

DIMENSION REDUCTION IN RECURRENT NETWORKS BY CANONICALIZATION

LYUDMILA GRIGORYEVA

Department of Statistics
University of Warwick
Coventry CV4 7AL, UK

JUAN-PABLO ORTEGA*

Division of Mathematical Sciences
Nanyang Technological University
21 Nanyang Link, Singapore 637371

(Communicated by Andrew D. Lewis)

ABSTRACT. Many recurrent neural network machine learning paradigms can be formulated using state-space representations. The classical notion of canonical state-space realization is adapted in this paper to accommodate semi-infinite inputs so that it can be used as a dimension reduction tool in the recurrent networks setup. The so-called input forgetting property is identified as the key hypothesis that guarantees the existence and uniqueness (up to system isomorphisms) of canonical realizations for causal and time-invariant input/output systems with semi-infinite inputs. Additionally, the notion of optimal reduction coming from the theory of symmetric Hamiltonian systems is implemented in our setup to construct canonical realizations out of input forgetting but not necessarily canonical ones. These two procedures are studied in detail in the framework of linear fading memory input/output systems. Finally, the notion of implicit reduction using reproducing kernel Hilbert spaces (RKHS) is introduced which allows, for systems with linear readouts, to achieve dimension reduction without the need to actually compute the reduced spaces introduced in the first part of the paper.

1. Introduction. State-space models are of widespread use in the construction of input/output systems in many application contexts. The Markovian nature of the state equation makes them particularly convenient in the construction of efficient simulation algorithms without preventing the possibility of encoding long-memory type behaviors. These models were first introduced in the context of systems and control theory [54, 55, 53, 51, 3, 52] and met spectacular success in all sorts of industrial, military, and scientific applications in relation to filtering, smoothing, and forecasting (see [53, 56, 45, 16, 86] and references therein for just a few examples).

2020 *Mathematics Subject Classification.* Primary: 68T05, 68T07; Secondary: 93-08, 93Bxx.

Key words and phrases. Recurrent neural network, reservoir computing, dimension reduction, state-space system, canonicalization, echo state network, ESN, linear recurrent network, machine learning, echo state property.

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).

*Corresponding author: Juan-Pablo Ortega.

More recently, these systems have reemerged in the context of the machine learning of dynamic processes as powerful recurrent network paradigms. The question of interest in this framework is the learning or the estimation of the parameters of a state-space system out of finite-length realizations of the input and output processes. This learning problem is, to some extent, just a reformulation of the non-linear identification problem that has been thoroughly studied in systems and control theory [105, 90, 13, 75, 73, 62] as well as in the theory of empirical processes [15].

Despite these similarities, there are new problems that need to be seriously addressed when using state-space systems in the machine learning context. For instance, much of the systems theory literature is dedicated to the characterization of the controllability question for invertible systems and formulated using a prescribed initial or final condition (see [18, 77, 49] for an in-depth study of the discrete-time case). However, in most machine learning situations, it is more appropriate to work using *semi-infinite temporal traces towards the past* in which the dependence on initial conditions disappears. This feature arises in the presence of time-invariant input/output systems and stationary stochastic processes, and it is a crucial element in the formulation of the fading memory property that pervades many modeling situations. Additionally, most systems that are considered in applications are subsystems of a Markovian system that, generically, exhibit a functional dependence on the infinite past.

Another distinctive feature of state-space models in the learning framework is the use of randomization. Early in the application of these models as recurrent networks, important difficulties were identified at the time of their training using classical gradient descent (backpropagation-type) methods having to do with bifurcation phenomena [14] in these intrinsic dynamical models. Recent progress in the regularization and training of recurrent structures (see, for instance [27, 82, 106], and references therein) solves to some extent some of these non-convergence problems. A different approach to circumvent this question, specially in data-intensive applications, is to use randomly generated state equations and to only train the time-independent observation equation that is selected out of a functionally simple (preferably linear) family. This revolutionary idea has its origin in static frameworks like, for instance, in the seminal works on random feature models [85] and Extreme Learning Machines [44]. This philosophy was extended to the dynamical context that we are interested in this paper under the names of *reservoir computing (RC)* [46, 47, 48] and *liquid state machines* [67, 66] and has proved to be very successful in a great variety of empirical classification and forecasting applications (see, for instance, [48, 103, 65, 102, 8, 32, 64, 83, 84]).

These empirical discoveries have motivated an intense activity in the theoretical front to understand, quantify, and optimize the information processing abilities of state-space systems. An important body of work has to do with the assessment of the memory and forecasting abilities of these constructions in terms of their architectures and dependence properties of the input signals [47, 101, 19, 42, 12, 2, 32, 31, 10, 33, 17, 26, 30, 104, 9, 71, 100, 22]. Additionally, memory capacities have been extensively compared with other related concepts like Fisher information-based criteria [97, 63, 96].

In a more learning theoretical note, much progress has been done in the last years in the understanding of the universal approximation and the generalization properties of this approach. By now, we can find in the literature many families of

state-space systems that have been proved to be universal approximants in different contexts. For example, when inputs are deterministic and uniformly bounded, universality has been proved for linear systems with polynomial observation equations [6, 36], state-affine systems (SAS) [36], the echo state networks (ESNs) [35, 24] introduced in [72, 73, 48], the so-called signature state-affine systems (SigSAS) [11] that encode in state-space form the truncation of Volterra series expansions, or the temporal convolutional networks [38]. These results have been extended to a stochastic setup in [25] and also exist in the context of the approximation of dynamical systems with a compact phase space [39, 40, 28, 29]. By now, risk [23] and approximation [21] bounds exist for some of these systems similar to those that can be formulated, for instance, in the context of shallow neural networks or other static machine learning paradigms.

In this paper we focus on another machine learning aspect of major importance in the practical use of reservoir computing and state-space systems, namely, ***dimension reduction***. Given a machine learning paradigm, the dimension reduction problem consists of generically finding a system with reduced complexity that exhibits equivalent or almost equivalent approximation properties. For example, in the feedforward neural networks context, there exist standard pruning techniques [41] that determine which neurons can be eliminated in a given network configuration when they are not relevant for a given approximation task. Other widespread strategies consist in using principal components analysis or random projections in the spirit of [50] (see [11] for a first step in the use of these techniques in reservoir computing).

In the framework of mechanical and controlled systems, dimension reduction is a classical and well-studied subject that goes back to Jacobi's elimination of the node in multi-body celestial mechanics in the nineteenth century. In that setup, dimension reduction is, most of the time, associated with the use of the conserved quantities associated to the symmetries of a given system and that are encoded in the level sets of a momentum map [59, 91, 92, 88]. Dimension reduction is generically obtained by restricting the dynamics to invariant manifolds and by projecting it onto the orbit space with respect to the residual symmetry that leaves those invariant. In the context of autonomous systems, this procedure is referred to as ***Marsden-Weinstein reduction*** [69]; see [80, 70] for self-contained presentations of this beautiful theory. Part of these mostly differential geometric techniques for dimension reduction has been extended to controlled systems. See, for instance, [98, 76, 37, 99, 4, 20, 78, 5] and references therein.

Many reservoir computing applications like, for instance, those in [48, 64, 83, 84] require the use of systems with state-space dimensions in the thousands that, generically, present no symmetries that could be used for reduction. This motivates the investigation of another natural dimension reduction related notion, this time only applicable to state-space systems, namely that of ***canonicalization***. The idea behind it is based on the observation that since the state-space representation of input/output systems is not unique, one should choose the most "economical" one in which "unused" states are dropped from the representation and those that are "undistinguishable" from a dynamical point of view are identified by the passage to a quotient space. These "optimal" state-space representations are called ***canonical realizations***, and in the context of forward looking systems it can be proved that they exist and are unique up to system isomorphisms. This result is usually called the ***Canonical Realization Theorem*** (see for instance [72, Chapter 2]).

The main goal of this paper is *extending these canonicalization results to the context of time-invariant and causal input/output systems with semi-infinite inputs* and, moreover, to obtain a Canonical Realization Theorem in this framework out of a reduction approach similar to the one introduced in [81, 79]. More explicitly, the paper contains two main canonicalization results:

- A **Canonical Realization Theorem** (Theorem 3.2) for input/output systems. This result shows that any **causal and time-invariant** filter that has the so-called **input forgetting property** admits a canonical state-space realization that is unique up to system isomorphisms. The input forgetting property (also referred to in the literature as the **unique steady-state property**) is a modeling feature that appears profusely in applications and that can be obtained out of the so-called **fading memory property** (see [6, 34] for a detailed discussion about these concepts). An important merit of Theorem 3.2 is identifying the input forgetting property as the key concept that leads to the availability of canonical realizations in the presence of semi-infinite inputs. Additionally, it constitutes a result of great generality as it provides a constructive procedure for the design of state-space realizations for a vast category of input/output systems; the price to pay for this generality is the potentially complicated nature of the representing state space or its infinite dimensional character (when such notion is well-defined).
- A **Canonicalization by Reduction Theorem** (Theorem 3.4). This result uses a reduction approach similar to the one introduced in [81, 79] in the context of symmetric Hamiltonian systems to construct a canonical realization for a state-space system that has the input forgetting property system by using a “reduced” version of it in a sense that will be introduced in detail later on.

These two results are illustrated and applied in detail in Section 4 in the context of linear fading memory filters. In particular, Theorem 4.5 shows that any linear, causal, time-invariant filter with semi-infinite inputs that has the fading memory property (that, as we shall see, implies the input forgetting property) admits a canonical linear state-space realization (possibly infinite dimensional). Additionally, this result also characterizes all the isomorphic canonical realizations of the given filter as a homogeneous manifold constructed using the general linear group of the state space. Finally, the Canonicalization by Reduction Theorem 3.4 in the linear setup yields Theorem 4.6, which fully characterizes how to construct a canonical linear realization by shrinking the linear state-space appropriately, for a given linear system that has the input forgetting property but that is not necessarily canonical.

The paper concludes with Section 5, where we introduce what we call **implicit reduction** using reproducing kernel Hilbert spaces (RKHS). The main goal of that section consists in circumventing the need of computing the reduced spaces introduced in the previous sections, which may be technically difficult, in order to achieve dimension reduction. As we show in those pages, the RKHS formulation of the estimation problem for state-space systems with linear readouts achieves exactly that as a consequence of the well-known Representer Theorem [74, page 117]. Section 6 concludes the paper.

2. Canonical systems with semi-infinite inputs. We briefly introduce a few definitions that make explicit the setup where we shall be working. The objects of interest in this paper are input/output systems determined by state-space systems. The symbols $\mathcal{Z}$ and $\mathcal{Y}$ will denote the **input** and the **output spaces**, respectively,

and $\mathcal{X}$ will be the **state space** of the system that will create the link between them. These three spaces are typically subsets of a Euclidean space or, more generally, finite or infinite dimensional manifolds; for the time being we shall assume no particular structure on them. A **discrete-time state-space system** is determined by the following two equations that put in relation sequences $\mathbf{z} \in \mathcal{Z}^{\mathbb{Z}}$, $\mathbf{y} \in \mathcal{Y}^{\mathbb{Z}}$, $\mathbf{x} \in \mathcal{X}^{\mathbb{Z}}$ in the three spaces that we just introduced:

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

$$(2)$$

for any $t \in \mathbb{Z}$. The map $F : \mathcal{X} \times \mathcal{Z} \rightarrow \mathcal{X}$ is called the **state map** and $h : \mathcal{X} \rightarrow \mathcal{Y}$ the **readout** or **observation** map. We shall sometimes denote a system by using the triple $(\mathcal{X}, F, h)$. The term **recurrent neural network (RNN)** is used sometimes in the literature to refer to state-space systems where the state map F in (1) is neural network-like, that is, it is a concatenation of compositions of a nonlinear activation function with an affine function of the states and the input. A particular case of RNNs are the **echo state networks** introduced in [72, 48] where one neural layer of this type is used (with random connectivity between neurons in [48]).

We focus on state-space systems of the type (1)-(2) that determine an **input/output** system. This happens in the presence of the so-called **echo state property (ESP)**, that is, when for any $\mathbf{z} \in \mathcal{Z}^{\mathbb{Z}}$ there exists a unique $\mathbf{y} \in \mathcal{Y}^{\mathbb{Z}}$ such that (1)-(2) hold. In that case, we talk about the **state-space filter** $U_h^F : \mathcal{Z}^{\mathbb{Z}} \rightarrow \mathcal{Y}^{\mathbb{Z}}$ associated to the state-space system $(\mathcal{X}, F, h)$ defined by:

$$U_h^F(\mathbf{z}) := \mathbf{y},$$

where $\mathbf{z} \in \mathcal{Z}^{\mathbb{Z}}$ and $\mathbf{y} \in \mathcal{Y}^{\mathbb{Z}}$ are linked by (1)-(2) via the ESP. If the ESP holds at the level of the state equation (1), we can define a **state filter** $U^F : \mathcal{Z}^{\mathbb{Z}} \rightarrow \mathcal{X}^{\mathbb{Z}}$ and, in that case, we have that

$$U_h^F := h \circ U^F.$$

It is easy to show that state and state-space filters are automatically causal and time-invariant (see [35, Proposition 2.1]) and hence it suffices to work with their restriction $U_h^F : \mathcal{Z}^{\mathbb{Z}_-} \rightarrow \mathcal{Y}^{\mathbb{Z}_-}$ to semi-infinite inputs and outputs. Moreover, U_h^F determines a state-space **functional** $H_h^F : \mathcal{Z}^{\mathbb{Z}} \rightarrow \mathcal{Y}$ as $H_h^F(\mathbf{z}) := U_h^F(\mathbf{z})_0$, for all $\mathbf{z} \in \mathcal{Z}^{\mathbb{Z}_-}$ (the same applies to U^F and H^F when the ESP holds at the level of the state equation). In the sequel we use the symbol $\mathbb{Z}_-$ to denote the negative integers including zero and $\mathbb{Z}^-$ without zero.

State-space morphisms. As we already mention in the introduction, a given input/output filter may have different state-space realizations. One way to construct them is by using the natural functors between state-space systems that we define below. Consider the state-space systems determined by the triples $(\mathcal{X}_i, F_i, h_i)$, $i \in \{1, 2\}$, with $F_i : \mathcal{X}_i \times \mathcal{Z} \rightarrow \mathcal{X}_i$ and $h_i : \mathcal{X}_i \rightarrow \mathcal{Y}$.

Definition 2.1. A map $f : \mathcal{X}_1 \rightarrow \mathcal{X}_2$ is a **morphism** between the systems $(\mathcal{X}_1, F_1, h_1)$ and $(\mathcal{X}_2, F_2, h_2)$ whenever it satisfies the following two properties:

- (i): **System equivariance:** $f(F_1(\mathbf{x}_1, \mathbf{z})) = F_2(f(\mathbf{x}_1), \mathbf{z})$, for all $\mathbf{x}_1 \in \mathcal{X}_1$ and $\mathbf{z} \in \mathcal{Z}$.
- (ii): **Readout invariance:** $h_1(\mathbf{x}_1) = h_2(f(\mathbf{x}_1))$, for all $\mathbf{x}_1 \in \mathcal{X}_1$.

When the map f has an inverse f^{-1} and this inverse is also a morphism between the systems determined by $(\mathcal{X}_2, F_2, h_2)$ and $(\mathcal{X}_1, F_1, h_1)$ and we say that f is a **system isomorphism** and that the systems $(\mathcal{X}_1, F_1, h_1)$ and $(\mathcal{X}_2, F_2, h_2)$ are **isomorphic**. We note that given a system $F_1 : \mathcal{X}_1 \times \mathcal{Z} \rightarrow \mathcal{X}_1$, $h_1 : \mathcal{X}_1 \rightarrow \mathcal{Y}$ and a bijection $f : \mathcal{X}_1 \rightarrow \mathcal{X}_2$, the map f is a system isomorphism with respect to the system $F_2 : \mathcal{X}_2 \times \mathcal{Z} \rightarrow \mathcal{X}_2$, $h_2 : \mathcal{X}_2 \rightarrow \mathcal{Y}$ defined by

$$F_2(\mathbf{x}_2, \mathbf{z}) := f(F_1(f^{-1}(\mathbf{x}_2), \mathbf{z})), \quad \text{for all } \mathbf{x}_2 \in \mathcal{X}_2, \mathbf{z} \in \mathcal{Z}, \quad (3)$$

$$h_2(\mathbf{x}_2) := h_1(f^{-1}(\mathbf{x}_2)), \quad \text{for all } \mathbf{x}_2 \in \mathcal{X}_2. \quad (4)$$

The proof of the following elementary result can be found in [22].

Proposition 2.2. *Let $(\mathcal{X}_i, F_i, h_i)$, $i \in \{1, 2\}$, be two systems with $F_i : \mathcal{X}_i \times \mathcal{Z} \rightarrow \mathcal{X}_i$ and $h_i : \mathcal{X}_i \rightarrow \mathcal{Y}$. Let $f : \mathcal{X}_1 \rightarrow \mathcal{X}_2$ be a map. Then:*

- (i): *If f is system equivariant and $\mathbf{x}^1 \in \mathcal{X}_1^{\mathbb{Z}^-}$ is a solution for the state system associated to F_1 and the input $\mathbf{z} \in \mathcal{Z}^{\mathbb{Z}^-}$, then so is $(f(\mathbf{x}_t^1))_{t \in \mathbb{Z}^-} \in \mathcal{X}_2^{\mathbb{Z}^-}$ for the system associated to F_2 and the same input.*
- (ii): *Suppose that the system determined by $(\mathcal{X}_2, F_2, h_2)$ has the echo state property and assume that the state system determined by F_1 has at least one solution for each element $\mathbf{z} \in \mathcal{Z}^{\mathbb{Z}^-}$. If f is a morphism between $(\mathcal{X}_1, F_1, h_1)$ and $(\mathcal{X}_2, F_2, h_2)$, then $(\mathcal{X}_1, F_1, h_1)$ has the echo state property and, moreover,*

$$U_{h_1}^{F_1} = U_{h_2}^{F_2}. \quad (5)$$

- (iii): *If f is a system isomorphism, then the implications in the previous two points are reversible, that is, the indices 1 and 2 can be exchanged.*

Reachability and observability. We just showed in the previous paragraph that system morphisms produce different state-space system realizations for a given input/output system. We now introduce dynamical properties that ensure that the reverse implication holds, that is, if we have two different state-space system realizations for a given input/output system we can ensure that there exists a system morphism between them. The following definitions are natural adaptations of the concepts with the same name in the context of forward-in-time systems [89, 61, 7].

The definition uses the following notation: if $\mathbf{z} \in \mathcal{Z}^{\mathbb{Z}^-}$ and $\tilde{\mathbf{z}} \in \mathcal{Z}^T$ for some $T \in \mathbb{N}$, then the symbol $\mathbf{z}\tilde{\mathbf{z}} \in \mathcal{Z}^{\mathbb{Z}^-}$ denotes the semi-infinite sequence obtained by concatenation of $\mathbf{z}$ and $\tilde{\mathbf{z}}$.

Definition 2.3. Let $(\mathcal{X}, F, h)$ be a state-space system with $F : \mathcal{X} \times \mathcal{Z} \rightarrow \mathcal{X}$ and $h : \mathcal{X} \rightarrow \mathcal{Y}$. Assume that $(\mathcal{X}, F, h)$ has the echo state property. Then, we say that $(\mathcal{X}, F, h)$ is:

- (i): **Reachable** (respectively, **strongly reachable**), when for any $\mathbf{y} \in \mathcal{Y}$ (respectively, $\mathbf{x} \in \mathcal{X}$) there exists $\mathbf{z} \in \mathcal{Z}^{\mathbb{Z}^-}$ such that $H_h^F(\mathbf{z}) = \mathbf{y}$ (respectively, $H^F(\mathbf{z}) = \mathbf{x}$).
- (ii): **Observable**, when it does not have indistinguishable states. Two distinct states $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{X}$ are called **indistinguishable** when there exist $\mathbf{z}_1, \mathbf{z}_2 \in \mathcal{Z}^{\mathbb{Z}^-}$ such that $\mathbf{x}_1 = H^F(\mathbf{z}_1)$, $\mathbf{x}_2 = H^F(\mathbf{z}_2)$ and, additionally, we have that $H_h^F(\mathbf{z}_1\tilde{\mathbf{z}}) = H_h^F(\mathbf{z}_2\tilde{\mathbf{z}})$, for any $\tilde{\mathbf{z}} \in \mathcal{Z}^T$ and any $T \in \mathbb{N}$.
- (iii): **Canonical**, when $(\mathcal{X}, F, h)$ is strongly reachable and observable.

Note that if the observation map h is surjective, then strong reachability implies reachability.

Proposition 2.4. *Let $(\mathcal{X}, F, h)$ and $(\bar{\mathcal{X}}, \bar{F}, \bar{h})$ be two systems that have the echo state property and yield the same time-invariant input/output system, that is, $H_h^F = H_{\bar{h}}^{\bar{F}}$. If $(\mathcal{X}, F, h)$ is strongly reachable and $(\bar{\mathcal{X}}, \bar{F}, \bar{h})$ is observable then there exists a unique system morphism $f : \mathcal{X} \rightarrow \bar{\mathcal{X}}$.*

Before we proceed with the proof of this proposition, we list in the following lemma three elementary properties of time-invariant state-space filters. In the proof we use the **time delay** operators $T_\tau : \mathcal{Z}^{\mathbb{Z}^-} \rightarrow \mathcal{Z}^{\mathbb{Z}^-}$ that, for any $\tau \in \mathbb{N}$, are defined as

$$T_\tau(\mathbf{z})_t := \mathbf{z}_{t-\tau}, \quad t \in \mathbb{Z}_-. \quad (6)$$

We recall that filters are called time-invariant when they commute with the time delay operators. Additionally, we will be using the notion of invertible state map. We recall that the map $F : \mathcal{X} \times \mathcal{Z} \rightarrow \mathcal{X}$ is **invertible** when for any $\mathbf{z} \in \mathcal{Z}$, the maps $F(\cdot, \mathbf{z}) : \mathcal{X} \rightarrow \mathcal{X}$ are injective and hence there exists a map $F^{-1} : \mathcal{X} \times \mathcal{Z} \rightarrow \mathcal{X}$ such that

$$F^{-1}(F(\mathbf{x}, \mathbf{z}), \mathbf{z}) = \mathbf{x}, \quad \text{for all } \mathbf{x} \in \mathcal{X} \text{ and } \mathbf{z} \in \mathcal{Z}. \quad (7)$$

Lemma 2.5. *Let $(\mathcal{X}, F, h)$ be a system that has the echo state property with input and output spaces $\mathcal{Z}$ and $\mathcal{Y}$, respectively, and $\tilde{\mathbf{z}} \in \mathcal{Z}$. Then*

$$U^F(\mathbf{z}\tilde{\mathbf{z}})_{-1} = H^F(\mathbf{z}) \quad \text{and} \quad H^F(\mathbf{z}\tilde{\mathbf{z}}) = F(H^F(\mathbf{z}), \tilde{\mathbf{z}}), \quad \text{for } \mathbf{z} \in \mathcal{Z}^{\mathbb{Z}^-}. \quad (8)$$

Additionally, if $\mathbf{z}_1, \mathbf{z}_2 \in \mathcal{Z}^{\mathbb{Z}^-}$ are such that $H^F(\mathbf{z}_1) = H^F(\mathbf{z}_2)$ then

$$H^F(\mathbf{z}_1\tilde{\mathbf{z}}) = H^F(\mathbf{z}_2\tilde{\mathbf{z}}), \quad \text{for any } \tilde{\mathbf{z}} \in \mathcal{Z}^T \text{ and any } T \in \mathbb{N}. \quad (9)$$

The converse holds when F is an invertible state map.

Proof of the Lemma. The identities in (8) are a consequence of the time-invariance of U^F . Indeed,

$$U^F(\mathbf{z}\tilde{\mathbf{z}})_{-1} = (T_1 \circ U^F(\mathbf{z}\tilde{\mathbf{z}}))_0 = (U^F(T_1(\mathbf{z}\tilde{\mathbf{z}})))_0 = U^F(\mathbf{z})_0 = H^F(\mathbf{z}).$$

As to the second equality in (8), by definition and the identity that we just proved:

$$H^F(\mathbf{z}\tilde{\mathbf{z}}) = U^F(\mathbf{z}\tilde{\mathbf{z}})_0 = F(U^F(\mathbf{z}\tilde{\mathbf{z}})_{-1}, \tilde{\mathbf{z}}) = F(H^F(\mathbf{z}), \tilde{\mathbf{z}}).$$

Concerning (9), let $\tilde{\mathbf{z}} = (\tilde{\mathbf{z}}_1, \dots, \tilde{\mathbf{z}}_T) \in \mathcal{Z}^T$. Then, by the hypothesis $H^F(\mathbf{z}_1) = H^F(\mathbf{z}_2)$ and the identity that we just proved:

$$H^F(\mathbf{z}_1\tilde{\mathbf{z}}_1) = F(H^F(\mathbf{z}_1), \tilde{\mathbf{z}}_1) = F(H^F(\mathbf{z}_2), \tilde{\mathbf{z}}_1) = H^F(\mathbf{z}_2\tilde{\mathbf{z}}_1).$$

Analogously,

$$H^F(\mathbf{z}_1\tilde{\mathbf{z}}_1\tilde{\mathbf{z}}_2) = F(H^F(\mathbf{z}_1\tilde{\mathbf{z}}_1), \tilde{\mathbf{z}}_2) = F(H^F(\mathbf{z}_2\tilde{\mathbf{z}}_1), \tilde{\mathbf{z}}_2) = H^F(\mathbf{z}_2\tilde{\mathbf{z}}_1\tilde{\mathbf{z}}_2).$$

Repeating this procedure T times yields (9). Suppose now that F is invertible and that (9) holds. In particular, we have that $H^F(\mathbf{z}_1\tilde{\mathbf{z}}) = H^F(\mathbf{z}_2\tilde{\mathbf{z}})$, for any $\tilde{\mathbf{z}} \in \mathcal{Z}$ which, by (8), implies that $F(H^F(\mathbf{z}_1), \tilde{\mathbf{z}}) = F(H^F(\mathbf{z}_2), \tilde{\mathbf{z}})$. If we now apply F^{-1} to both sides of this equality we have by (7) that $H^F(\mathbf{z}_1) = H^F(\mathbf{z}_2)$, as required. $\square$

These facts can be used to prove that any system that has the echo state property at the level of the state equation can be restricted to a smaller state space where it becomes strongly reachable. Additionally, they also imply that invertible state maps and injective readouts determine observable state-space systems.

Corollary 2.6. *Let $(\mathcal{X}, F, h)$ be a system with input space $\mathcal{Z}$ that has the echo state property at the level of the state equation. Then there exists a subset $\mathcal{X}' \subset \mathcal{X}$ such that F restricts to a map (denoted with the same symbol) $F : \mathcal{X}' \times \mathcal{Z} \rightarrow \mathcal{X}'$ and, moreover, $(\mathcal{X}', F, h)$ is strongly reachable.*

Additionally, if the map $F : \mathcal{X} \times \mathcal{Z} \rightarrow \mathcal{X}$ is invertible, then the system $(\mathcal{X}, F, h)$ is necessarily observable for any readout map $h : \mathcal{X} \rightarrow \mathcal{Y}$ that is injective when restricted to $\mathcal{X}' := H^F(\mathcal{Z}^{\mathbb{Z}^-})$.

Proof of the Corollary. First, the ESP at the level of the state equation implies the existence of a state functional $H^F : \mathcal{Z}^{\mathbb{Z}^-} \rightarrow \mathcal{X}$. Define $\mathcal{X}' := H^F(\mathcal{Z}^{\mathbb{Z}^-})$. The relation (8) implies that F restricts to a map $F : \mathcal{X}' \times \mathcal{Z} \rightarrow \mathcal{X}'$ because for any $\mathbf{x}' = H^F(\mathbf{z}') \in \mathcal{X}'$, with $\mathbf{z}' \in \mathcal{Z}^{\mathbb{Z}^-}$ and any $\tilde{\mathbf{z}} \in \mathcal{Z}$,

$$F(\mathbf{x}', \tilde{\mathbf{z}}) = H^F(\mathbf{z}'\tilde{\mathbf{z}}) \in \mathcal{X}'.$$

The restricted state map obviously also has the ESP at the state level and has as associated functional the map with restricted codomain $H^F : \mathcal{Z}^{\mathbb{Z}^-} \rightarrow \mathcal{X}'$, which proves that $(\mathcal{X}', F, h)$ is strongly reachable.

Consider now a system $(\mathcal{X}, F, h)$ such that F is invertible and h is injective when restricted to $\mathcal{X}' := H^F(\mathcal{Z}^{\mathbb{Z}^-})$. Let $\mathbf{z}_1, \mathbf{z}_2 \in \mathcal{Z}^{\mathbb{Z}^-}$ be such that $H_h^F(\mathbf{z}_1\tilde{\mathbf{z}}) = H_h^F(\mathbf{z}_2\tilde{\mathbf{z}})$ for any $\tilde{\mathbf{z}} \in \mathcal{Z}^T$ and any $T \in \mathbb{N}$. The injectivity of h implies that $H^F(\mathbf{z}_1\tilde{\mathbf{z}}) = H^F(\mathbf{z}_2\tilde{\mathbf{z}})$. Since the converse of (9) holds by the invertibility of F , we have then that $H^F(\mathbf{z}_1) = H^F(\mathbf{z}_2)$ and we can hence conclude that the system does not have indistinguishable states and it is hence observable. $\square$

Proof of Proposition 2.4. Using the hypothesis on the strong reachability of $(\mathcal{X}, F, h)$, we know that for any $\mathbf{x} \in \mathcal{X}$ there exists $\mathbf{z} \in \mathcal{Z}^{\mathbb{Z}^-}$ such that $H^F(\mathbf{z}) = \mathbf{x}$. Define:

$$\begin{aligned} f : \mathcal{X} &\longrightarrow \overline{\mathcal{X}} \\ \mathbf{x} = H^F(\mathbf{z}) &\longmapsto f(\mathbf{x}) := H^{\overline{F}}(\mathbf{z}). \end{aligned}$$

We now show that this map is well-defined and that it is the unique system morphism in the statement of the proposition.

(i): f is well-defined: given $\mathbf{x} \in \mathcal{X}$, let $\mathbf{z}_1, \mathbf{z}_2 \in \mathcal{Z}^{\mathbb{Z}^-}$ be such that $\mathbf{x} = H^F(\mathbf{z}_1) = H^F(\mathbf{z}_2)$. We now show that $H^{\overline{F}}(\mathbf{z}_1) = H^{\overline{F}}(\mathbf{z}_2)$, necessarily. By contradiction, suppose that $\mathbf{x}_1 := H^{\overline{F}}(\mathbf{z}_1)$, $\mathbf{x}_2 := H^{\overline{F}}(\mathbf{z}_2)$, and that $\mathbf{x}_1 \neq \mathbf{x}_2$. As by hypothesis $(\overline{\mathcal{X}}, \overline{F}, \overline{h})$ is observable, there exists $\tilde{\mathbf{z}} \in \mathcal{Z}^T$, for some $T \in \mathbb{N}$, such that

$$H_h^{\overline{F}}(\mathbf{z}_1\tilde{\mathbf{z}}) \neq H_h^{\overline{F}}(\mathbf{z}_2\tilde{\mathbf{z}}). \quad (10)$$

However, the equality $H^F(\mathbf{z}_1) = H^F(\mathbf{z}_2)$ and (9) in Lemma 2.5 imply that $H^F(\mathbf{z}_1\tilde{\mathbf{z}}) = H^F(\mathbf{z}_2\tilde{\mathbf{z}})$ and hence $H_h^F(\mathbf{z}_1\tilde{\mathbf{z}}) = H_h^F(\mathbf{z}_2\tilde{\mathbf{z}})$. The hypothesis $H_h^F = H_h^{\overline{F}}$ implies that $H_h^{\overline{F}}(\mathbf{z}_1\tilde{\mathbf{z}}) = H_h^{\overline{F}}(\mathbf{z}_2\tilde{\mathbf{z}})$ which contradicts (10).

(ii): f is system equivariant: Let $\mathbf{x} \in \mathcal{X}$, $\mathbf{z} \in \mathcal{Z}^{\mathbb{Z}^-}$, and $\tilde{\mathbf{z}} \in \mathcal{Z}$, be such that $\mathbf{x} = H^F(\mathbf{z})$. Then, by (8) in Lemma 2.5 we have that

$$f(F(\mathbf{x}, \tilde{\mathbf{z}})) = f(F(H^F(\mathbf{z}), \tilde{\mathbf{z}})) = f(H^F(\mathbf{z}\tilde{\mathbf{z}})) = H^{\overline{F}}(\mathbf{z}\tilde{\mathbf{z}}) = \overline{F}(H^{\overline{F}}(\mathbf{z}), \tilde{\mathbf{z}}) = \overline{F}(f(\mathbf{x}), \tilde{\mathbf{z}}),$$

as required.

(iii): f is readout invariant: using the same elements as in the previous point:

$$h(\mathbf{x}) = h(H^F(\mathbf{z})) = H_h^F(\mathbf{z}) = H_h^{\overline{F}}(\mathbf{z}) = \overline{h}(H^{\overline{F}}(\mathbf{z})) = \overline{h}(f(\mathbf{x})),$$

as required.

(iv): **f is unique:** Let $\bar{f} : \mathcal{X} \rightarrow \bar{\mathcal{X}}$ be another system morphism. Let $\mathbf{x} = H^F(\mathbf{z}) \in \mathcal{X}$ arbitrary. We first show that the sequence

$$(\bar{f}(H^F(T_{-t}(\mathbf{z}))), \mathbf{z}_t)_{t \in \mathbb{Z}_-} \in (\bar{\mathcal{X}} \times \mathcal{Z})^{\mathbb{Z}_-}$$

is a solution of the system associated to $\bar{F}$. Indeed, for any $t \in \mathbb{Z}_-$, and by (8) and the system equivariance of $\bar{f}$:

$$\begin{aligned} \bar{f}(H^F(T_{-t}(\mathbf{z}))) &= \bar{f}(F(U^F(T_{-t}(\mathbf{z})))_{-1}, \mathbf{z}_t)) \\ &= \bar{f}(F(U^F(T_{-(t-1)}(\mathbf{z})))_0, \mathbf{z}_t)) = \bar{F}(\bar{f}(H^F(T_{-(t-1)}(\mathbf{z}))), \mathbf{z}_t), \end{aligned}$$

as required. Now, since $((H^{\bar{F}}(T_{-t}(\mathbf{z}))), \mathbf{z}_t)_{t \in \mathbb{Z}_-} \in (\bar{\mathcal{X}} \times \mathcal{Z})^{\mathbb{Z}_-}$ is also a solution for the system associated to $\bar{F}$ that, by hypothesis, has the echo state property, we necessarily have that:

$$\bar{f}(\mathbf{x}) = \bar{f}(H^F(\mathbf{z})) = H^{\bar{F}}(\mathbf{z}) = f(\mathbf{x}),$$

which proves the uniqueness of the morphism f . $\square$

Corollary 2.7. *If the two systems $(\mathcal{X}, F, h)$ and $(\bar{\mathcal{X}}, \bar{F}, \bar{h})$ in the statement of Proposition 2.4 are canonical then they are necessarily system isomorphic.*

Proof. By Proposition 2.4, the maps $f : \mathcal{X} \rightarrow \bar{\mathcal{X}}$ and $\bar{f} : \bar{\mathcal{X}} \rightarrow \mathcal{X}$ defined by $f(\mathbf{x}) := H^{\bar{F}}(\mathbf{z})$, with $\mathbf{x} = H^F(\mathbf{z})$, and $\bar{f}(\bar{\mathbf{x}}) := H^F(\bar{\mathbf{z}})$, with $\bar{\mathbf{x}} = H^{\bar{F}}(\bar{\mathbf{z}})$, for $\mathbf{z}, \bar{\mathbf{z}} \in \mathcal{Z}^{\mathbb{Z}_-}$, are well-defined system morphisms. Then, for any $\mathbf{x} = H^F(\mathbf{z}) \in \mathcal{X}$ and $\bar{\mathbf{x}} = H^{\bar{F}}(\bar{\mathbf{z}}) \in \bar{\mathcal{X}}$ we can verify that

$$\bar{f} \circ f(\mathbf{x}) = \bar{f}(H^{\bar{F}}(\mathbf{z})) = H^F(\mathbf{z}) = \mathbf{x}, \quad \text{and} \quad f \circ \bar{f}(\bar{\mathbf{x}}) = f(H^F(\bar{\mathbf{z}})) = H^{\bar{F}}(\bar{\mathbf{z}}) = \bar{\mathbf{x}},$$

which shows that $\bar{f} = f^{-1}$ and $f = \bar{f}^{-1}$, as required. $\square$

3. Canonical Realization Theorems. In this section we propose two results in connection with the state-space system realization of input/output systems. The first result shows that any causal and time-invariant input/output system with discrete semi-infinite inputs admits a canonical state-space realization that is unique up to system isomorphisms. As we shall see later on in the examples in Section 4, there is no guarantee that this realization takes place in a finite dimensional space. In a second result, we show that given any state-space system that satisfies the echo state property, we can always associate to it a canonical state-space realization (also unique up to system isomorphisms) that generates the same input/output system. This new canonical system is obtained from the original one by a procedure that we will generically call **reduction** and is defined on a new state space whose dimension (whenever that term is well-defined) is equal or smaller.

Apart from the causality and time-invariance, there is another dynamical feature that is needed to ensure the existence of these canonical realizations, namely, the **input forgetting property** (see [46]).

Definition 3.1. Let $\mathcal{Z}$ be a set, $(\mathcal{Y}, d)$ a metric space, and let $U : \mathcal{Z}^{\mathbb{Z}_-} \rightarrow \mathcal{Y}^{\mathbb{Z}_-}$ be a causal and time-invariant filter. We say that U has the **input forgetting property** whenever for any $\mathbf{u}, \mathbf{v} \in \mathcal{Z}^{\mathbb{Z}_-}$ and any $\mathbf{z} \in \mathcal{Z}^{\mathbb{N}^+}$:

$$\lim_{t \rightarrow \infty} d(H_U(\mathbf{u}\tilde{\mathbf{z}}_t), H_U(\mathbf{v}\tilde{\mathbf{z}}_t)) = 0, \quad (11)$$

where $\tilde{\mathbf{z}}_t := (\mathbf{z}_1, \dots, \mathbf{z}_t) \in \mathcal{Z}^t$, $t \in \mathbb{N}^+$ and $H_U : \mathcal{Z}^{\mathbb{Z}^-} \rightarrow \mathcal{Y}$ is the functional associated to U and defined by $H_U(\mathbf{z}) = U(\mathbf{z})_0$.

This property is also referred to in the literature as the **unique steady-state property** (see [6]) and is usually obtained as a consequence of various continuity properties like the **fading memory property** (see, for instance, [34, Theorem 24] and the definition later on in Section 4).

Theorem 3.2 (Canonical realization of input/output systems). *Let $\mathcal{Z}$ be a set, $(\mathcal{Y}, d)$ a metric space, and let $U : \mathcal{Z}^{\mathbb{Z}^-} \rightarrow \mathcal{Y}^{\mathbb{Z}^-}$ be a causal and time-invariant input/output system that has the input forgetting property. Then, there exists a canonical state-space system $(\mathcal{X}, F, h)$ such that $U = U_h^F$. This canonical realization of U is unique up to system isomorphisms.*

Proof. We start by defining the so-called **Nerode equivalence relation** in $\mathcal{Z}^{\mathbb{Z}^-}$ with respect to the functional $H_U : \mathcal{Z}^{\mathbb{Z}^-} \rightarrow \mathcal{Y}$ determined by U via the assignment $H_U(\mathbf{z}) = U(\mathbf{z})_0$. We say that two elements $\mathbf{z}_1, \mathbf{z}_2 \in \mathcal{Z}^{\mathbb{Z}^-}$ are Nerode equivalent and write $\mathbf{z}_1 \sim_I \mathbf{z}_2$, whenever $H_U(\mathbf{z}_1 \tilde{\mathbf{z}}) = H_U(\mathbf{z}_2 \tilde{\mathbf{z}})$, for all $\tilde{\mathbf{z}} \in \mathcal{Z}^T$ and all $T \in \mathbb{N}$. Define $\mathcal{X} := \mathcal{Z}^{\mathbb{Z}^-} / \sim_I$, where the right-hand side of this equality stands for the set of equivalence classes in $\mathcal{Z}^{\mathbb{Z}^-}$ determined by the equivalence relation $\sim_I$, and denote by $[\mathbf{z}] \in \mathcal{X}$ the class that contains the element $\mathbf{z} \in \mathcal{Z}^{\mathbb{Z}^-}$.

Define now the system $(\mathcal{X}, F, h)$, with $F : \mathcal{X} \times \mathcal{Z} \rightarrow \mathcal{X}$ and $h : \mathcal{X} \rightarrow \mathcal{Y}$ given by

$$F([\mathbf{z}], \tilde{\mathbf{z}}) := [\mathbf{z}\tilde{\mathbf{z}}] \quad \text{and} \quad h([\mathbf{z}]) := H_U(\mathbf{z}). \quad (12)$$

We now show that this system is well-defined, it has the echo state property, and that it is a canonical realization of U . If that is the case, the uniqueness up to system isomorphisms follows from Corollary 2.7. We proceed point by point:

(i): **$(\mathcal{X}, F, h)$ is well-defined:** First of all, $F : \mathcal{X} \times \mathcal{Z} \rightarrow \mathcal{X}$ is well-defined because if $\mathbf{z}_1, \mathbf{z}_2 \in \mathcal{Z}^{\mathbb{Z}^-}$ are such that $\mathbf{z}_1 \sim_I \mathbf{z}_2$ then, by definition,

$$H_U(\mathbf{z}_1 \tilde{\mathbf{z}}) = H_U(\mathbf{z}_2 \tilde{\mathbf{z}}), \quad \text{for all } \tilde{\mathbf{z}} \in \mathcal{Z}^T \text{ and all } T \in \mathbb{N}. \quad (13)$$

In particular, for any $\hat{\mathbf{z}} \in \mathcal{Z}$, we have that $F([\mathbf{z}_1], \hat{\mathbf{z}}) = F([\mathbf{z}_2], \hat{\mathbf{z}})$ because $[\mathbf{z}_1 \hat{\mathbf{z}}] = [\mathbf{z}_2 \hat{\mathbf{z}}]$, as (13) also implies that $H_U(\mathbf{z}_1 \hat{\mathbf{z}} \tilde{\mathbf{z}}) = H_U(\mathbf{z}_2 \hat{\mathbf{z}} \tilde{\mathbf{z}})$ for all $\tilde{\mathbf{z}} \in \mathcal{Z}^T$ and all $T \in \mathbb{N}$. The map $h : \mathcal{X} \rightarrow \mathcal{Y}$ is also well-defined because if we consider $\mathbf{z}_1, \mathbf{z}_2 \in \mathcal{Z}^{\mathbb{Z}^-}$ that, as above, $\mathbf{z}_1 \sim_I \mathbf{z}_2$, the equality (13) implies, in particular, that $H_U(\mathbf{z}_1) = H_U(\mathbf{z}_2)$ and hence $h([\mathbf{z}_1]) = H_U(\mathbf{z}_1) = H_U(\mathbf{z}_2) = h([\mathbf{z}_2])$.

(ii): **The system $(\mathcal{X}, F, h)$ has the echo state property:** Given $\mathbf{z} \in \mathcal{Z}^{\mathbb{Z}^-}$, we first show that the sequence $([T_{-t}(\mathbf{z})], \mathbf{z}_t)_{t \in \mathbb{Z}_-} \in (\mathcal{X} \times \mathcal{Z})^{\mathbb{Z}_-}$ is a solution of the state system $(\mathcal{X}, F)$. This is so because, for any $t \in \mathbb{Z}_-$, we have

$$F([T_{-(t-1)}(\mathbf{z})], \mathbf{z}_t) = [T_{-(t-1)}(\mathbf{z})\mathbf{z}_t] = [T_{-t}(\mathbf{z})].$$

We now show that this solution is unique. Suppose that $(\mathbf{x}_t, \mathbf{z}_t)_{t \in \mathbb{Z}_-} \in (\mathcal{X} \times \mathcal{Z})^{\mathbb{Z}_-}$ is also a solution for $(\mathcal{X}, F)$ with respect to the same input sequence. Since the quotient map $\mathcal{Z}^{\mathbb{Z}^-} \rightarrow \mathcal{Z}^{\mathbb{Z}^-} / \sim_I$ is surjective, for any $t \in \mathbb{Z}_-$ there exists an element $\bar{\mathbf{z}}_t \in \mathcal{Z}^{\mathbb{Z}^-}$ such that $\mathbf{x}_t = [\bar{\mathbf{z}}_t]$. The solution condition on $(\mathbf{x}_t, \mathbf{z}_t)_{t \in \mathbb{Z}_-}$ implies that, also for any $t \in \mathbb{Z}_-$, $[\bar{\mathbf{z}}_t] = F([\bar{\mathbf{z}}_{t-1}], \mathbf{z}_t) = [\bar{\mathbf{z}}_{t-1}\mathbf{z}_t]$ and hence $H_U(\bar{\mathbf{z}}_t \tilde{\mathbf{z}}) = H_U(\bar{\mathbf{z}}_{t-1}\mathbf{z}_t \tilde{\mathbf{z}})$, for all $\tilde{\mathbf{z}} \in \mathcal{Z}^T$ and all $T \in \mathbb{N}$. If we use recursively this identity, we can show that

$$H_U(\bar{\mathbf{z}}_t \tilde{\mathbf{z}}) = H_U(\bar{\mathbf{z}}_{t-1}\mathbf{z}_t \tilde{\mathbf{z}}) = H_U(\bar{\mathbf{z}}_{t-2}\mathbf{z}_{t-1}\mathbf{z}_t \tilde{\mathbf{z}}) = \dots = H_U(\bar{\mathbf{z}}_{t-\tau}\mathbf{z}_{t-(\tau-1)} \dots \mathbf{z}_{t-1}\mathbf{z}_t \tilde{\mathbf{z}}),$$

for all $\tilde{\mathbf{z}} \in \mathcal{Z}^T$ and all $\tau, T \in \mathbb{N}$. These equalities imply that for any $t \in \mathbb{Z}_-$ and $\tau \in \mathbb{N}$:

$$d(H_U(\bar{\mathbf{z}}_t \tilde{\mathbf{z}}), H_U(T_{-t}(\mathbf{z}) \tilde{\mathbf{z}})) = d(H_U(\bar{\mathbf{z}}_{t-\tau} \mathbf{z}_{t-(\tau-1)} \cdots \mathbf{z}_{t-1} \mathbf{z}_t \tilde{\mathbf{z}}), H_U(T_{-t}(\mathbf{z}) \tilde{\mathbf{z}})).$$

Now, since by hypothesis U satisfies the input forgetting property, we can take a limit on τ on the right-hand side of this equality and conclude that

$$\begin{aligned} d(H_U(\bar{\mathbf{z}}_t \tilde{\mathbf{z}}), H_U(T_{-t}(\mathbf{z}) \tilde{\mathbf{z}})) &= \lim_{\tau \rightarrow \infty} d(H_U(\bar{\mathbf{z}}_{t-\tau} \mathbf{z}_{t-(\tau-1)} \cdots \mathbf{z}_{t-1} \mathbf{z}_t \tilde{\mathbf{z}}), H_U(T_{-t}(\mathbf{z}) \tilde{\mathbf{z}})) \\ &= 0, \end{aligned}$$

which implies that $H_U(\bar{\mathbf{z}}_t \tilde{\mathbf{z}}) = H_U(T_{-t}(\mathbf{z}) \tilde{\mathbf{z}})$ and hence that $\mathbf{x}_t = [\bar{\mathbf{z}}_t] = [T_{-t}(\mathbf{z})]$, as required.

- (iii): **$(\mathcal{X}, F, h)$ is a state-space realization of U :** Since in the previous point we proved that $(\mathcal{X}, F, h)$ has the echo state property, we can associate to it a system filter $U_h^F : \mathcal{Z}^{\mathbb{Z}_-} \rightarrow \mathcal{Y}^{\mathbb{Z}_-}$. We also showed that for any input $\mathbf{z} \in \mathcal{Z}^{\mathbb{Z}_-}$ the sequence $([T_{-t}(\mathbf{z})], \mathbf{z}_t)_{t \in \mathbb{Z}_-} \in (\mathcal{X} \times \mathcal{Z})^{\mathbb{Z}_-}$ is the unique solution of the state system $(\mathcal{X}, F)$ which proves that the state filter $U^F : \mathcal{Z}^{\mathbb{Z}_-} \rightarrow \mathcal{X}^{\mathbb{Z}_-}$ is given by

$$U^F(\mathbf{z})_t = [T_{-t}(\mathbf{z})]. \quad (14)$$

Consequently, for any $t \in \mathbb{Z}_-$, we have that

$$U_h^F(\mathbf{z})_t = h([T_{-t}(\mathbf{z})]) = H_U(T_{-t}(\mathbf{z})) = U(\mathbf{z})_t, \quad (15)$$

which implies that $U_h^F = U$.

- (iv): **$(\mathcal{X}, F, h)$ is canonical:** Since for any $\mathbf{z} \in \mathcal{Z}^{\mathbb{Z}_-}$ the equality (14) guarantees that $H^F(\mathbf{z}) = [\mathbf{z}]$, we can immediately conclude that $(\mathcal{X}, F, h)$ is strongly reachable. Let now $\mathbf{x}_1 = H^F(\mathbf{z}_1) = [\mathbf{z}_1]$ and $\mathbf{x}_2 = H^F(\mathbf{z}_2) = [\mathbf{z}_2]$ be two indistinguishable states, that is, for any $\tilde{\mathbf{z}} \in \mathbb{Z}^T$ and any $T \in \mathbb{N}$, we have that $H_h^F(\mathbf{z}_1 \tilde{\mathbf{z}}) = H_h^F(\mathbf{z}_2 \tilde{\mathbf{z}})$. The equality (15) evaluated at $t = 0$ implies that in that case $H_U(\mathbf{z}_1 \tilde{\mathbf{z}}) = H_U(\mathbf{z}_2 \tilde{\mathbf{z}})$, necessarily, and hence we can conclude that $[\mathbf{z}_1] = [\mathbf{z}_2]$, which is equivalent to $\mathbf{x}_1 = \mathbf{x}_2$, as required. $\square$

Remark 3.3. It is easy to see that Theorem 3.2 remains valid when the spaces $\mathcal{Z}^{\mathbb{Z}_-}$ and $\mathcal{Y}^{\mathbb{Z}_-}$ are replaced by time-invariant subsets $\mathcal{V}_{\mathcal{Z}} \subset \mathcal{Z}^{\mathbb{Z}_-}$ and $\mathcal{V}_{\mathcal{Y}} \subset \mathcal{Y}^{\mathbb{Z}_-}$, respectively, that additionally are also invariant with respect to the concatenation with finite sequences that was used in the definition of the Nerode equivalence relation. The time invariance is defined by the property $T_\tau(\mathcal{V}_{\mathcal{Z}}) \subset \mathcal{V}_{\mathcal{Z}}$ and $T_\tau(\mathcal{V}_{\mathcal{Y}}) \subset \mathcal{V}_{\mathcal{Y}}$, for any $\tau \in \mathbb{N}$.

The canonicalization theorem that we just proved provides a canonical state-space realization for any input-forgetting, causal, and time-invariant filter by using as state-space the set of equivalence classes in the space of semi-infinite input sequences with respect to the Nerode equivalence. If that filter happens to be already given in a state-space form, we shall show in the next theorem that a canonical realization can be constructed for it by **reducing** the given state-space.

The reduction procedure that we propose next is reminiscent of the *optimal reduction* method introduced in [81, 79] in the context of symmetric Hamiltonian systems and consists in two steps. First, given a (generically non-canonical) state-space system $(\mathcal{X}, F, h)$ with $\mathcal{Z}$ and $\mathcal{Y}$ as input and output spaces, respectively, and

that satisfies the echo state property, we restrict the state equation to the subset $\mathcal{X}_R \subset \mathcal{X}$ of **reachable states** defined by

$$\mathcal{X}_R := \{\mathbf{x} \in \mathcal{X} \mid \mathbf{x} = H^F(\mathbf{z}) \text{ for some } \mathbf{z} \in \mathcal{Z}^{\mathbb{Z}^-}\}. \quad (16)$$

Note that $\mathcal{X}_R$ is the state subspace already introduced in Corollary 2.6.

In a second step, we can define in $\mathcal{X}_R$ the Nerode equivalence relation $\sim_S$ that in the previous theorem was formulated in the space of semi-infinite input sequences. More explicitly, given $\mathbf{x}_1 = H^F(\mathbf{z}_1), \mathbf{x}_2 = H^F(\mathbf{z}_2) \in \mathcal{X}_R$, for some $\mathbf{z}_1, \mathbf{z}_2 \in \mathcal{Z}^{\mathbb{Z}^-}$, we say that these two states are **Nerode equivalent** and, as before, we denote

$$\mathbf{x}_1 \sim_S \mathbf{x}_2 \text{ whenever } H_h^F(\mathbf{z}_1 \tilde{\mathbf{z}}) = H_h^F(\mathbf{z}_2 \tilde{\mathbf{z}}), \text{ for all } \tilde{\mathbf{z}} \in \mathcal{Z}^T \text{ and all } T \in \mathbb{N}. \quad (17)$$

Notice that this definition of Nerode equivalent states is equivalent to the so-called indistinguishable states which is introduced in part (ii) of Definition 2.3.

The symbol $[\mathbf{x}] \in \mathcal{X}_R / \sim_S$ denotes the equivalence class that contains the element $\mathbf{x} \in \mathcal{X}_R$. We emphasize that this relation is well-defined since it does not depend on the elements $\mathbf{z}_1, \mathbf{z}_2 \in \mathcal{Z}^{\mathbb{Z}^-}$ used to define $\mathbf{x}_1$ and $\mathbf{x}_2$ because of (9) in Lemma 2.5.

In the next theorem will show that $(\mathcal{X}, F, h)$ naturally projects to a system on the quotient $\mathcal{X}_R / \sim_S$ that has the echo state property if $(\mathcal{X}, F, h)$ is input-forgetting and, more importantly, is canonical.

Theorem 3.4 (Canonicalization by reduction). *Let $\mathcal{Z}$ be a set, $(\mathcal{Y}, d)$ a metric space, and let $(\mathcal{X}, F, h)$ be a state-space system that has $\mathcal{Z}$ and $\mathcal{Y}$ as input and output spaces, respectively. Suppose that $(\mathcal{X}, F)$ has the echo state property and that the state-space filter $U_h^F : \mathcal{Z}^{\mathbb{Z}^-} \rightarrow \mathcal{Y}^{\mathbb{Z}^-}$ has the input forgetting property. Let $\mathcal{X}_R \subset \mathcal{X}$ be the set of reachable states defined in (16) and $\bar{\mathcal{X}} := \mathcal{X}_R / \sim_S$ the quotient set with respect to the Nerode equivalence relation $\sim_S$ defined in (17).*

The state-space system $(\mathcal{X}, F, h)$ drops to another system $(\bar{\mathcal{X}}, \bar{F}, \bar{h})$ with the same input and output spaces, with states in the quotient space $\bar{\mathcal{X}}$, and maps $\bar{F} : \bar{\mathcal{X}} \times \mathcal{Z} \rightarrow \bar{\mathcal{X}}$ and $\bar{h} : \bar{\mathcal{X}} \rightarrow \mathcal{Y}$ defined by:

$$\begin{cases} \bar{F}([\mathbf{x}], \mathbf{z}) &:= [F(\mathbf{x}, \mathbf{z})], \\ \bar{h}([\mathbf{x}]) &:= h(\mathbf{x}). \end{cases} \quad (18)$$

*The state-space system $(\bar{\mathcal{X}}, \bar{F}, \bar{h})$ has the echo state property and it is a canonical realization of U_h^F . We refer to $(\bar{\mathcal{X}}, \bar{F}, \bar{h})$ as the **canonical reduced realization** of $(\mathcal{X}, F, h)$.*

Proof. We first show that the reduced state and readout maps $\bar{F}$ and $\bar{h}$ in (18) are well-defined. Concerning $\bar{F}$, we show first that the restriction of F to $\mathcal{X}_R \times \mathcal{Z}$ maps into $\mathcal{X}_R$. Indeed, let $\mathbf{x} \in \mathcal{X}_R$ arbitrary and let $\mathbf{z} \in \mathcal{Z}^{\mathbb{Z}^-}$ be such that $\mathbf{x} = H^F(\mathbf{z})$. Then, for any $\tilde{\mathbf{z}} \in \mathcal{Z}$, by (8) in Lemma 2.5, we have that

$$F(\mathbf{x}, \tilde{\mathbf{z}}) = F(H^F(\mathbf{z}), \tilde{\mathbf{z}}) = H^F(\mathbf{z}\tilde{\mathbf{z}}) \in \mathcal{X}_R.$$

This guarantees that $F : \mathcal{X} \times \mathcal{Z} \rightarrow \mathcal{X}$ restricts to a map $F_R : \mathcal{X}_R \times \mathcal{Z} \rightarrow \mathcal{X}_R$ that we now show drops to $\bar{F} : \bar{\mathcal{X}} \times \mathcal{Z} \rightarrow \bar{\mathcal{X}}$ by proving that if $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{X}_R$ are such that $\mathbf{x}_1 \sim_S \mathbf{x}_2$, then $F_R(\mathbf{x}_1, \mathbf{z}) \sim_S F_R(\mathbf{x}_2, \mathbf{z})$, for all $\mathbf{z} \in \mathcal{Z}$. Indeed, if $\mathbf{x}_1 \sim_S \mathbf{x}_2$, by definition (17), $H_h^F(\mathbf{z}_1 \tilde{\mathbf{z}}) = H_h^F(\mathbf{z}_2 \tilde{\mathbf{z}})$, for all $\tilde{\mathbf{z}} \in \mathcal{Z}^T$ and all $T \in \mathbb{N}$, where $\mathbf{x}_1 = H^F(\mathbf{z}_1), \mathbf{x}_2 = H^F(\mathbf{z}_2) \in \mathcal{X}_R$, for some $\mathbf{z}_1, \mathbf{z}_2 \in \mathcal{Z}^{\mathbb{Z}^-}$. Now, by (8) and for all $\mathbf{z} \in \mathcal{Z}$, $F_R(\mathbf{x}_1, \mathbf{z}) = H^F(\mathbf{z}_1 \mathbf{z})$, $F_R(\mathbf{x}_2, \mathbf{z}) = H^F(\mathbf{z}_2 \mathbf{z})$ and since by (17) $H_h^F(\mathbf{z}_1 \mathbf{z}\tilde{\mathbf{z}}) = H_h^F(\mathbf{z}_2 \mathbf{z}\tilde{\mathbf{z}})$, for all $\tilde{\mathbf{z}} \in \mathcal{Z}^T$ and all $T \in \mathbb{N}$, we can conclude that $F_R(\mathbf{x}_1, \mathbf{z}) \sim_S F_R(\mathbf{x}_2, \mathbf{z})$, as required. In order to show that $\bar{h}$ is well-defined, consider first the restriction $h_R := h|_{\mathcal{X}_R} : \mathcal{X}_R \rightarrow \mathcal{Y}$ as well as two elements $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{X}_R$ as

above such that $\mathbf{x}_1 \sim_S \mathbf{x}_2$. Taking now for $\tilde{\mathbf{z}}$ the empty sequence in the definition of the equivalence relation $\sim_S$, we have that:

$$h_R(\mathbf{x}_1) = h(H^F(\mathbf{z}_1)) = H_h^F(\mathbf{z}_1) = H_h^F(\mathbf{z}_2) = h(H^F(\mathbf{z}_2)) = h_R(\mathbf{x}_2),$$

which proves that h_R drops to the map $\bar{h}$ in the statement and it is hence well-defined.

We now show that the reduced system $(\bar{\mathcal{X}}, \bar{F}, \bar{h})$ has the echo state property by following a scheme similar to part (ii) in the proof of Theorem 3.2. First of all, it is easy to see that if $(\mathbf{x}_t, \mathbf{z}_t)_{t \in \mathbb{Z}_-}$ is the unique solution of the system $(\mathcal{X}, F)$ (that by hypothesis satisfies the echo state property) associated to $\mathbf{z} := (\mathbf{z}_t)_{t \in \mathbb{Z}_-}$, then $([\mathbf{x}_t], \mathbf{z}_t)_{t \in \mathbb{Z}_-}$ is a solution of the system $(\bar{\mathcal{X}}, \bar{F})$ associated to $\mathbf{z}$. We now show that that solution is unique. Suppose that $([\bar{\mathbf{x}}_t], \mathbf{z}_t)_{t \in \mathbb{Z}_-}$ is another solution of $(\bar{\mathcal{X}}, \bar{F})$ for the same input $\mathbf{z}$. For any $t \in \mathbb{Z}_-$, let $\bar{\mathbf{z}}_t \in \mathcal{Z}^{\mathbb{Z}_-}$ be such that $\bar{\mathbf{x}}_t = H^F(\bar{\mathbf{z}}_t)$. The solution condition on $([\bar{\mathbf{x}}_t], \mathbf{z}_t)_{t \in \mathbb{Z}_-}$ implies that, also for any $t \in \mathbb{Z}_-$,

$$[H^F(\bar{\mathbf{z}}_t)] = [\bar{\mathbf{x}}_t] = \bar{F}([\bar{\mathbf{x}}_{t-1}], \mathbf{z}_t) = [F(\bar{\mathbf{x}}_{t-1}, \mathbf{z}_t)] = [H^F(\bar{\mathbf{z}}_{t-1} \mathbf{z}_t)],$$

which by (17), implies that for all $\tilde{\mathbf{z}} \in \mathcal{Z}^T$ and all $T \in \mathbb{N}$ one has $H_h^F(\bar{\mathbf{z}}_t \tilde{\mathbf{z}}) = H_h^F(\bar{\mathbf{z}}_{t-1} \mathbf{z}_t \tilde{\mathbf{z}})$, necessarily. If we use recursively this identity, we can show that

$$H_h^F(\bar{\mathbf{z}}_t \tilde{\mathbf{z}}) = H_h^F(\bar{\mathbf{z}}_{t-1} \mathbf{z}_t \tilde{\mathbf{z}}) = H_h^F(\bar{\mathbf{z}}_{t-2} \mathbf{z}_{t-1} \mathbf{z}_t \tilde{\mathbf{z}}) = \cdots = H_h^F(\bar{\mathbf{z}}_{t-\tau} \mathbf{z}_{t-(\tau-1)} \cdots \mathbf{z}_{t-1} \mathbf{z}_t \tilde{\mathbf{z}}),$$

for all $\tilde{\mathbf{z}} \in \mathcal{Z}^T$ and all $\tau, T \in \mathbb{N}$. These equalities imply that for any $t \in \mathbb{Z}_-$ and $\tau \in \mathbb{N}$:

$$d(H_h^F(\bar{\mathbf{z}}_t \tilde{\mathbf{z}}), H_h^F(T_{-t}(\mathbf{z}) \tilde{\mathbf{z}})) = d(H_h^F(\bar{\mathbf{z}}_{t-\tau} \mathbf{z}_{t-(\tau-1)} \cdots \mathbf{z}_{t-1} \mathbf{z}_t \tilde{\mathbf{z}}), H_h^F(T_{-t}(\mathbf{z}) \tilde{\mathbf{z}})).$$

Now, since by hypothesis H_h^F has the input forgetting property, we can take a limit on τ on the right-hand side of this equality and conclude that

$$\begin{aligned} d(H_h^F(\bar{\mathbf{z}}_t \tilde{\mathbf{z}}), H_h^F(T_{-t}(\mathbf{z}) \tilde{\mathbf{z}})) &= \lim_{\tau \rightarrow \infty} d(H_h^F(\bar{\mathbf{z}}_{t-\tau} \mathbf{z}_{t-(\tau-1)} \cdots \mathbf{z}_{t-1} \mathbf{z}_t \tilde{\mathbf{z}}), H_h^F(T_{-t}(\mathbf{z}) \tilde{\mathbf{z}})) \\ &= 0, \end{aligned}$$

which implies that $H_h^F(\bar{\mathbf{z}}_t \tilde{\mathbf{z}}) = H_h^F(T_{-t}(\mathbf{z}) \tilde{\mathbf{z}})$ and hence that $[\bar{\mathbf{x}}_t] = [H^F(\bar{\mathbf{z}}_t)] = [H^F(T_{-t}(\mathbf{z}))] = [\mathbf{x}_t]$, as required.

Finally, the fact that $([\mathbf{x}_t], \mathbf{z}_t)_{t \in \mathbb{Z}_-}$ is the unique solution of $(\bar{\mathcal{X}}, \bar{F})$ associated to $\mathbf{z}$ when $(\mathbf{x}_t, \mathbf{z}_t)_{t \in \mathbb{Z}_-}$ is the unique solution of $(\mathcal{X}, F)$ amounts to the equality $H_h^{\bar{F}} = H_h^F$. Consequently, $(\bar{\mathcal{X}}, \bar{F}, \bar{h})$ is a realization for the filter associated to $(\mathcal{X}, F, h)$ and it is trivially canonical. $\square$

Since Theorems 3.2 and 3.4 produce two different canonical realizations of a given system and we know by Corollary 2.7 that those realizations are unique up to system isomorphisms, we can conclude the non-trivial statement that the two sets of classes $\mathcal{Z}^{\mathbb{Z}_-} / \sim_I$ and $\mathcal{X}_R / \sim_S$ in the space of semi-infinite input sequences and on the space of reachable states, respectively, are isomorphic quotient spaces. We frame that result in the next corollary.

Corollary 3.5. *Let $\mathcal{Z}$ be a set, $(\mathcal{Y}, d)$ a metric space, and let $(\mathcal{X}, F, h)$ be a state-space system that has $\mathcal{Z}$ and $\mathcal{Y}$ as input and output spaces, respectively. Suppose that $(\mathcal{X}, F)$ has the echo state property and that the state-space system filter $U_h^F : \mathcal{Z}^{\mathbb{Z}_-} \rightarrow \mathcal{Y}^{\mathbb{Z}_-}$ has the input forgetting property. Let $\mathcal{X}_R / \sim_S$ be the reduced state-space defined in (17) and let $\mathcal{Z}^{\mathbb{Z}_-} / \sim_I$ be the quotient space defined in the proof*

of Theorem 3.2. These two quotient spaces are isomorphic. The isomorphism is implemented by the map:

$$\begin{aligned} f : \mathcal{Z}^{\mathbb{Z}_-} / \sim_I &\longrightarrow \mathcal{X}_R / \sim_S \\ [\mathbf{z}] &\longmapsto [H^F(\mathbf{z})]. \end{aligned}$$

4. Realization and canonicalization of linear filters. In this section we study the realization and canonicalization problem for linear, time-invariant, and causal filters that satisfy the so-called fading memory property. In order to explicitly define the input spaces and this property we first consider the **supremum norm** $\|\cdot\|_\infty$ in the space of semi-infinite sequences $\mathbb{R}^{\mathbb{Z}_-}$ in $\mathbb{R}$ defined by

$$\|\mathbf{z}\|_\infty := \sup_{t \in \mathbb{Z}_-} \{|z_t|\}, \quad \text{for any } \mathbf{z} \in \mathbb{R}^{\mathbb{Z}_-}. \quad (19)$$

Let $(\ell_-^\infty(\mathbb{R}), \|\cdot\|_\infty)$ be the Banach space formed by the elements in $\mathbb{R}^{\mathbb{Z}_-}$ that have a finite supremum norm. We define now a **weighting sequence** $w : \mathbb{N} \rightarrow (0, 1]$ as a strictly decreasing sequence with zero limit such that $w_0 = 1$. Given an element $\mathbf{z} \in \ell_-^\infty(\mathbb{R})$, we define its **w -weighted norm** $\|\cdot\|_w$ by

$$\|\mathbf{z}\|_w := \sup_{t \in \mathbb{Z}_-} \{|\mathbf{z}_t|w_{-t}\}.$$

Consider now a linear, time-invariant, and causal filter $U : \ell_-^\infty(\mathbb{R}) \rightarrow \ell_-^\infty(\mathbb{R})$. We say that the functional $H_U : \ell_-^\infty(\mathbb{R}) \rightarrow \mathbb{R}$ associated to U has the so-called **fading memory property (FMP)** with respect to the weighting sequence w whenever for any $\epsilon > 0$, there exists $\delta(\epsilon) > 0$ such that if $\mathbf{z} \in \ell_-^\infty(\mathbb{R})$ is such that $\|\mathbf{z}\|_w < \delta(\epsilon)$ then $|H_U(\mathbf{z})| < \epsilon$, necessarily.

The Convolution Theorem (see [6, Theorem 5]) shows that H_U has the FMP if and only if the filter U has a convolution representation, that is, there exists an element $\Psi \in \ell_-^1(\mathbb{R})$ such that

$$U(\mathbf{z})_t = \sum_{j \in \mathbb{Z}_-} \Psi_j z_{t+j} =: (\Psi * \mathbf{z})_t, \quad \text{for any } \mathbf{z} \in \ell_-^\infty(\mathbb{R}), t \in \mathbb{Z}_-.$$

In such case, it is easy to see that $U : \ell_-^\infty(\mathbb{R}) \rightarrow \ell_-^\infty(\mathbb{R})$ is a bounded linear operator and that its operator norm $\|U\|_\infty$ satisfies that $\|U\|_\infty \leq \|\Psi\|_1 =: \sum_{j \in \mathbb{Z}_-} |\Psi_j| < \infty$. Additionally, as we already mentioned after Definition 3.1, the FMP implies the input forgetting property that we used in the main results in Section 3. This fact has been proved in the literature in general non-linear setups (see [6, Theorem 6] and [34, Theorem 6]).

The following result collects all the facts that we just mentioned. Before we proceed with the statement, we extend the definition of the time delay operator $T_\tau : \mathcal{Z}^{\mathbb{Z}_-} \rightarrow \mathcal{Z}^{\mathbb{Z}_-}$ defined in (6) for any $\tau \in \mathbb{N}$, to accommodate any $\tau \in \mathbb{Z}$ by setting, for any $\tau < 0$:

$$T_\tau(\mathbf{z}) := (\mathbf{z}, \underbrace{\mathbf{0}, \dots, \mathbf{0}}_{-\tau \text{ times}}), \quad \mathbf{z} \in \mathcal{Z}^{\mathbb{Z}_-}. \quad (20)$$

Proposition 4.1. *Let $U : \ell_-^\infty(\mathbb{R}) \rightarrow \ell_-^\infty(\mathbb{R})$ be a linear, time-invariant, and causal filter such that $H_U : \ell_-^\infty(\mathbb{R}) \rightarrow \mathbb{R}$ has the fading memory property with respect to a weighting sequence w . Then, there exists a unique element $\Psi \in \ell_-^1(\mathbb{R})$ such that $U(\mathbf{z}) = \Psi * \mathbf{z}$ for any $\mathbf{z} \in \ell_-^\infty(\mathbb{R})$. Moreover, U is a bounded linear automorphism of $\ell_-^\infty(\mathbb{R})$ such that $\|U\|_\infty \leq \|\Psi\|_1 < \infty$ and it has the input forgetting property.*

Proof. In view of the references quoted above, it just remains to be shown that the element $\Psi \in \ell^1_-(\mathbb{R})$ that provides the convolution representation is unique, which is due to the fact that its components are uniquely determined by the impulse response of U , that is, for any $t \in \mathbb{Z}_-$

$$\Psi_t = H_U(\mathbf{e}_t) = (\Psi * \mathbf{e}_t)_0, \quad \text{where } \mathbf{e}_t := (\dots, 0, \underbrace{1}_{t \text{ entry}}, 0, \dots, 0) \in \ell^\infty_-(\mathbb{R}).$$

It is worth pointing out that the fact that U satisfies the input forgetting property can be easily proved in this case using the convolution representation. Let $\mathbf{u}, \mathbf{v} \in \ell^\infty_-(\mathbb{R})$, $\mathbf{z} \in \mathbb{R}^{\mathbb{N}^+}$, and denote $\tilde{\mathbf{z}}_t := (z_1, \dots, z_t) \in \mathbb{R}^t$, for any $t \in \mathbb{N}^+$. It is easy to see using (20) that

$$\mathbf{u}\tilde{\mathbf{z}}_t = T_{-t}(\mathbf{u}) + (\dots, 0, z_1, z_2, \dots, z_t) \quad \text{and} \quad \mathbf{v}\tilde{\mathbf{z}}_t = T_{-t}(\mathbf{v}) + (\dots, 0, z_1, z_2, \dots, z_t).$$

We now use these equalities with the convolution representation of U and the linearity of T_{-t} and show that:

$$\begin{aligned} |H_U(\mathbf{u}\tilde{\mathbf{z}}_t) - H_U(\mathbf{v}\tilde{\mathbf{z}}_t)| &= |U(\mathbf{u}\tilde{\mathbf{z}}_t)_0 - U(\mathbf{v}\tilde{\mathbf{z}}_t)_0| \\ &= |\Psi * T_{-t}(\mathbf{u} - \mathbf{v})| = \left| \sum_{j=-\infty}^{-t} \Psi_j (u - v)_{j+t} \right| \\ &\leq \sum_{j=-\infty}^{-t} |\Psi_j| \|\mathbf{u} - \mathbf{v}\|_\infty = \left(\|\Psi\|_1 - \sum_{j=-t+1}^0 |\Psi_j| \right) \|\mathbf{u} - \mathbf{v}\|_\infty \xrightarrow{t \rightarrow \infty} 0. \end{aligned} \quad (21)$$

□

The proposition that we just proved shows, in particular, that FMP linear, causal, and time invariant filters satisfy the hypotheses of the Canonical Realization Theorem 3.2 and hence they always have a canonical state-space realization that, as we show later in Theorem 4.5, is linear even though the state space may be an infinite dimensional vector space. We emphasize that the fading memory property plays a crucial role in the result that we just proved since, in general, the Convolution Theorem does not hold in its absence (see the counterexample in Section A3 of [6]).

Before we proceed with that theorem, we first state a result that lists important properties of finite-dimensional linear state-space realizations that are needed in the sequel.

Proposition 4.2 (Linear state-space realizations with semi-infinite inputs). *Let $N \in \mathbb{N}$, let $A \in \mathbb{M}_N$ be a diagonalizable matrix, $\mathbf{C} \in \mathbb{R}^N$, $W \in \mathbb{M}_{1,N}$, and consider the linear state-space system $(\mathcal{V}, F, h)$ defined by $\mathcal{V} = \mathbb{R}^N$ and*

$$\begin{cases} F(\mathbf{x}, z) := A\mathbf{x} + \mathbf{C}z, \\ h(\mathbf{x}) := W\mathbf{x}. \end{cases} \quad (22)$$

$$(23)$$

- (i): *The state equation associated to (22) has a unique solution in $\ell^\infty(\mathbb{R}^N)$ for each input in $\ell^\infty_-(\mathbb{R})$ (we call this property the $(\ell^\infty_-(\mathbb{R}^N), \ell^\infty_-(\mathbb{R}))$ -ESP) if and only if $\rho(A) < 1$, where $\rho(A)$ stands for the spectral radius of A .*
- (ii): *In the remainder of this proposition suppose that $\rho(A) < 1$. Then, there exists a state filter $U^F : \ell^\infty_-(\mathbb{R}) \rightarrow \ell^\infty_-(\mathbb{R}^N)$ and a corresponding state-space filter*

$U_h^F : \ell^\infty(\mathbb{R}) \longrightarrow \ell^\infty(\mathbb{R})$ given by

$$U^F(\mathbf{z})_t := \sum_{j=0}^{\infty} A^j \mathbf{C} z_{t-j} \quad \text{and} \quad U_h^F(\mathbf{z})_t := W \sum_{j=0}^{\infty} A^j \mathbf{C} z_{t-j}, \quad \text{respectively.} \quad (24)$$

The state-space filter U_h^F has the input forgetting property.

(iii): The set $\mathcal{V}_R \subset \mathcal{V}$ of reachable states defined in (16) of $(\mathcal{V}, F, h)$ is given by

$$\mathcal{V}_R = \text{span} \{ \mathbf{C}, A\mathbf{C}, A^2\mathbf{C}, \dots, A^{N-1}\mathbf{C} \}. \quad (25)$$

(iv): Given $\mathbf{x} \in \mathcal{V}_R$, the set of indistinguishable states of $\mathbf{x}$ in $\mathcal{V}_R$ is given by the coset

$$I_{F,h}^{\mathbf{x}} := \mathbf{x} + I_{F,h} \quad \text{with} \quad I_{F,h} := \bigcap_{i=0}^{N-1} \ker W A^i. \quad (26)$$

The state-space system $(\mathcal{V}, F, h)$ is hence observable if and only if $I_{F,h} = \{\mathbf{0}\}$. This condition is equivalent to the maximality of the rank of the observability matrix $O(A, W)$ defined by

$$O(A, W) = \begin{pmatrix} W \\ W A \\ \vdots \\ W A^{N-1} \end{pmatrix}, \quad \text{that is, } \text{rank } O(A, W) = N \text{ if and only if } I_{F,h} = \{\mathbf{0}\}. \quad (27)$$

Remark 4.3. The dimension of $\mathcal{V}_R$ in (25) coincides with the rank of the **controlability** or **reachability matrix** $R(A, \mathbf{C})$ defined by

$$R(A, \mathbf{C}) := (\mathbf{C} \mid A\mathbf{C} \mid \dots \mid A^{N-1}\mathbf{C}).$$

When this rank is maximal, the linear system $(\mathcal{V}, F, h)$ is strongly reachable in the sense of the Definition 2.3 and also in the control theoretical sense (see [52, 89]). It has been shown in [22] that if A is diagonalizable then $R(A, \mathbf{C})$ has maximal rank if and only if all the eigenvalues in the spectrum $\sigma(A)$ of A are distinct and in the linear decomposition $\mathbf{C} = \sum_{i=1}^N c_i \mathbf{v}_i$, with $\{\mathbf{v}_1, \dots, \mathbf{v}_N\}$ a basis of eigenvectors of A , all the coefficients c_i , with $i \in \{1, \dots, N\}$, are non-zero.

The dimension of $\mathcal{V}_R$ also coincides with the so-called **memory capacity** [47] of the recurrent network associated to (22)-(23). This fact has been recently proved in [22].

Proof of the Proposition. (i) and (ii) We first show that if $\rho(A) \geq 1$ then $(\mathcal{V}, F)$ cannot have the $(\ell^\infty(\mathbb{R}^N), \ell^\infty(\mathbb{R}))$ -ESP. Let $\lambda \geq 1$ be one of the elements in the spectrum $\sigma(A)$ and let $\mathbf{v} \in \mathbb{R}^N$ be an associated norm-one eigenvector. Let $\mathbf{x}^\lambda \in \ell^\infty(\mathbb{R}^N)$ be defined by $\mathbf{x}_t^\lambda := \lambda^t \mathbf{v}$, $t \in \mathbb{Z}_-$. It is clear that as $\lambda \geq 1$ then $\|\mathbf{x}^\lambda\|_\infty = \|\mathbf{v}\| = 1$. Moreover, $\mathbf{x}^\lambda$ is a solution of the system associated to F with zero input because for any $t \in \mathbb{Z}_-$ we have

$$F(\mathbf{x}_{t-1}^\lambda, 0) = A\mathbf{x}_{t-1}^\lambda = \lambda^{t-1} A\mathbf{v} = \lambda^t \mathbf{v} = \mathbf{x}_t^\lambda.$$

Since $\mathbf{0} \in \ell^\infty(\mathbb{R}^N)$ is also a solution for the same input, then $(\mathcal{V}, F)$ does not have the ESP. What we just proved is equivalent to stating that if $(\mathcal{V}, F)$ has the $(\ell^\infty(\mathbb{R}^N), \ell^\infty(\mathbb{R}))$ -ESP then $\rho < 1$ necessarily.

Conversely, suppose that $\rho(A) < 1$. We now show that first, for any $\mathbf{z} \in \ell^\infty(\mathbb{R})$ the sequence $\mathbf{x} \in \ell^\infty(\mathbb{R}^N)$ whose terms $\mathbf{x}_t$ are defined by

$$\mathbf{x}_t := \sum_{j=0}^{\infty} A^j \mathbf{C} z_{t-j} \quad (28)$$

is a solution of $(\mathcal{V}, F)$ for the input $\mathbf{z}$ and second, that this solution is unique. In order to show that (28) is a solution, we first recall that by Gelfand's formula (see [60]) $\lim_{k \rightarrow \infty} \|A^k\|^{1/k} = \rho(A) < 1$, which implies the existence of a number $k_0 \in \mathbb{N}$ such that $\|A^k\| < 1$, for all $k \geq k_0$. Consequently, the infinite sum

$$\sum_{j=0}^{\infty} A^j = \mathbb{I}_N + A + \cdots + A^{k_0-1} + \sum_{j=1}^{\infty} \sum_{i=0}^{k_0-1} A^{jk_0} A^i = \sum_{j=0}^{\infty} \sum_{i=0}^{k_0-1} A^{jk_0} A^i \quad (29)$$

converges in operator norm because as $\|A^{jk_0} A^i\| \leq \|A^{k_0}\|^j \|A^i\|$ for all $j \in \mathbb{N}$, $i \in \{0, \dots, k_0-1\}$ then (29) implies that

$$\left\| \sum_{j=0}^{\infty} A^j \right\| \leq \sum_{i=0}^{k_0-1} \frac{\|A^i\|}{1 - \|A^{k_0}\|} < \infty. \quad (30)$$

This inequality, (29), and (28) imply that

$$\begin{aligned} \|\mathbf{x}_t\| &\leq \left\| \sum_{j=0}^{\infty} \sum_{i=0}^{k_0-1} A^{jk_0} A^i \mathbf{C} z_{t-(jk_0+i)} \right\| \leq \left(\sum_{j=0}^{\infty} \sum_{i=0}^{k_0-1} \|A^{k_0}\|^j \|A^i\| \right) \|\mathbf{C}\| \|\mathbf{z}\|_{\infty} \\ &= \sum_{i=0}^{k_0-1} \frac{\|A^i\|}{1 - \|A^{k_0}\|} \|\mathbf{C}\| \|\mathbf{z}\|_{\infty}, \end{aligned}$$

which shows that the series in (28) are convergent and also that

$$\|\mathbf{x}\|_{\infty} \leq \sum_{i=0}^{k_0-1} \frac{\|A^i\|}{1 - \|A^{k_0}\|} \|\mathbf{C}\| \|\mathbf{z}\|_{\infty} < \infty.$$

The fact that $\mathbf{x} \in \ell^\infty(\mathbb{R}^N)$ is a solution of $(\mathcal{V}, F)$ for the input $\mathbf{z} \in \ell^\infty(\mathbb{R})$ is a straightforward verification. Suppose now that $\bar{\mathbf{x}} \in \ell^\infty(\mathbb{R}^N)$ is another solution of $(\mathcal{V}, F)$ for the same input, that is, $\bar{\mathbf{x}}_t = A\bar{\mathbf{x}}_{t-1} + \mathbf{C}z_t$, for all $t \in \mathbb{Z}_-$. This implies that $\mathbf{x} - \bar{\mathbf{x}} \in \ell^\infty(\mathbb{R}^N)$ is a solution of $(\mathcal{V}, F)$ for the zero input and hence

$$\mathbf{x}_t - \bar{\mathbf{x}}_t = A(\mathbf{x}_{t-1} - \bar{\mathbf{x}}_{t-1}), \quad \text{for all } t \in \mathbb{Z}_-. \quad (31)$$

Using the same decomposition as in (29), we have that for any $l \in \mathbb{N}$ there exists $j \in \mathbb{N}$ and $i \in \{1, \dots, k_0-1\}$ such that $A^l = A^{jk_0} A^i$. Hence, by iterating (31) we have that $\mathbf{x}_t - \bar{\mathbf{x}}_t = A^l(\mathbf{x}_{t-l} - \bar{\mathbf{x}}_{t-l})$ and therefore

$$\|\mathbf{x}_t - \bar{\mathbf{x}}_t\| \leq \|A^{jk_0} A^i\| \|\mathbf{x}_{t-l} - \bar{\mathbf{x}}_{t-l}\| \leq \|A^{k_0}\|^j \|A^i\| \|\mathbf{x} - \bar{\mathbf{x}}\|_{\infty}.$$

Taking the limit $j \rightarrow \infty$ in this inequality, we obtain that $\|\mathbf{x}_t - \bar{\mathbf{x}}_t\| = 0$, for all $t \in \mathbb{Z}_-$, which guarantees that $\mathbf{x} = \bar{\mathbf{x}}$, as required.

Finally, we show that when $\rho(A) < 1$ then the filter U_h^F in (24) has the input forgetting property. Notice first that (24) amounts to a convolution representation for U_h^F , that is, $U_h^F(\mathbf{z}) = \Psi * \mathbf{z}$, for any $\mathbf{z} \in \ell^\infty(\mathbb{R})$, where

$\Psi_{-j} = WA^j \mathbf{C}$, $j \in \mathbb{N}$. If we show that $\Psi \in \ell^1_-(\mathbb{R})$, then an argument similar to (21) proves that U_h^F has the input forgetting property. This is the case because by (30)

$$\|\Psi\|_1 = \left\| \sum_{j=0}^{\infty} WA^j \mathbf{C} \right\| \leq \|W\| \left\| \sum_{j=0}^{\infty} A^j \right\| \|\mathbf{C}\| \leq \|W\| \|\mathbf{C}\| \sum_{i=0}^{k_0-1} \frac{\|A^i\|}{1 - \|A^{k_0}\|} < \infty.$$

- (iii) First of all, since by (24) the state functional $H^F : \ell^\infty_-(\mathbb{R}) \rightarrow \mathbb{R}^N$ is linear and given by $H^F(\mathbf{z}) := \sum_{j=0}^{\infty} A^j \mathbf{C} z_{-j}$, we can immediately conclude that the reachable set $\mathcal{V}_R = H^F(\ell^\infty_-(\mathbb{R})) \subseteq \mathbb{R}^N$ is a vector subspace of $\mathbb{R}^N$. We now establish (25) by double inclusion. The inclusion $\mathcal{V}_R \supseteq \text{span}\{\mathbf{C}, A\mathbf{C}, A^2\mathbf{C}, \dots, A^{N-1}\mathbf{C}\}$ is proved by applying H^F to inputs of the form

$$\mathbf{z}^j := (\dots, 0, \underbrace{1}_{(-j)\text{th entry}}, 0, \dots, 0) \in \ell^\infty_-(\mathbb{R}), \quad \text{with } j \in \{0, \dots, N-1\}.$$

Conversely, let $\mathcal{V}_R^l$ be the reachable set associated to the truncated functional $H_l^F(\mathbf{z}) := \sum_{j=0}^l A^j \mathbf{C} z_{-j}$, $l \in \mathbb{N}$. It is obvious that $\mathcal{V}_R^{N-1} \subset \text{span}\{\mathbf{C}, A\mathbf{C}, A^2\mathbf{C}, \dots, A^{N-1}\mathbf{C}\}$. We now prove by induction that

$$\mathcal{V}_R^{N+i} \subset \text{span}\{\mathbf{C}, A\mathbf{C}, A^2\mathbf{C}, \dots, A^{N-1}\mathbf{C}\},$$

for all $i \in \mathbb{N}$. First, by the Cayley-Hamilton Theorem [43, Theorem 2.4.3.2] there exist constants $\{\alpha_0, \dots, \alpha_{N-1}\}$ not all zero such that

$$A^N = \alpha_0 \mathbb{I}_N + \alpha_1 A + \dots + \alpha_{N-1} A^{N-1} \quad (32)$$

and hence $H_N^F(\mathbf{z}) := \sum_{i=0}^{N-1} \alpha_i A^i \mathbf{C} z_{-N} + \sum_{j=0}^{N-1} A^j \mathbf{C} z_{-j}$, which shows that $\mathcal{V}_R^N \subset \text{span}\{\mathbf{C}, A\mathbf{C}, A^2\mathbf{C}, \dots, A^{N-1}\mathbf{C}\}$. In order to prove the induction step, suppose that the inequality $\mathcal{V}_R^{N+(i-1)} \subset \text{span}\{\mathbf{C}, A\mathbf{C}, A^2\mathbf{C}, \dots, A^{N-1}\mathbf{C}\}$ holds for a certain $i \in \mathbb{N}$. Again, using (32), we have that

$$\begin{aligned} H_{N+i}^F(\mathbf{z}) &= A^N A^i \mathbf{C} z_{-(N+i)} + \sum_{j=0}^{N+(i-1)} A^j \mathbf{C} z_{-j} \\ &= \sum_{l=0}^{N-1} \alpha_l A^{l+i} \mathbf{C} z_{-(N+i)} + \sum_{j=0}^{N+(i-1)} A^j \mathbf{C} z_{-j} \\ &= \sum_{j=0}^{i-1} A^j \mathbf{C} z_{-j} + \sum_{j=i}^{N+i-1} A^j \mathbf{C} (z_{-j} + \alpha_{j-i} z_{-(N+i)}) = H_{N+(i-1)}^F(\bar{\mathbf{z}}), \end{aligned}$$

with

$$\begin{aligned} \bar{\mathbf{z}} &= (\dots, z_{-(N+i)}, z_{-(N+i-1)} + \alpha_{N-1} z_{-(N+i)}, \dots, \\ &\quad z_{-i-1} + \alpha_1 z_{-(N+i)}, z_{-i} + \alpha_0 z_{-(N+i)}, z_{-(i-1)}, \dots, z_{-1}, z_0), \end{aligned}$$

which shows that $\mathcal{V}_R^{N+i} \subset \mathcal{V}_R^{N+(i-1)} \subset \text{span}\{\mathbf{C}, A\mathbf{C}, A^2\mathbf{C}, \dots, A^{N-1}\mathbf{C}\}$ and hence proves the induction step. This inclusion also implies that

$$\mathcal{V}_R = \bigcup_{l=0}^{\infty} \mathcal{V}_R^l \subset \text{span}\{\mathbf{C}, A\mathbf{C}, A^2\mathbf{C}, \dots, A^{N-1}\mathbf{C}\},$$

as required.

- (iv) Let $\mathbf{x}^1, \mathbf{x}^2 \in \mathcal{V}_R$ be two indistinguishable states of $(\mathcal{V}, F, h)$. By definition, this implies that there exist $\mathbf{z}^1, \mathbf{z}^2 \in \ell^\infty(\mathbb{R})$ such that $\mathbf{x}^1 = H^F(\mathbf{z}^1)$, $\mathbf{x}^2 = H^F(\mathbf{z}^2)$, and that for any $\tilde{\mathbf{z}} \in \mathbb{R}^T$ and any $T \in \mathbb{N}$ we have that $H_h^F(\mathbf{z}^1 \tilde{\mathbf{z}}) = H_h^F(\mathbf{z}^2 \tilde{\mathbf{z}})$. By (24), this is equivalent to

$$W \left(\sum_{j=0}^{\infty} A^{j+T} \mathbf{C} z_{-j}^1 + \sum_{j=0}^{T-1} A^j \mathbf{C} \tilde{z}_{T-j} \right) = W \left(\sum_{j=0}^{\infty} A^{j+T} \mathbf{C} z_{-j}^2 + \sum_{j=0}^{T-1} A^j \mathbf{C} \tilde{z}_{T-j} \right),$$

which amounts to

$$W A^T \sum_{j=0}^{\infty} A^j \mathbf{C} z_{-j}^1 = W A^T \sum_{j=0}^{\infty} A^j \mathbf{C} z_{-j}^2,$$

and is in turn equivalent to the relation $W A^T (\mathbf{x}^1 - \mathbf{x}^2) = 0$, for all $T \in \mathbb{N}$ or, analogously, to $\mathbf{x}^1 - \mathbf{x}^2 \in \bigcap_{j=0}^{\infty} \ker W A^j$. In order to conclude the proof, it hence suffices to show that

$$\bigcap_{j=0}^{\infty} \ker W A^j = \bigcap_{j=0}^{N-1} \ker W A^j.$$

The inclusion $\bigcap_{j=0}^{\infty} \ker W A^j \subseteq \bigcap_{j=0}^{N-1} \ker W A^j$ is obvious. Conversely, we show by induction that

$$\bigcap_{j=0}^{N-1} \ker W A^j \subseteq \bigcap_{j=0}^{N-1+i} \ker W A^j \quad \text{for all } i \in \mathbb{N}. \quad (33)$$

The initialization step is proved using the Cayley-Hamilton Theorem as formulated in (32). Indeed:

$$\bigcap_{j=0}^N \ker W A^j = \ker \left(\sum_{j=0}^{N-1} \alpha_j W A^j \right) \cap \left(\bigcap_{j=0}^{N-1} \ker W A^j \right),$$

which obviously implies that $\bigcap_{j=0}^{N-1} \ker W A^j \subseteq \bigcap_{j=0}^N \ker W A^j$. In order to prove the induction step, suppose that (33) holds for a given $i \in \mathbb{N}$. Given that

$$\bigcap_{j=0}^{N+i} \ker W A^j = \left(\bigcap_{j=0}^{N-1+i} \ker W A^j \right) \cap \ker W A^{N+i},$$

by the induction hypothesis we just need to show that $\bigcap_{j=0}^{N-1} \ker W A^j \subseteq \ker W A^{N+i}$. This inclusion is easily established by using again the Cayley-Hamilton Theorem, which implies that $\ker W A^{N+i} = \ker \sum_{j=0}^{N-1} \alpha_i W A^{i+j}$. The inclusion then follows from the induction hypothesis.

Finally, the statement (27) follows in a straightforward manner from observing that:

$$\ker O(A, W) = I_{F,h}.$$

□

Before we use Theorem 3.2 in order to show that fading memory linear filters admit a linear canonical state-space realization, we motivate that result with an elementary example that hints how such construction may be obtained.

Example 4.4. Canonical realization of finite-memory linear filters. Consider the finite-memory linear filter

$$U(\mathbf{z})_t = \sum_{j=0}^{-N+1} \Psi_j z_{t+j}, \quad \text{with } \Psi \in \mathbb{R}^N, \mathbf{z} \in \ell_-^\infty(\mathbb{R}), t \in \mathbb{Z}_-,$$

and some $N \in \mathbb{N}$. Using the definition of the Nerode equivalence $\sim_I$ on the input space introduced in the proof of Theorem 3.2, it is easy to see that $\mathbf{z}^1, \mathbf{z}^2 \in \ell_-^\infty(\mathbb{R})$ are such that $\mathbf{z}^1 \sim_I \mathbf{z}^2$ if and only if $(z_{-N+1}^1, \dots, z_0^1) = (z_{-N+1}^2, \dots, z_0^2)$ and hence $\ell_-^\infty(\mathbb{R})/\sim_I$ can be identified in this case with $\mathbb{R}^N$ via the map

$$\begin{aligned} \phi: \ell_-^\infty(\mathbb{R})/\sim_I &\longrightarrow \mathbb{R}^N \\ [\mathbf{z}] &\longmapsto (z_{-N+1}, \dots, z_0). \end{aligned}$$

With this identification, $\ell_-^\infty(\mathbb{R})/\sim_I$ inherits the vector space structure of $\ell_-^\infty(\mathbb{R})$ and, moreover, the canonical state-space realization (12) introduced in Theorem 3.2 is given by

$$\begin{cases} F((z_{-N+1}, \dots, z_0), \tilde{z}) = (z_{-N+2}, \dots, z_0, \tilde{z}), \\ h((z_{-N+1}, \dots, z_0)) = \sum_{j=0}^{-N+1} \Psi_j z_j, \end{cases}$$

or, in matrix form:

$$\begin{cases} F((z_{-N+1}, \dots, z_0), \tilde{z}) = \underbrace{\begin{pmatrix} 0 & 1 & & 0 \\ & \ddots & \ddots & \\ 0 & & \ddots & 1 \\ & & & 0 \end{pmatrix}}_A \begin{pmatrix} z_{-N+1} \\ z_{-N+2} \\ \vdots \\ z_0 \end{pmatrix} + \underbrace{\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}}_C \tilde{z}, \\ h((z_{-N+1}, \dots, z_0)) = (\Psi_{-N+1}, \dots, \Psi_0) \begin{pmatrix} z_{-N+1} \\ z_{-N+2} \\ \vdots \\ z_0 \end{pmatrix}. \end{cases} \quad (34)$$

By Theorem 3.2, this realization of U is canonical. An observation that will be key in the next result is that the nilpotent matrix A in (34) is the projection onto the quotient space $\ell_-^\infty(\mathbb{R})/\sim_I$ of the time delay operator $T_{-1}: \ell_-^\infty(\mathbb{R}) \rightarrow \ell_-^\infty(\mathbb{R})$ and that the input vector $\mathbf{C}$ is a matrix expression for the projected version of the inclusion

$$\begin{aligned} i_0: \mathbb{R} &\hookrightarrow \ell_-^\infty(\mathbb{R}) \\ z &\longmapsto (\dots, 0, z). \end{aligned} \quad (35)$$

Theorem 4.5 (Canonical realization of linear fading memory filters). *Let $U: \ell_-^\infty(\mathbb{R}) \rightarrow \ell_-^\infty(\mathbb{R})$ be a linear, causal, and time-invariant filter such that the associated functional $H_U: \ell_-^\infty(\mathbb{R}) \rightarrow \mathbb{R}$ has the fading memory property. Then:*

- (i): *The quotient space $\mathcal{V} := \ell_-^\infty(\mathbb{R})/\sim_I$ has a natural vector space structure inherited from $\ell_-^\infty(\mathbb{R})$. Let $\pi_{\sim_I}: \ell_-^\infty(\mathbb{R}) \rightarrow \mathcal{V} := \ell_-^\infty(\mathbb{R})/\sim_I$ be the canonical projection. The time delay operator T_{-1} the inclusion in (35), and the functional H_U can be projected to the linear maps $\mathcal{A} := [T_{-1}] \in L(\mathcal{V}, \mathcal{V})$,*

$\mathcal{C} := [i_0] \in L(\mathbb{R}, \mathcal{V})$, and $\mathcal{W} := [H_U] \in L(\mathcal{V}, \mathbb{R})$, uniquely determined by the equalities:

$$\mathcal{A} \circ \pi_{\sim_I} = \pi_{\sim_I} \circ T_{-1}, \quad \mathcal{C} = \pi_{\sim_I} \circ i_0, \quad \text{and} \quad \mathcal{W} \circ \pi_{\sim_I} = H_U. \quad (36)$$

The state-space system $(\mathcal{V}, \mathcal{F}, \mathcal{h})$ with $\mathcal{F}(\mathbf{v}, z) := \mathcal{A}(\mathbf{v}) + \mathcal{C}(z)$ and $\mathcal{h}(\mathbf{v}) := \mathcal{W}(\mathbf{v})$ is a canonical linear realization of U .

(ii): Consider the action of the group $\text{GL}(\mathcal{V})$ of all the linear automorphisms of $\mathcal{V}$ and its action Φ on the product $(L(\mathcal{V}, \mathcal{V}) \times L(\mathbb{R}, \mathcal{V}) \times L(\mathcal{V}, \mathbb{R}))$ via the map

$$\begin{aligned} \Phi : \quad \text{GL}(\mathcal{V}) \times (L(\mathcal{V}, \mathcal{V}) \times L(\mathbb{R}, \mathcal{V}) \times L(\mathcal{V}, \mathbb{R})) &\longrightarrow (L(\mathcal{V}, \mathcal{V}) \times L(\mathbb{R}, \mathcal{V}) \times L(\mathcal{V}, \mathbb{R})) \\ (B, (\mathcal{A}, \mathcal{C}, \mathcal{W})) &\longmapsto (BAB^{-1}, BC, \mathcal{W}B^{-1}). \end{aligned}$$

All the canonical representations of U are given by the orbit of the triple $(\mathcal{A}, \mathcal{C}, \mathcal{W})$ introduced in part (i). The space of canonical representations is isomorphic to the homogeneous manifold $\text{GL}(\mathcal{V})/\text{GL}(\mathcal{V})_{(\mathcal{A}, \mathcal{C}, \mathcal{W})}$, with $\text{GL}(\mathcal{V})_{(\mathcal{A}, \mathcal{C}, \mathcal{W})}$ the isotropy subgroup of the element $(\mathcal{A}, \mathcal{C}, \mathcal{W})$.

(iii): If the canonical realization in (ii) is finite-dimensional, then there exists $N \in \mathbb{N}$ such that $\mathcal{V} \simeq \mathbb{R}^N$, where this isomorphism is implemented by a choice of basis in $\mathcal{V}$. There are also matrices $A \in \mathbb{M}_N$, $\mathbf{C} \in \mathbb{R}^N$, $W \in \mathbb{M}_{1,N}$ that express in that basis $\mathcal{A}$, $\mathcal{C}$, and $\mathcal{W}$, respectively. Let $(\mathbb{R}^N, F, h)$ be the system corresponding to $(\mathcal{V}, \mathcal{F}, \mathcal{h})$ in that basis. Then:

(a): $\rho(A) < 1$.

(b): The set of reachable states of $(\mathbb{R}^N, F, h)$ coincides with

$$\mathbb{R}^N = \text{span} \{ \mathbf{C}, A\mathbf{C}, A^2\mathbf{C}, \dots, A^{N-1}\mathbf{C} \}.$$

(c): $I_{F,h} := \bigcap_{i=1}^{N-1} \ker WA^i = \{ \mathbf{0} \}$.

(d): $U(\mathbf{z})_t = \sum_{j=0}^{\infty} WA^j \mathbf{C} z_{t-j}$, with $\mathbf{z} \in \ell_{-}^{\infty}(\mathbb{R})$, for all $t \in \mathbb{Z}_{-}$.

(e): Let $\Psi \in \ell_{-}^1(\mathbb{R})$ be the unique element such that $U(\mathbf{z}) = \Psi * \mathbf{z}$ for any $\mathbf{z} \in \ell_{-}^{\infty}(\mathbb{R})$. Then, $\Psi_{-j} = WA^j \mathbf{C}$, for any $j \in \mathbb{N}$.

Proof. (i) Since by Proposition 4.1 the fading memory property implies the input forgetting property, any linear filter that satisfies the hypotheses in the statement satisfies too those in Theorem 3.2 and consequently has a unique (up to system isomorphism) canonical state-space realization. We shall now study the realization introduced in the proof of that theorem and shall also see that it has the linear form stated in part (ii). First of all, recall that by Proposition 4.1 there exists a unique element $\Psi \in \ell_{-}^1(\mathbb{R})$ such that $U(\mathbf{z}) = \Psi * \mathbf{z}$, for any $\mathbf{z} \in \ell_{-}^{\infty}(\mathbb{R})$. Using this convolution representation and the properties of infinite series it is obvious to prove that if $\mathbf{z}^1 \sim_I \bar{\mathbf{z}}^1$ and $\mathbf{z}^2 \sim_I \bar{\mathbf{z}}^2$, then for any $\lambda \in \mathbb{R}$ we have that $\lambda \mathbf{z}^1 + \mathbf{z}^2 \sim_I \lambda \bar{\mathbf{z}}^1 + \bar{\mathbf{z}}^2$. This implies that the sum and multiplication by scalars in $\ell_{-}^{\infty}(\mathbb{R})$ drop to the quotient space $\mathcal{V} := \ell_{-}^{\infty}(\mathbb{R}) / \sim_I$, making it into a vector space.

Also, using the convolution representation of U it is easy to prove that both the time delay operator T_{-1} , the inclusion in (35), and the functional H_U can be naturally projected to the linear maps in (36) which can be used to rewrite the canonical realization proposed in (12) as

$$F([\mathbf{z}], \tilde{z}) := [\mathbf{z}\tilde{z}] = \pi_{\sim_I}(T_{-1}(\mathbf{z}) + i_0(\tilde{z})) = \mathcal{A}([\mathbf{z}]) + \mathcal{C}(\tilde{z}) \text{ and } h([\mathbf{z}]) := H_U(\mathbf{z}) = \mathcal{W}([\mathbf{z}]),$$

as required.

(ii) is a consequence of Corollary 2.7 and the equalities (3)-(4). Finally, (iii) is a corollary of the characterization in Proposition 4.2. $\square$

In the previous theorem we showed that as a Corollary of the Canonical Realization Theorem 3.2, any fading memory linear filter admits a canonical linear

state-space realization. We now show that the Canonicalization by Reduction Theorem 3.4 implies that any linear state-space system that has the echo state property and the fading memory property can be reduced to a canonical system that is also linear and has the same linear filter associated. The proof is a straightforward consequence of Theorem 3.4 and of Proposition 4.2.

Theorem 4.6 (Canonicalization by reduction of linear state-space systems). *Let $(\mathbb{R}^N, F, h)$ be the linear system determined by the maps $F(\mathbf{x}, z) := A\mathbf{x} + \mathbf{C}z$ and $h(\mathbf{x}) := W\mathbf{x}$, with $A \in \mathbb{M}_N$ such that $\rho(A) < 1$, $\mathbf{C} \in \mathbb{R}^N$, $W \in \mathbb{M}_{1,N}$, and with inputs $\mathbf{z} \in \ell_-^\infty(\mathbb{R})$. Denote by $U_h^F : \ell_-^\infty(\mathbb{R}) \rightarrow \ell_-^\infty(\mathbb{R})$ the associated linear input forgetting filter given by (24). Let $\mathcal{V}_R \subset \mathbb{R}^N$ and $I_{F,h} \subset \mathbb{R}^N$ be the subspaces defined in (25) and (26), respectively. Then U_h^F has a canonical linear realization $(\mathcal{V}, \bar{F}, \bar{h})$ on the quotient vector space $\mathcal{V} := \mathcal{V}_R / I_{F,h}$ given by the maps:*

$$\bar{F}([\mathbf{x}], z) := [A]([\mathbf{x}]) + [\mathbf{C}]z, \quad (37)$$

$$\bar{h}([\mathbf{x}]) := [W]([\mathbf{x}]), \quad (38)$$

where if $\pi : \mathcal{V}_R \rightarrow \mathcal{V}_R / I_{F,h}$ and $i : \mathcal{V}_R \hookrightarrow \mathbb{R}^N$ are the canonical projection and inclusion, respectively, the linear maps $[A] \in L(\mathcal{V}, \mathcal{V})$, $[W] \in L(\mathcal{V}, \mathbb{R})$, and $[C] \in \mathcal{V}$ in (37)-(38) uniquely determined by the relations

$$[A] \circ \pi = \pi \circ A \circ i, \quad [\mathbf{C}] = \pi(\mathbf{C}), \quad \text{and} \quad [W] \circ \pi = \pi \circ W \circ i.$$

5. Implicit reduction using RKHS. An important drawback of the dimension reduction techniques proposed in the previous sections is the need to compute and characterize various reachable sets and quotient spaces, which may be complicated and hence may reduce the practical value of the results that we propose. A situation where these problems may be circumvented is the case is when the readout $h : \mathcal{X} \rightarrow \mathcal{Y}$ in the observation equation (2) is linear. This situation is practically relevant since various state-space systems that satisfy this condition have been shown to exhibit universal approximation properties. It is the case, for instance, of state-affine systems [36] and the widely used echo state networks [35, 24].

The way we proceed in that setup consists in associating to any state-space system $F : \mathcal{X} \times \mathcal{Z} \rightarrow \mathcal{X}$ that satisfies the echo state property, a reproducing kernel Hilbert space (RKHS) $\mathbb{H}$ (see, for instance, Chapter 6 in [74] or [87] for a general presentation of kernel methods) using the state functional $H^F : \mathcal{Z}^{\mathbb{Z}^-} \rightarrow \mathcal{X}$ as a feature map. We shall then show that when the state space $\mathcal{X}$ is a finite dimensional Hilbert space, then $\mathbb{H}$ is isometrically isomorphic to the linear span $\overline{\mathcal{X}_R}$ given by

$$\overline{\mathcal{X}_R} := \text{span} \{ \mathcal{X}_R \} = \text{span} \{ H^F(\mathbf{z}) \mid \mathbf{z} \in \mathcal{Z}^{\mathbb{Z}^-} \} \quad (39)$$

of the set of reachable states $\mathcal{X}_R$.

The importance of this characterization is in the fact that it allows us to show, using the classical Representer Theorem [74, page 117], that the search for an optimal readout with respect to the regularized empirical risk minimization associated to any loss can be reduced to the search for a readout defined on the smaller space $\overline{\mathcal{X}_R}$ without having to actually compute it. We call this procedure **implicit reduction**.

The RKHS associated to a state system. Let $F : \mathcal{X} \times \mathcal{Z} \rightarrow \mathcal{X}$ be a state equation such that the pair $(\mathcal{X}, \langle \cdot, \cdot \rangle_{\mathcal{X}})$ is a finite dimensional Hilbert space and F has the echo state property. Let $H^F : \mathcal{Z}^{\mathbb{Z}^-} \rightarrow \mathcal{X}$ be the corresponding state functional.

Define the *kernel map*

$$\begin{aligned} K : \mathcal{Z}^{\mathbb{Z}-} \times \mathcal{Z}^{\mathbb{Z}-} &\longrightarrow \mathbb{R} \\ (\mathbf{z}, \mathbf{z}') &\longmapsto \langle H^F(\mathbf{z}), H^F(\mathbf{z}') \rangle_{\mathcal{X}}. \end{aligned} \quad (40)$$

The map K is obviously symmetric and positive semidefinite in the sense that for any $a_i \in \mathbb{R}$, $\mathbf{z}_i \in \mathcal{Z}^{\mathbb{Z}-}$, $i \in \{1, \dots, n\}$, we have that $\sum_{i,j=1}^n a_i a_j K(\mathbf{z}_i, \mathbf{z}_j) \geq 0$. Let $(\mathbb{H}, \langle \cdot, \cdot \rangle_{\mathbb{H}})$ be the corresponding RKHS given by

$$\mathbb{H} := \overline{\text{span} \{K_{\mathbf{z}} := K(\mathbf{z}, \cdot) : \mathcal{Z}^{\mathbb{Z}-} \longrightarrow \mathbb{R} \mid \mathbf{z} \in \mathcal{Z}^{\mathbb{Z}-}\}} \quad (41)$$

made out of finite linear combinations of elements of the type $K_{\mathbf{z}}$, $\mathbf{z} \in \mathcal{Z}^{\mathbb{Z}-}$, together with all the limits of Cauchy sequences with respect to the metric induced by the inner product obtained as the linear extension of

$$\langle K_{\mathbf{z}}, K_{\mathbf{z}'} \rangle_{\mathbb{H}} = K(\mathbf{z}, \mathbf{z}') = \langle H^F(\mathbf{z}), H^F(\mathbf{z}') \rangle_{\mathcal{X}}, \quad \mathbf{z}, \mathbf{z}' \in \mathcal{Z}^{\mathbb{Z}-}. \quad (42)$$

Note that in this setup, the reservoir functional H^F with respect to the kernel K and the elements in $\mathbb{H}$ can be written as $K_{\mathbf{z}}(\cdot) = \langle H^F(\mathbf{z}), H^F(\cdot) \rangle_{\mathcal{X}}$.

Proposition 5.1. *Let $(\mathcal{X}, \langle \cdot, \cdot \rangle_{\mathcal{X}})$ be a finite dimensional Hilbert space and let $F : \mathcal{X} \times \mathcal{Z} \longrightarrow \mathcal{X}$ be a state equation that satisfies the echo state property. Let $(\mathbb{H}, \langle \cdot, \cdot \rangle_{\mathbb{H}})$ be the associated RKHS introduced in (41). Then*

$$\mathbb{H} = \{ \langle \mathbf{W}, H^F(\cdot) \rangle_{\mathcal{X}} \mid \mathbf{W} \in \overline{\mathcal{X}_R} \}. \quad (43)$$

Moreover, for any $\mathbf{W}_1, \mathbf{W}_2 \in \overline{\mathcal{X}_R}$, we have that

$$\langle \langle \mathbf{W}_1, H^F(\cdot) \rangle_{\mathcal{X}}, \langle \mathbf{W}_2, H^F(\cdot) \rangle_{\mathcal{X}} \rangle_{\mathbb{H}} = \langle \mathbf{W}_1, \mathbf{W}_2 \rangle_{\mathcal{X}}, \quad (44)$$

and the map

$$\begin{aligned} \Psi : (\overline{\mathcal{X}_R}, \langle \cdot, \cdot \rangle_{\mathcal{X}}) &\longrightarrow (\mathbb{H}, \langle \cdot, \cdot \rangle_{\mathbb{H}}) \\ \mathbf{W} &\longmapsto \langle \mathbf{W}, H^F(\cdot) \rangle_{\mathcal{X}} =: H_{\mathbf{W}}^F(\cdot) \end{aligned} \quad (45)$$

is an isometric isomorphism.

Proof. We first establish the identity (43) by double inclusion. In order to show that $\mathcal{H} \subset \{ \langle \mathbf{W}, H^F(\cdot) \rangle_{\mathcal{X}} \mid \mathbf{W} \in \overline{\mathcal{X}_R} \}$ consider the element

$$f = \sum_{i=1}^n a_i K_{\mathbf{z}_i} \in \mathbb{H}, \quad \text{for some } a_1, \dots, a_n \in \mathbb{R}. \quad (46)$$

Then $f(\cdot) = \sum_{i=1}^n \langle a_i H^F(\mathbf{z}_i), H^F(\cdot) \rangle_{\mathcal{X}}$. Hence, it is clear that if we set $\mathbf{W} := \sum_{i=1}^n a_i H^F(\mathbf{z}_i)$ we can then obviously write that $f(\cdot) = \langle \mathbf{W}, H^F(\cdot) \rangle_{\mathcal{X}}$, as required. More generally, what we just showed also proves that for any sequence $\{f_n\}_{n \in \mathbb{N}}$ of elements like (46) there are elements $\mathbf{W}_n \in \overline{\mathcal{X}_R}$ such that $f_n(\cdot) = \langle \mathbf{W}_n, H^F(\cdot) \rangle_{\mathcal{X}}$. If we assume that $\{f_n\}_{n \in \mathbb{N}}$ is Cauchy then $\|f_n - f_m\|_{\mathbb{H}} \rightarrow 0$ as $n, m \rightarrow \infty$. This in turn implies that for any $\mathbf{z} \in \mathcal{Z}^{\mathbb{Z}-}$ we have that

$$|f_n(\mathbf{z}) - f_m(\mathbf{z})| = |\langle K_{\mathbf{z}}, f_n - f_m \rangle_{\mathbb{H}}| \leq \|f_n - f_m\|_{\mathbb{H}} \|K_{\mathbf{z}}\|_{\mathbb{H}} \longrightarrow 0 \quad \text{as } n, m \rightarrow \infty,$$

which guarantees that $|\langle \mathbf{W}_n - \mathbf{W}_m, H^F(\mathbf{z}) \rangle_{\mathcal{X}}| \rightarrow 0$ and hence that $|\langle \mathbf{W}_n - \mathbf{W}_m, \mathbf{v} \rangle_{\mathcal{X}}| \rightarrow 0$ as $n, m \rightarrow \infty$, for any $\mathbf{v} \in \overline{\mathcal{X}_R}$. Now, since any vector $\mathbf{x} \in \mathcal{X}$ can be uniquely decomposed as $\mathbf{x} = \mathbf{v} + \mathbf{v}^{\perp}$ with $\mathbf{v} \in \overline{\mathcal{X}_R}$ and $\mathbf{v}^{\perp} \in \overline{\mathcal{X}_R}^{\perp}$, we also have that

$$|\langle \mathbf{W}_n - \mathbf{W}_m, \mathbf{x} \rangle_{\mathcal{X}}| = |\langle \mathbf{W}_n - \mathbf{W}_m, \mathbf{v} \rangle_{\mathcal{X}}| \rightarrow 0 \quad \text{as } n, m \rightarrow \infty, \text{ for any } \mathbf{x} \in \overline{\mathcal{X}_R}.$$

Given that $\mathcal{X}$ is finite dimensional, we can conclude that weak and strong convergence coincide and hence that $\|\mathbf{W}_n - \mathbf{W}_m\|_{\mathcal{X}} \rightarrow 0$ as $n, m \rightarrow \infty$. Since $\mathcal{X}$ is

complete then so is $\overline{\mathcal{X}_R}$ and hence there exists $\mathbf{W} \in \overline{\mathcal{X}_R}$ such that $\lim_{n \rightarrow \infty} \mathbf{W}_n = \mathbf{W}$. It is easy to see that this implies that

$$f(\cdot) := \lim_{n \rightarrow \infty} f_n(\cdot) = \langle \mathbf{W}, H^F(\cdot) \rangle_{\mathcal{X}},$$

as required. In order to prove the converse inclusion, note first that by definition, for any $\mathbf{W} \in \overline{\mathcal{X}_R}$ there exist $\mathbf{z}_1, \dots, \mathbf{z}_n \in \mathcal{Z}^{\mathbb{Z}^-}$ and $a_1, \dots, a_n \in \mathbb{R}$ such that $\mathbf{W} = \sum_{i=1}^n a_i H^F(\mathbf{z}_i)$. It is hence easy to see that

$$\langle \mathbf{W}, H^F(\cdot) \rangle_{\mathcal{X}} = \sum_{i=1}^n a_i \langle H^F(\mathbf{z}_i), H^F(\cdot) \rangle_{\mathcal{X}} = \sum_{i=1}^n a_i K_{\mathbf{z}_i}(\cdot),$$

which is an element in $\mathbb{H}$, as required.

We now show the identity (44). Let $\mathbf{W}_1, \mathbf{W}_2 \in \overline{\mathcal{X}_R}$ and let $\mathbf{W}_1 = \sum_{i=1}^{n_1} a_i^1 H^F(\mathbf{z}_i^1)$ and $\mathbf{W}_2 = \sum_{j=1}^{n_2} a_j^2 H^F(\mathbf{z}_j^2)$ two representations of the two vectors according to the definition of $\overline{\mathcal{X}_R}$. Then, it is easy to see that

$$\begin{aligned} & \langle \langle \mathbf{W}_1, H^F(\cdot) \rangle_{\mathcal{X}}, \langle \mathbf{W}_2, H^F(\cdot) \rangle_{\mathcal{X}} \rangle_{\mathbb{H}} \\ &= \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} a_i^1 a_j^2 \langle \langle H^F(\mathbf{z}_i^1), H^F(\cdot) \rangle_{\mathcal{X}}, \langle H^F(\mathbf{z}_j^2), H^F(\cdot) \rangle_{\mathcal{X}} \rangle_{\mathbb{H}} \\ &= \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} a_i^1 a_j^2 \langle K_{\mathbf{z}_i^1}, K_{\mathbf{z}_j^2} \rangle_{\mathbb{H}} = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} a_i^1 a_j^2 \langle H^F(\mathbf{z}_i^1), H^F(\mathbf{z}_j^2) \rangle_{\mathbb{X}} = \langle \mathbf{W}_1, \mathbf{W}_2 \rangle_{\mathcal{X}}. \end{aligned}$$

Finally, we show that the map Ψ in (45) is an isometric isomorphism. First, it is clear that the map is linear, the equality (43) guarantees that Ψ is onto, and (44) that it is an isometry. In order to show injectivity, suppose that $\mathbf{W} \in \overline{\mathcal{X}_R}$ is such that $\Psi(\mathbf{W})(\cdot) = \langle \mathbf{W}, H^F(\cdot) \rangle_{\mathcal{X}} = 0$. If we use a representation for $\mathbf{W}$ of the type $\mathbf{W} = \sum_{i=1}^n a_i H^F(\mathbf{z}_i)$ we can write that

$$\langle \mathbf{W}, \mathbf{W} \rangle_{\mathcal{X}} = \left\langle \mathbf{W}, \sum_{i=1}^n a_i H^F(\mathbf{z}_i) \right\rangle_{\mathcal{X}} = \sum_{i=1}^n a_i \Psi(\mathbf{W})(\mathbf{z}_i) = 0,$$

which guarantees that $\mathbf{W} = 0$, as required. $\square$

Estimation of the empirical risk minimizing readout. A common estimation problem that appears when using in practice systems of the form (1)-(2) and where the readout is linear is finding the readout vector $\mathbf{W} \in \mathcal{X}$ that minimizes the empirical risk associated to a prescribed loss function $L : \mathcal{Y} \times \mathcal{Y} \rightarrow \mathbb{R}$ with respect to a finite sample of input/output observations. This is typically how one proceeds in reservoir computing (see the introduction section) where the state equation is fixed and only a linear observation equation is subjected to training. In that particular case and if a quadratic loss is used, the estimation problem reduces itself to a (eventually regularized) regression problem with as many covariates as the dimension of the state space $\mathcal{X}$, which is in most cases very large. It is in this context that for quadratic or more general losses, the possibility of reducing the dimensionality of the estimation problem to the dimension of $\overline{\mathcal{X}_R}$ using the RKHS technology that we just introduced may prove computationally advantageous.

To be more specific, in the next proposition we will show two main facts as a consequence of the RKHS formulation of the estimation problem. First, that even though the optimization problem that provides the optimal readout is originally formulated in the space $\mathcal{X}$, it can be reduced to the dimensionally smaller $\overline{\mathcal{X}_R}$.

Second, the Representer Theorem [74, page 117] shows that *the optimal readout is in the “span of the data”*; this is the well-known “kernelization trick” that in our case is computationally relevant in the presence of state spaces of dimension larger than the sample size. An important observation is that this second result *yields automatically a solution in the span $\overline{\mathcal{X}_R}$ of the reachable set $\mathcal{X}_R$ without actually having to compute it.*

We now introduce the different elements that are necessary for the statement of the Proposition. First, we will assume that the output space $\mathcal{Y}$ is one-dimensional, the state system $F : \mathcal{X} \times \mathcal{Y} \rightarrow \mathcal{X}$ is fixed and satisfies the ESP, and we are provided with a finite sample $\{(\mathbf{Z}_{-i}, Y_{-i})\}_{i \in \{0, \dots, n-1\}}$ of size n of input/output observations. For each time step $i \in \{0, \dots, n-1\}$ we define the truncated training sample for the input stochastic process $\mathbf{Z}$ as

$$\mathbf{Z}_{-i}^{-n+1} := (\dots, \mathbf{0}, \mathbf{0}, \mathbf{Z}_{-n+1}, \dots, \mathbf{Z}_{-i-1}, \mathbf{Z}_{-i}),$$

that we use to define the **training error** or the **empirical risk** $\hat{R}_n(H_{\mathbf{W}}^F)$ associated to the loss $L : \mathcal{Y} \times \mathcal{Y} \rightarrow \mathbb{R}$ for the system $H_{\mathbf{W}}^F(\cdot) = \langle \mathbf{W}, H^F(\cdot) \rangle_{\mathcal{X}}$ with readout vector $\mathbf{W} \in \mathcal{X}$ as

$$\hat{R}_n(H_{\mathbf{W}}^F) = \frac{1}{n} \sum_{i=0}^{n-1} L(\langle \mathbf{W}, H^F(\mathbf{Z}_{-i}^{-n+1}) \rangle_{\mathcal{X}}, \mathbf{Y}_{-i}).$$

Proposition 5.2. *Let $F : \mathcal{X} \times \mathcal{Y} \rightarrow \mathcal{X}$ be a state system that satisfies the ESP and let $L : \mathcal{Y} \times \mathcal{Y} \rightarrow \mathbb{R}$ be a loss function with respect to the one-dimensional output space $\mathcal{Y}$. Let $\{(\mathbf{Z}_{-i}, Y_{-i})\}_{i \in \{0, \dots, n-1\}}$ be a sample of size n of input/output observations. Let $\Omega : \mathbb{R}^+ \rightarrow \mathbb{R}$ be a strictly increasing function. Then:*

$$\begin{aligned} \min_{\mathbf{W} \in \mathcal{X}} \left\{ \hat{R}_n(H_{\mathbf{W}}^F) + \Omega(\|\mathbf{W}\|_{\mathcal{X}}^2) \right\} &= \min_{\mathbf{W} \in \overline{\mathcal{X}_R}} \left\{ \hat{R}_n(H_{\mathbf{W}}^F) + \Omega(\|\mathbf{W}\|_{\mathcal{X}}^2) \right\} \quad (47) \\ &= \min_{H_{\mathbf{W}}^F \in \mathbb{H}} \left\{ \hat{R}_n(H_{\mathbf{W}}^F) + \Omega(\|H_{\mathbf{W}}^F\|_{\mathbb{H}}^2) \right\}, \quad (48) \end{aligned}$$

where $\mathbb{H}$ is the RKHS introduced in (41). The minimum in (48) is realized by an element in $\mathbb{H}$ of the form

$$\sum_{i=0}^{n-1} \alpha_i K_{\mathbf{Z}_{-i}^{-n+1}} = \left\langle \sum_{i=0}^{n-1} \alpha_i H^F(\mathbf{Z}_{-i}^{-n+1}), H^F(\cdot) \right\rangle_{\mathcal{X}}, \quad \text{for some } \alpha_0, \alpha_1, \dots, \alpha_{n-1} \in \mathbb{R}. \quad (49)$$

The element $\sum_{i=0}^{n-1} \alpha_i H^F(\mathbf{Z}_{-i}^{-n+1}) \in \overline{\mathcal{X}_R}$ is the minimizer of the terms in (47).

Proof. Given that any $\mathbf{W} \in \mathcal{X}$ can be uniquely decomposed as $\mathbf{W} = \mathbf{W}_R + \mathbf{W}_R^\perp$ with $\mathbf{W}_R \in \mathcal{X}_R$ and $\mathbf{W}_R^\perp \in \mathcal{X}_R^\perp$, we can write that

$$\begin{aligned} \hat{R}_n(H_{\mathbf{W}}^F) + \Omega(\|\mathbf{W}\|_{\mathcal{X}}^2) &= \hat{R}_n(H_{\mathbf{W}_R}^F) + \Omega(\|\mathbf{W}_R\|_{\mathcal{X}}^2 + \|\mathbf{W}_R^\perp\|_{\mathcal{X}}^2) \\ &\geq \hat{R}_n(H_{\mathbf{W}_R}^F) + \Omega(\|\mathbf{W}_R\|_{\mathcal{X}}^2), \end{aligned}$$

where in the last inequality we used that, by hypothesis, Ω is strictly increasing. This inequality implies that

$$\min_{\mathbf{W} \in \mathcal{X}} \left\{ \hat{R}_n(H_{\mathbf{W}}^F) + \Omega(\|\mathbf{W}\|_{\mathcal{X}}^2) \right\} \geq \min_{\mathbf{W} \in \overline{\mathcal{X}_R}} \left\{ \hat{R}_n(H_{\mathbf{W}}^F) + \Omega(\|\mathbf{W}\|_{\mathcal{X}}^2) \right\}.$$

However, given that $\overline{\mathcal{X}_R} \subseteq \mathcal{X}$ the converse inequality also obviously holds, which proves the equality (47). The relation (48) is a consequence of (43) and also of the

fact that by (44)

$$\|H_F^{\mathbf{W}}\|_{\mathbb{H}}^2 = \langle \langle \mathbf{W}, H^F(\cdot) \rangle_{\mathcal{X}}, \langle \mathbf{W}, H^F(\cdot) \rangle_{\mathcal{X}} \rangle_{\mathbb{H}} = \|\mathbf{W}\|_{\mathcal{X}}^2.$$

Finally, the statement (49) is a straightforward consequence of the Representer Theorem [74, page 117]. $\square$

6. Conclusions. In this paper we have extended the classical notion of canonical state-space realization to accommodate semi-infinite inputs so that it can be used as a dimension reduction tool in the framework of recurrent networks. We have formulated two main results that identify the so-called input forgetting property (introduced in Definition 3.1) as the key hypothesis that guarantees the existence and uniqueness (up to system isomorphisms) of canonical realizations for causal and time-invariant input/output systems with semi-infinite inputs.

The first result (Theorem 3.2) shows that any causal and time-invariant filter with semi-infinite inputs that has the input forgetting property admits a canonical state-space realization that is unique up to system isomorphisms. The second one (Theorem 3.4) uses a reduction approach similar to the one introduced in [81, 79] in the context of symmetric Hamiltonian systems to construct a canonical realization for a state-space system that has the input forgetting property system by using an “optimally reduced” version of it, in the sense of those references. These two results have been illustrated and applied in detail in Section 4 in the context of linear fading memory filters.

The contributions in this paper should be considered just as a first step in the full understanding of this problem as, in comparison with the classical theory of forward-looking input-driven state-space systems, there are many deficiencies in the level of comprehension of several important mathematical issues. We now list a few of them that are part of our research agenda and that will be studied in forthcoming works:

- ***The geometric nature of reachable sets by semi-infinite inputs*** (see the definition in (16)). Reachable sets are central objects in the context of continuous-time forward looking systems in connection with the notion of controllability (see [89, 61, 7, 5] and references therein). From the geometric viewpoint, this important application question has given rise to the notions of generalized foliation and distribution [93, 94, 95, 58]. Some of these results have a discrete-time counterpart (see, for instance, [18, 49, 1]) but the situation is mostly unknown when it comes to semi-infinite inputs. Some partial information [68] can be obtained by using the recent theory of nonautonomous dynamical systems [57].
- ***The geometric nature of the canonical state spaces obtained by reduction*** (see the definition in Theorem 3.4). Again, in other contexts like the reduction of symmetric Hamiltonian systems or control systems, this is a very well studied question (see [80, 70] for the autonomous case or [98, 76, 5] for the control case). The semi-infinite inputs framework presents new mathematical challenges that need to be addressed with innovative tools.
- ***The geometric nature of the canonical realization state-spaces in Theorem 3.2.*** In the linear case treated in Section 4 we were able to easily pinpoint the vector space structure of the quotient space $\mathcal{X} := \mathcal{Z}^{\mathbb{Z}_-} / \sim_I$ and to comfortably work with it. In more general nonlinear situations it is very difficult to answer even elementary questions (like the dimension) about the

canonical state-space $\mathcal{X}$ even when we impose strong regularity assumptions on the original input space $\mathcal{Z}^{\mathbb{Z}^-}$.

- Even in the linear case, there is, as far as we know, no **readily usable characterization of the situations in which the canonical realizations introduced in Theorem 4.5 are finite dimensional**. Such criterion is necessary for the practical implementation of this result.

Acknowledgments. 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). The authors thank the hospitality and the generosity of the FIM at ETH Zurich and the Division of Mathematical Sciences of the Nanyang Technological University, Singapore, where a significant portion of the results in this paper was obtained.

REFERENCES

- [1] A. C. Antoulas, *Mathematical System Theory. The Influence of R. E. Kalman*, Springer-Verlag, Berlin, 1991.
- [2] P. Barančok and I. Farkas, *Memory capacity of input-driven echo state networks at the edge of chaos*, in *Artificial Neural Networks and Machine Learning – ICANN 2014*, Lecture Notes in Computer Science, 8681, Springer, Cham, 2014, 41–48.
- [3] L. E. Baum and T. Petrie, *Statistical inference for probabilistic functions of finite state Markov chains*, *Ann. Math. Statist.*, **37** (1966), 1554–1563.
- [4] G. Blankenstein and T. S. Ratiu, *Singular reduction of implicit Hamiltonian systems*, *Rep. Math. Phys.*, **53** (2004), 211–260.
- [5] A. M. Bloch, *Nonholonomic Mechanics and Control*, 2nd edition, Interdisciplinary Applied Mathematics, 24, Springer, New York, 2015.
- [6] S. Boyd and L. O. Chua, *Fading memory and the problem of approximating nonlinear operators with Volterra series*, *IEEE Trans. Circuits and Systems*, **32** (1985), 1150–1161.
- [7] F. Bullo and A. D. Lewis, *Geometric Control of Mechanical Systems. Modeling, Analysis, and Design for Simple Mechanical Control Systems*, Applied Mathematics, 49, Springer-Verlag, New York, 2005.
- [8] P. Buteneers, D. Verstraeten, B. V. Nieuwenhuyse, D. Stroobandt and R. Raedt, et al., *Real-time detection of epileptic seizures in animal models using reservoir computing*, *Epilepsy Res.*, **103** (2013), 124–134.
- [9] A. S. Charles, D. Yin and C. J. Rozell, *Distributed sequence memory of multidimensional inputs in recurrent networks*, *J. Mach. Learn. Res.*, **18** (2017), 37pp.
- [10] R. Couillet, G. Wainrib, H. Sevi and H. Tiomoko Ali, *The asymptotic performance of linear echo state neural networks*, *J. Mach. Learn. Res.*, **17** (2016), 35pp.
- [11] C. Cuchiero, L. Gonon, L. Grigoryeva, J.-P. Ortega and J. Teichmann, *Discrete-time signatures and randomness in reservoir computing*, *IEEE Trans. Neural Netw. Learn. Syst.*, (2021), 1–10.
- [12] J. Dambre, D. Verstraeten, B. Schrauwen and S. Massar, *Information processing capacity of dynamical systems*, *Scientific Reports*, **2** (2012).
- [13] H. Dang Van Mien and D. Normand-Cyrot, *Nonlinear state affine identification methods: Applications to electrical power plants*, *Automatica*, **20** (1984), 175–188.
- [14] K. Doya, *Bifurcations in the learning of recurrent neural networks*, *Proc. IEEE Internat. Symposium on Circuits Syst.*, San Diego, CA, 1992.
- [15] M. Duffo, *Random Iterative Models*, Applications of Mathematics (New York), 34, Springer-Verlag, Berlin, 1997.
- [16] J. Durbin and S. J. Koopman, *Time Series Analysis by State Space Methods*, Oxford Statistical Science Series, 38, Oxford University Press, Oxford, 2012.
- [17] I. Farkas, R. Bosák and P. Gergel', *Computational analysis of memory capacity in echo state networks*, *Neural Networks*, **83** (2016), 109–120.
- [18] M. Fliess and D. Normand-Cyrot, *A group-theoretic approach to discrete-time non-linear controllability*, 20th IEEE Conference on Decision and Control including the Symposium on Adaptive Processes, San Diego, CA, 1981.

- [19] S. Ganguli, D. Huh and H. Sompolinsky, [Memory traces in dynamical systems](#), *PNAS*, **105** (2008), 18970–18975.
- [20] F. Gay-Balmaz and T. S. Ratiu, [Clebsch optimal control formulation in mechanics](#), *J. Geom. Mech.*, **3** (2011), 41–79.
- [21] L. Gonon, L. Grigoryeva and J.-P. Ortega, Approximation error estimates for random neural networks and reservoir systems, preprint, [arXiv:2002.05933](#).
- [22] L. Gonon, L. Grigoryeva and J.-P. Ortega, [Memory and forecasting capacities of nonlinear recurrent networks](#), *Phys. D*, **414** (2020), 13pp.
- [23] L. Gonon, L. Grigoryeva and J.-P. Ortega, Risk bounds for reservoir computing, *J. Mach. Learn. Res.*, **21** (2020), 61pp.
- [24] L. Gonon and J.-P. Ortega, [Fading memory echo state networks are universal](#), *Neural Networks*, **138** (2021), 10–13.
- [25] L. Gonon and J.-P. Ortega, [Reservoir computing universality with stochastic inputs](#), *IEEE Trans. Neural Netw. Learn. Syst.*, **31** (2020), 100–112.
- [26] A. Goudarzi, S. Marzen, P. Banda, G. Feldman, M. R. Lakin, C. Teuscher and D. Stefanovic, Memory and information processing in recurrent neural networks, preprint, [arXiv:1604.06929](#).
- [27] A. Graves, A.-R. Mohamed and G. Hinton, [Speech recognition with deep recurrent neural networks](#), IEEE International Conference on Acoustics, Speech and Signal Processing, Vancouver, BC, Canada, 2013.
- [28] L. Grigoryeva, A. Hart and J.-P. Ortega, [Chaos on compact manifolds: Differentiable synchronizations beyond the Takens theorem](#), *Phys. Rev. E*, **103** (2021), 12pp.
- [29] L. Grigoryeva, A. Hart and J.-P. Ortega, Learning strange attractors with reservoir systems, preprint, [arXiv:2108.05024](#).
- [30] L. Grigoryeva, J. Henriques, L. Larger and J.-P. Ortega, [Nonlinear memory capacity of parallel time-delay reservoir computers in the processing of multidimensional signals](#), *Neural Comput.*, **28** (2016), 1411–1451.
- [31] L. Grigoryeva, J. Henriques, L. Larger and J.-P. Ortega, [Optimal nonlinear information processing capacity in delay-based reservoir computers](#), *Scientific Rep.*, **5** (2015), 1–11.
- [32] L. Grigoryeva, J. Henriques, L. Larger and J.-P. Ortega, [Stochastic time series forecasting using time-delay reservoir computers: Performance and universality](#), *Neural Networks*, **55** (2014), 59–71.
- [33] L. Grigoryeva, J. Henriques and J.-P. Ortega, [Reservoir computing: Information processing of stationary signals](#), IEEE International Conference on Computational Science and Engineering (CSE), Paris, France, 2016.
- [34] L. Grigoryeva and J.-P. Ortega, Differentiable reservoir computing, *J. Mach. Learn. Res.*, **20** (2019), 62pp.
- [35] L. Grigoryeva and J.-P. Ortega, [Echo state networks are universal](#), *Neural Networks*, **108** (2018), 495–508.
- [36] L. Grigoryeva and J.-P. Ortega, Universal discrete-time reservoir computers with stochastic inputs and linear readouts using non-homogeneous state-affine systems, *J. Mach. Learn. Res.*, **19** (2018), 40pp.
- [37] J. W. Grizzle and S. I. Marcus, [The structure of nonlinear control systems possessing symmetries](#), *IEEE Trans. Automat. Control*, **30** (1985), 248–258.
- [38] J. Hanson and M. Raginsky, Universal approximation of input-output maps by temporal convolutional nets, preprint, [arXiv:1906.09211](#).
- [39] A. Hart, J. Hook and J. Dawes, [Embedding and approximation theorems for echo state networks](#), *Neural Networks*, **128** (2020), 234–247.
- [40] A. G. Hart, J. L. Hook and J. H. P. Dawes, [Echo state networks trained by Tikhonov least squares are \$L^2\(\mu\)\$ approximators of ergodic dynamical systems](#), *Phys. D*, **421** (2021), 9pp.
- [41] S. Haykin, *Neural Networks and Learning Machines*, Pearson, Addison Wesley, 2009.
- [42] M. Hermans and B. Schrauwen, [Memory in linear recurrent neural networks in continuous time.](#), *Neural Networks*, **23** (2010), 341–355.
- [43] R. A. Horn and C. R. Johnson, *Matrix Analysis*, 2nd edition, Cambridge University Press, Cambridge, 2013.
- [44] G.-B. Huang, Q.-Y. Zhu and C.-K. Siew, [Extreme learning machine: Theory and applications](#), *Neurocomputing*, **70** (2006), 489–501.
- [45] C. E. Hutchinson, [The Kalman filter applied to aerospace and electronic systems](#), *IEEE Trans. Aerospace Electron. Syst.*, **AES-20** (1984), 500–504.

- [46] H. Jaeger, *The “Echo State” Approach to Analysing and Training Recurrent Neural Networks with an Erratum Note*, GMD Report, German National Research Center for Information Technology, 2010.
- [47] H. Jaeger, Short term memory in echo state networks, *Fraunhofer Institute for Autonomous Intelligent Systems*, **152**.
- [48] H. Jaeger and H. Haas, [Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication](#), *Science*, **304** (2004), 78–80.
- [49] B. Jakubczyk and E. D. Sontag, [Controllability of nonlinear discrete-time systems: A Lie-algebraic approach](#), *SIAM J. Control Optim.*, **28** (1990), 1–33.
- [50] W. B. Johnson and J. Lindenstrauss, [Extensions of Lipschitz mappings into a Hilbert space](#), in *Conference in Modern Analysis and Probability*, Contemp. Math., 26, Amer. Math. Soc., Providence, RI, 1984, 189–206.
- [51] R. E. Kalman, [Canonical structure of linear dynamical systems](#), *Proc. Nat. Acad. Sci. U.S.A.*, **48** (1962), 596–600.
- [52] R. E. Kalman, [Lectures on controllability and observability](#), in *Controllability and Observability*, C.I.M.E. Summer Sch., 46, Springer, Heidelberg, 2010, 1–149.
- [53] R. E. Kalman, [A new approach to linear filtering and prediction problems](#), *Trans. ASME Ser. D. J. Basic Engrg.*, **82** (1960), 35–45.
- [54] R. E. Kalman and J. E. Bertram, [General synthesis procedure for computer control of single-loop and multiloop linear systems \(an optimal sampling system\)](#), *Trans. Amer. Inst. Electrical Engineers, Part II: Appl. Industry*, **77** (1959), 602–609.
- [55] R. E. Kalman and J. E. Bertram, [A unified approach to the theory of sampling systems](#), *J. Franklin Inst.*, **267** (1959), 405–436.
- [56] R. E. Kalman and R. S. Bucy, [New results in linear filtering and prediction theory](#), *Trans. ASME Ser. D. J. Basic Engrg.*, **83** (1961), 95–108.
- [57] P. E. Kloeden and M. Rasmussen, *Nonautonomous Dynamical Systems*, Mathematical Surveys and Monographs, 176, American Mathematical Society, Providence, RI, 2011.
- [58] I. Kolář, P. W. Michor and J. Slovák, *Natural Operations in Differential Geometry*, Springer-Verlag, Berlin, 1993.
- [59] B. Kostant, Orbits, symplectic structures and representation theory, in *Proc. U.S.-Japan Seminar in Differential Geometry*, Nippon Hyoronsha, Tokyo, 1966.
- [60] P. D. Lax, *Functional Analysis*, Pure and Applied Mathematics, Wiley-Interscience, New York, 2002.
- [61] A. Lewis, A brief on controllability of nonlinear systems, 2002.
- [62] A. Lindquist and G. Picci, *Linear Stochastic Systems. A Geometric Approach to Modeling, Estimation and Identification*, Series in Contemporary Mathematics, 1, Springer, Heidelberg, 2015.
- [63] L. Livi, F. M. Bianchi and C. Alippi, [Determination of the edge of criticality in echo state networks through Fisher information maximization](#), *IEEE Trans. Neural Netw. Learn. Syst.*, **29** (2018), 706–717.
- [64] Z. Lu, B. R. Hunt and E. Ott, [Attractor reconstruction by machine learning](#), *Chaos*, **28** (2018), 9pp.
- [65] M. Lukoševičius and H. Jaeger, [Reservoir computing approaches to recurrent neural network training](#), *Comput. Sci. Rev.*, **3** (2009), 127–149.
- [66] W. Maass, [Liquid state machines: Motivation, theory, and applications](#), in *Computability in Context*, Imp. Coll. Press, London, 2011, 275–296.
- [67] W. Maass, T. Natschläger and H. Markram, [Real-time computing without stable states: A new framework for neural computation based on perturbations](#), *Neural Comput.*, **14** (2002), 2531–2560.
- [68] G. Manjunath, P. Tiño and H. Jaeger, Theory of input driven dynamical systems, *ESANN 2012 Proceedings, 20th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*, 1–12.
- [69] J. Marsden and A. Weinstein, [Reduction of symplectic manifolds with symmetry](#), *Rep. Mathematical Phys.*, **5** (1974), 121–130.
- [70] J. E. Marsden, G. Misiołek, J.-P. Ortega, M. Perlmutter and T. S. Ratiu, *Hamiltonian Reduction by Stages*, Lecture Notes in Mathematics, 1913, Springer, Berlin, 2007.
- [71] S. Marzen, [Difference between memory and prediction in linear recurrent networks](#), *Phys. Rev. E*, **96** (2017), 1–7.

- [72] M. B. Matthews, *On the Uniform Approximation of Nonlinear Discrete-Time Fading-Memory Systems Using Neural Network Models*, Ph.D thesis, ETH Zürich, 1992. Available from: <https://www.research-collection.ethz.ch:443/handle/20.500.11850/140592>.
- [73] M. B. Matthews and G. S. Moschytz, *The identification of nonlinear discrete-time fading-memory systems using neural network models*, *IEEE Trans. Circuits Syst. II: Analog and Digital Signal Process.*, **41** (1994), 740–751.
- [74] M. Mohri, A. Rostamizadeh and A. Tawalkar, *c*, 2nd edition, Adaptive Computation and Machine Learning, MIT Press, Cambridge, MA, 2018.
- [75] K. S. Narendra and K. Parthasarathy, *Identification and control of dynamical systems using neural networks*, *IEEE Trans. Neural Networks*, **1** (1990), 4–27.
- [76] H. Nijmeijer and A. van der Schaft, *Controlled invariance for nonlinear systems*, *IEEE Trans. Automat. Control*, **27** (1982), 904–914.
- [77] D. Normand-Cyrot, *Théorie et Pratique des Systèmes Non Linéaires en Temps Discret*, Ph.D thesis, Université Paris-Sud, 1983.
- [78] T. Ohsawa, *Symmetry reduction of optimal control systems and principal connections*, *SIAM J. Control Optim.*, **51** (2013), 96–120.
- [79] J.-P. Ortega, *The symplectic reduced spaces of a Poisson action*, *C. R. Math. Acad. Sci. Paris*, **334** (2002), 999–1004.
- [80] J.-P. Ortega and T. S. Ratiu, *Momentum Maps and Hamiltonian Reduction*, Progress in Mathematics, 222, Birkhäuser Boston, Inc., Boston, MA, 2004.
- [81] J.-P. Ortega and T. S. Ratiu, *The optimal momentum map*, in *Geometry, Mechanics, and Dynamics*, Springer, New York, 2002, 329–362.
- [82] R. Pascanu, C. Gulcehre, K. Cho and Y. Bengio, How to construct deep recurrent neural networks, preprint, [arXiv:1312.6026](https://arxiv.org/abs/1312.6026).
- [83] J. Pathak, B. Hunt, M. Girvan, Z. Lu and E. Ott, *Model-free prediction of large spatiotemporally chaotic systems from data: A reservoir computing approach*, *Phys. Rev. Lett.*, **120** (2018).
- [84] J. Pathak, A. Wikner, R. Fussell, S. Chandra, B. R. Hunt, M. Girvan and E. Ott, *Hybrid forecasting of chaotic processes: Using machine learning in conjunction with a knowledge-based model*, *Chaos*, **28** (2018), 9pp.
- [85] A. Rahimi and B. Recht, Random features for large-scale kernel machines, *Advances in Neural Information*. Available from: <http://people.eecs.berkeley.edu/~brecht/papers/07.rah.rec.nips.pdf>.
- [86] S. Särkkä, *Bayesian Filtering and Smoothing*, Institute of Mathematical Statistics Textbooks, 3, Cambridge University Press, Cambridge, 2013.
- [87] B. Schölkopf and A. J. Smola, *Learning with Kernels*, MIT Press, 2002.
- [88] S. Smale, *Topology and mechanics. I.*, *Invent. Math.*, **10** (1970), 305–331.
- [89] E. D. Sontag, *Mathematical Control Theory. Deterministic Finite-Dimensional Systems*, 2nd edition, Texts in Applied Mathematics, 6, Springer-Verlag, New York, 1998.
- [90] E. D. Sontag, *Realization theory of discrete-time nonlinear systems. I. The bounded case*, *IEEE Trans. Circuits and Systems*, **26** (1979), 342–356.
- [91] J.-M. Souriau, Quantification géométrique, *Comm. Math. Phys.*, **1** (1966), 374–398.
- [92] J.-M. Souriau, *Structure des Systèmes Dynamiques*, Dunod, Paris, 1970.
- [93] P. Stefan, *Accessibility and foliations with singularities*, *Bull. Amer. Math. Soc.*, **80** (1974), 1142–1145.
- [94] P. Stefan, *Accessible sets, orbits, and foliations with singularities*, *Proc. London Math. Soc.* (3), **29** (1974), 699–713.
- [95] H. J. Sussmann, *Orbits of families of vector fields and integrability of distributions*, *Trans. Amer. Math. Soc.*, **180** (1973), 171–188.
- [96] P. Tiño, *Asymptotic Fisher memory of randomized linear symmetric Echo State Networks*, *Neurocomputing*, **298** (2018), 4–8.
- [97] P. Tino and A. Rodan, *Short term memory in input-driven linear dynamical systems*, *Neurocomputing*, **112** (2013), 58–63.
- [98] A. van der Schaft, *Symmetries and conservation laws for Hamiltonian systems with inputs and outputs: A generalization of Noether’s theorem*, *Systems Control Lett.*, **1** (1981/82), 108–115.
- [99] A. J. van der Schaft, *Symmetries in optimal control*, *SIAM J. Control Optim.*, **25** (1987), 245–259.

- [100] P. Verzelli, C. Alippi and L. Livi, [Echo state networks with self-normalizing activations on the hyper-sphere](#), *Scientific Rep.*, **9** (2019).
- [101] O. L. White, D. D. Lee and H. Sompolinsky, [Short-term memory in orthogonal neural networks](#), *Phys. Rev. Lett.*, **92** (2004).
- [102] F. Wyffels and B. Schrauwen, [A comparative study of Reservoir Computing strategies for monthly time series prediction](#), *Neurocomputing*, **73** (2010), 1958–1964, <http://www.sciencedirect.com/science/article/pii/S0925231210000962>.
- [103] F. Wyffels, B. Schrauwen and D. Stroobandt, Using reservoir computing in a decomposition approach for time series prediction, 2008.
- [104] F. Xue, Q. Li and X. Li, [The combination of circle topology and leaky integrator neurons remarkably improves the performance of echo state network on time series prediction.](#), *PLoS one*, **12** (2017).
- [105] K. I. Yared and N. R. Sandell, Maximum likelihood identification of state space models for linear dynamic systems, *Electronic Systems Laboratory, Dept. of Electrical Engineering and Computer Science, Massachusetts Institute of Technology*, **R-814**. Available from: <http://dspace.mit.edu/handle/1721.1/1297>.
- [106] W. Zaremba, I. Sutskever and O. Vinyals, Recurrent neural network regularization, preprint, [arXiv:1409.2329](#).

Received July 2020; 1st revision August 2021; 2nd revision October 2021; early access November 2021.

E-mail address: lyudmila.grigoryeva@warwick.ac.uk

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