follows V_n stands for any open convex subset of the Banach space $\left(\ell^w_-(\mathbb{R}^n),\|\cdot\|_w\right)$.

Lemma 5.6 *In the hypotheses of the theorem, consider the map*

$$\begin{array}{rccc}\mathcal{F}: & \ell^w_-(\mathbb{R}^N)\times V_n & \longrightarrow & (\mathbb{R}^N)^{\mathbb{Z}_-}\\ & (\mathbf{x},\mathbf{z}) & \longmapsto & (\mathcal{F}(\mathbf{x},\mathbf{z}))_t:=F(\mathbf{x}_{t-1},\mathbf{z}_t),\end{array}\tag{5.8}$$

where V_n is an open convex subset of $\ell^w_-(\mathbb{R}^n)$. Then,

(i) *$\mathcal{F}$ is Lipschitz continuous with constant L_FL_w and maps into $\ell^w_-(\mathbb{R}^N)$.*

(ii) *If F is of class $C^r(\mathbb{R}^N\times\mathbb{R}^n)$, $r\geq 1$, suppose that*

$$L_{F,r}:=\sup_{(\mathbf{x},\mathbf{z})\in\mathbb{R}^N\times\mathbb{R}^n}\{|||D^rF(\mathbf{x},\mathbf{z})|||\}<+\infty.\tag{5.9}$$

and let w' be any weighting sequence such that

$$c_{w',w^r}=\sup_{t\in\mathbb{Z}_-}\left\{\frac{w'_{-t}}{w^r_{-t}}\right\}<+\infty.\tag{5.10}$$

Then the map $\mathcal{F}$ is a functor between the sets

$$\mathcal{F}:\ell^w_-(\mathbb{R}^N)\times V_n\subset\ell^w_-(\mathbb{R}^N)\oplus\ell^w_-(\mathbb{R}^n)\longrightarrow\ell^{w'}_-(\mathbb{R}^N)$$

and is differentiable of order r and of class $C^{r-1}(\ell^w_-(\mathbb{R}^N)\times V_n)$. Moreover,

$$|||D^r\mathcal{F}(\mathbf{x},\mathbf{z})|||_{w,w'}\leq L_{F,r}L^r_wc_{w',w^r},\ \textit{for all } (\mathbf{x},\mathbf{z})\in\ell^w_-(\mathbb{R}^N)\times V_n\tag{5.11}$$

and the map $D^{r-1}\mathcal{F}:\ell^w_-(\mathbb{R}^N)\times V_n\longrightarrow L^{r-1}(\ell^w_-(\mathbb{R}^n)\oplus\ell^w_-(\mathbb{R}^N),\ell^{w'}_-(\mathbb{R}^N))$ is Lipschitz continuous with Lipschitz constant $L_{F,r}L^r_wc_{w',w^r}$.

(iii) *The linear map $D_x\mathcal{F}(\mathbf{x}^0,\mathbf{z}^0):(\ell^w_-(\mathbb{R}^N),\|\cdot\|_w)\longrightarrow\left(\ell^w_-(\mathbb{R}^N),\|\cdot\|_w\right)$ is a contraction with constant $L_{F_x}(\mathbf{x}^0,\mathbf{z}^0)L_w<1$.*

These results also hold when the spaces $\left(\ell^w_-(\mathbb{R}^n),\|\cdot\|_w\right)$ and $\left(\ell^w_-(\mathbb{R}^N),\|\cdot\|_w\right)$ are replaced by $\left(\ell^\infty_-(\mathbb{R}^n),\|\cdot\|_\infty\right)$ and $\left(\ell^\infty_-(\mathbb{R}^N),\|\cdot\|_\infty\right)$, respectively. In that case, the statement is obtained by taking as the sequences w and w' the constant sequence w^ι given by $w^\iota_t:=1$, for all $t\in\mathbb{N}$. The inequality (5.11) holds true with $L_w=c_{w',w^r}=1$.

Proof of the lemma. **(i)** Notice first that, as we pointed out in (4.6), and using the notation in Lemma 2.5,

$$\mathcal{F}=\prod_{t\in\mathbb{Z}_-}F_t,\quad\text{where}\quad F_t:=F\circ p_t\circ(T_1\times\mathrm{id}_{V_n}):\ell^w_-(\mathbb{R}^N)\times V_n\longrightarrow\mathbb{R}^N.\tag{5.12}$$

Also, the hypothesis (5.1), the mean value theorem, and the convexity of the set

$$p_t\circ(T_1\times\mathrm{id}_{V_n})\left(\ell^w_-(\mathbb{R}^N)\times V_n\right)$$

imply that F is a Lipschitz function with constant L_F. A development identical to (4.7) guarantees that the maps F_t are Lipschitz and that $L_F L_w / w_{-t}$ is a Lipschitz constant of F_t, $t \in \mathbb{Z}_-$. Given that the sequence $c_{\mathcal{F}} := (L_F L_w / w_{-t})_{t \in \mathbb{Z}_-}$ is such that $\|c_{\mathcal{F}}\|_w = L_F L_w < +\infty$ and $\mathcal{F} = \prod_{t \in \mathbb{Z}_-} F_t$, the part **(ii)** of Lemma 2.5 guarantees that $\mathcal{F}$ is Lipschitz continuous and that $L_F L_w$ is a Lipschitz constant of $\mathcal{F}$.

Since by hypothesis the reservoir system has a solution $(\mathbf{x}^0, \mathbf{z}^0) \in \ell_-^w(\mathbb{R}^N) \times V_n$, we have that $\mathcal{F}(\mathbf{x}^0, \mathbf{z}^0) = \mathbf{x}^0 \in \ell_-^w(\mathbb{R}^N)$. This implies that $\mathcal{F}$ maps into $\ell_-^w(\mathbb{R}^N)$ since the Lipschitz condition that we just proved shows that for any $(\mathbf{x}, \mathbf{z}) \in \ell_-^w(\mathbb{R}^N) \times V_n$

$$\|\mathcal{F}(\mathbf{x}, \mathbf{z})\|_w \leq L_F L_w \left\|(\mathbf{x}, \mathbf{z}) - (\mathbf{x}^0, \mathbf{z}^0)\right\|_{w \oplus w} + \|\mathcal{F}(\mathbf{x}^0, \mathbf{z}^0)\|_w,$$

which shows that $\|\mathcal{F}(\mathbf{x}, \mathbf{z})\|_w < +\infty$ and hence that $\mathcal{F}(\mathbf{x}, \mathbf{z}) \in \ell_-^w(\mathbb{R}^N)$.

(ii) The expression (5.12), the chain rule, the finiteness of L_w, and the linearity of p_t and T_1 imply that for any $(\mathbf{x}, \mathbf{z}) \in \ell_-^w(\mathbb{R}^N) \times V_n$:

$$D^r F_t(\mathbf{x}, \mathbf{z}) = D^r F(\mathbf{x}_{t-1}, \mathbf{z}_t) \circ (p_t \circ (T_1 \times \mathrm{id}_{V_n}), \ldots, p_t \circ (T_1 \times \mathrm{id}_{V_n})) : \left(\ell_-^w(\mathbb{R}^N) \oplus \ell_-^w(\mathbb{R}^n)\right)^r \longrightarrow \mathbb{R}^N. \tag{5.13}$$

We now prove (5.11). Notice first that for $\mathbf{u} = \left(\mathbf{u}^1, \ldots, \mathbf{u}^r\right) = \left((\mathbf{u}_\mathbf{x}^1, \mathbf{u}_\mathbf{z}^1), \ldots, (\mathbf{u}_\mathbf{x}^r, \mathbf{u}_\mathbf{z}^r)\right) \in \left(\ell_-^w(\mathbb{R}^N) \oplus \ell_-^w(\mathbb{R}^n)\right)^r$ we can write using (5.1) and Lemma 3.1:

$$\begin{aligned}
\|D^r F_t(\mathbf{x}, \mathbf{z}) \cdot \mathbf{u}\| &= \left\|D^r F(\mathbf{x}_{t-1}, \mathbf{z}_t) \circ \left(p_t \circ (T_1 \times \mathrm{id}_{V_n})(\mathbf{u}^1), \ldots, p_t \circ (T_1 \times \mathrm{id}_{V_n})(\mathbf{u}^r)\right)\right\| \\
&\leq |||D^r F(\mathbf{x}_{t-1}, \mathbf{z}_t)||| \left\|p_t \circ (T_1 \times \mathrm{id}_{V_n})(\mathbf{u}^1)\right\| \cdots \|p_t \circ (T_1 \times \mathrm{id}_{V_n})(\mathbf{u}^r)\| \\
&\leq \frac{L_{F,r}}{w_{-t}^r} \left\|(T_1(\mathbf{u}_\mathbf{x}^1), \mathbf{u}_\mathbf{z}^1)\right\|_{w \oplus w} \cdots \|(T_1(\mathbf{u}_\mathbf{x}^r), \mathbf{u}_\mathbf{z}^r)\|_{w \oplus w} \\
&\leq \frac{L_{F,r}}{w_{-t}^r} \left(\left\|T_1(\mathbf{u}_\mathbf{x}^1)\right\|_w + \left\|\mathbf{u}_\mathbf{z}^1\right\|_w\right) \cdots (\|T_1(\mathbf{u}_\mathbf{x}^r)\|_w + \|\mathbf{u}_\mathbf{z}^r\|_w) \\
&\leq \frac{L_{F,r}}{w_{-t}^r} \left(L_w \left\|\mathbf{u}_\mathbf{x}^1\right\|_w + \left\|\mathbf{u}_\mathbf{z}^1\right\|_w\right) \cdots (L_w \|\mathbf{u}_\mathbf{x}^r\|_w + \|\mathbf{u}_\mathbf{z}^r\|_w) \\
&\leq \frac{L_{F,r} L_w^r}{w_{-t}^r} \left(\left\|\mathbf{u}_\mathbf{x}^1\right\|_w + \left\|\mathbf{u}_\mathbf{z}^1\right\|_w\right) \cdots (\|\mathbf{u}_\mathbf{x}^r\|_w + \|\mathbf{u}_\mathbf{z}^r\|_w) = \frac{L_{F,r} L_w^r}{w_{-t}^r} \left\|\mathbf{u}^1\right\|_{w \oplus w} \cdots \|\mathbf{u}^r\|_{w \oplus w},
\end{aligned}$$

which shows that

$$|||D^r F_t(\mathbf{x}, \mathbf{z})|||_w \leq \frac{L_{F,r} L_w^r}{w_{-t}^r}. \tag{5.14}$$

Since, as we saw in part **(i)** $\mathcal{F}$ maps into $\ell_-^w(\mathbb{R}^N)$, and by Lemma 2.4 $\ell_-^w(\mathbb{R}^N) \subset \ell_-^{w^r}(\mathbb{R}^N)$, then $\mathcal{F}$ also maps into $\ell_-^{w^r}(\mathbb{R}^N)$. Additionally, since the sequence $c^r := \left(L_{F,r} L_w^r / w_{-t}^r\right)_{t \in \mathbb{Z}_-}$ is such that $\|c^r\|_{w^r} = L_{F,r} L_w^r < +\infty$, the part **(iii)** of Lemma 2.5 guarantees that the map

$$\mathcal{F} : \ell_-^w(\mathbb{R}^N) \times V_n \longrightarrow \ell_-^{w^r}(\mathbb{R}^N)$$

is differentiable of order r and that

$$|||D^r \mathcal{F}(\mathbf{x}, \mathbf{z})|||_{w, w^r} \leq L_{F,r} L_w^r < +\infty. \tag{5.15}$$

This argument can be reproduced with the power sequence w^r replaced by any other sequence w' that satisfies (5.10), in which case, it is easy to see that $\ell_-^{w^r}(\mathbb{R}^N) \subset \ell_-^{w'}(\mathbb{R}^N)$, and we can conclude the differentiability of the map $\mathcal{F} : \ell_-^w(\mathbb{R}^N) \times V_n \longrightarrow \ell_-^{w'}(\mathbb{R}^N)$ for which the relation (5.15) is replaced by

$$|||D^r \mathcal{F}(\mathbf{x}, \mathbf{z})|||_{w, w'} \leq L_{F,r} L_w^r c_{w', w^r} < +\infty. \tag{5.16}$$

The rest of the statement is a consequence of part **(iii)** of Lemma 2.5 applied in this setup.

(iii) A computation similar to the one that was used to establish (5.13) leads to the following expression for the partial derivatives $D_x\mathcal{F}$ of $\mathcal{F}$:

$$D_x^r\mathcal{F}(\mathbf{x},\mathbf{z}) = \prod_{t\in\mathbb{Z}_-} D_x^r F_t(\mathbf{x},\mathbf{z}) = \prod_{t\in\mathbb{Z}_-} D_x^r F(\mathbf{x}_{t-1},\mathbf{z}_t)\circ(p_t\circ T_1,\ldots,p_t\circ T_1)\,. \tag{5.17}$$

Using this expression for $r=1$ and Lemma 3.1 we can write, for any $\mathbf{u}\in\ell_-^w(\mathbb{R}^N)$,

$$\begin{aligned}\left\|D_x\mathcal{F}(\mathbf{x}^0,\mathbf{z}^0)\cdot\mathbf{u}\right\|_w &= \sup_{t\in\mathbb{Z}_-}\left\{\|D_xF(\mathbf{x}^0_{t-1},\mathbf{z}^0_t)\circ(p_t\circ T_1)\,(\mathbf{u})\|w_{-t}\right\}\\ &\leq L_{F_x}(\mathbf{x}^0,\mathbf{z}^0)\sup_{t\in\mathbb{Z}_-}\left\{\|\!|p_t|\!\|_w\|T_1(\mathbf{u})\|_w w_{-t}\right\}\\ &\leq L_{F_x}(\mathbf{x}^0,\mathbf{z}^0)\sup_{t\in\mathbb{Z}_-}\left\{\frac{1}{w_{-t}}\|T_1(\mathbf{u})\|_w w_{-t}\right\}\leq L_{F_x}(\mathbf{x}^0,\mathbf{z}^0)L_w\left\|\mathbf{u}\right\|_w,\end{aligned}$$

as required. ▼

We now proceed with the proof of the theorem in which we obtain the persistence result as a consequence of the Implicit Function Theorem and of the Lemma 5.6 that we just proved. Using the same notation as in that result we define the map

$$\begin{array}{rccc}\mathcal{G}: & \ell_-^w(\mathbb{R}^N)\times\ell_-^w(\mathbb{R}^n) & \longrightarrow & \ell_-^w(\mathbb{R}^N)\\ & (\mathbf{x},\mathbf{z}) & \longmapsto & \mathcal{F}(\mathbf{x},\mathbf{z})-\mathbf{x},\end{array}$$

or equivalently, $\mathcal{G}=\mathcal{F}-\pi_N$, where $\pi_N:\ell_-^w(\mathbb{R}^N)\times\ell_-^w(\mathbb{R}^n)\longrightarrow\ell_-^w(\mathbb{R}^N)$ is just the projection onto the first factor.

Notice that by construction and the hypothesis on the point $(\mathbf{x}^0,\mathbf{z}^0)$ we have that

$$\mathcal{G}(\mathbf{x}^0,\mathbf{z}^0)=\mathbf{0}. \tag{5.18}$$

Since the projection π_N is linear and by Lemma 5.6 $\mathcal{F}$ is Lipschitz continuous and differentiable of order 1, then so is $\mathcal{G}=\mathcal{F}-\pi_N$. This implies in particular that the partial derivative $D_x\mathcal{G}(\mathbf{x}^0,\mathbf{z}^0):\ell_-^w(\mathbb{R}^N)\longrightarrow\ell_-^w(\mathbb{R}^N)$ is a bounded operator that we now set to prove that it is an isomorphism. We proceed in two stages that show how the hypotheses in the statement of the theorem imply that this linear map is both injective and surjective.

The partial derivative $D_x\mathcal{G}(\mathbf{x}^0,\mathbf{z}^0):\ell_-^w(\mathbb{R}^N)\longrightarrow\ell_-^w(\mathbb{R}^N)$ is injective. Notice first that,

$$D_x\mathcal{G}(\mathbf{x}^0,\mathbf{z}^0)\cdot\mathbf{u}=D_x\mathcal{F}(\mathbf{x}^0,\mathbf{z}^0)\cdot\mathbf{u}-\mathbf{u},\quad\text{for any }\mathbf{u}\in\ell_-^w(\mathbb{R}^N).$$

Consequently, the points $\mathbf{u}\in\ell_-^w(\mathbb{R}^N)$ such that $D_x\mathcal{G}(\mathbf{x}^0,\mathbf{z}^0)\cdot\mathbf{u}=\mathbf{0}$ coincide with the fixed points of the map $D_x\mathcal{F}(\mathbf{x}^0,\mathbf{z}^0):\ell_-^w(\mathbb{R}^N)\longrightarrow\ell_-^w(\mathbb{R}^N)$. Since by part **(iii)** of Lemma 5.6 $D_x\mathcal{F}(\mathbf{x}^0,\mathbf{z}^0)$ is a contracting linear map in $\ell_-^w(\mathbb{R}^N)$ it has hence only zero as unique fixed point and the claim follows.

The partial derivative $D_x\mathcal{G}(\mathbf{x}^0,\mathbf{z}^0):\ell_-^w(\mathbb{R}^N)\longrightarrow\ell_-^w(\mathbb{R}^N)$ is surjective. We prove that for any $\mathbf{v}\in\ell_-^w(\mathbb{R}^N)$ there exists $\mathbf{u}\in\ell_-^w(\mathbb{R}^N)$ such that $D_x\mathcal{G}(\mathbf{x}^0,\mathbf{z}^0)\cdot\mathbf{u}=\mathbf{v}$. By the definition of $\mathcal{F}$ in (5.8) and the expression of its partial derivative in (5.17), this equation is equivalent to the recursions,

$$\mathbf{v}_t=D_xF(\mathbf{x}^0_{t-1},\mathbf{z}^0_t)\cdot\mathbf{u}_{t-1}-\mathbf{u}_t,\quad\text{for all }t\in\mathbb{Z}_-. \tag{5.19}$$

This equation has a unique solution given by the series

$$\mathbf{u}_t=-\mathbf{v}_t+\sum_{j=1}^{\infty}D_xF(\mathbf{x}^0_{t-1},\mathbf{z}^0_t)\cdot D_xF(\mathbf{x}^0_{t-2},\mathbf{z}^0_{t-1})\cdots D_xF(\mathbf{x}^0_{t-j},\mathbf{z}^0_{t-j+1})(-\mathbf{v}_{t-j}),\quad t\in\mathbb{Z}_-. \tag{5.20}$$

Indeed, it is straightforward to show that (5.20) satisfies (5.19). It remains then to be shown that the sequence $\mathbf{u}$ determined by (5.20) belongs to $\ell^w_-(\mathbb{R}^N)$. In order to do so we first show that the series in (5.20) is convergent by proving that for any $t \in \mathbb{Z}_-$, the sequence $\{S_n\}_{n\in\mathbb{N}^+}$ defined by

$$S_n := \sum_{j=1}^{n} D_xF(\mathbf{x}^0_{t-1}, \mathbf{z}^0_t) \cdot D_xF(\mathbf{x}^0_{t-2}, \mathbf{z}^0_{t-1}) \cdots D_xF(\mathbf{x}^0_{t-j}, \mathbf{z}^0_{t-j+1})(-\mathbf{v}_{t-j})w_{-t}, \tag{5.21}$$

is a Cauchy sequence. This is so because for any $m, n \in \mathbb{N}^+$, $m \geq n$,

$$\begin{aligned}
&\|S_m - S_n\| = \\
&\left\| \sum_{j=n+1}^{m} D_xF(\mathbf{x}^0_{t-1}, \mathbf{z}^0_t) \cdot D_xF(\mathbf{x}^0_{t-2}, \mathbf{z}^0_{t-1}) \cdots D_xF(\mathbf{x}^0_{t-j}, \mathbf{z}^0_{t-j+1})(-\mathbf{v}_{t-j}) \right\| w_{-(t-j)} \frac{w_{-(t-j+1)}}{w_{-(t-j)}} \cdots \frac{w_{-t}}{w_{-(t-1)}} \\
&\leq \sum_{j=n+1}^{m} |||D_xF(\mathbf{x}^0_{t-1}, \mathbf{z}^0_t)||| \cdot |||D_xF(\mathbf{x}^0_{t-2}, \mathbf{z}^0_{t-1})||| \cdots |||D_xF(\mathbf{x}^0_{t-j}, \mathbf{z}^0_{t-j+1})||| \, \|-\mathbf{v}_{t-j}\| \, w_{-(t-j)} L_w^j \\
&\leq \sum_{j=n+1}^{m} L_{F_x}(\mathbf{x}^0, \mathbf{z}^0)^j L_w^j \|\mathbf{v}\|_w = \frac{\left(L_{F_x}(\mathbf{x}^0, \mathbf{z}^0)L_w\right)^{n+1} - \left(L_{F_x}(\mathbf{x}^0, \mathbf{z}^0)L_w\right)^{m+1}}{1 - L_{F_x}(\mathbf{x}^0, \mathbf{z}^0)L_w} \|\mathbf{v}\|_w,
\end{aligned} \tag{5.22}$$

which can be made as small as we want because the sequence $\{\left(L_{F_x}(\mathbf{x}^0, \mathbf{z}^0)L_w\right)^j\}_{j\in\mathbb{N}^+}$ is convergent and hence Cauchy due to the hypothesis $L_{F_x}(\mathbf{x}^0, \mathbf{z}^0)L_w < 1$. This implies that $\{S_n\}_{n\in\mathbb{N}^+}$ is convergent and hence so is the series that defines $\mathbf{u}_t$ in (5.20).

It remains to be shown that the sequence $\mathbf{u} := (\mathbf{u}_t)_{t\in\mathbb{Z}_-}$ defined by (5.20) is an element of $\ell^w_-(\mathbb{R}^N)$. Following the same strategy that we used to construct the inequalities (5.22) it is easy to see that

$$\|\mathbf{u}_t\| \, w_{-t} \leq \frac{1}{1 - L_{F_x}(\mathbf{x}^0, \mathbf{z}^0)L_w} \|\mathbf{v}\|_w, \text{ for all } t \in \mathbb{Z}_-.$$

Consequently,

$$\|\mathbf{u}\|_w = \sup_{t\in\mathbb{Z}_-} \{\|\mathbf{u}_t\| \, w_{-t}\} \leq \frac{1}{1 - L_{F_x}(\mathbf{x}^0, \mathbf{z}^0)L_w} \|\mathbf{v}\|_w < +\infty,$$

as required.

The partial derivative $D_x\mathcal{G}(\mathbf{x}^0, \mathbf{z}^0) : \ell^w_-(\mathbb{R}^N) \longrightarrow \ell^w_-(\mathbb{R}^N)$ is a linear homeomorphism. This fact is a consequence of the Banach Isomorphism Theorem (see for instance [Abra 88]) that states that any continuous linear isomorphism of Banach spaces has necessarily a continuous inverse.

Using all the facts that we just proved, we can invoke the the Implicit Function Theorem as formulated in [Sche 97, page 671] (see also [Ver 74]) to show the existence of two open neighborhoods $\widetilde{V}_{\mathbf{x}^0}$ and $\widetilde{V}_{\mathbf{z}^0}$ of $\mathbf{x}^0$ and $\mathbf{z}^0$ in $\ell^w_-(\mathbb{R}^N)$ and $\ell^w_-(\mathbb{R}^n)$, respectively, and a unique Lipschitz continuous map $\widetilde{U^F} : (\widetilde{V}_{\mathbf{z}^0}, \|\cdot\|_w) \longrightarrow (\widetilde{V}_{\mathbf{x}^0}, \|\cdot\|_w)$ that is differentiable at $\mathbf{z}^0$ and satisfies

$$\mathcal{G}(\widetilde{U^F}(\mathbf{z}), \mathbf{z}) = \mathbf{0}, \text{ for all } \mathbf{z} \in \widetilde{V}_{\mathbf{z}^0},$$

which is equivalent to $\mathcal{F}(\widetilde{U^F}(\mathbf{z}), \mathbf{z}) = U^F(\mathbf{z})$. In view of the identities (4.1) this means, in other words, that $\widetilde{U^F}$ is the unique reservoir filter with inputs in $\widetilde{V}_{\mathbf{z}^0}$ associated to the reservoir system determined by F. This filter is clearly causal and its Lipschitz continuity implies that it has the fading memory property.

We conclude the proof by showing that the filter $\widetilde{U^F}$ can be extended to a time-invariant filter U^F defined on the time-invariant saturations $V_{\mathbf{x}^0}$ and $V_{\mathbf{z}^0}$ of the sets $\widetilde{V}_{\mathbf{x}^0}$ and $\widetilde{V}_{\mathbf{z}^0}$, respectively, and that has the properties listed in the statement. Indeed, define

$$V_{\mathbf{x}^0} := \bigcup_{t\in\mathbb{Z}_-} T_{-t}\left(\widetilde{V}_{\mathbf{x}^0}\right) \quad \text{and} \quad V_{\mathbf{z}^0} := \bigcup_{t\in\mathbb{Z}_-} T_{-t}\left(\widetilde{V}_{\mathbf{z}^0}\right).$$

The sets $V_{\mathbf{x}^0}$ and $V_{\mathbf{z}^0}$ are by construction time-invariant and open by the openness of the maps T_{-t} that we established in part **(ii)** of Lemma 3.1. Define now the map $U^F : V_{\mathbf{z}^0} \longrightarrow V_{\mathbf{x}^0}$ as

$$U^F\left(T_{-t}(\mathbf{z})\right) := T_{-t}\left(\widetilde{U^F}(\mathbf{z})\right), \quad \text{for some} \quad t\in\mathbb{Z}_- \text{ and } \mathbf{z}\in\widetilde{V}_{\mathbf{z}^0}. \tag{5.23}$$

We first show that U^F is well-defined and time-invariant. Let $t_1, t_2 \in \mathbb{Z}_-$ and $\mathbf{z}_1, \mathbf{z}_2 \in \widetilde{V}_{\mathbf{z}^0}$ be such that $T_{-t_1}(\mathbf{z}_1) = T_{-t_2}(\mathbf{z}_2)$. Let us now show that

$$U^F\left(T_{-t_1}(\mathbf{z}_1)\right) = U^F\left(T_{-t_2}(\mathbf{z}_2)\right). \tag{5.24}$$

Indeed, for any $t \in \mathbb{Z}_-$, the definition (5.23) and the causality of $\widetilde{U^F}$ imply that

$$\begin{aligned}\left(U^F\left(T_{-t_1}(\mathbf{z}_1)\right)\right)_t &= \left(T_{-t_1}\left(\widetilde{U^F}(\mathbf{z}_1)\right)\right)_t \\ &= \widetilde{U^F}(\mathbf{z}_1)_{t+t_1} = \widetilde{U^F}(\mathbf{z}_2)_{t+t_2} = \left(T_{-t_2}\left(\widetilde{U^F}(\mathbf{z}_2)\right)\right)_t = \left(U^F\left(T_{-t_2}(\mathbf{z}_2)\right)\right)_t,\end{aligned}$$

which proves (5.24). The time-invariance of U^F, as defined in (5.23), is straightforward.

We conclude by showing that U^F is differentiable at all the points of the form $T_{-t}(\mathbf{z}^0)$, $t \in \mathbb{Z}_-$ and that it is locally Lipschitz continuous on $V_{\mathbf{z}^0}$. Since differentiability is a local property, it suffices to prove this property for the restriction of U^F to open sets. Before we do that, we note that since by part **(ii)** of Lemma 3.1 the map $T_{-t} : \widetilde{V}_{\mathbf{z}^0} \longrightarrow T_{-t}\left(\widetilde{V}_{\mathbf{z}^0}\right)$ is a submersion, the Local Onto Theorem (see [Abra 88, Theorem 3.5.2]) guarantees that for $\mathbf{z}' := T_{-t}(\mathbf{z}^0) \in T_{-t}\left(\widetilde{V}_{\mathbf{z}^0}\right)$ there exists an open neighborhood $V_{\mathbf{z}'} \subset T_{-t}\left(\widetilde{V}_{\mathbf{z}^0}\right)$ and a smooth section $\sigma_{\mathbf{z}'} : V_{\mathbf{z}'} \longrightarrow \widetilde{V}_{\mathbf{z}^0}$ of T_{-t} that satisfies that

$$\sigma_{\mathbf{z}'}(\mathbf{z}') = \mathbf{z}^0 \quad \text{and} \quad T_{-t}\circ\sigma_{\mathbf{z}'} = \mathrm{id}_{V_{\mathbf{z}'}}. \tag{5.25}$$

The section $\sigma_{\mathbf{z}'}$ allows us to write down the restriction $U^F|_{V_{\mathbf{z}'}}$ of U^F to the open subset $V_{\mathbf{z}'}$ as

$$U^F|_{V_{\mathbf{z}'}}(\mathbf{z}) = T_{-t}\circ\widetilde{U^F}\left(\sigma_{\mathbf{z}'}(\mathbf{z})\right), \quad \text{for all} \quad \mathbf{z}\in V_{\mathbf{z}'}. \tag{5.26}$$

This is so because by (5.25) we have that $\mathbf{z} = T_{-t}\left(\sigma_{\mathbf{z}'}(\mathbf{z})\right)$, with $\sigma_{\mathbf{z}'}(\mathbf{z}) \in \widetilde{V}_{\mathbf{z}^0}$, as well as by (5.23). Consequently, since by (5.26) the restriction $U^F|_{V_{\mathbf{z}'}}$ is a composition of Lipschitz continuous functions then so is $U^F|_{V_{\mathbf{z}'}}$. The differentiability of U^F at the point $\mathbf{z}' = T_{-t}(\mathbf{z}^0)$ can also be concluded using (5.26) by invoking the differentiability of T_{-t} and $\sigma_{\mathbf{z}'}$ on their domains and the differentiability of $\widetilde{U^F}$ at $\sigma_{\mathbf{z}'}(\mathbf{z}') = \mathbf{z}^0$. ■

The previous theorem proves that when the persistence condition (5.2) is satisfied at a preexisting solution of a reservoir system then this system has a unique fading memory (and differentiable) filter associated for neighboring inputs. In the next results we show that a global version of that condition ensures first, that globally defined reservoir filters exist, and second, that those filters are differentiable and hence have the fading memory property.

Theorem 5.7 (Characterization of global reservoir filter differentiability) *Let $F:\mathbb{R}^N\times\mathbb{R}^n\longrightarrow\mathbb{R}^N$ be a reservoir map of class $C^1(\mathbb{R}^N\times\mathbb{R}^n)$ and let w be a weighting sequence with finite inverse decay ratio L_w.*

(i) *Suppose that F satisfies (5.1) and define*

$$L_{F_x}:=\sup_{(\mathbf{x},\mathbf{z})\in\mathbb{R}^N\times\mathbb{R}^n}\{|||D_xF(\mathbf{x},\mathbf{z})|||\}\quad\text{and}\quad L_{F_z}:=\sup_{(\mathbf{x},\mathbf{z})\in\mathbb{R}^N\times\mathbb{R}^n}\{|||D_zF(\mathbf{x},\mathbf{z})|||\}.$$

If the reservoir system (1.1) associated to F has a solution $(\mathbf{x}^0,\mathbf{z}^0)\in\ell_-^w(\mathbb{R}^N)\times\ell_-^w(\mathbb{R}^n)$, that is, $\mathbf{x}_t^0=F(\mathbf{x}_{t-1}^0,\mathbf{z}_t^0)$, for all $t\in\mathbb{Z}_-$, and

$$L_{F_x}L_w<1 \tag{5.27}$$

then it has the echo state property and hence determines a unique causal and time-invariant reservoir filter $U^F:(\ell_-^w(\mathbb{R}^n),\|\cdot\|_w)\longrightarrow(\ell_-^w(\mathbb{R}^N),\|\cdot\|_w)$. Moreover, U^F is differentiable and Lipschitz continuous on $\ell_-^w(\mathbb{R}^n)$ with Lipschitz constant L_{U^F} given by

$$L_{U^F}:=\frac{L_{F_z}}{1-L_{F_x}L_w}\quad\text{and}\quad |||DU^F(\mathbf{z})|||_w\leq\frac{L_{F_z}}{1-L_{F_x}L_w},\quad\text{for any }\mathbf{z}\in\ell_-^w(\mathbb{R}^n). \tag{5.28}$$

The filter U^F has hence the fading memory property.

(ii) *Conversely, let $V_n\subset\ell_-^w(\mathbb{R}^n)$ be an open and time-invariant subset of $\ell_-^w(\mathbb{R}^n)$ and assume that the reservoir system (1.1) associated to F has a unique causal and time-invariant reservoir filter $U^F:V_n\longrightarrow\ell_-^w(\mathbb{R}^N)$ that is differentiable at $\mathbf{z}^0\in\ell_-^w(\mathbb{R}^n)$. Then,*

$$\rho\left(\left(\prod_{t\in\mathbb{Z}_-}D_xF\left(U^F(\mathbf{z}^0)_{t-1},\mathbf{z}_t^0\right)\right)\circ T_1\right)<1, \tag{5.29}$$

where ρ stands for the spectral radius. This in turn implies that

$$\lim_{k\to+\infty}\left(|||D_xF\left(U^F(\mathbf{z}^0)_{-1},\mathbf{z}_0^0\right)\circ\cdots\circ D_xF\left(U^F(\mathbf{z}^0)_{-k},\mathbf{z}_{-k+1}^0\right)|||\frac{1}{w_k}\right)=0. \tag{5.30}$$

Examples 5.8 We briefly examine the form that the hypotheses of Theorem 5.7 take for the three families of reservoir systems that we analyzed in Section 4.1:

(i) Linear reservoir maps. In this case, for any $\mathbf{x}\in\mathbb{R}^N$ and $\mathbf{z}\in\mathbb{R}^n$,

$$DF(\mathbf{x},\mathbf{z})=(A\,|\,\mathbf{c}),\quad D_xF(\mathbf{x},\mathbf{z})=A,\quad\text{and}\quad D_zF(\mathbf{x},\mathbf{z})=\mathbf{c}.$$

Consequently $L_F=|||(A\,|\,\mathbf{c})|||$, $L_{F_x}=|||A|||$, $L_{F_z}=|||\mathbf{c}|||$. The condition (5.1) is always satisfied and in this case the sufficient differentiability condition (5.27) amounts to $|||A|||L_w<1$ that, as we saw in (4.19), is the same as the sufficient condition for the FMP to hold.

(ii) Echo state networks (ESN). Consider an ESN constructed using a squashing function σ that satisfies that $L_\sigma:=\sup_{x\in\mathbb{R}}\{|\sigma'(x)|\}<+\infty$. In this case, for any $\mathbf{x}\in\mathbb{R}^N$ and $\mathbf{z}\in\mathbb{R}^n$,

$$\begin{aligned}DF(\mathbf{x},\mathbf{z})&=D\boldsymbol{\sigma}(A\mathbf{x}+\mathbf{c}\mathbf{z}+\boldsymbol{\zeta})\circ(A\,|\,\mathbf{c}),\\D_xF(\mathbf{x},\mathbf{z})&=D\boldsymbol{\sigma}(A\mathbf{x}+\mathbf{c}\mathbf{z}+\boldsymbol{\zeta})\circ A,\\D_zF(\mathbf{x},\mathbf{z})&=D\boldsymbol{\sigma}(A\mathbf{x}+\mathbf{c}\mathbf{z}+\boldsymbol{\zeta})\circ\mathbf{c}.\end{aligned}$$

Notice that $|||D\boldsymbol{\sigma}(\mathbf{x})||| < L_\sigma < +\infty$, for any $\mathbf{x} \in \mathbb{R}^N$, and hence

$$\begin{array}{rcl} |||DF(\mathbf{x},\mathbf{z})||| & \leq & L_\sigma |||\,(A\,|\,\mathbf{c})||| < +\infty, \\ |||D_xF(\mathbf{x},\mathbf{z})||| & \leq & L_\sigma |||A||| < +\infty, \\ |||D_zF(\mathbf{x},\mathbf{z})||| & \leq & L_\sigma |||\mathbf{c}||| < +\infty, \end{array}$$

for any $\mathbf{x} \in \mathbb{R}^N$ and $\mathbf{z} \in \mathbb{R}^n$. This implies, in particular, that in this case

$$L_F < +\infty, \quad L_{F_x} < +\infty, \quad L_{F_z} < +\infty,$$

and the sufficient differentiability condition (5.27) is implied by the inequality

$$|||A|||L_\sigma L_w < 1. \tag{5.31}$$

(iii) Non-homogeneous state-affine systems (SAS). A straightforward computations shows that for any $\mathbf{x} \in \mathbb{R}^N$ and $\mathbf{z} \in \mathbb{R}^n$,

$$\begin{array}{rcl} DF(\mathbf{x},\mathbf{z}) & = & (p(\mathbf{z}), Dp(\mathbf{z})(\cdot)\mathbf{x} + Dq(\mathbf{z})(\cdot))\,, \\ D_xF(\mathbf{x},\mathbf{z}) & = & p(\mathbf{z}), \\ D_zF(\mathbf{x},\mathbf{z}) & = & Dp(\mathbf{z})(\cdot)\mathbf{x} + Dq(\mathbf{z})(\cdot). \end{array} \tag{5.32}$$

As we already pointed out, for regular SAS defined by nontrivial polynomials the norm $|||p(z)|||$ is not bounded in $\mathbb{R}^n$ and hence $L_{F_x} = \sup_{(\mathbf{x},\mathbf{z})\in\mathbb{R}^N\times\mathbb{R}^n} \{|||D_xF(\mathbf{x},\mathbf{z})|||\} = \sup_{\mathbf{z}\in\mathbb{R}^n} \{|||p(\mathbf{z})|||\} = M_p$ is not finite; the same applies to L_F, which implies that in this case neither (5.1) nor (5.27) can be satisfied.

This is not the case for trigonometric SAS for which the norms of the derivatives in (5.32) are bounded on their domains which, in particular, implies that $L_F < +\infty$, $L_{F_x} < +\infty$, and $L_{F_z} < +\infty$. Moreover, the sufficient differentiability condition (5.27) in this case reads

$$M_p L_w < 1.$$

Proof of the theorem. **(i)** We start with a lemma that shows how condition (5.27) guarantees the existence of a globally defined filter $U^F : (\ell^w_-(\mathbb{R}^n), \|\cdot\|_w) \longrightarrow (\ell^w_-(\mathbb{R}^N), \|\cdot\|_w)$.

Lemma 5.9 *Let $F : \mathbb{R}^N \times \mathbb{R}^n \longrightarrow \mathbb{R}^N$ be a reservoir map of class $C^1(\mathbb{R}^N \times \mathbb{R}^n)$ and let w be a weighting sequence with finite inverse decay ratio L_w. The reservoir map F is a contraction on the first entry if and only if*

$$L_{F_x} < 1. \tag{5.33}$$

Moreover, whenever conditions (5.1) and (5.27) are satisfied and $(\mathbf{x}^0, \mathbf{z}^0) \in (\mathbb{R}^N)^{\mathbb{Z}_-} \times (\mathbb{R}^n)^{\mathbb{Z}_-}$ is a solution of the reservoir system determined by F, then there exists a unique causal, time-invariant, and fading memory filter $U^F : (\ell^w_-(\mathbb{R}^n), \|\cdot\|_w) \longrightarrow (\ell^w_-(\mathbb{R}^N), \|\cdot\|_w)$.

Proof of the lemma. We first show that F is a contraction on the first entry if and only if $L_{F_x} < 1$. Suppose first that F is a contraction with contraction rate $0 < c < 1$. Then for any $(\mathbf{x},\mathbf{z}) \in \mathbb{R}^N \times \mathbb{R}^n$ and any $\mathbf{u} \in \mathbb{R}^N$, the partial derivative $D_xF(\mathbf{x},\mathbf{z}) : \mathbb{R}^N \longrightarrow \mathbb{R}^N$ satisfies that

$$\|D_xF(\mathbf{x},\mathbf{z}) \cdot \mathbf{u}\| = \lim_{t\to 0} \frac{\|F(\mathbf{x}+t\mathbf{u},\mathbf{z}) - F(\mathbf{x},\mathbf{z})\|}{t} \leq \lim_{t\to 0} \frac{ct\,\|\mathbf{u}\|}{t} = c\,\|\mathbf{u}\|\,,$$

which implies that $|||D_xF(\mathbf{x},\mathbf{z})||| \leq c$ and hence

$$L_{F_x} := \sup_{(\mathbf{x},\mathbf{z})\in D_N\times D_n} \{|||D_xF(\mathbf{x},\mathbf{z})|||\} \leq c < 1.$$

Conversely, suppose that $L_{F_x} < 1$. Since F is of class $C^1(\mathbb{R}^N \times \mathbb{R}^n)$, the mean value theorem guarantees that for any $(\mathbf{x}^1, \mathbf{z}), (\mathbf{x}^2, \mathbf{z}) \in \mathbb{R}^N \times \mathbb{R}^n$:

$$\left\| F(\mathbf{x}^1, \mathbf{z}) - F(\mathbf{x}^2, \mathbf{z}) \right\| \leq \sup_{(\mathbf{x},\mathbf{z}) \in \mathbb{R}^N \times \mathbb{R}^n} \{ |||D_x F(\mathbf{x}, \mathbf{z})||| \} \left\| \mathbf{x}^1 - \mathbf{x}^2 \right\| = L_{F_x} \left\| \mathbf{x}^1 - \mathbf{x}^2 \right\|,$$

and F is hence is a contraction on the first entry.

Suppose now that conditions (5.1) and (5.27) are satisfied and that $(\mathbf{x}^0, \mathbf{z}^0) \in (\mathbb{R}^N)^{\mathbb{Z}_-} \times (\mathbb{R}^n)^{\mathbb{Z}_-}$ is a solution of the reservoir system determined by F. Notice first that since $L_w > 1$ then the condition (5.27) implies that $L_{F_x} < 1$, necessarily, and hence, as we just proved, F is a contraction on the first entry with constant L_{F_x}. Additionally, as (5.1) is satisfied, the mean value theorem implies that F is Lipschitz continuous with constant L_F. All these facts allow us to invoke part **(ii)** of Theorem 4.1 to conclude the existence of the filter U^F in the statement, since in this situation, the condition (4.3) coincides with (5.27). ▼

The proof of the first part of the theorem can now be obtained by applying Theorem 5.1 to each point of the form

$$(U^F(\mathbf{z}), \mathbf{z}) \in \ell_-^w(\mathbb{R}^N) \times \ell_-^w(\mathbb{R}^n)$$

for which, according to its statement, there exist open neighborhoods $V_{U^F(\mathbf{z})}$ and $V_{\mathbf{z}}$ of $U^F(\mathbf{z})$ and $\mathbf{z}$ in $\ell_-^w(\mathbb{R}^N)$ and $\ell_-^w(\mathbb{R}^n)$, as well as a unique locally defined causal reservoir filter $\widetilde{U}_F : V_{\mathbf{z}} \longrightarrow V_{U^F(\mathbf{z})}$ associated to F. The uniqueness feature implies that $\widetilde{U}_F = U^F|_{V_{\mathbf{z}}}$. Moreover, since $\widetilde{U}_F$ is differentiable at $\mathbf{z}$ and we can repeat this construction for any point $\mathbf{z} \in \ell_-^w(\mathbb{R}^n)$ we can conclude that U^F is differentiable at any point in $\ell_-^w(\mathbb{R}^n)$.

Finally, the Lipschitz continuity on $\ell_-^w(\mathbb{R}^n)$ of U^F is a consequence of the mean value theorem, the inequality (5.5), and the fact that

$$\sup_{\mathbf{z} \in \ell_-^w(\mathbb{R}^n)} \left\{ \left|\left|\left| DU^F(\mathbf{z}) \right|\right|\right|_w \right\} \leq \sup_{\mathbf{z} \in \ell_-^w(\mathbb{R}^n)} \left\{ \frac{L_{F_z}(U^F(\mathbf{z}), \mathbf{z})}{1 - L_{F_x}(U^F(\mathbf{z}), \mathbf{z}))L_w} \right\} \leq \frac{L_{F_z}}{1 - L_{F_x} L_w},$$

which proves (5.28).

(ii) First of all, the existence of the filter $U^F : V_n \longrightarrow \ell_-^w(\mathbb{R}^N)$ and its differentiability at $\mathbf{z}^0 \in V_n$ imply that for any $\mathbf{u} \in \ell_-^w(\mathbb{R}^n)$ and $t \in \mathbb{Z}_-$ it satisfies (4.1) as well as (5.4), that is,

$$(DU^F(\mathbf{z}^0) \cdot \mathbf{u})_t = D_x F\left(U^F(\mathbf{z}^0)_{t-1}, \mathbf{z}_t^0\right) \cdot \left(DU^F(\mathbf{z}^0) \cdot \mathbf{u}\right)_{t-1} + D_z F(U(\mathbf{z}^0)_{t-1}, \mathbf{z}_t^0) \cdot \mathbf{u}_t.$$

This identity can be rewritten in terms of operators on sequences as

$$DU^F(\mathbf{z}^0) = \left(\prod_{t \in \mathbb{Z}_-} D_x F\left(U^F(\mathbf{z}^0)_{t-1}, \mathbf{z}_t^0\right) \right) \circ T_1 \circ DU^F(\mathbf{z}^0) + \prod_{t \in \mathbb{Z}_-} D_z F\left(U^F(\mathbf{z}^0)_{t-1}, \mathbf{z}_t^0\right),$$

or equivalently as

$$\left(\mathbb{I}_{\ell_-^w(\mathbb{R}^N)} - \left(\prod_{t \in \mathbb{Z}_-} D_x F\left(U^F(\mathbf{z}^0)_{t-1}, \mathbf{z}_t^0\right) \right) \circ T_1 \right) DU^F(\mathbf{z}^0) = \prod_{t \in \mathbb{Z}_-} D_z F\left(U^F(\mathbf{z}^0)_{t-1}, \mathbf{z}_t^0\right). \tag{5.34}$$

This identity determines $DU^F(\mathbf{z}^0)$ that by hypothesis exists if and only if the operator on the left hand side is invertible, which is in turn equivalent to the condition (5.29). We finally show that (5.29) implies (5.30).

We first notice that by Gelfand's formula [Lax 02, page 195] the condition (5.29) is equivalent to

$$\lim_{k\to+\infty} \left|\!\left|\!\left| \left((\prod_{t\in\mathbb{Z}_-} D_xF\left(U^F(\mathbf{z}^0)_{t-1}, \mathbf{z}^0_t\right)) \circ T_1 \right)^k \right|\!\right|\!\right|_w = 0.$$

This in turn implies that for any $\mathbf{u} \in \ell^w_-(\mathbb{R}^n)$, we have that

$$\lim_{k\to+\infty} \left\| \left((\prod_{t\in\mathbb{Z}_-} D_xF\left(U^F(\mathbf{z}^0)_{t-1}, \mathbf{z}^0_t\right)) \circ T_1 \right)^k (\mathbf{u}) \right\|_w = 0,$$

or, equivalently, that

$$\lim_{k\to+\infty} \left(\sup_{t\in\mathbb{Z}_-} \left\{ \left\| \left(D_xF\left(U^F(\mathbf{z}^0)_{t-1}, \mathbf{z}^0_t\right) \circ \cdots \circ D_xF\left(U^F(\mathbf{z}^0)_{t-k}, \mathbf{z}^0_{t-k+1}\right)\right) (\mathbf{u}_{t-k}) \right\| \right\} \right) = 0. \tag{5.35}$$

If we now take vectors $\mathbf{u} \in \ell^w_-(\mathbb{R}^n)$ in (5.35) of the form $\mathbf{u}_t := \widetilde{\mathbf{u}}/w_{-t}$, $t \in \mathbb{Z}_-$, with $\widetilde{\mathbf{u}} \in \mathbb{R}^n$ such that $\|\widetilde{\mathbf{u}}\| = 1$, and we take the supremum in (5.35) with respect to all those vectors $\widetilde{\mathbf{u}}$, we obtain that

$$\lim_{k\to+\infty} \left(\sup_{t\in\mathbb{Z}_-} \left\{ \left|\!\left|\!\left| D_xF\left(U^F(\mathbf{z}^0)_{t-1}, \mathbf{z}^0_t\right) \circ \cdots \circ D_xF\left(U^F(\mathbf{z}^0)_{t-k}, \mathbf{z}^0_{t-k+1}\right) \right|\!\right|\!\right| \frac{w_{-t}}{w_{-(t-k)}} \right\} \right) = 0. \tag{5.36}$$

Given that

$$\begin{aligned} &\left|\!\left|\!\left| D_xF\left(U^F(\mathbf{z}^0)_{-1}, \mathbf{z}^0_0\right) \circ \cdots \circ D_xF\left(U^F(\mathbf{z}^0)_{-k}, \mathbf{z}^0_{-k+1}\right) \right|\!\right|\!\right| \frac{1}{w_k} \\ &\qquad \leq \sup_{t\in\mathbb{Z}_-} \left\{ \left|\!\left|\!\left| D_xF\left(U^F(\mathbf{z}^0)_{t-1}, \mathbf{z}^0_t\right) \circ \cdots \circ D_xF\left(U^F(\mathbf{z}^0)_{t-k}, \mathbf{z}^0_{t-k+1}\right) \right|\!\right|\!\right| \frac{w_{-t}}{w_{-(t-k)}} \right\}, \end{aligned}$$

the condition (5.30) follows. ■

Remark 5.10 We recall here an example that we introduced in Section 4.1 to show that, as it was already the case with the FMP condition (4.3) in Theorem 4.1, the differentiability condition (5.27) is sufficient but not necessary. Indeed, consider a linear system with matrix A given by

$$A = \begin{pmatrix} 0 & a \\ 0 & 0 \end{pmatrix}, \quad \text{with} \quad a > 0.$$

Given that $|||A||| = a$, the reservoir map determined by A is not necessarily a contraction on the first entry. Nevertheless, the nilpotency of A implies that the reservoir system associated to (4.17) always has a solution for any input $\mathbf{z} \in (\mathbb{R}^2)^{\mathbb{Z}_-}$ and hence has the ESP and induces a filter $U : (\mathbb{R}^2)^{\mathbb{Z}_-} \longrightarrow (\mathbb{R}^2)^{\mathbb{Z}_-}$ given by $U(\mathbf{z})_t := \mathbf{z}_t + A\mathbf{z}_{t-1}$, $t \in \mathbb{Z}_-$ or, equivalently, $U = \mathbb{I}_{(\mathbb{R}^2)^{\mathbb{Z}_-}} + (\prod_{t\in\mathbb{Z}_-} A) \circ T_1$. Consider now any weighting sequence w with finite inverse decay ratio L_w. Then the restriction of U to $\ell^w_-(\mathbb{R}^2)$ always maps into $\ell^w_-(\mathbb{R}^2)$, has the FMP, and it is differentiable. Indeed, it is easy to show using the linearity of the filter that $U = DU(\mathbf{z})$ for any $\mathbf{z} \in \ell^w_-(\mathbb{R}^2)$ and that

$$|||U|||_w = |||DU(\mathbf{z})|||_w \leq (1 + aL_w). \tag{5.37}$$

Note that in this case $L_{F_x} = |||A||| = a$ and as (5.37) shows the differentiability of U with respect to any weighting sequence with finite L_w, we can conclude that the condition (5.27) is not necessary for filter differentiability.

The following corollary puts together the previous theorem and a condition on the readout map that guarantees that the filter associated to the resulting reservoir system is differentiable.

Corollary 5.11 *Consider a reservoir system determined by a reservoir map $F : \mathbb{R}^N \times \mathbb{R}^n \longrightarrow \mathbb{R}^N$ of class $C^1(\mathbb{R}^N \times \mathbb{R}^n)$ and by a readout map $h : \mathbb{R}^N \longrightarrow \mathbb{R}^d$ that is also of class $C^1(\mathbb{R}^N)$. Assume, additionally that F satisfies the hypotheses in part* **(i)** *of Theorem 5.7 and that h is such that*

$$c_h := \sup_{\mathbf{x}\in\mathbb{R}^N} \{|||Dh(\mathbf{x})|||\} < +\infty, \tag{5.38}$$

and the sequence $\mathbf{y}^0 := \left(h\left(\mathbf{x}^0\right)\right)_{t\in\mathbb{Z}_-} = \left(h\left(U^F(\mathbf{z}^0)\right)\right)_{t\in\mathbb{Z}_-} \in \ell^w_-(\mathbb{R}^d)$. Then, the reservoir filter $U_h^F : (\ell^w_-(\mathbb{R}^n), \|\cdot\|_w) \longrightarrow (\ell^w_-(\mathbb{R}^d), \|\cdot\|_w)$ is differentiable at each point in its domain and it hence has the fading memory property.

Proof. Define first the map

$$\mathcal{H} := \prod_{t\in\mathbb{Z}_-} h \circ p_t : \ell^w_-(\mathbb{R}^N) \longrightarrow (\mathbb{R}^d)^{\mathbb{Z}_-}. \tag{5.39}$$

Given that $U_h^F = \mathcal{H} \circ U^F$ and by Theorem 5.7 the filter U^F is differentiable then it suffices to prove that $\mathcal{H}$ is differentiable. This is a consequence of part **(iii)** in Lemma 2.5 and the hypothesis (5.38). Indeed, let $H_t := h \circ p_t$, $t \in \mathbb{Z}_-$, and notice that by the first part of Lemma 3.1

$$\sup_{\mathbf{x}\in\ell^w_-(\mathbb{R}^N)} \{|||DH_t(\mathbf{x})|||\} \leq \sup_{\mathbf{x}_t\in\mathbb{R}^N} \{|||Dh(\mathbf{x}_t)|||\} \cdot \sup_{\mathbf{x}\in\ell^w_-(\mathbb{R}^N)} \{|||p_t(\mathbf{x})|||\} \leq \frac{c_h}{w_{-t}}.$$

Now, as $\left\|(c_h/w_{-t})_{t\in\mathbb{Z}_-}\right\|_w = c_h < +\infty$ and by hypothesis $\mathcal{H}(\mathbf{x}^0) \in \ell^w_-(\mathbb{R}^d)$ it follows from Lemma 2.5 that $\mathcal{H}$ maps into $\ell^w_-(\mathbb{R}^d)$ and that it is differentiable, as required. ■

In some occasions it is important to determine if a given filter is invertible. The differentiability of reservoir filters associated to reservoir systems associated to differentiable reservoir and readout maps that we established in the previous result allows us to use the inverse function theorem to formulate a sufficient invertibility condition. As we see in the next statement, this criterion can be written down entirely in terms of the derivatives of the reservoir and the readout maps.

Corollary 5.12 *Consider a reservoir system determined by a reservoir map $F : \mathbb{R}^N \times \mathbb{R}^n \longrightarrow \mathbb{R}^N$ and a readout map $h : \mathbb{R}^N \longrightarrow \mathbb{R}^d$ that are of class $C^1(\mathbb{R}^N \times \mathbb{R}^n)$ and $C^1(\mathbb{R}^N)$, respectively, and additionally satisfy the conditions spelled out in the statement of Corollary 5.11. Let $\mathbf{z} \in \ell^w_-(\mathbb{R}^n)$, $\mathbf{x} := U^F(\mathbf{z}) \in \ell^w_-(\mathbb{R}^N)$, and $\mathbf{y} := U_h^F(\mathbf{z}) \in \ell^w_-(\mathbb{R}^d)$, and suppose that the map*

$$D\mathcal{H}(\mathbf{x}) \circ \left(\mathbb{I}_{\ell^w_-(\mathbb{R}^N)} - \left(\prod_{t\in\mathbb{Z}_-} D_x F\left(\mathbf{x}_{t-1}, \mathbf{z}_t\right)\right) \circ T_1\right)^{-1} \circ \left(\prod_{t\in\mathbb{Z}_-} D_z F\left(\mathbf{x}_{t-1}, \mathbf{z}_t\right)\right) : \ell^w_-(\mathbb{R}^n) \longrightarrow \ell^w_-(\mathbb{R}^d) \tag{5.40}$$

is a linear homeomorphism (continuous linear bijection with continuous inverse) with $\mathcal{H}$ as defined in (5.39). Then there exist open neighborhoods $V_{\mathbf{z}} \subset \ell^w_-(\mathbb{R}^n)$ and $V_{\mathbf{y}} \subset \ell^w_-(\mathbb{R}^N)$ of $\mathbf{z}$ and $\mathbf{y}$, respectively, such that the restriction of the filter $U_h^F|_{V_{\mathbf{z}}} : V_{\mathbf{z}} \longrightarrow V_{\mathbf{y}}$ has an inverse $\left(U_h^F|_{V_{\mathbf{z}}}\right)^{-1}$. When the condition (5.40) is satisfied for all the solutions $(\mathbf{z}, U^F(\mathbf{z}))$ of the reservoir system determined by F then the reservoir filter U_h^F admits a global inverse $\left(U_h^F\right)^{-1} : U_h^F\left(\ell^w_-(\mathbb{R}^n)\right) \longrightarrow \ell^w_-(\mathbb{R}^n)$.

Proof. It is a straightforward consequence of the inverse function theorem as formulated in [Sche 97, page 670] (see also [Ver 74]) applied to the Fréchet derivative of $U_h^F = \mathcal{H} \circ U^F$ at the point $\mathbf{z} \in \ell^w_-(\mathbb{R}^n)$. It is easy to see using the chain rule and (5.34) (which is in turn a consequence of (5.4)) that this derivative coincides with the operator in (5.40) whose invertibility we require. ■

5.2 The local versus the global echo state property

Theorem 5.1 emphasizes the local nature of both the echo state and the fading memory properties by providing a sufficient condition that ensures the existence of a locally defined causal and time-invariant filter around a given solution that is shown to have the FMP. In contrast with this local approach, Theorem 5.7 characterizes the existence of a globally defined differentiable filter associated to a given reservoir system, that hence satisfies the FMP and the ESP for any input.

Even though the conditions in Theorems 5.1 and 5.7 are very alike, the latter is much stronger than the former. In the following paragraphs we illustrate with a family of ESNs of the type introduced in Section 4.1 how it is possible to be in violation of the global condition of Theorem 5.7 and nevertheless to find solutions of such reservoir systems around which one can locally define FMP reservoir filters. This example illustrates how *the ESP and the FMP are structural features of a reservoir system when considered globally but are mostly input dependent when considered only locally.* This important observation has already been noticed in [Manj 13] where, using tools coming from the theory of non-autonomous dynamical systems, sufficient conditions have been formulated (see, for instance, [Manj 13, Theorem 2]) that ensure the ESP in connection to a given specific input. The differentiability conditions that we impose to our reservoir systems allow us to draw similar conclusions and, additionally, to automatically conclude the FMP of the resulting locally defined reservoir filters.

Consider the one-dimensional echo state map $F : \mathbb{R} \times \mathbb{R} \longrightarrow \mathbb{R}$, where

$$F(x, z) := \sigma(ax + z), \quad \text{with} \quad a \in \mathbb{R} \quad \text{and} \quad \sigma(x) := \frac{x}{\sqrt{1 + x^2}}. \tag{5.41}$$

The sigmoid function σ in this expression has been chosen so that we can provide algebraic expressions in the following developments. Similar conclusions could nevertheless be drawn using other popular squashing functions.

The function σ maps the real line into the interval $[-1, 1]$ and it is easy to see, using the notation introduced in the examples 5.8, that $L_\sigma := \sup_{x\in\mathbb{R}}\{|\sigma'(x)|\} = 1$. Moreover, the one-dimensional character of the system makes that, in this case,

$$L_{F_x} = |a|. \tag{5.42}$$

Consequently, by Lemma 5.9, the reservoir map F is a contraction on the first entry if and only if $|a| < 1$, in which case, by Theorem 4.1, the associated ESN has the ESP and the FMP with respect to any input in $\ell_-^w(\mathbb{R}^n)$, where w is a weighting sequence that satisfies

$$|a|L_w < 1. \tag{5.43}$$

The FMP holds with respect to any sequence w if we consider uniformly bounded inputs by Corollary 4.5. Moreover, a well-known result for ESNs due to H. Jaeger (see [Jaeg 10, Proposition 3]) shows that the ESP cannot be satisfied whenever

$$|a| > 1. \tag{5.44}$$

Additionally, the global sufficient differentiability condition (5.27) in Theorem 5.7 states that the condition (5.43) also ensures that the ESP filter is also differentiable.

We now prove using Theorem 5.1 *the existence of locally defined FMP filters associated to this ESN in a neighborhood of certain inputs, even when condition* (5.44) *is satisfied which, as we already mentioned, prevents the global existence of such objects.*

Notice first that the solutions of the equation $\sigma(ax) = x$, $x \in \mathbb{R}$, are characterized by the relation

$$a^2x^2 = x^2(a^2x^2 + 1) \tag{5.45}$$

that has as solutions

$$\begin{cases} x_0 & = & 0, \\ x_a^{\pm} & = & \pm\frac{\sqrt{a^2-1}}{a}, \end{cases}$$

where the solutions in the second line obviously exist and are different from the first one only when $|a| > 1$, a condition that we assume holds true in the rest of the section. The condition (5.45) implies that the constant sequences $(\mathbf{x}_0, \mathbf{z}_0)$ and $(\mathbf{x}_a^{\pm}, \mathbf{z}_0)$ defined by

$$(\mathbf{x}_0, \mathbf{z}_0)_t := (x_0, 0) \quad \text{and} \quad (\mathbf{x}_a^{\pm}, \mathbf{z}_0)_t := (x_a^{\pm}, 0), \quad \text{for any} \quad t \in \mathbb{Z}_-,$$

are solutions of the reservoir system determined by F. Moreover, in the notation of Theorem 5.1, it is easy to see that

$$L_{F_x}(\mathbf{x}_0, \mathbf{z}_0) = |a| > 1 \quad \text{and} \quad L_{F_x}(\mathbf{x}_a^{\pm}, \mathbf{z}_0) = \frac{1}{a^2} < 1.$$

The persistence condition (5.2) in that result implies that for any weighting sequence that satisfies

$$\frac{L_w}{a^2} < 1,$$

there exist open time-invariant neighborhoods $V_{\mathbf{x}_a^{\pm}}$ and $V_{\mathbf{z}^0}$ of $\mathbf{x}_a^{\pm}$ and $\mathbf{z}^0$ in $\ell_-^w(\mathbb{R}^N)$ and $\ell_-^w(\mathbb{R}^n)$, respectively, such that the reservoir system associated to F with inputs in $V_{\mathbf{z}^0}$ has the echo state property and hence determines a unique causal, time-invariant, and FMP reservoir filter $U^F : (V_{\mathbf{z}^0}, \|\cdot\|_w) \longrightarrow (V_{\mathbf{x}^0}, \|\cdot\|_w)$.

5.3 Remote past input independence and the state forgetting property for unbounded inputs

In Section 3.1 we saw how fading memory filters presented with uniformly bounded inputs exhibit what we called the uniform input forgetting property. An analysis of the proof of the main result in that section, namely Theorem 3.6, shows that the compactness of the space of inputs guaranteed the existence of a modulus of continuity for the filter, which ensured the validity of the input forgetting property and, moreover, it made it uniform. In the context of reservoir systems, we saw in Theorems 4.1 and 5.7 that there are very weak hypotheses that, even when the inputs are not uniformly bounded, guarantee that the associated reservoir filters are Lipschitz and hence have a modulus of continuity. This allows us to prove an input forgetting property in that more general context.

Theorem 5.13 (Input forgetting property for FMP reservoir filters) *Let $F : D_N \times D_n \longrightarrow D_N$ be a reservoir map where $D_n \subset \mathbb{R}^n$, $D_N \subset \mathbb{R}^N$, $n, N \in \mathbb{N}^+$. Assume that the hypotheses of Theorem 4.1 part* **(ii)** *or 5.7 part* **(i)** *are satisfied. Let $U^F : (V_n, \|\cdot\|_w) \longrightarrow ((D_N)^{\mathbb{Z}_-} \cap \ell_-^w(\mathbb{R}^N), \|\cdot\|_w)$ be the associated causal and time-invariant reservoir filter ($V_n \subset (D_n)^{\mathbb{Z}_-} \cap \ell_-^w(\mathbb{R}^n)$ under the hypotheses of Theorem 4.1; $V_n = \ell_-^w(\mathbb{R}^n)$ and $D_N = \mathbb{R}^N$ under the hypotheses of Theorem 5.7). Then, for any $\mathbf{u}, \mathbf{v} \in \ell_-^w(\mathbb{R}^n)$ and $\mathbf{z} \in (D_N)^{\mathbb{Z}_-}$ we have that*

$$\lim_{t \to +\infty} \left\| U^F(\mathbf{u}, \mathbf{z})_t - U^F(\mathbf{v}, \mathbf{z})_t \right\| = 0. \tag{5.46}$$

Proof. It mimics the proof of Theorem 3.6 using as modulus of continuity the map $\omega_{U^F}(t) := L_{U^F} t$, $t \geq 0$, where L_{U^F} is the Lipschitz constant whose existence is ensured by the hypotheses of Theorem 4.1 or 5.7 and given by (4.4) or by (5.28). ■

Remark 5.14 As we saw in Remark 4.4, Theorem 4.1 can be extended to continuous reservoir systems with inputs and outputs in $\ell_-^{p,w}(\mathbb{R}^n)$ and $\ell_-^{p,w}(\mathbb{R}^N)$, respectively. In particular, we saw that the resulting filters are Lipschitz and hence have a non-trivial modulus of continuity. This implies that a result analogous to Theorem 5.13 can be proved for such systems that hence could also be referred to as fading memory from a dynamical point of view.

When filters are differentiable, there is one more way to measure how they forget inputs simply by looking at their partial derivatives with respect to past input components. The result is a differential input forgetting property that, unlike Theorem 5.13, can be formulated in a uniform way even when the inputs are not uniformly bounded.

Theorem 5.15 (Differential uniform input forgetting property) *Assume that the hypotheses of Theorem 5.7* **(i)** *are satisfied. Let $D_{z_t^i} H^F(\mathbf{z}) \in \mathbb{R}^N$ be the partial derivative of the reservoir functional $H^F : \ell_-^w(\mathbb{R}^n) \longrightarrow \mathbb{R}^N$ with respect to the i-th component of the t-th entry of $\mathbf{z} \in \ell_-^w(\mathbb{R}^n)$. Then, there exists a monotonously decreasing sequence w^F with zero limit such that, for any $t \in \mathbb{Z}_-$,*

$$\left\| D_{z_t^i} H^F(\mathbf{z}) \right\| \leq w_{-t}^F, \quad \text{for any} \quad \mathbf{z} \in \ell_-^w(\mathbb{R}^n) \quad \text{and} \quad i \in \{1, \ldots, n\}. \tag{5.47}$$

Proof. Let $\mathbf{e}^{i,t} := \left(\ldots, \mathbf{0}, \mathbf{e}^i, \mathbf{0}, \ldots, \mathbf{0}\right) \in \ell_-^w(\mathbb{R}^n)$, where the vector $\mathbf{e}^i$ is the canonical vector in $\mathbb{R}^n$ and it is placed in the t-th position. Then, since $\left\|\mathbf{e}^{i,t}\right\|_w = w_{-t}$, we have by (5.28) and for any $\mathbf{z} \in \ell_-^w(\mathbb{R}^n)$ that

$$\left\| D_{z_t^i} H^F(\mathbf{z}) \right\| = \left\| DH^F(\mathbf{z}) \cdot \mathbf{e}^{i,t} \right\| \leq \left|\left|\left| DH^F(\mathbf{z}) \right|\right|\right|_w \left\|\mathbf{e}^{i,t}\right\|_w \leq \left|\left|\left| p_0 \circ DU^F(\mathbf{z}) \right|\right|\right|_w w_{-t} \leq \frac{L_{F_z}}{1 - L_{F_x} L_w} w_{-t},$$

which proves (5.47) by setting

$$w_t^F := \frac{L_{F_z}}{1 - L_{F_x} L_w} w_t, \quad t \in \mathbb{N}. \quad \blacksquare$$

Apart from the filters that reservoir maps define when they have the echo state property, we can also use this object to define controlled forward-looking dynamical systems and flows. Indeed, given $F : D_N \times D_n \longrightarrow D_N$ a reservoir map, we denote by $U^F : (D_n)^{\mathbb{N}^+} \times D_N \longrightarrow (D_N)^{\mathbb{N}^+}$ the ***reservoir flow*** associated to F that is uniquely determined by the recurrence relations:

$$\begin{cases} U^F(\mathbf{z}, \mathbf{x}_0)_1 &= F(\mathbf{x}_0, \mathbf{z}_1) \quad \text{with} \quad \mathbf{z} \in (D_n)^{\mathbb{N}^+}, \ \mathbf{x}_0 \in D_N, \\ U^F(\mathbf{z}, \mathbf{x}_0)_t &= F(U^F(\mathbf{z}, \mathbf{x}_0)_{t-1}, \mathbf{z}_t), \quad t > 1. \end{cases} \tag{5.48}$$

The value $\mathbf{x}_0 \in D_N$ is called the ***initial condition*** of the ***path*** $U^F(\mathbf{z}, \mathbf{x}_0) \in (D_N)^{\mathbb{N}^+}$ associated to the ***input*** or ***control sequence*** $\mathbf{z} \in (D_n)^{\mathbb{N}^+}$.

As we saw in Theorems 4.1 and 5.7, the contracting property on the first component in a reservoir map is much related to the ESP and the FMP of the resulting reservoir filter and, in passing, (see Theorem 5.13) to the input forgetting property. The next result shows that something similar happens with reservoir flows associated to contracting reservoir maps as they *forget* the influence of initial conditions that are used to create the paths. This feature is referred to as the ***state forgetting property*** in [Jaeg 10].

Theorem 5.16 (State forgetting property for contracting reservoir flows) *Let $F : D_N \times D_n \longrightarrow D_N$ be a reservoir map where $D_n \subset \mathbb{R}^n$, $D_N \subset \mathbb{R}^N$, $n, N \in \mathbb{N}^+$, and suppose that F is a contraction on the first component. Given an input sequence $\mathbf{z} \in (D_n)^{\mathbb{N}^+}$, the reservoir flow $U^F : (D_n)^{\mathbb{N}^+} \times D_N \longrightarrow (D_N)^{\mathbb{N}^+}$ associated to F satisfies that:*

$$\lim_{t \to +\infty} \left\| U^F(\mathbf{z}, \mathbf{x}_0)_t - U^F(\mathbf{z}, \overline{\mathbf{x}}_0)_t \right\| = 0, \quad \text{for any} \quad \mathbf{x}_0, \overline{\mathbf{x}}_0 \in D_N. \tag{5.49}$$

Proof. Let $c < 1$ be the contraction constant of F. Using the recursions (5.48) that define the reservoir flow we can write that for any $t > 1$:

$$\begin{aligned} \left\| U^F(\mathbf{z}, \mathbf{x}_0)_t - U^F(\mathbf{z}, \overline{\mathbf{x}}_0)_t \right\| &= \left\| F(U^F(\mathbf{z}, \mathbf{x}_0)_{t-1}, \mathbf{z}_t) - F(U^F(\mathbf{z}, \overline{\mathbf{x}}_0)_{t-1}, \mathbf{z}_t) \right\| \\ &\leq c \left\| U^F(\mathbf{z}, \mathbf{x}_0)_{t-1} - U^F(\mathbf{z}, \overline{\mathbf{x}}_0)_{t-1} \right\| \leq \cdots \leq c^{t-1} \left\| U^F(\mathbf{z}, \mathbf{x}_0)_1 - U^F(\mathbf{z}, \overline{\mathbf{x}}_0)_1 \right\| \\ &\leq c^{t-1} \left\| F(\mathbf{x}_0, \mathbf{z}_1) - F(\overline{\mathbf{x}}_0, \mathbf{z}_1) \right\|. \end{aligned}$$

Taking limits $t \to +\infty$ on both sides of this inequality yields (5.49). ■

5.4 Analytic reservoir filters associated to analytic reservoir maps

The results in Section 5.1 characterized the conditions under which reservoir maps of class C^1 yield differentiable reservoir filters with respect to inputs and outputs in weighted sequence spaces. This setup is convenient because it is able to accommodate unbounded signals and allows for an elegant encoding of the fading memory property. However, due to a phenomenon similar to the one already observed in Remark 3.10, one cannot immediately obtain higher order differentiable reservoir filters out of higher order differentiable reservoir maps because, as we showed in Proposition 3.9, one needs roughly speaking to modify the weighted norm in the target of the map that defines the filter. This makes impossible the application in a higher order differentiability context of the Implicit Function Theorem, which is the main tool used in the results in the previous section. That is why in the following paragraphs we deal with analytic reservoir maps (as real valued functions) and we study the analyticity of the associated reservoir filters with respect to the supremum norm, as opposed to the weighted norms that we considered in the previous section.

Using the supremum norm implies that filter differentiability in that context, when one manages to establish it, ensures filter continuity and not the fading memory property. In exchange, analyticity allows us to construct Taylor series expansions that, as we see later on, are discrete-time Volterra series representations.

The next result is the analytic analog of the Local Persistence Theorem 5.1 formulated using the supremum norm that proves that analytic reservoir maps have locally defined analytic reservoir filters associated around constant solutions.

Theorem 5.17 (Local persistence of the ESP, continuity, and analyticity) *Let $F : \mathbb{R}^N \times \mathbb{R}^n \longrightarrow \mathbb{R}^N$ be a reservoir map. Suppose that F is analytic and that the corresponding reservoir system (1.1) has a constant solution $(\mathbf{x}^0, \mathbf{z}^0) \in \mathbb{R}^N \times \mathbb{R}^n$, that is, $\mathbf{x}^0 = F(\mathbf{x}^0, \mathbf{z}^0)$. Suppose, additionally, that for all $r \geq 1$,*

$$L_{F,r} := \sup_{(\mathbf{x},\mathbf{z}) \in \mathbb{R}^N \times \mathbb{R}^n} \{ |||D^r F(\mathbf{x}, \mathbf{z})||| \} < +\infty. \tag{5.50}$$

Suppose that

$$L_{F_x}(\mathbf{x}^0, \mathbf{z}^0) := \left|\left|\left| D_x F(\mathbf{x}^0, \mathbf{z}^0) \right|\right|\right| < 1. \tag{5.51}$$

Then, there exist open time-invariant neighborhoods $V_{\mathbf{x}^0}$ and $V_{\mathbf{z}^0}$ of $\mathbf{x}^0$ and $\mathbf{z}^0$ in $\ell_-^\infty(\mathbb{R}^N)$ and $\ell_-^\infty(\mathbb{R}^n)$, respectively, such that the reservoir system associated to F with inputs in $V_{\mathbf{z}^0}$ has the echo state property and hence determines a unique causal, time-invariant, and analytic (and hence continuous) reservoir filter $U^F : (V_{\mathbf{z}^0}, \|\cdot\|_\infty) \longrightarrow (V_{\mathbf{x}^0}, \|\cdot\|_\infty)$.

Proof. It follows the same scheme as that of Theorem 5.1. In the following paragraphs we just hint the additional facts that need to be taken into account in order to adapt that proof to this setup.

The first complementary fact has to do with the second part of Lemma 5.6 which, using the hypothesis (5.50) allows us to conclude that the map $\mathcal{F} : \ell_-^\infty(\mathbb{R}^N) \times \ell_-^\infty(\mathbb{R}^n) \longrightarrow \ell_-^\infty(\mathbb{R}^N)$ defined in (4.6) is smooth. Additionally, it can be easily seen that it is also analytic and that the radii of convergence ρ_F and $\rho_{\mathcal{F}}$ of the Taylor series expansions of F and $\mathcal{F}$ around $(\mathbf{x}^0, \mathbf{z}^0)$ and the associated constant sequence (that we denote with the same symbol) satisfy

$$\rho_F \leq \rho_{\mathcal{F}}. \tag{5.52}$$

Indeed, (5.13) implies that the Taylor series expansion of $\mathcal{F}$ around the constant sequence $(\mathbf{x}^0, \mathbf{z}^0)$ can be written, for any $\mathbf{u}^r := (\mathbf{u},\ldots,\mathbf{u}) = ((\mathbf{u_x},\mathbf{u_z}),\ldots,(\mathbf{u_x},\mathbf{u_z})) \in \left(\ell^\infty_-(\mathbb{R}^N)\oplus\ell^\infty_-(\mathbb{R}^n)\right)^r$, as

$$\mathcal{F}(\mathbf{x}^0,\mathbf{z}^0)+\sum_{r=1}^{\infty}\frac{1}{r!}D^r\mathcal{F}(\mathbf{x}^0,\mathbf{z}^0)\cdot(\mathbf{u}-(\mathbf{x}^0,\mathbf{z}^0))^r = \prod_{t\in\mathbb{Z}_-}\left(F_t(\mathbf{x}^0,\mathbf{z}^0)+\sum_{r=1}^{\infty}\frac{1}{r!}D^rF_t(\mathbf{x}^0,\mathbf{z}^0)\cdot(\mathbf{u}-(\mathbf{x}^0,\mathbf{z}^0))^r\right)$$
$$=\prod_{t\in\mathbb{Z}_-}\left(F(\mathbf{x}^0,\mathbf{z}^0)+\sum_{r=1}^{\infty}\frac{1}{r!}D^rF(\mathbf{x}_{t-1},\mathbf{z}_t)\circ(p_t\circ(T_1\times\mathrm{id}),\ldots,p_t\circ(T_1\times\mathrm{id}))\cdot(\mathbf{u}-(\mathbf{x}^0,\mathbf{z}^0))^r\right). \tag{5.53}$$

Suppose now that $\mathbf{u}=(\mathbf{u_x},\mathbf{u_z})\in\ell^\infty_-(\mathbb{R}^N)\oplus\ell^\infty_-(\mathbb{R}^n)$ is chosen such that

$$\|\mathbf{u}\|_\infty=\|\mathbf{u_x}\|_\infty+\|\mathbf{u_z}\|_\infty<\rho_F. \tag{5.54}$$

Lemma 3.1 implies that for any $t\in\mathbb{Z}_-$, we have in that case that

$$\|p_t\circ(T_1\times\mathrm{id})(\mathbf{u})\|\leq\|\mathbf{u}\|_\infty<\rho_F$$

and hence we can conclude that all the series labeled by $t\in\mathbb{Z}_-$ in each of the factors that make up the last term of (5.53) converge for all the elements $\mathbf{u}\in\ell^\infty_-(\mathbb{R}^N)\oplus\ell^\infty_-(\mathbb{R}^n)$ that satisfy (5.54). This implies that such elements are inside the radius of convergence of the Taylor series expansion of $\mathcal{F}$ around the constant sequence $(\mathbf{x}^0,\mathbf{z}^0)$ and hence (5.52) holds which, as ρ_F is nontrivial by hypothesis, proves that $\mathcal{F}$ is analytic.

The rest of the proof can be obtained by mimicking that of Theorem 5.1 where, as it is customary, we replace the weighting sequence w by the constant sequence w^ι given by $w^\iota_t:=1$, for all $t\in\mathbb{N}$, and L_w is replaced by the constant 1.

A technical modification is needed at the time of invoking the Implicit Function Theorem. In Theorem 5.1 we used a version that requires only first order differentiability as hypothesis and produces Lipschitz continuous implicitly defined functions. In this case we can prove that the function $\mathcal{G}$ is analytic and hence it can be shown that the implicitly defined local filter $\widetilde{U^F}:(\widetilde{V}_{\mathbf{z}^0},\|\cdot\|_w)\longrightarrow(\widetilde{V}_{\mathbf{x}^0},\|\cdot\|_w)$ is analytic by invoking, for instance, [Vale 88, page 175], and references therein. ∎

6 The Volterra series representation of analytic filters and a universality theorem

In this section we study the Taylor series expansions of analytic causal and time-invariant filters that, as we prove in the next result, coincide with the so called discrete-time Volterra series representations. A very similar result has been formulated in [Sand 98a, Sand 99] for analytic filters with respect to the supremum norm and with inputs with a finite past. The next result extends that statement and characterizes the inputs for which an analytic time-invariant fading memory filter with respect to a weighted norm admits a Volterra series representation with semi-infinite past inputs. This generalized result allows this series representation for inputs that are not necessarily bounded. Additionally, we use the causality and time-invariance hypotheses to show that the corresponding Volterra series representations have time-independent coefficients.

Theorem 6.1 *Let w be a weighting sequence and let $U:B_{\|\cdot\|_w}(\mathbf{z}^0,M)\subset\ell^w_-(\mathbb{R})\longrightarrow B_{\|\cdot\|_w}(U(\mathbf{z}^0),L)\subset\ell^w_-(\mathbb{R}^N)$ be a causal and time-invariant analytic filter, for some time-invariant $\mathbf{z}^0\in\ell^{1,w}_-(\mathbb{R})$ (that is, $T_{-t}(\mathbf{z}^0)=\mathbf{z}^0$, for all $t\in\mathbb{Z}_-$) and $M,L>0$. Then, for any element in the domain that satisfies*

$$\mathbf{z}\in B_{\|\cdot\|_w}(\mathbf{z}^0,M)\cap\ell^{1,w}_-(\mathbb{R}),\quad \textit{that is}\quad \sum_{t\in\mathbb{Z}_-}|z_t|w_{-t}<+\infty, \tag{6.1}$$

there exists a unique expansion

$$U(\mathbf{z})_t = U(\mathbf{z}^0)_t + \sum_{j=1}^{\infty} \sum_{m_1=-\infty}^{0} \cdots \sum_{m_j=-\infty}^{0} g_j(m_1, \ldots, m_j)(z_{m_1+t} - z^0_{m_1+t}) \cdots (z_{m_j+t} - z^0_{m_j+t}), \quad t \in \mathbb{Z}_-, \tag{6.2}$$

where the maps $g_j : \mathbb{Z}_-^j \longrightarrow \mathbb{R}^N$, $j \geq 1$, are uniquely determined by the derivatives of the functional $H_U : B_{\|\cdot\|_w}(\mathbf{z}^0, M) \subset \ell^w_-(\mathbb{R}) \longrightarrow \mathbb{R}^N$ associated to U (that by Proposition 3.9 is analytic) via the relation

$$g_j(m_1, \ldots, m_j) := \frac{1}{j!} D^j H(\mathbf{z}^0)\left(e_{m_1}, \ldots, e_{m_j}\right) \quad \text{with} \quad (e_n)_t := \begin{cases} 1 & \text{if } t = n, \\ 0 & \text{otherwise.} \end{cases} \tag{6.3}$$

Moreover, for any $p \in \mathbb{N}^+$, we have that

$$\left\| U(\mathbf{z})_t - U(\mathbf{z}^0)_t - \sum_{j=1}^{p} \sum_{m_1=-\infty}^{0} \cdots \sum_{m_j=-\infty}^{0} g_j(m_1, \ldots, m_j)(z_{m_1+t} - z^0_{m_1+t}) \cdots (z_{m_j+t} - z^0_{m_j+t}) \right\| \\ \leq \frac{L}{w_{-t}} \left(1 - \frac{\|\mathbf{z}\|_w}{M}\right)^{-1} \left(\frac{\|\mathbf{z}\|_w}{M}\right)^{p+1}. \tag{6.4}$$

These statements also hold true when $\ell^w_-(\mathbb{R})$ and $\ell^w_-(\mathbb{R}^N)$ are replaced by $\ell^\infty_-(\mathbb{R})$ and $\ell^\infty_-(\mathbb{R}^N)$, respectively. In that case, the relation (6.2) holds whenever $\mathbf{z} \in B_{\|\cdot\|_\infty}(\mathbf{z}^0, M) \cap \ell^1_-(\mathbb{R}^n)$ and the inequality (6.4) is obtained by taking as the sequence w the constant sequence w^ι given by $w^\iota_t := 1$, for all $t \in \mathbb{N}$.

Remark 6.2 The error estimate (6.4) can be reformulated in terms of the weighted norm of the sequence

$$\mathbf{R}_p(\mathbf{z}) := \left(U(\mathbf{z})_t - U(\mathbf{z}^0)_t - \sum_{j=1}^{p} \sum_{m_1=-\infty}^{0} \cdots \sum_{m_j=-\infty}^{0} g_j(m_1, \ldots, m_j)(z_{m_1+t} - z^0_{m_1+t}) \right)_{t \in \mathbb{Z}_-},$$

as

$$\|\mathbf{R}_p(\mathbf{z})\|_w \leq L \left(1 - \frac{\|\mathbf{z}\|_w}{M}\right)^{-1} \left(\frac{\|\mathbf{z}\|_w}{M}\right)^{p+1}. \tag{6.5}$$

Proof. Since by hypothesis U is analytic in $B_{\|\cdot\|_w}(\mathbf{z}^0, M)$ then

$$U(\mathbf{z}) = U(\mathbf{z}^0) + \sum_{j=1}^{\infty} \frac{1}{j!} D^j U(\mathbf{z}^0)(\underbrace{\mathbf{z} - \mathbf{z}^0, \ldots, \mathbf{z} - \mathbf{z}^0}_{j \text{ times}}), \quad \text{for any} \quad \mathbf{z} \in B_{\|\cdot\|_w}(\mathbf{z}^0, M). \tag{6.6}$$

We now show that for the elements that satisfy (6.1) the series expansion (6.6) amounts to the discrete-time Volterra series expansion (6.2). Let $m \in \mathbb{Z}_-$ and let $\delta_m \in \ell^w_-(\mathbb{R})$ be the sequence defined by

$$(\delta_m)_t := \begin{cases} \frac{1}{w_{-m}} & \text{if } t = m, \\ 0 & \text{otherwise.} \end{cases} \tag{6.7}$$

Note that $\|\delta_m\|_w = 1$ for all $m \in \mathbb{Z}_-$. Moreover, for any $\mathbf{z} \in \ell^w_-(\mathbb{R})$ we can write

$$\mathbf{z} - \mathbf{z}^0 = \sum_{t \in \mathbb{Z}_-} \widetilde{z}_t \delta_t, \quad \text{with} \quad \widetilde{z}_t = (z_t - z^0_t) w_{-t},$$

and hence by the multilinearity of the derivatives $D^jU(\mathbf{z}^0)(\mathbf{z}-\mathbf{z}^0,\ldots,\mathbf{z}-\mathbf{z}^0)$ and the causality of the filter U we have that

$$D^jU(\mathbf{z}^0)(\mathbf{z}-\mathbf{z}^0,\ldots,\mathbf{z}-\mathbf{z}^0)_t=\sum_{m_1=-\infty}^{t}\cdots\sum_{m_j=-\infty}^{t}\widetilde{z}_{m_1}\cdots\widetilde{z}_{m_j}D^jU(\mathbf{z}^0)(\delta_{m_1},\ldots,\delta_{m_j})_t,\ \text{for all } t\in\mathbb{Z}_-. \tag{6.8}$$

We first show that for the elements that satisfy (6.1) the sum in the right hand side of (6.8) is finite. Indeed, for any $t\in\mathbb{Z}_-$:

$$\begin{aligned}
&\left|\sum_{m_1=-\infty}^{t}\cdots\sum_{m_j=-\infty}^{t}\widetilde{z}_{m_1}\cdots\widetilde{z}_{m_j}D^jU(\mathbf{z}^0)(\delta_{m_1},\ldots,\delta_{m_j})_t\right|\\
&\qquad=\left|\sum_{m_1=-\infty}^{t}\cdots\sum_{m_j=-\infty}^{t}\widetilde{z}_{m_1}\cdots\widetilde{z}_{m_j}\frac{1}{w_{-t}}w_{-t}D^jU(\mathbf{z}^0)(\delta_{m_1},\ldots,\delta_{m_j})_t\right|\\
&\qquad\leq\sum_{m_1=-\infty}^{t}\cdots\sum_{m_j=-\infty}^{t}\left|\widetilde{z}_{m_1}\cdots\widetilde{z}_{m_j}\right|\frac{1}{w_{-t}}\|D^jU(\mathbf{z}^0)(\delta_{m_1},\ldots,\delta_{m_j})\|_w\\
&\qquad\leq\sum_{m_1=-\infty}^{t}\cdots\sum_{m_j=-\infty}^{t}\frac{\left|\widetilde{z}_{m_1}\cdots\widetilde{z}_{m_j}\right|}{w_{-t}}\left|\!\left|\!\left|D^jU(\mathbf{z}^0)\right|\!\right|\!\right|_w\|(\delta_{m_1},\ldots,\delta_{m_j})\|_w\\
&\qquad=j\frac{\left|\!\left|\!\left|D^jU(\mathbf{z}^0)\right|\!\right|\!\right|_w}{w_{-t}}\sum_{m_1=-\infty}^{t}\cdots\sum_{m_j=-\infty}^{t}\left|\widetilde{z}_{m_1}\right|\cdots\left|\widetilde{z}_{m_j}\right|\\
&\qquad=j\frac{\left|\!\left|\!\left|D^jU(\mathbf{z}^0)\right|\!\right|\!\right|_w}{w_{-t}}\left(\sum_{m=-\infty}^{t}\left|z_m-z_m^0\right|w_{-m}\right)^j<+\infty,
\end{aligned}\tag{6.9}$$

where the last equality is a consequence of, for example, [Apos 74, Theorem 8.44], and the last inequality follows from two facts. First, as U is analytic, it is in particular smooth and hence $\left|\!\left|\!\left|D^jU(\mathbf{z}^0)\right|\!\right|\!\right|_w<+\infty$ for all $j\in\mathbb{Z}_-$. Second, since by hypothesis $\mathbf{z},\mathbf{z}^0\in\ell_-^{1,w}(\mathbb{R}^n)$ then $\mathbf{z}-\mathbf{z}^0\in\ell_-^{1,w}(\mathbb{R}^n)$ and hence $\sum_{m=-\infty}^{t}\left|z_m-z_m^0\right|w_{-m}<+\infty$.

We now show that (6.6) can be rewritten as (6.2). Notice first that for any $t,m\in\mathbb{Z}_-$ such that $m\leq t$, the sequences (6.7) satisfy

$$T_{-t}\left(\delta_m\right)=\frac{w_{-(m-t)}}{w_{-m}}\delta_{m-t}. \tag{6.10}$$

Second, the time-invariance of U and of the sequence $\mathbf{z}^0$, imply that for any $j\in\mathbb{N}^+$, $t\in\mathbb{Z}_-$, and $\mathbf{z}^1,\ldots,\mathbf{z}^j\in\ell_-^w(\mathbb{R})$, we have that

$$T_{-t}\left(D^jU(\mathbf{z}^0)\left(\mathbf{z}^1,\ldots,\mathbf{z}^j\right)\right)=D^jU(T_{-t}\left(\mathbf{z}^0\right))\left(T_{-t}\left(\mathbf{z}^1\right),\ldots,T_{-t}\left(\mathbf{z}^j\right)\right)=D^jU(\mathbf{z}^0)\left(T_{-t}\left(\mathbf{z}^1\right),\ldots,T_{-t}\left(\mathbf{z}^j\right)\right).$$

These two relations imply that for any $t\in\mathbb{Z}_-$

$$\begin{aligned}
D^jU(\mathbf{z}^0)(\delta_{m_1},\ldots,\delta_{m_j})_t&=\left(T_{-t}\left(D^jU(\mathbf{z}^0)(\delta_{m_1},\ldots,\delta_{m_j})\right)\right)_0=D^jU(\mathbf{z}^0)(T_{-t}(\delta_{m_1}),\ldots,T_{-t}(\delta_{m_j}))_0\\
&=D^jU(\mathbf{z}^0)(\delta_{m_1-t},\ldots,\delta_{m_j-t})_0\frac{w_{-(m_1-t)}}{w_{-m_1}}\cdots\frac{w_{-(m_j-t)}}{w_{-m_j}}.
\end{aligned}$$

If we substitute this relation in the summands of (6.8), we obtain that

$$\widetilde{z}_{m_1}\cdots\widetilde{z}_{m_j} D^jU(\mathbf{z}^0)(\delta_{m_1},\ldots,\delta_{m_j})_t \\ = (z_{m_1}-z^0_{m_1})\cdots(z_{m_j}-z^0_{m_j})\cdot w_{-(m_1-t)}\cdots w_{-(m_j-t)}\cdot D^jU(\mathbf{z}^0)(\delta_{m_1-t},\ldots,\delta_{m_j-t})_0. \quad (6.11)$$

Define now

$$g_j(n_1,\ldots,n_j) := w_{-n_1}\cdots w_{-n_j}\frac{1}{j!}D^jU(\mathbf{z}^0)(\delta_{n_1},\ldots,\delta_{n_j})_0 \\ = w_{-n_1}\cdots w_{-n_j}\frac{1}{j!}D^jH(\mathbf{z}^0)(\delta_{n_1},\ldots,\delta_{n_j})_0 = \frac{1}{j!}D^jH(\mathbf{z}^0)(e_{n_1},\ldots,e_{n_j})_0, \quad (6.12)$$

where $e_m \in \ell^w_-(\mathbb{R})$ is the sequence defined in (6.3). If we make the change of variables $n_i := m_i - t$ in (6.11), we use (6.12), and we insert the resulting expression in (6.8) and subsequently in (6.6) we obtain (6.2). The uniqueness of this series expansion follows from the same argument as in [Sand 99, Theorem 1].

We now prove the error estimates (6.4) with the same strategy as in [Sand 99]. Using the Cauchy bounds for analytic functions (see, for instance, the last expression in [Hill 57, page 112]) and the analyticity hypothesis on $U : B_{\|\cdot\|_w}(\mathbf{z}^0, M) \subset \ell^w_-(\mathbb{R}) \longrightarrow B_{\|\cdot\|_w}(U(\mathbf{z}^0), L) \subset \ell^w_-(\mathbb{R}^N)$, we have that for any $j \in \mathbb{N}^+$ and $t \in \mathbb{Z}_-$

$$\|D^jU(\mathbf{z}^0)(\mathbf{z},\ldots,\mathbf{z})_t\| = \|p_t\circ D^jU(\mathbf{z}^0)(\mathbf{z},\ldots,\mathbf{z})\| \leq \|\!|p_t|\!\|_w \|D^jU(\mathbf{z}^0)(\mathbf{z},\ldots,\mathbf{z})\|_w \leq \frac{j!L}{w_{-t}}\left(\frac{\|\mathbf{z}\|_w}{M}\right)^j, \quad (6.13)$$

where we also used the first part of Lemma 3.1. Now, as we saw in the previous paragraphs,

$$\left\|U(\mathbf{z})_t - U(\mathbf{z}^0)_t - \sum_{j=1}^{p}\sum_{m_1=-\infty}^{0}\cdots\sum_{m_j=-\infty}^{0} g_j(m_1,\ldots,m_j)(z_{m_1+t}-z^0_{m_1+t})\cdots(z_{m_j+t}-z^0_{m_j+t})\right\| \\ \leq \left\|U(\mathbf{z})_t - U(\mathbf{z}^0)_t - \sum_{j=1}^{p}\frac{1}{j!}D^jU(\mathbf{z}^0)(\mathbf{z}-\mathbf{z}^0,\ldots,\mathbf{z}-\mathbf{z}^0)\right\| = \sum_{j=p+1}^{\infty}\frac{1}{j!}D^jU(\mathbf{z}^0)(\mathbf{z}-\mathbf{z}^0,\ldots,\mathbf{z}-\mathbf{z}^0) \\ \leq \frac{L}{w_{-t}}\sum_{j=p+1}^{\infty}\left(\frac{\|\mathbf{z}\|_w}{M}\right)^j \leq \frac{L}{w_{-t}}\left(1-\frac{\|\mathbf{z}\|_w}{M}\right)^{-1}\left(\frac{\|\mathbf{z}\|_w}{M}\right)^{p+1}, \quad (6.14)$$

where the inequalities in the last line follow from (6.13). ■

6.1 Finite discrete-time Volterra series are universal in the fading memory category

In this section we combine the Volterra series representation Theorem 6.1 with previous universality results in [Grig 18b] to show that any fading memory filter with uniformly bounded inputs can be arbitrarily well approximated with a Volterra series with finite terms of the type in (6.2). This result provides an alternative proof of a Volterra series universality theorem that was stated for the first time in [Boyd 85, Theorems 3 and 4]. In particular, this result shows that any time-invariant and causal fading memory filter can be uniformly approximated by a finite memory filter.

Theorem 6.3 (Universality of finite discrete-time Volterra series) *Let $M, L > 0$ and let $K_M \subset (\mathbb{R})^{\mathbb{Z}_-}$, $K_L \subset \left(\mathbb{R}^d\right)^{\mathbb{Z}_-}$ be as in (1.3). Let $U : K_M \longrightarrow K_L$ be a causal and time-invariant fading memory* *filter. Then, for any* $\epsilon > 0$ *there exist* $\mathbf{x}^0 \in K_L$ *and* $J \in \mathbb{N}^+$ *such that for any* $j \in \{1, \ldots, J\}$ *there exist* j *numbers* $M_1^j, \ldots, M_j^j \in \mathbb{N}^+$ *and maps* $g_j : \mathbb{Z}_-^j \longrightarrow \mathbb{R}$ *such that the filter determined by the finite Volterra series given by*

$$V(\mathbf{z})_t = \mathbf{x}_t^0 + \sum_{j=1}^{J} \sum_{m_1=-M_1^j}^{0} \cdots \sum_{m_j=-M_j^j}^{0} g_j(m_1, \ldots, m_j) z_{m_1+t} \cdots z_{m_j+t} \tag{6.15}$$

is such that

$$|||U - V|||_\infty = \sup_{\mathbf{z} \in K_M} \{\|U(\mathbf{z}) - V(\mathbf{z})\|\} < \epsilon.$$

Proof. The Corollary 11 in [Grig 18b] guarantees that for any $\epsilon > 0$ there exists a linear reservoir system with polynomial readout $h \in \mathbb{R}[\mathbf{x}]$ and nilpotent connectivity matrix $A \in \mathbb{M}_N$, determined by the expressions

$$\begin{cases} \mathbf{x}_t = A\mathbf{x}_{t-1} + \mathbf{c}\mathbf{z}_t, & A \in \mathbb{M}_N, \mathbf{c} \in \mathbb{M}_{N,n}, \qquad (6.16) \\ y_t = h(\mathbf{x}_t), & h \in \mathbb{R}[\mathbf{x}], \qquad (6.17) \end{cases}$$

such that it has an associated reservoir filter $U_h^{A,\mathbf{c}} : K_M \longrightarrow K_L$ that satisfies

$$\left|\left|\left| U - U_h^{A,\mathbf{c}} \right|\right|\right|_\infty < \epsilon. \tag{6.18}$$

Let $J = \deg(h) + 1$ and assume that A is nilpotent of index p. In order to prove the theorem it suffices to show that the Volterra series expansion in (6.2) corresponding to $U_h^{A,\mathbf{c}}$ has an expression of the type (6.15). If that is the case, the statement in (6.18) proves the theorem.

Indeed, recall (see, for instance, [Grig 18b, Corollary 11]) that the functional $H_h^{A,\mathbf{c}}$ associated to the filter $U_h^{A,\mathbf{c}}$ is given by

$$H_h^{A,\mathbf{c}}(\mathbf{z}) = h\left(\sum_{j=0}^{p-1} A^j(\mathbf{c}z_{-j})\right),$$

which is a composition of the polynomial h with the functional $H^{A,\mathbf{c}}$ associated to the reservoir equation (6.16) given by the linear operator

$$H^{A,\mathbf{c}}(\mathbf{z}) := \sum_{j=0}^{p-1} A^j(\mathbf{c}z_{-j}). \tag{6.19}$$

It is easy to see that $H^{A,\mathbf{c}} : (\ell_-^\infty(\mathbb{R}), \|\cdot\|_\infty) \longrightarrow \mathbb{R}^N$ has a finite operator norm $\left|\left|\left|H^{A,\mathbf{c}}\right|\right|\right|_\infty$ and that $\left|\left|\left|H^{A,\mathbf{c}}\right|\right|\right|_\infty \leq |||\mathbf{c}|||/(1 - |||A|||)$, with $|||\mathbf{c}|||$ and $|||A|||$ the top singular values of $\mathbf{c}$ and A, respectively. Moreover, it is easy to see that for any $j \in \mathbb{N}^+$, $\mathbf{z} \in K_M$, and $\mathbf{v}_1, \ldots, \mathbf{v}_j \in \ell_-^\infty(\mathbb{R})$, we have

$$D^j H_h^{A,\mathbf{c}}(\mathbf{z})(\mathbf{v}_1, \ldots, \mathbf{v}_j) = D^j h(H^{A,\mathbf{c}}(\mathbf{z}))\left(H^{A,\mathbf{c}}(\mathbf{v}_1), \ldots, H^{A,\mathbf{c}}(\mathbf{v}_j))\right),$$

which shows that $H_h^{A,\mathbf{c}} : (\ell_-^\infty(\mathbb{R}), \|\cdot\|_\infty) \longrightarrow \mathbb{R}^d$ is everywhere analytic. Using this expression and (6.3) we define

$$g_j(m_1, \ldots, m_j) := \frac{1}{j!} D^j h(\mathbf{0})\left(H^{A,\mathbf{c}}(e_{m_1}), \ldots, H^{A,\mathbf{c}}(e_{m_j})\right). \tag{6.20}$$

As h has finite degree then $D^j h(\mathbf{0}) = 0$ for any $j > \deg(h) + 1 = J$. Moreover, since the sum in (6.19) is finite by the nilpotency of A it is clear that $g_j(m_1, \ldots, m_j)$ in (6.20) is nonzero as long as $1 \leq j \leq \deg(h) + 1 = J$ and $-(p-1) \leq m_1, \ldots, m_j \leq 0$. If we define $M_1^j, \ldots, M_j^j := p - 1$ then the Taylor series expansion of $U^{A,\mathbf{c}}(\mathbf{z})$ coincides with (6.15).

We emphasize that in this case this expansion is valid for any $\mathbf{z} \in K_M$ by the finiteness of the number of terms in the sum and that the condition (6.1) is hence not necessary. ■

7 Appendices

7.1 Time invariance of the solutions of a reservoir system

The filters studied in this paper are those determined by reservoir systems of the type introduced in (1.1)–(1.2). As we already pointed out, in that case we can associate unique reservoir filters U^F and U_h^F to the reservoir map F and the reservoir system, respectively, whenever (1.1) satisfies the echo state property. In that case, it has been shown in [Grig 18a, Proposition 2.1] that both U^F and U_h^F are necessarily causal and time-invariant. We complement this fact with a similar elementary statement that does not require the echo state property or the existence reservoir filters.

Lemma 7.1 *Let $(\mathbf{x}^0, \mathbf{z}^0) \in \left(\mathbb{R}^N\right)^{\mathbb{Z}_-} \times \left(\mathbb{R}^n\right)^{\mathbb{Z}_-}$ be a solution of the reservoir system determined by the map $F : \mathbb{R}^N \times \mathbb{R}^n \longrightarrow \mathbb{R}^N$. Then, for any $\tau \in \mathbb{Z}_-$, the pair $(T_\tau(\mathbf{x}^0), T_\tau(\mathbf{z}^0)) \in \left(\mathbb{R}^N\right)^{\mathbb{Z}_-} \times \left(\mathbb{R}^n\right)^{\mathbb{Z}_-}$ is also a solution.*

Proof. By hypothesis, for any $t \in \mathbb{Z}_-$ we have that

$$F(\mathbf{x}^0_{t-1}, \mathbf{z}^0_t) = \mathbf{x}^0_t,$$

and hence

$$F\left(T_\tau(\mathbf{x}^0)_{t-1}, T_\tau(\mathbf{z}^0)_t\right) = F(\mathbf{x}^0_{t-\tau-1}, \mathbf{z}^0_{t-\tau}) = \mathbf{x}^0_{t-\tau} = T_\tau(\mathbf{x}^0)_t, \quad \text{as required.} \quad \blacksquare$$

Acknowledgments: We thank Lukas Gonon for fruitful discussions. The authors acknowledge partial financial support of the French ANR "BIPHOPROC" project (ANR-14-OHRI-0002-02) as well as the hospitality of the Centre Interfacultaire Bernoulli of the Ecole Polytechnique Fédérale de Lausanne during the program "Stochastic Dynamical Models in Mathematical Finance, Econometrics, and Actuarial Sciences" that made possible the collaboration that lead to some of the results included in this paper. LG acknowledges partial financial support of the Graduate School of Decision Sciences of the Universität Konstanz. JPO acknowledges partial financial support coming from the Research Commission of the Universität Sankt Gallen and the Swiss National Science Foundation (grant number 200021_175801/1).

References

[Abra 88] R. Abraham, J. E. Marsden, and T. S. Ratiu. *Manifolds, Tensor Analysis, and Applications*. Vol. 75, Applied Mathematical Sciences. Springer-Verlag, 1988.

[Apos 74] T. Apostol. *Mathematical Analysis*. Addison Wesley, second Ed., 1974.

[Boyd 85] S. Boyd and L. Chua. "Fading memory and the problem of approximating nonlinear operators with Volterra series". *IEEE Transactions on Circuits and Systems*, Vol. 32, No. 11, pp. 1150–1161, nov 1985.

[Chos 03] P. Chossat, D. Lewis, J.-P. Ortega, and T. S. Ratiu. "Bifurcation of relative equilibria in mechanical systems with symmetry". *Advances in Applied Mathematics*, Vol. 31, pp. 10–45, 2003.

[Croo 07] N. Crook. "Nonlinear transient computation". *Neurocomputing*, Vol. 70, pp. 1167–1176, 2007.

[Gono 18] L. Gonon and J.-P. Ortega. “Reservoir computing universality with stochastic inputs”. *Conditionally accepted in IEEE Transactions on Neural Networks and Learning Systems*, 2018.

[Grig 18a] L. Grigoryeva and J.-P. Ortega. “Echo state networks are universal”. *Neural Networks*, Vol. 108, pp. 495–508, 2018.

[Grig 18b] L. Grigoryeva and J.-P. Ortega. “Universal discrete-time reservoir computers with stochastic inputs and linear readouts using non-homogeneous state-affine systems”. *Journal of Machine Learning Research*, Vol. 19, No. 24, pp. 1–40, 2018.

[Guna 15] H. Gunawan, S. Konca, and M. Idris. “p-summable sequence spaces with inner products”. *Bitlis Eren University Journal of Science and Technology*, Vol. 5, No. 1, pp. 1–9, 2015.

[Hill 57] E. Hille and R. S. Phillips. *Functional Analysis and Semi-Groups*. American Mathematical Society, 1957.

[Jaeg 04] H. Jaeger and H. Haas. “Harnessing Nonlinearity: Predicting Chaotic Systems and Saving Energy in Wireless Communication”. *Science*, Vol. 304, No. 5667, pp. 78–80, 2004.

[Jaeg 10] H. Jaeger. “The ’echo state’ approach to analysing and training recurrent neural networks with an erratum note”. Tech. Rep., German National Research Center for Information Technology, 2010.

[Lax 02] P. Lax. *Functional Analysis*. Wiley-Interscience, 2002.

[Lind 15] A. Lindquist and G. Picci. *Linear Stochastic Systems*. Springer-Verlag, 2015.

[Luko 09] M. Lukoševičius and H. Jaeger. “Reservoir computing approaches to recurrent neural network training”. *Computer Science Review*, Vol. 3, No. 3, pp. 127–149, 2009.

[Maas 02] W. Maass, T. Natschläger, and H. Markram. “Real-time computing without stable states: a new framework for neural computation based on perturbations”. *Neural Computation*, Vol. 14, pp. 2531–2560, 2002.

[Maas 11] W. Maass. “Liquid state machines: motivation, theory, and applications”. In: S. S. Barry Cooper and A. Sorbi, Eds., *Computability In Context: Computation and Logic in the Real World*, Chap. 8, pp. 275–296, 2011.

[Manj 13] G. Manjunath and H. Jaeger. “Echo state property linked to an input: exploring a fundamental characteristic of recurrent neural networks”. *Neural Computation*, Vol. 25, No. 3, pp. 671–696, 2013.

[Mont 97a] J. A. Montaldi. “Persistance d’orbites périodiques relatives dans les systèmes hamiltoniens symétriques”. *C. R. Acad. Sci. Paris Sér. I Math.*, Vol. 324, pp. 553–558, 1997.

[Mont 97b] J. A. Montaldi. “Persistence and stability of relative equilibria”. *Nonlinearity*, Vol. 10, pp. 449–466, 1997.

[Munk 14] J. Munkres. *Topology*. Pearson, second Ed., 2014.

[Orte 97] J.-P. Ortega and T. S. Ratiu. “Persistence and smoothness of critical relative elements in Hamiltonian systems with symmetry”. *Comptes Rendus de l’Académie des Sciences - Series I - Mathematics*, Vol. 325, No. 10, pp. 1107–1111, nov 1997.

[Prie 88] M. B. Priestley. *Non-linear and Non-stationary Time Series Analysis.* Academic Press, 1988.

[Reki 15] A. Rekic-Vukovic, N. Okicic, and E. Dunjakovic. "On weighted Banach sequence spaces". *Advances in Mathematics: Scientific Journal*, Vol. 4, No. 2, pp. 127–138, 2015.

[Rugh 81] W. J. Rugh. *Nonlinear System Theory. The Volterra/Wiener Approach.* The Johns Hopkins University Press, 1981.

[Sand 03] I. W. Sandberg. "Notes of fading-memory conditions". *Circuits, Systems, and Signal Processing*, Vol. 22, No. 1, pp. 43–55, 2003.

[Sand 98a] I. W. Sandberg. "Time-delay neural networks, Volterra series, and rates of approximation". *Circuits, Systems, and Signal Processing*, Vol. 17, No. 5, pp. 653–655, 1998.

[Sand 98b] I. W. Sandberg. "A note on representation theorems for linear discrete-space systems". *Circuits, Systems, and Signal Processing*, Vol. 17, No. 6, pp. 703–708, 1998.

[Sand 98c] I. W. Sandberg. "A representation theorem for linear discrete-space systems". *Mathematical Problems in Engineering*, Vol. 4, pp. 369–375, 1998.

[Sand 99] I. W. Sandberg. "Bounds for discrete-time Volterra series representations". *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, Vol. 46, No. 1, pp. 135–139, 1999.

[Sche 80] M. Schetzen. *The Volterra and Wiener Theories of Nonlinear Systems.* Wiley, 1980.

[Sche 97] E. Schechter. *Handbook of Analysis and its Foundations.* Vol. 91, Academic Press, 1997.

[Ster 10] S. Sternberg. *Dynamical Systems.* Dover, 2010.

[Vale 88] T. Valent. *Boundary Value Problems of Finite Elasticity.* Springer Verlag, 1988.

[Ver 74] P. Ver Eecke. "Sur le calcul différentiel dans les espaces vectoriels topologiques". *Cahiers de topologie et géométrie différentielle catégoriques*, Vol. 15, No. 3, pp. 293–339, 1974.

[Vers 07] D. Verstraeten, B. Schrauwen, M. D'Haene, and D. Stroobandt. "An experimental unification of reservoir computing methods". *Neural Networks*, Vol. 20, pp. 391–403, 2007.

[Volt 30] V. Volterra. *Theory of Functionals and of Integral and Integro-Differential Equations.* Blackie & Son Limited, Glasgow, 1930.

[Wien 58] N. Wiener. *Nonlinear Problems in Random Theory.* The Technology Press of MIT, 1958.